Eigenvalue Decomposition of a Parahermitian Matrix: Extraction of Analytic Eigenvalues

Stephan Weiss, *Senior Member, IEEE*, Ian K. Proudler, and Fraser K. Coutts, *Member, IEEE*

***Abstract*—An analytic parahermitian matrix admits an eigenvalue decomposition (EVD) with analytic eigenvalues and eigenvectors except in the case of multiplexed data. In this paper, we propose an iterative algorithm for the estimation of the analytic eigenvalues. Since these are generally transcendental, we find a polynomial approximation with a defined error. Our approach operates in the discrete Fourier transform (DFT) domain and for every DFT length generates a maximally smooth association through EVDs evaluated in DFT bins; an outer loop iteratively grows the DFT order and is shown, in general, to converge to the analytic eigenvalues. In simulations, we compare our results to existing approaches.**

I. Introduction

Most broadband array problems, in which data received by M sensors is held in a zero-mean vector $\mathbf{x}[n] \in \mathbb{C}^M$ with time index $n \in \mathbb{Z}$, can be formulated using polynomial matrix algebra. This includes, e.g., beamforming [1], [2], blind source separation [3], multichannel coding [4], [5], and MIMO system design [6]. Central to this approach is the space-time covariance matrix $\mathbf{R}[\tau] = \mathcal{E}\{\mathbf{x}[n]\mathbf{x}^{\mathrm{H}}[n-\tau]\}$, where $\mathcal{E}\{\cdot\}$ is the expectation operator, $\{\cdot\}^{\mathrm{H}}$ the Hermitian transposition, and $\tau \in \mathbb{Z}$ a lag parameter. Since $\mathbf{R}[\tau]$ contains auto- and cross-correlation sequences, we have $\mathbf{R}[\tau] = \mathbf{R}^{\mathrm{H}}[-\tau]$. Its z-transform, the cross spectral density (CSD) matrix $\boldsymbol{R}(z) = \sum_\tau \mathbf{R}[\tau]z^{-\tau}$, $z \in \mathbb{C}$, satisfies the parahermitian property [7] $\boldsymbol{R}(z) = \boldsymbol{R}^{\mathrm{P}}(z)$, with $\boldsymbol{R}^{\mathrm{P}}(z) \equiv \boldsymbol{R}^{\mathrm{H}}(1/z^*)$.

For narrowband array problems, optimal solutions are often based on the diagonalisation, via an eigenvalue decomposition (EVD), of the narrowband covariance matrix $\mathbf{R}[0]$. For an analogous broadband problem, we would like to diagonalise $\mathbf{R}[\tau]$ for every lag value τ, or equivalently the CSD matrix $\boldsymbol{R}(z)$ for all values of z. The McWhirter decomposition, or polynomial EVD (PEVD) [8], is an approximate factorisation $\boldsymbol{R}(z) \approx \boldsymbol{U}(z)\boldsymbol{D}(z)\boldsymbol{U}^{\mathrm{P}}(z)$ that uses Laurent polynomial matrices $\boldsymbol{U}(z)$ and $\boldsymbol{D}(z)$. The Laurent polynomial matrix $\boldsymbol{D}(z)$ is diagonal and contains the M approximate polynomial eigenvalues $d_m(z)$, $m = 1, \ldots, M$. On the unit circle these eigenvalues are spectrally majorised [9] such that $d_m(\mathrm{e}^{\mathrm{j}\Omega}) \geq d_{m+1}(\mathrm{e}^{\mathrm{j}\Omega})$ $\forall \Omega$ and $m = 1, \ldots, (M-1)$. The polynomial matrix of eigenvectors $\boldsymbol{U}(z)$ is paraunitary [7], such that $\boldsymbol{U}^{-1}(z) = \boldsymbol{U}^{\mathrm{P}}(z)$. Several algorithms exist for calculating the PEVD that are proven to converge to a diagonal $\boldsymbol{D}(z)$, including the sequential best rotation (SBR2) [5], [8], [10], [11]) and sequential matrix diagonalisation (SMD, [12], [13]) families of algorithms. In general, spectral majorisation is encouraged but for the case of SBR2 it has been shown that spectral majorisation is guaranteed [14].

In [15], [16], we have shown that for an analytic parahermitian $\boldsymbol{R}(z)$, a parahermitian matrix EVD (PhEVD)

$$\boldsymbol{R}(z) = \boldsymbol{Q}(z)\boldsymbol{\Lambda}(z)\boldsymbol{Q}^{\mathrm{P}}(z) \tag{1}$$

exists with an analytic paraunitary $\boldsymbol{Q}(z)$ and an analytic diagonal $\boldsymbol{\Lambda}(z) = \mathrm{diag}\{\lambda_1(z), \ldots \lambda_M(z)\}$, unless $\boldsymbol{R}(z)$ arises from multiplexed data [16]. The factors $\boldsymbol{\Lambda}(z)$ and $\boldsymbol{Q}(z)$ are absolutely convergent but generally transcendental functions in z, i.e. (infinite) Laurent series. Hence any algorithm for calculating the PhEVD will most likely only be able to approximate the result (again using Laurent polynomials).

If the underlying eigenvalues $\lambda_m(z)$, $m = 1, \ldots, M$, are not majorised, the PEVD will produce a spectrally majorised solution that is a frequency dependent permutation of the analytic PhEVD solution, as shown in the example in Fig. 1(a) and (b). In this case, profound consequences arise for the PEVD: the eigenvalues in $\boldsymbol{D}(z)$ of the PEVD may need to approximate non-differentiable functions, in which case the PEVD eigenvectors in $\boldsymbol{U}(z)$ must approximate discontinuous functions [15]. Therefore, much higher approximation orders are generally required to model the PEVD factors $\boldsymbol{D}(z)$ and $\boldsymbol{U}(z)$ compared to approximations of the PhEVD factors $\boldsymbol{\Lambda}(z)$ and $\boldsymbol{Q}(z)$ in (1). High polynomial orders are undesirable in terms of the cost of both calculating the PEVD/PhEVD and implementing its factors in, e.g., [1]–[6].

Although [15], [16] proved the existence of analytic eigenvalues, no algorithm was proposed. We are therefore interested in algorithms that can approximate the analytic solution in (1). Because of analyticity of both $\boldsymbol{R}(z)$ and its desired factors, it suffices to consider such functions on the unit circle, $\boldsymbol{R}(\mathrm{e}^{\mathrm{j}\Omega}) = \boldsymbol{R}(z)|_{z=\mathrm{e}^{\mathrm{j}\Omega}}$, since a solution in z may be obtained by re-parameterisation. The key factor here is that analyticity implies smoothness. Related efforts have been undertaken for general matrices $\mathbf{A}(t)$ that depend on a real parameter $t \in \mathbb{R}$ on some interval. For the EVD, singular value (SVD) and QR decompositions, [17] has shown that smooth factorisations exist for matrices $\mathbf{A}(t)$ that are q-times differentiable but not necessarily analytic, with the analytic factorisation case covered in [18]. Specifically, an analytic SVD for an analytic $\mathbf{A}(t)$ has been defined in [19], [20], which uses the arc length or minimum total variation of an interpolation across SVDs of the discretised $\mathbf{A}(t)$ as an optimisation criterion

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1 and the MOD University Defence Research Collaboration in Signal Processing.

S. Weiss and I.K. Proudler are with the Dept. of EEE, University of Strathclyde, Glasgow G1 1XW, Scotland (e-mail {stephan.weiss,ian.proudler}@strath.ac.uk.

F.K. Coutts is with the Institute for Digital Communications (IDCOM), University of Edinburgh, Scotland (e-mail fraser.coutts@ed.ac.uk).

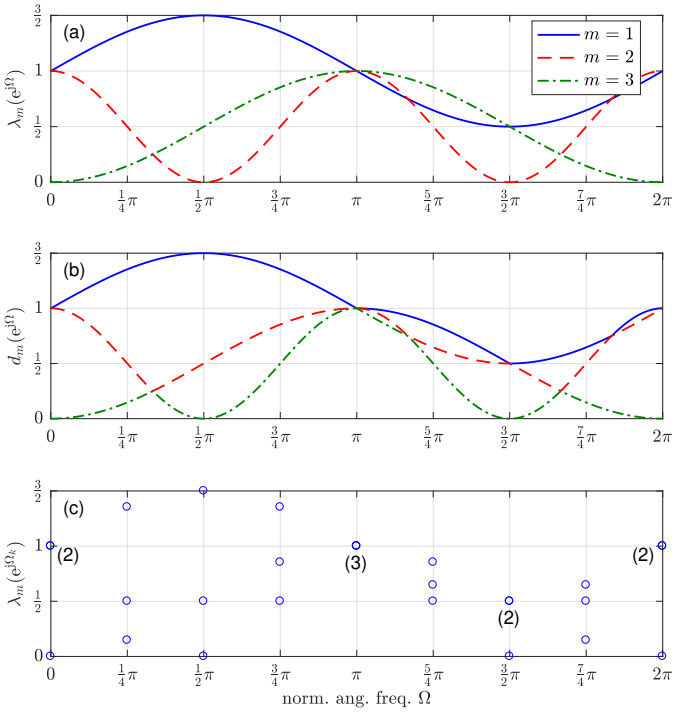

Fig. 1. (a) Analytic eigenvalues of the PhEVD of a matrix $\boldsymbol{R}(z)$, evaluated on the unit circle, $z = \mathrm{e}^{\mathrm{j}\Omega}$, (b) spectrally majorised eigenvalues as approximated by the PEVD, and (c) eigenvalues evaluated in $K = 8$ DFT bins.

for smoothness [20]. A similar differential equation-based approach in [21] is solved via a Runge-Kutta method. In [22], a sample-based 'continuation algorithm' employs prediction/correction steps to weave smooth functions through discrete samples along the t-axis. More recently, Chebyshev interpolation is utilised in [23] to determine an analytic SVD of $\boldsymbol{A}(t)$. Note however, that an analytic decomposition of $\boldsymbol{A}(t)$ does not imply that such a decomposition exists for a matrix $\boldsymbol{B}(\mathrm{e}^{\mathrm{j}\Omega})$ that depends periodically on $\Omega \in \mathbb{R}$ [16]. If it does exist, then the cyclic nature offers unique exploitation in terms of interpolation and efficient implementation that are not utilised in [20]–[23].

In the context of discrete time systems, [24], [25] have investigated the QR decomposition of a general matrix $\boldsymbol{B}(z)$ on the unit circle. The assumption there is that the number of sample points along the unit circle is sufficiently high to permit an interpolated solution. Finally, Tohidian *et al.* [26] have chosen a DFT-based approach to a polynomial eigenvalue decomposition; operating with a set DFT length, they perform a bin-wise EVD and aim to approximate either spectrally majorised or smooth eigenvalues. The approaches in [24]–[26] do not explicitly exploit analyticity of the factors. Particularly for [26], no convergence proof is stated, and there is no mechanism for calculating what DFT length might be sufficient.

DFT-based approaches that treat the frequency bins as independent somewhat negate the advantage of spectral coherence that is guaranteed by time-domain PEVD algorithms such as [5], [8], [10]–[13]. Therefore, spectral coherence has to be re-introduced across bins similarly to [20]–[23]. In the case of the smooth polynomial EVD in [26], this is accomplished by monitoring the orthogonality of eigenvectors across bins. Since eigenvectors are not unique [15] and can form an arbitrary C-fold basis at a C-fold algebraic multiplicity of eigenvalues, this can be misleading [27].

In this paper, building on the work in [27], we present an approach to extract analytic eigenvalues from an analytic CSD matrix $\boldsymbol{R}(z)$. We exploit the absolute convergence of analytic functions twofold: firstly we utilise infinite differentiability to define a smoothness criterion that can establish a coherent association across DFT bins based on [28], [29]. Secondly, we iteratively increase the DFT length and thereby approximation order until a sufficiently low truncation error is achieved. Different from [26], we operate on the eigenvalues only and leave aside the potentially more perturbed eigenvectors in determining associations. Unlike [27], the algorithm uses a computationally cheaper and more robust evaluation of a smoothness metric as defined in [29], and a modified cost function for which convergence can be proven.

Below, Sec. II defines some important properties of the PhEVD, followed by the definition of a smoothness metric in Sec. III. The proposed algorithm is outlined in Sec. IV, which for a given DFT length finds a maximally smooth association across the frequency bins, and then iteratively increases the DFT length until a correct association with a sufficiently small approximation error is found. A maximum likelihood sequence estimation approach for this algorithm is presented in Sec. V, which is benchmarked in Sec. VI.

II. Parahermitian Matrix EVD

A. Space-Time Covariance Model

The space-time covariance $\mathbf{R}[\tau]$ is based on the measurement vector $\mathbf{x}[n] = [x_1[n], \ldots, x_M[n]]^{\mathrm{T}}$, which is assumed to emerge from the source model in Fig. 2; this contains a convolutive mixing system $\mathbf{H}[n] \in \mathbb{C}^{M \times L_\mathrm{s}}$ that combines L_s mutually independent source signals $s_\ell[n]$, $\ell = 1, \ldots, L_\mathrm{s}$. This is a generic signal model which includes the case where some or all of the sources are spectrally majorised. We exclude the case of mutually dependent signals since, as with the ordinary EVD, being correlated they would be seen by the PhEVD as one signal. The z-transform $\boldsymbol{H}(z) = \sum_n \mathbf{H}[n] z^{-n}$: $\mathbb{C} \to \mathbb{C}^{M \times L_\mathrm{s}}$ is a matrix of transfer functions. In the following we represent such a z-transform pair using the notation $\mathbf{H}[n] \circ\!\!-\!\!\bullet\, \boldsymbol{H}(z)$. The input signals to the mixing system, $s_\ell[n]$, can be tied to zero mean unit variance uncorrelated Gaussian sources $u_\ell[n]$ via innovation filters $g_\ell[n] \circ\!\!-\!\!\bullet\, G_\ell(z)$ [30], such that the power spectral density of $s_\ell[n]$ is given by $\Gamma_\ell(z) = G_\ell(z) G_\ell^{\mathrm{P}}(z)$ if evaluated on the unit circle. With $\boldsymbol{G}(z) = \mathrm{diag}\{G_1(z), \ldots, G_{L_\mathrm{s}}(z)\}$,

$$\boldsymbol{R}(z) = \boldsymbol{H}(z)\, \boldsymbol{G}(z)\, \boldsymbol{G}^{\mathrm{P}}(z) \boldsymbol{H}^{\mathrm{P}}(z) \tag{2}$$

is the CSD matrix based on the source model in Fig. 2. We assume that we are given an error-free CSD matrix based on such a source model, but note that there might be issues when using estimated covariance matrices from finite data [60].

If the system components of this model are causal and stable, then the CSD matrix $\boldsymbol{R}(z)$ in (2) will be analytic and parahermitian. If evaluated on the unit circle, $\boldsymbol{R}(\mathrm{e}^{\mathrm{j}\Omega}) = \boldsymbol{R}(z)|_{z=\mathrm{e}^{\mathrm{j}\Omega}}$ is a self-adjoint operator such that $\boldsymbol{R}(\mathrm{e}^{\mathrm{j}\Omega}) =$

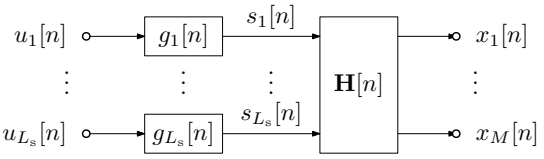

Fig. 2. Source model for measurement vector $\mathbf{x}[n]$.

$\boldsymbol{R}^{\mathrm{H}}(\mathrm{e}^{\mathrm{j}\Omega})$, i.e., when evaluated at any particular frequency $\Omega_0 \in \mathbb{R}$, $\boldsymbol{R}(\mathrm{e}^{\mathrm{j}\Omega_0})$ is a Hermitian matrix. It is also positive semi-definite.

B. Parahermitian Matrix EVD

As long as the source model in Fig. 2 cannot be tied to any multiplexing operations [16], for an analytic $\boldsymbol{R}(z)$ the following can be stated for its eigenvalues in (1).

Theorem 1 (Eigenvalues of a parahermitian matrix): For an analytic, parahermitian, and non-multiplexed $\boldsymbol{R}(z) : \mathbb{C} \to \mathbb{C}^{M\times M}$, there exist M unique analytic eigenvalues.

Proof. This has been proven in [16], based on analysis in [15] extending results for analytic self-adjoint matrices in [31]. ■

C. Polynomial Approximation of PhEVD

Even if $\boldsymbol{R}(z)$ is a Laurent polynomial, in most cases, both $\boldsymbol{Q}(z)$ and $\boldsymbol{\Lambda}(z)$ in (1) will be (infinite) Laurent series, and represent algebraic or even transcendental functions. We are therefore interested in approximating $\boldsymbol{Q}(z)$ and $\boldsymbol{\Lambda}(z)$ by Laurent polynomial factors $\hat{\boldsymbol{Q}}^{(N)}(z)$ and $\hat{\boldsymbol{\Lambda}}^{(N)}(z)$ of order N. Given its profound consequences for our algorithmic approach later, we state the following relatively trivial and well-known theorem for the approximate eigenvalues $\hat{\boldsymbol{\Lambda}}^{(N)}(z)$.

Theorem 2 (Laurent polynomial approximation): For $|z| = 1$, the best Nth-order approximation $\hat{\boldsymbol{\Lambda}}^{(N)}(z)$ of an analytic $\boldsymbol{\Lambda}(z)$ in the least squares sense is obtained by truncating $\boldsymbol{\Lambda}(z)$ to the required order.

Proof. See Appendix A. ■

Further, the Fourier transform $\hat{\lambda}^{(N)}(\mathrm{e}^{\mathrm{j}\Omega}) \bullet\!\!-\!\!\circ\, \hat{\lambda}^{(N)}[\tau]$ will converge uniformly to $\lambda(\mathrm{e}^{\mathrm{j}\Omega})$ [32], i.e., the approximation error decreases to an arbitrarily small value at every frequency $\Omega \in \mathbb{R}$ given a sufficient approximation order N.

D. EVD at Sample Points of $\boldsymbol{R}(z)$

In order to build a DFT domain algorithm, we take K discrete sample points[1] of $\boldsymbol{R}(z)$ along the unit circle, resulting in $\mathbf{R}_k = \boldsymbol{R}(z)|_{z=\mathrm{e}^{\mathrm{j}2\pi k/K}}$, $k = 0,\ldots,(K-1)$, which is Hermitian according to Sec. II-A. The EVD $\mathbf{R}_k = \mathbf{Q}_k\mathbf{\Lambda}_k\mathbf{Q}_k^{\mathrm{H}}$ provides a diagonal $\mathbf{\Lambda}_k$,

$$\mathbf{\Lambda}_k = \mathrm{diag}\left\{ F_{1,k}^{(K)}, \ldots, F_{M,k}^{(K)} \right\} , \tag{3}$$

whereby we assume that the eigenvalues $F_{\mu,k}^{(K)}$, $\mu = 1,\ldots,M$, are majorised, i.e. arranged in descending order, such that $F_{\mu,k}^{(K)} \geq F_{\mu+1,k}^{(K)}$ $\forall \mu = 1,\ldots,(M-1)$. Because of this ordering, we find that in general the $F_{m,k}^{(K)}$ consist of samples from different $\lambda_m(z)$ i.e.

$$F_{\mu(m,k),k}^{(K)} = \lambda_m(z)|_{z=\mathrm{e}^{\mathrm{j}2\pi k/K}} \tag{4}$$

where $\mu(m,k)$ is a suitable permutation, and $\lambda_m(z)$ the mth eigenvalue of $\boldsymbol{R}(z)$ in (1). Since in (4), $\mu(m,k) = m$ cannot be assumed, the challenge in interpolating from the sample points $F_{\mu,k}^{(K)}$ to produce an approximation of the true, analytic eigenvalues is to find the correct mapping or permutation $\mu(m,k)$ of indices at each sample point. We will identify this permutation using the infinitely differentiability property of analytic functions via a metric for smoothness. Note that some constructed pathological cases exist where infinitely differentiable aperiodic functions are not analytic [33]. We are not aware of any periodic, real-valued, infinitely differentiable functions that are not analytic; if they exist, then in such pathological cases this approach may not work.

III. Measuring Smooth Association Across Frequency Bins

We are given sample points $F_{\mu,k}^{(K)}$, $\mu = 1,\ldots,M$ via (4) that are majorised in every bin $k = 0,\ldots,(K-1)$. The aim is to create associations across frequency bins, i.e. to find a sequence $\{F_{m,k}^{(K,a)}\}$ by reordering the sample points, on a bin-by-bin basis, in $\{F_{m,k}^{(K)}\}$ such that the interpolating functions $F_m(\mathrm{e}^{\mathrm{j}\Omega})$, with $F_m(\mathrm{e}^{\mathrm{j}\Omega_k}) = F_{m,k}^{(K,\mathrm{a})}$, $\Omega_k = \frac{2\pi k}{K}$, $k = 0,\ldots,(K-1)$, $m = 1,\ldots,M$, are maximally smooth. The best association will be found by suitably measuring the smoothness of an interpolation. For simplicity, in the following we only work with a single function, and omit the spatial index m or μ and the superscript 'a' for 'associated'.

Based on the definition of a Dirichlet kernel in Sec. III-A and of a maximally smooth interpolation in Sec. III-B, we derive a metric in Sec. III-C. This material is based on earlier work in [28], [29], but specifically addresses interpolation that reflects properties of eigenvalues, i.e. symmetry in the time domain or real-valuedness in the Fourier domain, and presents findings on improved numerical stability and computational complexity in a compacted and more rigorous form.

A. Dirichlet Interpolation

Ultimately, we will measure smoothness via the derivative of a smooth, continuous function $F^{(K)}(\mathrm{e}^{\mathrm{j}\Omega})$ that interpolates K sample points $F_k^{(K)}$. Typically, for periodic problems, the sampling function or interpolant [34], [35] relies on the periodic sinc function or Dirichlet kernel [36]

$$P_K(\mathrm{e}^{\mathrm{j}\Omega}) = \frac{\sin\frac{K}{2}\Omega}{\sin\frac{1}{2}\Omega} . \tag{5}$$

Its inverse Fourier transform $p_K[\tau] \circ\!\!-\!\!\bullet\, P_K(\mathrm{e}^{\mathrm{j}\Omega})$,

$$p_K[\tau] = \begin{cases} 1, & |\tau| \leq \frac{K-1}{2} , \\ 0, & \text{otherwise} , \end{cases} \tag{6}$$

is a rectangular window centred at the origin. In the definition of the zero-centred rectangular window, K in (5) and (6) must

[1] The term 'samples' refers to discrete values in the time domain, while 'sample points' indicates discrete values taken on the unit circle.

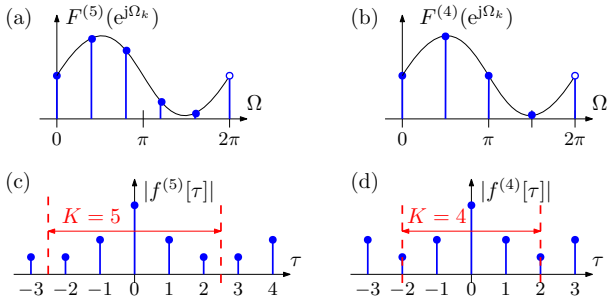

Fig. 3. Example for taking an (a) odd and (b) even number (K) of sample points of a real valued function on the unit circle, and their respective K-periodic time domain equivalents in (c) and (d), respectively.

be odd: for even K, the window in (6) becomes untethered from the sampling grid, while (5) becomes 4π-periodic [34].

While odd values of K match the odd support length of an auto-correlation-type sequence, one might prefer to rely on power-of-two-point fast Fourier transforms (FFT) to obtain the sample points. In order to admit general values $K \in \mathbb{N}$, the window in (6) is typically shifted to the interval $\tau = 0, \ldots, (K-1)$ [34], [35]. Its Fourier transform is $P_K(e^{j\Omega})e^{-j\frac{1}{2}(K-1)\Omega}$, i.e. it is phase-shifted with respect to (5) and therefore complex-valued. Since here the sample points are taken from a power spectral density-type function, we would like the interpolant $F^{(K)}(e^{j\Omega})$ to reflect its properties, i.e. be real-valued, and potentially also be symmetric in the case that real-valued data generates $\boldsymbol{R}(z)$. In order to satisfy these properties, we require an interpolant to be both real valued and symmetric. In the following section we will explore how this can be accomplished without a time or phase shift.

B. Modified Dirichlet Interpolation

To find a suitable interpolant, consider an analytic function[2] $\Lambda(e^{j\Omega}) \bullet\!\!-\!\!\circ\, \lambda(\tau)$ that generates the sample points $F_k^{(K)} = \lambda(e^{j\Omega_k})$, $\Omega_k = \frac{2\pi}{K}k$, $k = 0, \ldots, (K-1)$. Let $f^{(K)}[\tau]$ be the inverse DFT of $F_k^{(K)}$. Because $F_k^{(K)}$ is discretised in the Fourier domain, $f^{(K)}[\tau]$ will be periodised w.r.t. a fundamental period of K samples in the time domain, such that

$$f^{(K)}[\tau] = \sum_{\nu=-\infty}^{\infty} \lambda[\tau - \nu K] \,. \tag{7}$$

The symmetry of $f^{(K)}[\tau]$ is most easily satisfied if K is odd, as shown in Fig. 3(a) and (c). If K is even, as illustrated by the example in Fig. 3(b), then the periodicity above enforces symmetry with respect to $\tau = K/2$ as shown in Fig. 3(d). This in turn requires that the images of $\lambda[\tau]$ in (7) overlap by at least one point.

We therefore define as a sampling function, or interpolant, the real valued symmetric function

$$\tilde{p}_K[\tau] = \begin{cases} p_K[\tau], & K \text{ odd}, \\ p_{K-1}[\tau] + \frac{1}{2}\delta[\tau - \frac{K}{2}] & \\ \qquad + \frac{1}{2}\delta[\tau + \frac{K}{2}], & K \text{ even}, \end{cases} \tag{8}$$

[2]The correspondence is a DFT between a function indexed by τ in the time domain and one indexed by k in the frequency domain.

where $p_N[\tau]$ is the rectangular window in (6); its length N must be odd. Since $\tilde{p}_K[\tau]$ is symmetric and real valued, its Fourier transform $\tilde{P}_K(e^{j\Omega}) \bullet\!\!-\!\!\circ\, \tilde{p}_K[\tau]$,

$$\tilde{P}_K(e^{j\Omega}) = \begin{cases} \frac{\sin\frac{K}{2}\Omega}{\sin\frac{1}{2}\Omega}, & K \text{ odd}, \\ \left(\frac{\sin\frac{K-1}{2}\Omega}{\sin\frac{1}{2}\Omega} + \cos\frac{K}{2}\Omega\right), & K \text{ even}, \end{cases} \tag{9}$$

therefore is also real valued and symmetric. It represents a Dirichlet kernel or periodic sinc when K is odd. In the case of K is even, it is a modified Dirichlet kernel[3].

Theorem 3 (Maximally smooth interpolation): The kernel $\frac{1}{K}\tilde{P}_K(e^{j\Omega})$ in (9) is the smoothest possible real valued and symmetric interpolant through the K sample points on the unit circle in the sense that $\tilde{p}_K[\tau] \circ\!\!-\!\!\bullet\, \tilde{P}_K(e^{j\Omega})$ is the shortest possible window to satisfy the Nyquist-K condition.

Proof. We require a continuous sampling function that interpolates the sampling points on the unit circle with the smoothest possible continuous function in Ω. Firstly, it is well-known that a sampling function, for K samples, must satisfy the Nyquist-K condition [7]. Secondly, the time-bandwidth product of the Fourier transform ties maximal smoothness in one domain to the shortest possible support in the other domain.

With respect to the sampling function, because we interpolate through sample points in the Fourier domain, the usual time- and frequency-domain meanings of a Nyquist-K system [7], [37] are reversed. If a function $F(e^{j\Omega})$ is sampled K-fold on the unit circle, such that $F_k^{(K)} = F(e^{j\Omega_k})$, $\Omega_k = 2\pi k/K$, then for $f[\tau] \circ\!\!-\!\!\bullet\, F(e^{j\Omega})$ and $f^{(K)}[\tau] \circ\!\!-\!\!\bullet\, F_k^{(K)}(e^{j\Omega})$,

$$f^{(K)}[\tau] = \sum_{\nu=-\infty}^{\infty} f[\tau - \nu K] \,. \tag{10}$$

With $F(e^{j\Omega}) = \tilde{P}_K(e^{j\Omega})$, i.e. $f[\tau] = \tilde{p}_K[\tau]$, it is easy to show, with (8), that due to the periodisation (10) becomes

$$\tilde{p}_K^{(K)}[\tau] = 1 \qquad \forall \tau \,. \tag{11}$$

Note that with the domains reversed, (11) is equivalent to the typical frequency domain definition of a Nyquist-K system. In the 'other' domain, the interpolating kernel exhibits $K-1$ regular zero crossings, as shown in the example in Fig. 4. Hence, the modified periodic sinc or Dirichlet window in (8) and (9) represents a Nyquist-K system.

It remains to be shown that $\tilde{p}_K[\tau]$ is the shortest possible window. If instead of $f[\tau] = \tilde{p}_K[\tau]$ we chose a different window of length $J < K$, then the shifted versions of $f[\tau]$ in (10) will be separated by $(K-J)$ zero values, and hence the Nyquist-K property (11) is not satisfied. This therefore necessitates the form in (8). ∎

The smoothest possible interpolation through K uniformly spaced sample points on the unit circle, $F_k^{(K)} \in \mathbb{R}$, $k = 0, \ldots, (K-1)$, can therefore be based on the modified

[3]In the signal processing literature, real valued Dirichlet kernels for even K are generally excluded; see e.g. explicitly [35] or implicitly [34], [37], [38]. Nonetheless, in trigonometric interpolation theory a 'modified Dirichlet kernel' [39] exists for even K, while here this term refers to $K \in \mathbb{N}$ via (9).

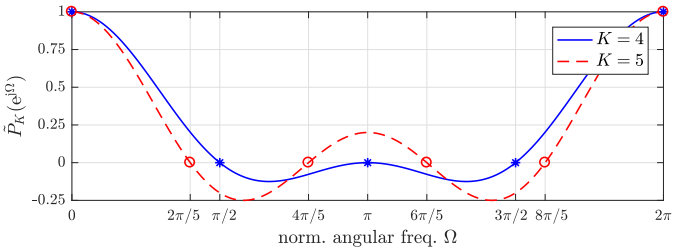

Fig. 4. Examples of modified Dirichlet kernels to provide a maximally smooth interpolation through $K = \{4, 5\}$ sample points.

Dirichlet kernel, s.t.

$$F^{(K)}(e^{j\Omega}) = \frac{1}{K}\sum_{k=0}^{K-1} F_k^{(K)} \tilde{P}_K(e^{j(\Omega-2\pi k/K)}) = \frac{1}{K}\sum_{k=0}^{K-1} F_k^{(K)} \sum_{\tau=-L_K}^{L_K} \tilde{p}_K[\tau] e^{-j(\Omega-2\pi k/K)\tau} ,$$

with $\tilde{p}_K[\tau]$ as defined in (8) and $L_K = (K-1)/2$ for K odd, and $L_K = K/2$ for K even. Using matrix and vector notation with

$$\mathbf{e}_K^{\mathrm{H}}(e^{j\Omega}) = \left[e^{j\Omega L_K},\ e^{j\Omega(L_K-1)},\ \cdots,\ e^{-j\Omega L_K}\right] \quad (12)$$

$$\mathbf{P}_K = \mathrm{diag}\{\tilde{p}_K[-L_K],\ \tilde{p}_K[-L_K+1],\ \ldots,\ \tilde{p}_K[L_K]\} \quad (13)$$

$$\mathbf{T}_K = \left[\begin{array}{cc} \mathbf{0}_{L_K\times(K-L_K)} & \mathbf{I}_{L_K} \\ \mathbf{I}_{L_K+1} & \mathbf{0}_{(L_K+1)\times(K-L_K-1)} \end{array}\right] \mathbf{W}_K^{\mathrm{H}} \quad (14)$$

$$\mathbf{f}_K^{\mathrm{T}} = \left[F_0^{(K)}, \ldots, F_{K-1}^{(K)}\right] , \quad (15)$$

where $\mathbf{W}_K$ is a K-point DFT matrix that is scaled to be unitary, and $\mathbf{T}_k \in \mathbb{Z}^{K\times K}$ for K odd and $\mathbf{T}_k \in \mathbb{Z}^{(K+1)\times K}$ for K even, the interpolation can be written as

$$F^{(K)}(e^{j\Omega}) = \frac{1}{\sqrt{K}}\mathbf{e}_K^{\mathrm{H}}(e^{j\Omega})\mathbf{P}_K\mathbf{T}_K\mathbf{f}_K . \quad (16)$$

C. Smoothness Metric

If a correct association has been made, the sample points $F_k^{(K)}$ belong to an analytic function, which will be infinitely differentiable. Therefore, we want to assess the smoothness of the interpolant $F^{(K)}(e^{j\Omega})$, and use the power in the pth derivative,

$$\chi_p = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\frac{\mathrm{d}^p}{\mathrm{d}\Omega^p} F^{(K)}(e^{j\Omega})\right|^2 \mathrm{d}\Omega , \quad (17)$$

as a discriminator. Employing the matrix notation of (16), the pth derivative of $F^{(K)}(e^{j\Omega})$ can be written as

$$\frac{\mathrm{d}^p}{\mathrm{d}\Omega^p} F^{(K)}(e^{j\Omega}) = \frac{1}{\sqrt{K}}\mathbf{e}_K^{\mathrm{H}}(e^{j\Omega})\mathbf{D}_K^p\mathbf{P}_K\mathbf{T}_K\mathbf{f}_K ,$$

where

$$\mathbf{D}_K = \mathrm{diag}\{-jL_K,\ -j(L_K-1),\ \ldots,\ jL_K\} . \quad (18)$$

We evaluate (17) using Parseval's theorem [28], [34], [35] so that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\mathbf{e}_K^{\mathrm{H}}(e^{j\Omega})\boldsymbol{\alpha}\right|^2 \mathrm{d}\Omega = \boldsymbol{\alpha}^{\mathrm{H}}\boldsymbol{\alpha} , \quad (19)$$

which only depends on $\boldsymbol{\alpha}$ but no longer depends on Ω. Rewriting (17), we can therefore assess the power in the derivatives of the smoothest possible interpolation through the sample points $F_k^{(N)}$ via

$$\chi_p = \|\mathbf{D}_K^p\mathbf{P}_K\mathbf{T}_K\mathbf{f}_K\|_2^2 . \quad (20)$$

Note that (20) may be written as

$$\chi_p = \mathbf{f}_K^{\mathrm{T}}\mathbf{C}_{p,K}\mathbf{f}_K , \quad (21)$$

where $\mathbf{C}_{p,K}$ is independent of the sample points contained in the vector $\mathbf{f}_K \in \mathbb{R}^K$.

Since we later evaluate the smoothness via the weighted inner product in (21), some properties of $\mathbf{C}_{p,K}$ are noteworthy. By construction, $\mathbf{C}_{p,K}$ is a Hermitian matrix and positive semi-definite. When multiplied out, it is easily demonstrated that

$$\mathbf{C}_{p,K} = \mathbf{W}_K\tilde{\mathbf{D}}_{p,K}\mathbf{W}_K^{\mathrm{H}} , \quad (22)$$

where

$$\tilde{\mathbf{D}}_{p,K} = \begin{cases} \mathrm{diag}\left\{0^{2p},\ 1^{2p},\ \ldots,\ L_K^{2p},\ L_K^{2p},\ \ldots, 1^{2p}\right\} , \\ \qquad\qquad K \text{ odd} , \quad (23a) \\ \mathrm{diag}\left\{0^{2p},\ 1^{2p},\ \ldots,\ (L_K-1)^{2p},\ \frac{1}{2}L_K^{2p},\right. \\ \left.\qquad (L_K-1)^{2p},\ \ldots,\ 1^{2p}\right\} , \quad K \text{ even} . \quad (23b) \end{cases}$$

Furthermore, because of the DFT matrix $\mathbf{W}_K$ in (22), $\mathbf{C}_{p,K}$ is circulant and contains as elements the inverse DFT of the diagonal elements of $\tilde{\mathbf{D}}_{p,K}$ [7]. We also find that (22) is real-valued because the diagonal elements form sequences that are symmetric w.r.t. their $(L_K^{2p}+1)$th entry, i.e. that $\mathbf{C}_{p,K} \in \mathbb{R}^{K\times K}$. Finally, since (22) also represents an eigenvalue decomposition, the eigenvalue matrix $\tilde{\mathbf{D}}_{p,K}$ reveals that $\mathbf{C}_{p,K}$ is rank deficient for $p > 0$.

If smoothness is measured by accumulating the cost χ_p up to the Pth derivative [28], then

$$\chi^{(P)} = \sum_{p=0}^{P}\chi_p = \mathbf{f}_K^{\mathrm{T}}\sum_{p=0}^{P}\mathbf{C}_{p,K}\mathbf{f}_K = \mathbf{f}_K^{\mathrm{T}}\mathbf{C}_K^{(P)}\mathbf{f}_K , \quad (24)$$

where $\mathbf{C}_K^{(P)} \in \mathbb{R}^{K\times K}$ inherits the properties of its constituent components $\mathbf{C}_{p,K}$. Note that $\mathbf{C}_{p,K}$ has rank $(K-1)$ for $p > 0$, but that $\mathbf{C}_{0,K}$, due to $\tilde{\mathbf{D}}_{p,K}$ in (23), is full rank and hence the cumulative matrix $\mathbf{C}_K^{(P)}$ is also full rank. Since the eigenvalues of the latter are $\sum_{p=0}^{P}\tilde{\mathbf{D}}_{p,K}$, the condition number of $\mathbf{C}_K^{(P)}$ is approximately proportional to L_K^{2p}.

IV. Iterative Extraction of Maximally Smooth Associations

A. General Approach

Driven by the smoothness metric in (24), the aim is now to assign the MK sample points $F_{m,k}^{(K)}$ in (3) to M interpolants represented each by their K sample points in $\mathbf{f}_{K,m} \in \mathbb{R}^K$, $m = 1, \ldots, M$. An example for the $M = 3$ eigenvalues in Fig. 1(b) with $K = 8$ is shown in Fig. 1(c). To associate the

Algorithm 1: Extraction of Analytic Eigenvalues

1: initialise K to exceed the support of $\boldsymbol{R}(z)$;
2: determine eigenvalues of $\boldsymbol{R}(\mathrm{e}^{\mathrm{j}\Omega_k})$, $k = 0, \ldots, (K-1)$;
3: determine smoothest association $F_{m,k}^{(K,\mathrm{a})}$;
4: **repeat**
5: $\quad K \leftarrow 2K$;
6: $\quad$ determine eigenvalues of $\boldsymbol{R}(\mathrm{e}^{\mathrm{j}\Omega_k})$, $k = 0, \ldots, (K-1)$;
7: $\quad$ determine smoothest association $F_{m,k}^{(K,\mathrm{a})}$;
8: $\quad$ from $F_{m,k}^{(K,\mathrm{a})}$ and $F_{m,k}^{(K/2,\mathrm{a})}$, determine metrics ξ_1, ξ_2;
9: **until** $(\xi_1 < \epsilon_1) \wedge (\xi_2 < \epsilon_2)$.

sample points such that the M interpolants collectively are as smooth as possible, we solve

$$\{F_{m,k}^{(K,\mathrm{a})}\} = \underset{\mathbf{f}_{K,m}}{\operatorname{argmin}} \sum_{m=1}^{M} \mathbf{f}_{K,m}^{\mathrm{T}} \mathbf{C}_K^{(P)} \mathbf{f}_{K,m} \;, \tag{25}$$

where the superscript $\{\cdot\}^{(a)}$ indicates that sample points are now ordered such that they associate with maximally smooth functions. For the DFT length $K \to \infty$ and the derivative order $P \to \infty$, ignoring the potential pathological cases mentioned in Sec. II-D, (25) will be proven to provide the association of sample points corresponding to the analytic functions.

Two challenges arise from (25). Firstly, $M!$ possible permutations have to be checked in each of the K bins, and it is left to Sec. V to replace an exhaustive search with a more scalable procedure. Secondly, for too small a value of K, (i) the number of sample points K may be insufficient to adequately represent the analytic functions, resulting in time-domain aliasing; and (ii) $F_{m,k}^{(K,\mathrm{a})}$ may not be correctly associated with the analytic functions.

The proposed procedure is outlined in Algorithm 1. For a given DFT length K, the algorithm finds the best association of eigenvalues across the DFT bins via (25); it then evaluates two metrics to test for the cases of time-domain aliasing and/or an incorrect association, and if necessary iteratively increases the size of the DFT. Across two successive DFT lengths $\frac{K}{2}$ and K, a metric ξ_1 checks the consistency of associations. If consistent, then a metric ξ_2 captures aliasing and correct association. In the time domain, ξ_2 assesses the difference between the tail ends of $\hat{\lambda}_m^{(K)}[\tau]$ and $\hat{\lambda}_m^{(K/2)}[\tau]$, which in the case of analytic functions according to Theorem 2 should go to zero as $K \to \infty$. In the frequency domain, ξ_2 exploits the fact that if $F_{m,k}^{(K,\mathrm{a})}$ are not associated with analytic functions, they can be sampled from piecewise analytic functions, which for some derivative $p < \infty$ will be discontinuous.

In the following we first discuss the metrics and then prove that the proposed algorithm converges to the correct solution.

B. Time vs DFT Domain Approximation and Metrics

Ultimately, we are interested in the time domain eigenvalue functions. From Theorem 1, we know that analytic eigenvalues $\lambda_m(z)$, $m = 1, \ldots, M$, exist for any analytic parahermitian matrix $\boldsymbol{R}(z)$ that does not arise from multiplexed data. Analyticity implies that their time domain equivalent functions $\lambda_m[\tau] \circ\!\!-\!\!\bullet\, \lambda_m(z)$ are absolutely convergent and Theorem 2 guarantees that a truncation $\hat{\lambda}_m^{(N)}[\tau] = \tilde{p}_N[\tau]\lambda_m[\tau]$ provides its best Nth order approximation. Furthermore, its Fourier transform $\hat{\lambda}_m^{(N)}(\mathrm{e}^{\mathrm{j}\Omega}) \bullet\!\!-\!\!\circ\, \hat{\lambda}_m^{(N)}[\tau]$, converges uniformly, such that

$$|\hat{\lambda}_m^{(N)}(\mathrm{e}^{\mathrm{j}\Omega}) - \lambda_m(\mathrm{e}^{\mathrm{j}\Omega})| < \epsilon \quad \forall \Omega$$

is achieved for an arbitrarily small ϵ by a sufficiently large N. All of this follows directly from the analyticity of $\boldsymbol{\Lambda}(z)$.

The same cannot be said for the K-th order frequency domain interpolation function. If $F_{m,k}^{(K,\mathrm{a})}$ are correctly associated, then these will match the analytic eigenvalue $\lambda_m(z)$ in the sample points, i.e.

$$F_{m,k}^{(K,\mathrm{a})} = \lambda_m(\mathrm{e}^{\mathrm{j}\Omega_k}),\ \Omega_k = 2\pi k/K,\ k = 0, \ldots, (K-1)\ .$$

Since $F_{m,k}^{(K,\mathrm{a})}$ is discrete, its time domain equivalent $f_m^{(K,\mathrm{a})}[\tau] \circ\!\!-\!\!\bullet\, F_{m,k}^{(K,\mathrm{a})}$,

$$f_m^{(K,\mathrm{a})}[\tau] = \sum_{\mu=-\infty}^{\infty} \lambda_m[\tau - K\mu] \;, \tag{26}$$

will be periodic analogously to (7). When the support of $\lambda_m[\tau]$ exceeds K — recalling that it is very likely an infinite Laurent series — then (26) results in time domain aliasing. If we isolate the fundamental period, indicated by $\hat{\{\cdot\}}$, using (8),

$$\hat{f}_m^{(K,\mathrm{a})}[\tau] = \tilde{p}_K[\tau] f_m^{(K,\mathrm{a})}[\tau] \;, \tag{27}$$

then $\hat{F}_m^{(K,\mathrm{a})}(\mathrm{e}^{\mathrm{j}\Omega}) \bullet\!\!-\!\!\circ\, \hat{f}_m^{(K,\mathrm{a})}[\tau]$ will match $\lambda_m(\mathrm{e}^{\mathrm{j}\Omega})$ in the sample points but may deviate at other frequencies. Thus, even if (25) provides the correct association, in the DFT-based approach, $\hat{F}_m^{(K,\mathrm{a})}(\mathrm{e}^{\mathrm{j}\Omega})$ does not converge uniformly towards $\lambda_m(\mathrm{e}^{\mathrm{j}\Omega})$ as K increases: in the sample points, the approximation error is zero and $\hat{F}_m^{(K,\mathrm{a})}(\mathrm{e}^{\mathrm{j}\Omega})$ converges instantly, while at other frequencies this is not the case.

For a sufficiently large K, i.e. closely spaced sample points, (25) will yield the correct association and our interpolant will have an estimation error (due to truncation and aliasing) that is less than some threshold. To find a small value of K that satisfies such a threshold, we iteratively increase the DFT size by a factor of two. While the specific factor for this increase is somewhat arbitrary, selecting powers of two as DFT length K is advantageous w.r.t. efficient DFT algorithms. Furthermore, when calculating the EVD in each frequency bin for a K-point DFT of $\mathbf{R}[\tau]$, half of the eigenvalues have already been calculated during the previous iteration with DFT length $\frac{K}{2}$, i.e.

$$F_{m,2k}^{(K)} = F_{m,k}^{(K/2)} \quad \forall m = 1, \ldots, M,\ k = 0, \ldots, (\tfrac{K}{2} - 1)\ .$$

To exploit this property, we assume in the rest of the paper that K is even.

Consider the following metric

$$\xi_1 = \sum_{m=1}^{M} \sum_{k=0}^{K/2-1} |F_{m,2k}^{(K,\mathrm{a})} - F_{m,k}^{(K/2,\mathrm{a})}|^2 \tag{28}$$

where $F_{m,k}^{(K/2,\mathrm{a})}$ and $F_{m,k}^{(K,\mathrm{a})}$ are the associated sets from (25) for the $\frac{K}{2}$- and K-point DFTs. If, for ϵ_1 a small constant

numerically close to zero, we have that $\xi_1 > \epsilon_1$ then the K-point DFT has produced an association that differs from the one for the $\frac{K}{2}$-point DFT. As a result, at least the association for $\frac{K}{2}$ is incorrect. Hence $\frac{K}{2}$, and possibly K, is too small.

For $\xi_1 \leq \epsilon_1$, the smooth associations for both $\frac{K}{2}$- and K-point DFT sets are identical, and there is at least a chance that the associations may be correct. If the associations are correct, then the difference between the interpolants,

$$\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega}) = \hat{F}_m^{(K,\mathrm{a})}(\mathrm{e}^{\mathrm{j}\Omega}) - \hat{F}_m^{(K/2,\mathrm{a})}(\mathrm{e}^{\mathrm{j}\Omega}) \,, \tag{29}$$

will converge to zero as K increases, since for sufficiently large K, time-domain aliasing becomes negligible, and subsequently Theorem 2 applies. Thus a correct association is a necessary condition for $\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega})$ to converge to zero as K increases.

To measure the difference between the interpolants, recall that $\frac{K}{2}$ is a power of two, i.e. even. The mth interpolants for both sets have coefficient vectors $\mathbf{f}_{m,K/2}$ and $\mathbf{f}_{m,K}$ similar to (15) but with a subscript $m = 1, \ldots, M$ as an index into the particular function. The two vectors are related by

$$\mathbf{f}_{m,K/2} = \underbrace{\left(\mathbf{I}_{K/2} \otimes [1 \;\; 0]\right)}_{\mathbf{M}_1} \mathbf{f}_{m,K} \,, \tag{30}$$

where $\otimes$ denotes the Kronecker product. Note that vectors of exponentials akin to (12) can be similarly related via

$$\mathbf{e}_{K/2}(\mathrm{e}^{\mathrm{j}\Omega}) = \underbrace{\left[\mathbf{0} \;\; \mathbf{I}_{K/2+1} \;\; \mathbf{0}\right]}_{\mathbf{M}_2} \mathbf{e}_K(\mathrm{e}^{\mathrm{j}\Omega}) \,, \tag{31}$$

with $\mathbf{0}$ a $(\frac{K}{2}+1) \times L_{\frac{K}{2}}$ zero matrix. With this, we can denote the pth derivative of the difference between the interpolants as

$$\begin{aligned}\frac{\mathrm{d}^p}{\mathrm{d}\Omega^p}\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega}) &= \mathbf{e}_K^{\mathrm{H}}(\mathrm{e}^{\mathrm{j}\Omega})\big(\sqrt{\tfrac{1}{K}}\mathbf{D}_K^p\mathbf{P}_K\mathbf{T}_K - \sqrt{\tfrac{2}{K}}\mathbf{M}_2^{\mathrm{T}} \\ &\qquad \cdot \mathbf{D}_{K/2}^p\mathbf{P}_{K/2}\mathbf{T}_{K/2}\mathbf{M}_1\big)\mathbf{f}_{m,K} \\ &= \mathbf{e}_K^{\mathrm{H}}(\mathrm{e}^{\mathrm{j}\Omega})\mathbf{A}_K\mathbf{f}_{m,K} \,, \end{aligned} \tag{32}$$

where $\mathbf{A}_K$ is defined by (32) and with $\mathbf{P}_K$, $\mathbf{T}_K$, and $\mathbf{D}_K$ according to (13), (14), and (18), and $\mathbf{M}_1$ and $\mathbf{M}_2$ as defined in (30) and (31).

Using (32) and (19), the power of the pth derivative across all M functions is

$$\xi_2^{(p)} = \frac{1}{2\pi}\sum_{m=1}^{M}\int_{-\pi}^{\pi}\left|\frac{\mathrm{d}^p}{\mathrm{d}\Omega^p}\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega})\right|^2\mathrm{d}\Omega = \sum_{m=1}^{M}\|\mathbf{A}_K\mathbf{f}_{m,K}\|_2^2, \tag{33}$$

which, for sufficiently large K, represents a measure of how much aliasing has occurred. We will also demonstrate that $\xi_2^{(p)}$ can also capture the case of (25) not providing an association with the analytic functions.

C. Convergence Proof

For $p \to \infty$, we now want to show that $\xi_2^{(p)} \to 0$ for increasing K iff $F_{m,2k}^{(K,\mathrm{a})}$ is correctly associated with the analytic eigenvalues. The 'if' part, or necessity, will be straightforward, and the 'only if' part, or sufficiency, is more complicated. We exploit the fact that clearly the samples $F_{m,k}^{(K,\mathrm{a})}$ imply the existence of some continuous functions $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$, $m = 1, \ldots, M$, that suitably interpolate the given samples. Using the argument that leads to (7) and (27), we have

$$\hat{f}_m^{(K,\mathrm{a})}[\tau] = \tilde{p}_K[\tau] \sum_{\nu=-\infty}^{\infty} \phi_m[\tau - \nu K] \,, \tag{34}$$

where $\phi_m[\tau] \circ\!\!-\!\!\bullet\, \Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$. Because there are infinitely many ways in which aliasing in (34) could generate $\hat{f}_m^{(K,\mathrm{a})}[\tau]$, the functions $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ are not unique but (34) nevertheless holds and, ultimately, our proof rests on evaluating $\hat{f}_m^{(K,\mathrm{a})}[\tau]$ and not $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$.

We define the time domain equivalent to the cost function in (29) as $\Theta_m^{(K)}[\tau] \circ\!\!-\!\!\bullet\, \Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega})$,

$$\begin{aligned}\Theta_m^{(K)}[\tau] &= \hat{f}_m^{(K,\mathrm{a})}[\tau] - \hat{f}_m^{(K/2,\mathrm{a})}[\tau] \\ &= \tilde{p}_K[\tau]\sum_{\nu=-\infty}^{\infty}\phi_m[\tau-\nu K] - \tilde{p}_{\frac{K}{2}}[\tau]\sum_{\mu=-\infty}^{\infty}\big\{\phi_m[\tau-2\mu\tfrac{K}{2}] \\ &\qquad + \phi_m[\tau-(2\mu+1)\tfrac{K}{2}]\big\} \qquad (35) \\ &= (\tilde{p}_K[\tau] - \tilde{p}_{\frac{K}{2}}[\tau])\sum_{\nu=-\infty}^{\infty}\phi_m[\tau-\nu K] \\ &\qquad -\tilde{p}_{\frac{K}{2}}[\tau]\sum_{\mu=-\infty}^{\infty}\phi_m[\tau-(2\mu+1)\tfrac{K}{2}] \,. \qquad (36)\end{aligned}$$

For the expansion of $\hat{f}_m^{(K/2,\mathrm{a})}[\tau]$ in (35), components for even and odd μ have been written explicitly. Thus, with $\frac{\mathrm{d}^p}{\mathrm{d}\Omega^p}\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega}) \bullet\!\!-\!\!\circ\, (-j\tau)^p\Theta_m^{(K)}[\tau]$ and using Parseval's theorem, the cost function $\xi_2^{(p)}$ in (33) becomes

$$\xi_2^{(p)} = \sum_{m=1}^{M}\sum_{\tau=-\lfloor K/2\rfloor}^{\lfloor K/2\rfloor}|(-\mathrm{j}\tau)^p\Theta_m^{(K)}[\tau]|^2 \,. \tag{37}$$

The two summation terms in (36), and hence (37), involve masked and translated versions of $\phi_m[\tau]$ such that only their tails are involved. In the case of a correct association, $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega}) = \lambda_m(\mathrm{e}^{\mathrm{j}\Omega})$ is analytic and $\phi_m[\tau] \circ\!\!-\!\!\bullet\, \Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ absolutely convergent; hence for $K \to \infty$, the tails of $\phi_m[\tau]$ decay at least exponentially, and $\xi_2^{(p)} \to 0$ which proves necessity.

For the sufficiency, note that the samples $F_{m,k}^{(K,\mathrm{a})}$ come from the analytic eigenvalue functions, but for a specific value $m = 1, \ldots, M$ do not necessarily stem from a single eigenvalue function. Hence $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ could be constructed piece-wise from segments of several eigenvalue functions. Whilst there might be pathological cases [33] where a function that is constructed from piece-wise analytical segments is infinitely differentiable, in general there will be a q such that the q-th differential is discontinuous. Note that since the functions $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ are not unique, we are allowed a great deal of choice in how they are constructed. Specifically, assume that $F_{m,k}^{(K,\mathrm{a})}$ is sampled from piece-wise analytic functions that contain a finite number of isolated 'switching' points; these points are the locations where we switch from one segment of an eigenvalue function to another.

Because the sampling points of $F_{m,k}^{(K/2,\mathrm{a})}$ are subsampled by a factor of two from those of $F_{m,k}^{(K,\mathrm{a})}$, we can say that both are sampled from the same piece-wise analytic functions. Since switching points can occur anywhere on an interval between two bin frequencies, we assign the switching point to even indexed frequency bins, such that they occur only at even integer multiples of $\frac{4\pi}{K}$, i.e., that for a switching point Ω_ℓ

$$\Omega_\ell \in \{\tfrac{4\pi k}{K},\ k = 0, \ldots, (\tfrac{K}{2} - 1)\} \ . \tag{38}$$

The switching points can create discontinuities in $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ or in some q-th order derivative of this function. The latter occurs if we switch at a point where the two eigenvalue functions have the same value and the same i-th order derivative for $1 \leq i < q$. Hence, for an incorrect association across the frequency bins and for some q where $0 \leq q < \infty$, the q-th order derivative of $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ can be constructed as a discontinuous function from a finite number of piece-wise analytic segments.

In the first instance, we state for the number of switching points:

Lemma 1 (Switching points of a discontinuous function): A piece-wise analytic function on the unit circle with discontinuities must possess at least two switching points.

Proof: See Appendix B. ■

In the following w.l.o.g., we assume that $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ possesses the first of Q potential switching points at $\Omega_1 = 0$, which coincides with $\Omega_{Q+1} = 2\pi$ (see Lemma 1). Therefore, we define

$$\Phi_m(\mathrm{e}^{\mathrm{j}\Omega}) = \sum_{\ell=1}^{Q} \lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega}) S_{m,\ell}(\mathrm{e}^{\mathrm{j}\Omega}) \ , \tag{39}$$

where $\lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega})$ is the analytic eigenvalue from which the ℓth segment of $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ arises. In (39), w.l.o.g. $\mu(1) = m$; $S_{m,\ell}(\mathrm{e}^{\mathrm{j}\Omega})$ is a 2π-periodic switching function, which on the interval $0 \leq \Omega \leq 2\pi$ is given by

$$S_{m,\ell}(\mathrm{e}^{\mathrm{j}\Omega}) = \begin{cases} 1 & \Omega_\ell \leq \Omega < \Omega_{\ell+1} \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

We first use (39) to show that the cost function in (33) will not converge to zero if $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ possesses a discontinuity, i.e. a $q = 0$-th order non-differentiability, at any of the switch points. Subsequently we show this is also true for $q > 0$.

Lemma 2 (Non-convergence for discontinuous functions): The derivative cost function $\xi_2^{(p)}$ in (33) for K samples of a discontinuous, piece-wise analytic function with a finite number of potentially discontinuous but piece-wise analytic segments remains finite as $K \to \infty$ and $p \geq 1$.

Proof: We will investigate the terms $(-\mathrm{j}\tau)\phi_m[\tau+T]$ that make up the cost function $\xi_2^{(p)}$ once (36) is inserted into (37). Their Fourier transform

$$\begin{aligned} \sum_{\tau=-\infty}^{\infty} (-\mathrm{j}\tau)\phi_m[\tau+T]\mathrm{e}^{-\mathrm{j}\Omega\tau} &= \frac{\mathrm{d}}{\mathrm{d}\Omega}\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})\mathrm{e}^{\mathrm{j}\Omega T} \\ &= \mathrm{j}T\mathrm{e}^{\mathrm{j}\Omega T}\Phi_m(\mathrm{e}^{\mathrm{j}\Omega}) + \mathrm{e}^{\mathrm{j}\Omega T}\dot{\Phi}_m(\mathrm{e}^{\mathrm{j}\Omega}) \end{aligned} \tag{40}$$

contains the first derivative $\dot{\Phi}_m(\mathrm{e}^{\mathrm{j}\Omega}) = \frac{\mathrm{d}}{\mathrm{d}\Omega}\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$,

$$\begin{aligned} \dot{\Phi}_m(\mathrm{e}^{\mathrm{j}\Omega}) &= \sum_{\ell=1}^{Q} \dot{\lambda}_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega}) S_{m,\ell}(\mathrm{e}^{\mathrm{j}\Omega}) \\ &\quad + \lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega}) \left[\bar{\delta}(\Omega - \Omega_\ell) - \bar{\delta}(\Omega - \Omega_{\ell+1})\right] \\ &= \sum_{\ell=1}^{Q} \dot{\lambda}_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega}) S_{m,\ell}(\mathrm{e}^{\mathrm{j}\Omega}) \\ &\quad + \left[\lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega_\ell}) - \lambda_{\mu(\ell-1)}(\mathrm{e}^{\mathrm{j}\Omega_\ell})\right] \bar{\delta}(\Omega - \Omega_\ell) \ . \end{aligned} \tag{41}$$

In (41), $\bar{\delta}(\Omega) = \sum_\nu \delta(\Omega - 2\pi\nu)$ is an impulse train[4], $\dot{\lambda}_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega}) = \frac{\mathrm{d}}{\mathrm{d}\Omega}\lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega})$ is analytic, and $\lambda_{\mu(0)}(\mathrm{e}^{\mathrm{j}\Omega}) = \lambda_{\mu(Q)}(\mathrm{e}^{\mathrm{j}\Omega})$ due to the cyclic nature of $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$. Therefore the expression (40) splits into three terms, whose time domain equivalents are

$$\begin{aligned} (-\mathrm{j}\tau)\phi[\tau+T] &= \mathrm{j}T \sum_{\ell=1}^{Q} \lambda_{\mu(\ell)}[\tau+T] * s_{m,\ell}[\tau+T] \\ &\quad + \sum_{\ell=1}^{Q} \dot{\lambda}_{\mu(\ell)}[\tau+T] * s_{m,\ell}[\tau+T] \\ &\quad + \sum_{\ell=1}^{Q} \left[\lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega_\ell}) - \lambda_{\mu(\ell-1)}(\mathrm{e}^{\mathrm{j}\Omega_\ell})\right] \cdot \mathrm{e}^{-\mathrm{j}\Omega_\ell(\tau+T)} \ , \end{aligned} \tag{42}$$

with $\dot{\lambda}_{\mu(\ell)}[\tau] \circ\!\!-\!\!\bullet\, \dot{\lambda}_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega})$ the derivative of the $\mu(\ell)$th eigenvalue, and $s_{m,\ell}[\tau] \circ\!\!-\!\!\bullet\, S_{m,\ell}(\mathrm{e}^{\mathrm{j}\Omega})$.

The first two terms in (42) represent convolutions between an absolutely convergent function, and the sinc function [41]

$$s_{m,\ell}[\tau] = \frac{\Omega_{\ell+1} - \Omega_\ell}{\pi}\delta[\tau] + \frac{\sin[(\Omega_{\ell+1} - \Omega_\ell)\tau]}{2\pi\tau} \ ,$$

which decays with $\frac{1}{\tau}$; therefore both terms decay. Recall from (38) that $\Omega_\ell = \frac{4\pi k}{K}$ and $T = \nu K$ ($k, \nu \in \mathbb{Z}$,) so that $\mathrm{e}^{-\mathrm{j}\Omega_\ell(\tau+T)} = \mathrm{e}^{-\mathrm{j}\Omega_\ell\tau}$. Then the last term in (42) contains non-decaying complex exponentials that are phase-aligned. Since in (37) both terms extend over an effective interval of $K/2$ each, the complex exponentials in (37) are mutually orthogonal for all $\ell = 1, \ldots, Q$ i.e. do not cancel.

Thus (37) is dominated by orthogonal complex exponentials unless $\lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega_\ell}) - \lambda_{\mu(\ell-1)}(\mathrm{e}^{\mathrm{j}\Omega_\ell}) = 0$ is satisfied for all ℓ, in which case $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ must be continuous. If discontinuities exist, then for some ℓ, $\lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega_\ell}) - \lambda_{\mu(\ell-1)}(\mathrm{e}^{\mathrm{j}\Omega_\ell}) \neq 0$, i.e. their amplitude does not vanish, and the cost function $\xi_2^{(1)}$ therefore does not converge. ■

Previously we argued that $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ could have a finite number of piecewise analytic segments but be discontinuous in the q-th order derivative for some q where $0 \leq q < \infty$. Lemma 2 considered the case when $q = 0$. Theorem 4 below considers the general case.

Theorem 4 (Convergence): The identified association $F_{\mu,k}^{(K,\mathrm{a})}$ across K bins is correct (i.e. represents sample points of the analytic eigenvalues) iff $\xi_2^{(p)}$, for $p \to \infty$ and sufficiently large value of K, is smaller than a specified, sufficiently small threshold ϵ_2, and the analytic eigenvalue functions are such that it is not possible to form a function

[4]We assume that a Heaviside function $u(t)$ can be differentiated as $\frac{\mathrm{d}}{\mathrm{d}t}u(t) = \delta(t)$, with $\delta(t)$ the Dirac impulse.

TABLE I
EXAMPLE FOR $\xi_2^{(p)}$ WHEN SAMPLING FROM THE NON-DIFFERENTIABLE FUNCTIONS IN FIG. 1(B) USING DIFFERENT DFT LENGTHS K.

p	$K=4$	$K=8$	$K=16$	$K=32$	$K=64$	Analytic $K=4$
1	1.12500	1.92614	2.52055	4.50133	7.51843	2.125
2	1.61e+1	2.84e+2	3.15e+3	1.01e+5	3.05e+6	32.125
3	2.56e+2	5.69e+4	9.31e+6	4.17e+9	1.9e+12	512.125

that is a piecewise combination of two eigenvalues and this function be infinitely differentiable.

Proof: From above, it is clear that $\xi_2^{(p)}$ necessarily must converge to zero with increasing K if the sample points in $F_{m,k}^{(K,\mathrm{a})}$ and $F_{m,k}^{(K/2,\mathrm{a})}$ are correctly associated with the analytic eigenvalues. It remains to be shown that convergence is also sufficient for a correct association.

If the association is not correct, we know that, for some q where $0 \leq q < \infty$, the q-th order derivative of $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ is discontinuous but consists of a finite number of piecewise analytic segments. If the functions $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ possess discontinuities only for $q > 0$, then $\frac{\mathrm{d}^r}{\mathrm{d}\Omega^r}\lambda_{\mu(\ell)}(\mathrm{e}^{\mathrm{j}\Omega_\ell}) = \frac{\mathrm{d}^r}{\mathrm{d}\Omega^r}\lambda_{\mu(\ell-1)}(\mathrm{e}^{\mathrm{j}\Omega_\ell})$ $\forall \ell = 1, \ldots, Q$ and $r = 0, \ldots, (q-1)$. The case $q = 0$ is covered by Lemma 2. Thus we can continue to differentiate $\dot{\Phi}_m(\mathrm{e}^{\mathrm{j}\Omega})$ in (41) q-fold until Dirac impulses appear in $\frac{\mathrm{d}^{q+1}}{\mathrm{d}\Omega^{q+1}}\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$. Consider the metric in (33). For $p > q$, using the chain rule we have that

$$\frac{\mathrm{d}^p}{\mathrm{d}\Omega^p}\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega}) = \frac{\mathrm{d}^{p-q-1}}{\mathrm{d}\Omega^{p-q-1}}\frac{\mathrm{d}^{q+1}}{\mathrm{d}\Omega^{q+1}}\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega}) \,,$$

where any of the functions $\frac{\mathrm{d}^q}{\mathrm{d}\Omega^q}\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ contributing to $\frac{\mathrm{d}^q}{\mathrm{d}\Omega^q}\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega})$ via (37) can have a finite number of discontinuities but otherwise piecewise analytic segments. Hence Lemma 2 applies and $\xi_2^{(p)}$ will diverge for $p \geq q+1$. Overall, therefore $\xi_2^{(p)} \to 0$ for $p \geq q+1$ and $K \to \infty$ iff the association is correct. ■

Example 1. When sampling from the analytic and spectrally majorised functions in Fig. 1(a) and (b) respectively, we obtain the metrics $\xi_2^{(p)}$ shown in Tab. I. For the spectrally majorised case, we sample from continuous but non-differentiable functions; the resulting values for $\xi_2^{(p)}$, given in the left columns of Tab. I, diverge as K increases. In contrast, when sampling from the analytic functions, $\xi_2^{(p)}$ has finite values for $K = 4$, displayed in the right-most column of Tab. I; for $K > 4$, in the analytic case $\xi_2^{(p)}$ drops close to machine precision.

Example 2. We show an example for $K = 64$ sample points taken from the spectrally majorised functions in Fig. 1(b). There, at $\Omega = \pi$, $d_i(\mathrm{e}^{\mathrm{j}\Omega})$, $i = 1, 3$ are continuous but not differentiable and thus discontinuous for $q = 1$, while $d_2(\mathrm{e}^{\mathrm{j}\Omega})$, is differentiable once and therefore discontinuous only for $q = 2$. The interpolants for $K = 64$ and $K/2 = 32$ are shown in Fig. 5(a). The resulting oscillations in the interpolants increase from $K/2$ to K, and more so for $m = 1, 3$ approximating functions that are discontinous already for $q = 1$. Thus, the differences $\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega})$ in (29) and their derivatives $\frac{\mathrm{d}^p}{\mathrm{d}\Omega^p}\Theta_m^{(K)}(\mathrm{e}^{\mathrm{j}\Omega})$ contribute energy to the metric ξ_2 in (33); this energy rises as the derivative order p increases, as Fig. 5(b) for $p = 0$ and Fig. 5(c) for $p = 1$ attest.

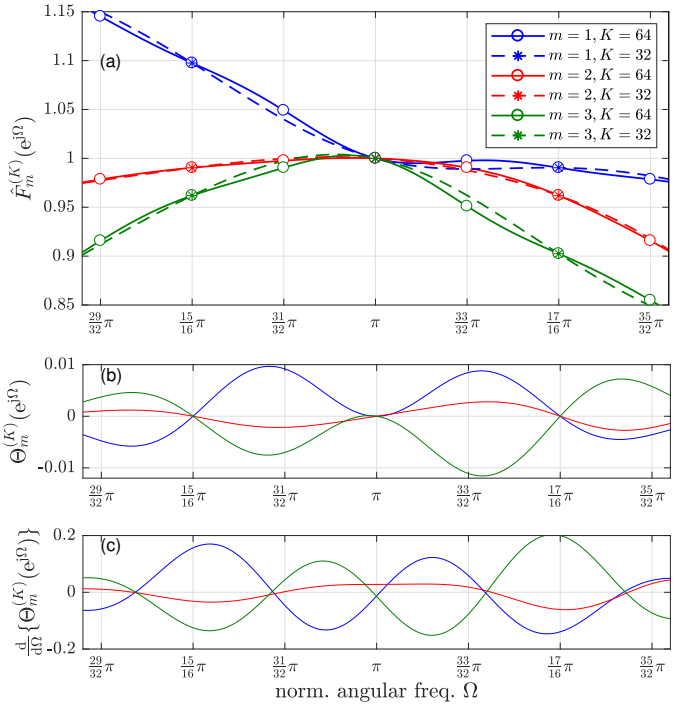

Fig. 5. (a) Interpolants for spectrally majorised association of sample points for the eigenvalues shown in Fig. 1(b) for $K = 32$ and $K = 64$; markers indicate the positions of sample points $F_{m,k}^{(K)}$; (b) difference between the interpolants in (a) according to (29), and (c) the first order differentiation of this difference contributing to (33).

V. IMPLEMENTATION ISSUES

For a practical realisation of the proposed procedure in Algorithm 1, we need to address three issues. Sec. V-A presents some consequences of having to work with a finite derivative order p. Sec. V-B discusses replacing the exhaustive search in (25) by an iterative Viterbi-type maximum sequence likelihood estimator (MLSE) [42], which evolves associations frequency bin by frequency bin. The MLSE requires a modification to the smoothness metric to allow for this iteration and this is presented in Sec. V-C. Sec. V-D defines the evolution of a set of paths that is ordered in terms of smoothness, and Sec. V-E discusses the pruning approach. The algorithm is summarised in Sec. V-F.

A. Finite Differentiation Order p

For the above convergence proof, specifically Theorem 4, we require $p > q$, i.e. the derivative power p in the metric to exceed the derivative q in which the piecewise analytic functions $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ are discontinuous. We generally do not know this value q, but if it is the case that $p < q$, at least we know that the $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ are continuous and differentiable up to order p. Then for a sufficient large p, the resulting assignment is, for practical purposes, a good approximation to the analytic functions.

B. Maximum Likelihood Sequence Estimation

Recall that the challenge is to reorder K sample points of M eigenvalues, $F_{m,k}^{(K)}$, into sample points $F_{m,k}^{(K,\mathrm{a})}$ that are maximally smoothly associated across K frequency bins. An

example for the eigenvalues in Fig. 1(a) with $K = 8$ was shown in Fig. 1(c).

Following the philosophy behind the Viterbi algorithm, the MSLE evaluation of (25) therefore operates on the $M \times K$ sample points $F_{m,k}^{(K)}$, $k = 0, \ldots, (K-1)$, $m = 1, \ldots, M$. Starting with the first bin and a single path $\mathbf{p}_{1,1} = 1$, a set of paths is grown from bin to bin. In each iteration, the fitness of each path is evaluated using a smoothness metric; since (24) requires a path of length K, we will derive a modified smoothness metric in Sec. V-C below for a reduced set of $1 < J \leq K$ sample points. At the end of an iteration, the set of paths is pruned to only retain the best $\rho_{\max}$ distinct paths. If path metrics exceed the smoothness of the spectrally majorised solution, then these are also pruned. After the Kth bin, the path $\mathbf{p}_{1,K}$ is the one with the lowest metric, indicating that the indexing captured in $\mathbf{p}_{1,K}$ will map $\{F_{m,k}^{(K)}\}$ to the set $\{F_{m,k}^{(K,a)}\}$ in (25) such that it has the smoothest association across all bins.

Convergence to the correct solution, i.e. the maximally smooth association across frequency bins, depends on the number of paths $\rho_{\max}$ that is retained. There is however a trade-off, and lower values for $\rho_{\max}$ that make the algorithm tractable may not always lead to the desired solution. However, as outlined in Algorithm 1, the metrics ξ_1 and ξ_2 will detect this, and rerun the association for (25) with an increased number of bins K. For a sufficiently high value of K, the analytic solution will be sufficiently smooth with respect to the sampling grid such that a decreasingly smaller number of retained paths, $\rho_{\max}$, will, in practice, suffice to identify the association belonging to the maximally smooth, analytic solution.

To iteratively evolve associations of eigenvalues across bins, we start with a set of possible associations for the first J, $1 \leq J < K$, bins. To follow the terminology of Viterbi's MLSE approach [42], we define each such possible and distinct association as one path. With respect to the example in Fig. 1(c), one path represents one way to draw M curves through the first J bins, using all available M sample points in each bin. This set of paths $\mathcal{P}_J$ is ordered, based on a smoothness metric that sums over the M Dirichlet interpolations that can be created from the associations in each path. The iteration then adds the $(J+1)$th bin into the consideration for associations, thus creating the set $\mathcal{P}_{J+1}$.

The approach terminates by extracting the smoothest possible path from set $\mathcal{P}_K$ once all K bins have been processed. As an initialisation, the initial set $\mathcal{P}_1$ only contains a single association; w.l.o.g. we select majorised eigenvalues in this first bin, which after completion only impacts on the ordering but not the smoothness of the extracted associations. Note that we do not attempt to order the resulting eigenvalues but merely make them analytic. In general, the EVD does not provide any information that would imply an ordering. Clearly they could be ordered using some metric if desired. Using total energy would result in an ordering in the same spirit as is often used for the conventional EVD.

The number of paths in each set $\mathcal{P}_J$, i.e. the cardinality $|\mathcal{P}_J|$, grows with J. Since adding a bin creates $M!$ new possibilities, we find that without further intervention, $|\mathcal{P}_J| \leq (M!)^{J-1}$, with equality in the case that none of the eigenvalues possesses an algebraic multiplicity greater than one. In order to retain a workable number of paths, the set $\mathcal{P}_J$ therefore requires pruning after each iteration.

C. Missing Sample Problem / Incomplete Sample Set

The proposed iterative MLSE algorithm adds a Jth bin to associations that have already been assessed over the previous $J-1$ bins, with $1 < J \leq K$. Therefore, we need to measure the smoothness of associations across these first J frequency bins. Since we do not yet know the associations of the remaining $K-J$ sample points, we assume that we select the values in those DFT bins such that they offer the smoothest possible interpolation. For the purpose of this section, we only consider a single function and hence omit the function index m. Thus, we denote the first J sample points of $F_k^{(K)}$, $k = 0, \ldots, (J-1)$, as a known parameter vector $\mathbf{f}_{J|K}$, and the remaining $K-J$ parameters as a vector of unknowns, $\mathbf{x} \in \mathbb{R}^{K-J}$, such that $\mathbf{f}_K^{\mathrm{T}} = [\mathbf{f}_{J|K}^{\mathrm{T}}, \mathbf{x}^{\mathrm{T}}]$. In other interpolation contexts, the challenge of determining $\mathbf{x}$ has been described as the 'missing samples problem' [43], [44].

The direct optimisation of

$$\chi^{(P)} = \min_{\mathbf{x}} [\mathbf{f}_{J|K}^{\mathrm{T}} \, \mathbf{x}^{\mathrm{T}}] \mathbf{C}_K^{(P)} \begin{bmatrix} \mathbf{f}_{J|K} \\ \mathbf{x} \end{bmatrix}, \tag{43}$$

can be performed by standard least-squares techniques [28], [29]. The matrix $\mathbf{C}_K^{(P)}$ in (43) is partitioned s.t.

$$\mathbf{C}_K^{(P)} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_2^{\mathrm{T}} & \mathbf{C}_4 \end{bmatrix}, \tag{44}$$

with $\mathbf{C}_1 \in \mathbb{R}^{J \times J}$ and all other matrices of appropriate dimensions, which leads to

$$\chi^{(P)} = \mathbf{f}_{J|K}^{\mathrm{T}} \left(\mathbf{C}_1 - \mathbf{C}_2 \mathbf{C}_4^{-1} \mathbf{C}_2^{\mathrm{T}} \right) \mathbf{f}_{J|K}, \tag{45}$$

involving the Schur complement [46] of $\mathbf{C}_K^{(P)}$. Since according to (22) and (23) $\mathbf{C}_i$, $i = 1, 2, 4$, only depend on a DFT matrix and the constants J and K, the Schur complement can be precomputed. Nonetheless, for $J \leq K$, the inversion of $\mathbf{C}_4$ can be computationally costly and challenging due to the poor conditioning of $\mathbf{C}_K^{(P)}$. Note this is the approach that was used in [27], [28] using a minimum variance distortionless response approach [47]. For this reason, the following theorem proposes an alternative route to the evaluation of (43).

Theorem 5 (Smoothness for Missing Samples Problem): The cost in (43) for the smoothest interpolation through $J \leq K$ sample points can be calculated via $\|\mathbf{R}_2 \mathbf{f}_{J|K}\|_2^2$, where $\mathbf{R}_2$ is the lower right-hand $J \times J$-dimensional partition of $\mathbf{R}$ obtained by the QR decomposition [46]

$$\mathbf{D}^{\frac{1}{2}} \mathbf{W}_K^{\mathrm{H}} = \mathbf{Q} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_2 \end{bmatrix}, \tag{46}$$

where $\mathbf{D} = \sum_p \tilde{\mathbf{D}}_{p,K}$.

Proof. We provide a shortened and simplified version of the analysis in [29], which avoids a Cholesky decomposition step. Recall that the matrix $\mathbf{C}_K^{(P)} \in \mathbb{R}^{K \times K}$ is symmetric and

circulant, and therefore Töplitz. As such, we can reorganise (44) as

$$\mathbf{C}_K^{(P)} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_2^{\mathrm{T}} & \mathbf{C}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_4 & \mathbf{C}_2^{\mathrm{T}} \\ \mathbf{C}_2 & \mathbf{C}_1 \end{bmatrix} \tag{47}$$

without affecting the overall matrix. With (47), (46), and the remaining submatrices of $\boldsymbol{R}$ in (46) of appropriate dimensions, we can equate

$$\begin{bmatrix} \mathbf{C}_4 & \mathbf{C}_2^{\mathrm{T}} \\ \mathbf{C}_2 & \mathbf{C}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1^{\mathrm{H}}\mathbf{R}_1 & \mathbf{R}_1^{\mathrm{H}}\mathbf{R}_{12} \\ \mathbf{R}_{12}^{\mathrm{H}}\mathbf{R}_1 & \mathbf{R}_{12}^{\mathrm{H}}\mathbf{R}_{12} + \mathbf{R}_2^{\mathrm{H}}\mathbf{R}_2 \end{bmatrix} .$$

Thus, the Schur complement $\mathbf{C}_1 - \mathbf{C}_2^{\mathrm{T}}\mathbf{C}_4^{-1}\mathbf{C}_2$ simplifies to

$$\begin{aligned} \mathbf{C}_1 - \mathbf{C}_2^{\mathrm{T}}\mathbf{C}_4^{-1}\mathbf{C}_2 &= \mathbf{R}_{12}^{\mathrm{H}}\mathbf{R}_{12} + \mathbf{R}_2^{\mathrm{H}}\mathbf{R}_2 \\ &\quad - \mathbf{R}_{12}^{\mathrm{H}}\mathbf{R}_1 \left(\mathbf{R}_1^{\mathrm{H}}\mathbf{R}_1\right)^{-1} \mathbf{R}_1^{\mathrm{H}}\mathbf{R}_{12} \\ &= \mathbf{R}_2^{\mathrm{H}}\mathbf{R}_2 \, . \end{aligned} \tag{48}$$

Since $\mathbf{C}_K^{(P)}$ is full rank, $\mathbf{R}_1$ in (48) is invertible. ■

Theorem 5 also holds even if only a particular derivative power is used to calculate smoothness according to (21), since the matrix $\mathbf{C}_{p,K}$ will have a rank of at least $(K-1)$, such that $\mathbf{R}_1$ remains invertible for $J < K$, while for $J = K$ the Schur complement is no longer required.

Compared to previous efforts in [27], [28], the approach via Theorem 5 and the QR decomposition of $\mathbf{D}^{\frac{1}{2}}\mathbf{W}_K^{\mathrm{H}}$ offers a lower cost and avoids conditioning problems associated with any matrix inversion by performing a well-conditioned QR decomposition. Further, since we operate with a fixed K but variable J, we only need to calculate the QR decomposition once, and only change its partitioning to determine the smoothness cost in (45). The matrix multiplication $\mathbf{R}_2\mathbf{f}_{J|K}$ can exploit the sparsity of the upper triangular $\mathbf{R}_2$.

Example. As an example for the efficacy of using the power in the derivatives of the interpolation, we take $K = 8$ samples of the raised cosine function $\lambda(\mathrm{e}^{\mathrm{j}\Omega}) = 1 + \cos\Omega$, whose power for every derivative $p \in \mathbb{N}$ is $\sigma^{(p)} = \frac{1}{2}$. Tab. II shows values for χ_p, $p = 1, \ldots, 5$ obtained from a limited sample set based on the J first sample points, with $J = 1, \ldots, K$. For a single sample point $J = 1$, the smoothest interpolation is a constant function, whose every derivative will be zero. If we use all sample points, $J = K$, in this case we have a complete representation of $\lambda(\mathrm{e}^{\mathrm{j}\Omega})$ and $\chi_p = \sigma^{(p)}$. For an incomplete sample set, $J < K$, although we can find interpolations that are smoother than the actual function we have $\chi_p \to \sigma^{(p)}$ for both $J \to K$ and $p \to \infty$. For a comparison to other interpolation methods, e.g. [44], that are computationally cheaper but inaccurate in assessing the smoothness as defined here, please see [45].

D. Set of Ordered Paths

Let $\mathbf{p}_{\rho,J}$ be a J-dimensional vector of integer indices $1, 2, \ldots, M!$ labelling the permutations that create an association for M sample points across the first $J < K$ bins. These labels are lexicographically ordered relative to the majorised sorting of eigenvalues within each bin. As such, $\mathbf{p}_{\rho,J}$ defines one possible association or path for these first $M \times J$ sample points, with $\rho \in \mathbb{N}$ the path index, $\rho = 1, \ldots, |\mathcal{P}_J|$.

TABLE II
POWER IN pTH DERIVATIVE FOR $K = 8$ BASED ON INTERPOLATION OVER $J = 1 \ldots K$ SAMPLE POINTS.

	χ_p				
J	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
1	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0671	0.1458	0.2440	0.3253	0.3752
3	0.1155	0.1788	0.2505	0.3271	0.3957
4	0.1160	0.1870	0.2729	0.3551	0.4169
5	0.1733	0.2698	0.3615	0.4279	0.4666
6	0.3388	0.4119	0.4584	0.4823	0.4930
7	0.4643	0.4861	0.4952	0.4985	0.4995
8	0.5000	0.5000	0.5000	0.5000	0.5000

TABLE III
POSSIBLE PERMUTATIONS OF EIGENVALUE INDICES WITHIN ONE BIN FOR $M = 3$.

label	indexing m		
1	1	2	3
2	1	3	2
3	2	1	3
4	2	3	1
5	3	1	2
6	3	2	1

Example. For $M = 3$, Tab. III shows the $M! = 6$ possible permutations as index combinations for the eigenvalues λ_m within any one bin. This table represents a code book, which links an integer label, forming one entry of $\mathbf{p}_{\rho,J}$, to a specific index permutation relative to a spectrally majorised ordering.

The set $\mathcal{P}_J$ is ordered in terms of the smoothness that its associations afford. If $\hat{F}_m^{(J|K)}(\mathrm{e}^{\mathrm{j}\Omega}, \mathbf{p}_{\rho,J})$, $m = 1, \ldots, M$ are the M functions interpolating the J sample points associated by path $\mathbf{p}_{\rho,J}$, and the vectors $\mathbf{f}_{m,J|K}(\mathbf{p}_{\rho,J}) \in \mathbb{R}^J$, $m = 1, \ldots, M$, hold the sampling points of each of these M functions, then the smoothness of the path $\mathbf{p}_{\rho,J}$ is defined as the cumulative smoothness of the functions $\hat{F}_m^{(J|K)}(\mathrm{e}^{\mathrm{j}\Omega}, \mathbf{p}_{\rho,J})$. Therefore the smoothness is measured as

$$\chi^{(P)}(\mathbf{p}_{\rho,J}) = \sum_{m=1}^{M} \|\mathbf{R}_2 \, \mathbf{f}_{m,J|K}(\mathbf{p}_{J,K})\|_2^2 \tag{49}$$

according to Theorem 5. Using (49), the paths in $\mathcal{P}_J$ are ordered such that

$$\chi^{(P)}(\mathbf{p}_{\rho,J}) \leq \chi^{(P)}(\mathbf{p}_{\rho+1,J}) \, , \qquad \rho = 1, \ldots, |\mathcal{P}_J| \, ,$$

i.e., given a specific maximum derivative order P, the paths are ordered in descending smoothness.

Example. Fig. 6 shows, for $J = 3$, the analytic solution that is given by the path $\mathbf{p}_{1,3} = [1, 1, 2]^{\mathrm{T}}$. The spectrally majorised solution has the path $\mathbf{p}_{2,3} = [1, 1, 1]^{\mathrm{T}}$. The modified Dirichlet interpolations arising from the two paths are shown in Fig. 6. If we take derivatives up to the $P = 5$th order to measure smoothness, we obtain $\chi^{(5)}(\mathbf{p}_{1,3}) = 0.4978$ and $\chi^{(5)}(\mathbf{p}_{2,3}) = 0.6700$. Note from the interpolated curves in Fig. 6 that arc length, as used to identify an analytic solution in [20], would here lead to a smaller value for the association in (b) compared to (a); similarly, if the derivative order is chosen very low, the path ordering based on smoothness may not be as intended, as e.g. $\chi^{(1)}([1, 1, 2]^{\mathrm{T}}) = 0.1649$

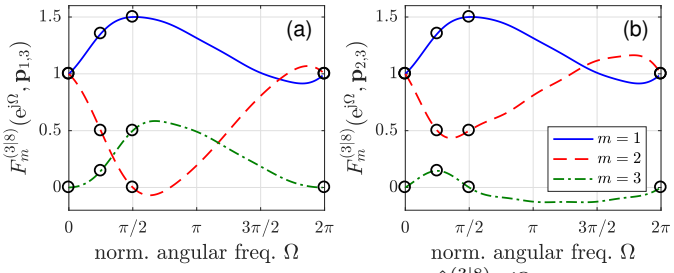

Fig. 6. Example for interpolated functions $\hat{F}_m^{(3|8)}(\mathrm{e}^{\mathrm{j}\Omega}, \mathbf{p}_{i,3})$ for two different associations (or 'paths'), (a) $\mathbf{p}_{1,3}$ and (b) $\mathbf{p}_{2,3}$ based on the first $J = 3$ of $K = 8$ sample points shown in Fig. 1(c).

while $\chi^{(1)}([1,\,1,\,1]^{\mathrm{T}}) = 0.0721$ would order the spectrally majorised association ahead of the analytic one.

E. Path Pruning

The motivation behind using a Viterbi-style MLSE approach is that it is deemed sufficient to only retain a maximum permissible number $\rho_{\max}$ of paths at any one time. This leads to a pruning of the set $\mathcal{P}_J$ to only retain the $\rho_{\max}$ best paths, such that all paths $\mathbf{p}_{\rho,J}$, $\rho > \rho_{\max}$ are discarded from the set and hence $|\mathcal{P}_J| \leq \rho_{\max}$. Note that the set of retained paths can be smaller than $\rho_{\max}$, since, e.g., $|\mathcal{P}_1| = 1$. Similarly, an algebraic multiplicity of eigenvalues greater than one will lead to a number of identical paths even though their labels differ; such paths must be purged from the set $\mathcal{P}_J$ in order to retain distinct solutions prior to any pruning. For example consider the entries in Tab. III for the case when $\lambda_2 = \lambda_3$ say. In this case, permutations '1' and '2' lead to the same sequence. Similarly the permutations '3' and '5' are equivalent, as are the permutations '4' and '6'.

The set of paths $\mathcal{P}_J$ can potentially be pruned further. The path $\mathbf{p}_{\mathrm{sm},K} = [1, 1, \ldots, 1]^{\mathrm{T}} \in \mathbb{Z}^K$ describes the spectrally majorised selection. Since we are looking for an association of sample points that is at least as smooth as that given by the spectrally majorised eigenvalues, we can demand that

$$\chi^{(P)}(\mathbf{p}_{\rho,J}) \leq \chi^{(P)}(\mathbf{p}_{\mathrm{sm},K}) \qquad \forall \quad \mathbf{p}_{\rho,J} \in \mathcal{P}_J \;, \tag{50}$$

and thus potentially shrink the set $\mathcal{P}_J$ further to satisfy (50).

F. Overall Algorithm

An implementation of Algorithm 1 is thus possible if the MLSE approach replaces the exhaustive search to solve (25). The correctness and precision of this association is iteratively checked by the metrics ξ_1 in (28) and $\xi_2^{(p)}$ in (33) until suitable thresholds are reached. Unfortunately, the MLSE does not scale particularly well with either the DFT length (related to the approximation order) or the spatial dimension of the parahermitian matrix. This issue would need to be addressed in scalable implementations by (i) local interpolations w.r.t. to the missing samples problem, and (ii) considering permutations only between neighbouring, ordered eigenvalues. This, however, is beyond the scope and space of this paper.

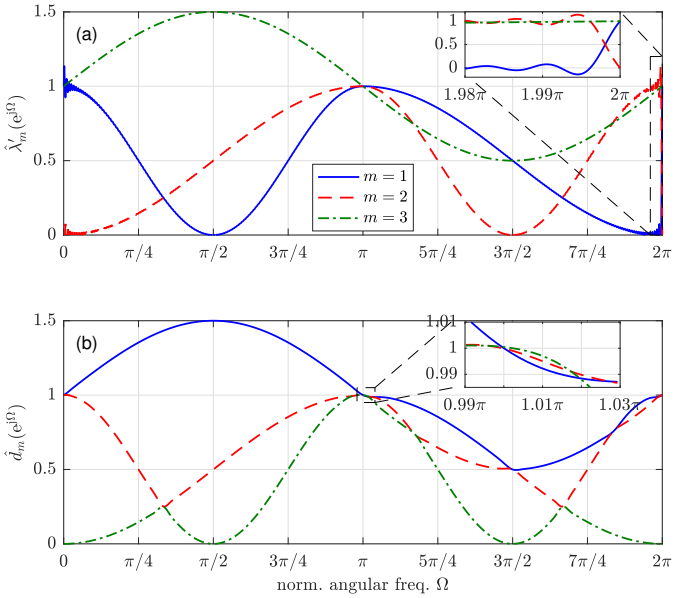

Fig. 7. Eigenvalues extracted by (a) the DFT-based approach in [26] and (b) SBR2 [5], [8].

VI. Simulations and Results

A. Numerical Example

The CSD matrix $\boldsymbol{R}(z)$ used to generate Fig. 1 possesses the analytic eigenvalues $\lambda_1(z) = -\mathrm{j}\frac{1}{4}z + 1 + \mathrm{j}\frac{1}{4}z^{-1}$, $\lambda_2(z) = \frac{1}{4}z^2 + \frac{1}{2} + \frac{1}{4}z^{-2}$, and $\lambda_3(z) = -\frac{1}{4}z + \frac{1}{2} - \frac{1}{4}z^{-1}$. While the approach in [26] often succeeds in extracting analytic eigenvalues if the DFT length is selected sufficiently long, this is not guaranteed to occur [48]. The association between bins in [26] is driven by the examination of the eigenvectors, and here the algebraic multiplicity of three at $\Omega = \pi$ misleads the selection. This results in two of the eigenvalues $\hat{\lambda}'_m(z)$, $m = 2, 3$ in Fig. 7(a) being incorrectly associated at $\Omega = \pi$, and subsequently having to approximate discontinuities at $\Omega = 2\pi$. In contrast, the proposed algorithm terminates after $i = 2$ iterations, with $K_2 = 256$ being the same DFT lengths as used for Fig. 7(a). By appropriately trimming trailing zeros [49]–[52], the extracted estimates $\hat{\lambda}_m(z)$, $m = 1, 2, 3$, match the above ground truth (i) in order $N = 4$, and (ii) close to machine precision with $\xi_2^{(3)} < 10^{-15}$.

For reference, Fig. 7(b) also includes results for the SBR2 algorithm [5], [8], which has been shown to converge towards spectrally majorised eigenvalues [14]. The approximation of non-differentiable points is evident on closer inspection, e.g. near the algebraic multiplicity at $\Omega = \pi$, which in this case requires a high approximation order with $N = 126$.

B. Ensemble Results with Known Ground Truth

Using the source model in Fig. 2 with $L_{\mathrm{s}} = M = 4$, randomised innovation filters $g_\ell[n]$ and paraunitary mixing matrices $\mathbf{H}[n]$ both of order $L = 0, \ldots, 12$, we build an ensemble of 13000 different CSD matrices $\boldsymbol{R}(z)$. The innovation filters are constructed from uncorrelated complex Gaussian coefficients, with the filters normalised such that the signals $s_\ell[n]$, $\ell = 1, \ldots, 4$, have unit power. Varying L results in orders for the eigenvalues $\lambda_m(z)$ that range from 0 to 24.

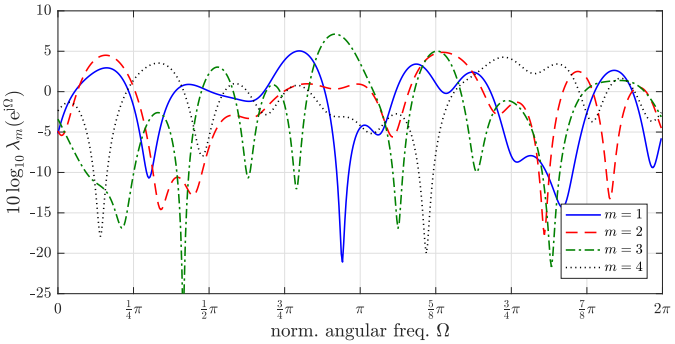

Fig. 8. Example for eigenvalues produced by the randomised source model for $L = 12$.

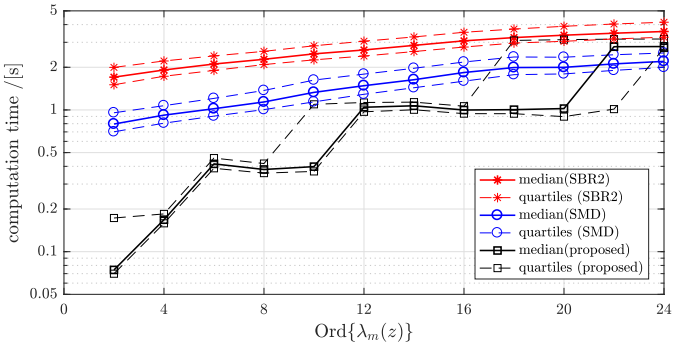

Fig. 10. Computational cost of the proposed method compared to SBR2 [8] and SMD [12] for the experiment in Fig. 9, showing the median, as well as the 25% and 75% percentiles.

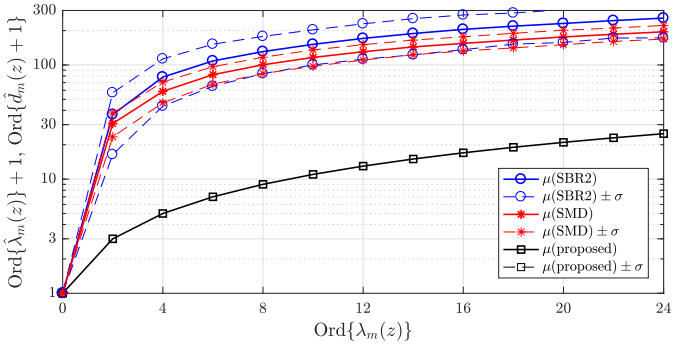

Fig. 9. Mean μ and standard deviation σ for the order $\mathrm{Ord}\{\cdot\}$ of the extracted eigenvalues with the proposed method, $\hat{\lambda}_m(z)$, and spectrally majorised values $\hat{d}_m(z)$ estimated by SBR2 [8] and SMD [12].

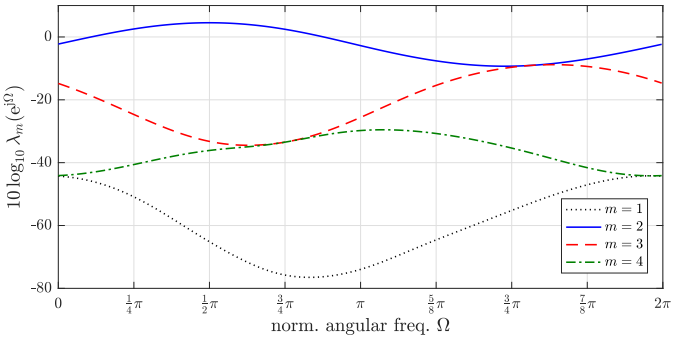

Fig. 11. Example for spectrally majorised eigenvalues $\lambda_m(\mathrm{e}^{\mathrm{j}\Omega})$, $m = 1 \ldots 4$ generated by the source model in [5].

An example for the eigenvalues produced by this approach is shown in Fig. 8.

The proposed approach yields estimates $\hat{\lambda}_m(z)$ with the correct polynomial order and $\xi_2^{(3)} < 10^{-15}$ in all cases, that match the correct values close to machine precision. In Fig. 9, these are compared to the orders of the estimated eigenvalues $\hat{d}_m(z)$ extracted by SBR2 [8] and SMD [12]. The graphs show the mean and standard deviation of the order of $\hat{d}_m(z)$ obtained across the ensemble. Both SBR2 and SMD show similar results, with SMD providing generally a lower order and a smaller spread of results compared to SBR2. Since both algorithms strive for spectral majorisation, their approximation of non-differentiable points leads to orders that are on average an order of magnitude above those of the analytic eigenvalues obtained with the proposed method, except for $L = 0$, where all approaches reduce to the standard EVD of a non-polynomial matrix. Also note that the proposed scheme extracts the correct order with zero standard deviation across the ensemble.

The execution time that is expended to reach the results in Fig. 9 under Matlab R2016b on an 1.1GHz Intel Celeron CPU is summarised in Fig. 10. The SBR2 algorithm is permitted a maximum of 500 iterations, and the faster converging SMD algorithm a maximum of 200 iterations. Despite its considerably higher per-iteration cost [12], SMD reaches its result generally faster than SBR2. For both algorithms, the cost depends on the difficulty to approximate a specific ensemble probe, resulting in a distribution of costs for each simulated order for the ground truth eigenvalues. For the proposed algorithm, the cost does not rise strictly monotonously with the order of the ground-truth eigenvalues, but depends on the DFT length that is required. As a result, the cost contains 'steps' roughly where a doubling of the DFT length occurs, subject to a small variability that can be observed via the quartiles of the cost distribution in Fig. 10. In general however, the proposed algorithm's lower approximation order noted in Fig. 9 translates into a significantly faster execution time.

To demonstrate how the proposed algorithm behaves if the ground truth parahermitian matrix possesses spectrally majorised eigenvalues, we modify the randomised source model of Fig. 2 akin to [12]. There, a gain control spectrally majorises the ground truth eigenvalues such that $\lambda_m(\mathrm{e}^{\mathrm{j}\Omega}) \geq \lambda_{m+1}(\mathrm{e}^{\mathrm{j}\Omega})$ $\forall\Omega$ and $m = 1 \ldots (M-1)$; limiting the radii of zeros in the innovation filter $G_\ell(z)$ in (2) contains the dynamic range of the functions. Nonetheless, the dynamic range of the eigenvalues can be significant, as shown for the example in Fig. 11, where $L_\mathrm{s} = 4$ and the order of $\lambda_m(z)$ is $\mathrm{Ord}\{\lambda_m(z)\} = 6$.

Fig. 12 shows the ensemble results for the order of the estimated eigenvalues. For SBR2, SMD, and the proposed approach, the time-domain coefficients of the eigenvalues are truncated at -100dB, which due to the high dynamic range and hence fast decay may lead to estimation orders that are lower than the ground truth. In this spectrally majorised case, SBR2 and SMD may be expected to perform similar to the proposed method, but rounding errors due to internal truncation of the estimated paraunitary matrices occur [5], [8], [12] and may lead to error propagation. This does not occur in the proposed approach, which only operates on the eigenvalues

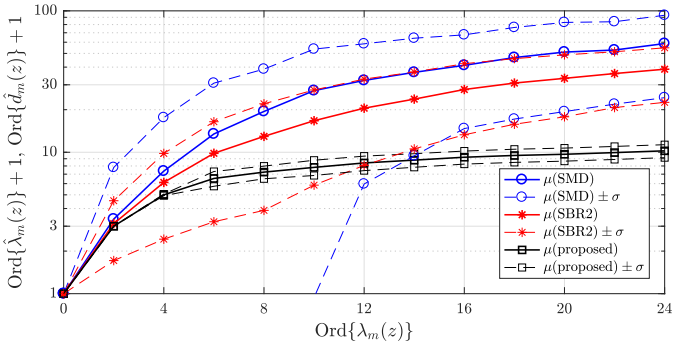

Fig. 12. Mean μ and standard deviation σ for the order $\mathrm{Ord}\{\cdot\}$ of the extracted eigenvalues with the proposed method, $\hat{\lambda}_m(z)$, and $\hat{d}_m(z)$ estimated by SBR2 [8] and SMD [12], for simulations where the ground-truth eigenvalues are spectrally majorised.

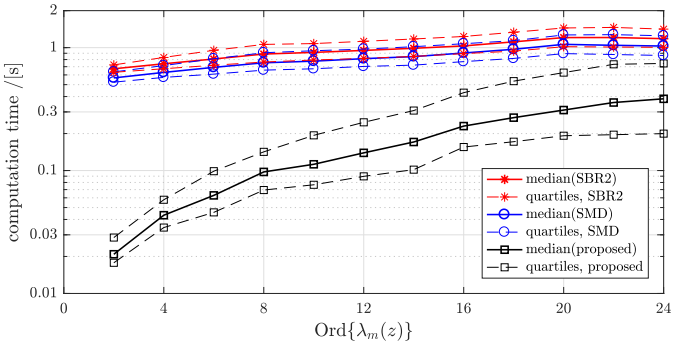

Fig. 13. Computational cost of the proposed method compared to SBR2 [8] and SMD [12] for the experiment in Fig. 12, showing the median, as well as the 25% and 75% percentiles.

and does not carry the paraunitary matrices. Even though the approximation order is not as impressively reduced as in the case of overlapping eigenvalues in Fig. 9, there is a significant advantage for the proposed approach evident from Fig. 12, particularly as the order of the ground truth eigenvalues grows.

The execution time for the ensemble results in Fig. 12 are contained in Fig. 13. Since the proposed approach typically extracts eigenvalues of lower order than SBR2 and SMD, there also is significant advantage in execution time. This is due to the moderate dimension $M = 4$; recall from Sec. V-F that the algorithm does not scale well w.r.t. the spatial dimension M and, to a lesser extent, the DFT length K.

VII. Conclusions

We have motivated the need to extract analytic eigenvalues from a parahermitian matrix using a DFT domain algorithm that relies on the eigenvalues only. For a given DFT length, a novel cost function drives the extraction of the smoothest possible interpolation across frequency bins. An algorithm with proven convergence iteratively adjusts the DFT lengths until a polynomial approximation is found that generates sufficiently small aliasing and truncation errors to satisfy a predefined approximation error. In practice, we employ an MLSE to extract the analytic eigenvalues; the algorithm still does not scale particularly well with the dimensions of a parahermitian matrix. Nonetheless, we have demonstrated that the algorithm can provide significant order reductions where the ground truth eigenvalues are not spectrally majorised; even in the case of spectrally majorised eigenvalues, the simulations suggest that the proposed algorithm is beneficial in reducing the order of the extracted Laurent polynomials w.r.t state-of-the-art algorithms.

Based on correctly associated eigenvalues, it is thereafter possible to extract analytic eigenspaces or eigenvectors [53], enabling significantly lower-order subspace decompositions than state-of-the-art algorithms such as SBR2 [5], [8] and SMD [12] for applications such as broadband angle of arrival estimation [54], broadband beamforming [2], source separation [3], speech enhancement [55], or MIMO precoding and equalisation [56]–[59]. Since eigenvectors generally possess a greater volatility compared to eigenvalues, particlarly when eigenvalues are closely spaced [46], [60], it will be important to base this extraction on associations derived from the more reliable eigenvalues as proposed and demonstrated in this paper.

Acknowledgements

Dr Jennifer Pestana from the University of Strathclyde's Dept. for Mathematics & Statistics provided very substantial input and feedback to this manuscript. We are immensely grateful for her contributions. We would also like to thank the anonymous reviewers for making several helpful suggestions that resulted in an improved manuscript.

References

[1] S. Redif, J. McWhirter, P. Baxter, and T. Cooper, "Robust broadband adaptive beamforming via polynomial eigenvalues," in *OCEANS*, Boston, MA, pp. 1–6, Sept. 2006.

[2] S. Weiss, S. Bendoukha, A. Alzin, F. Coutts, I. Proudler, and J. Chambers, "MVDR broadband beamforming using polynomial matrix techniques," in *23rd EUSIPCO*, Nice, France, pp. 839–843, Sept. 2015.

[3] S. Redif, S. Weiss, and J. McWhirter, "Relevance of polynomial matrix decompositions to broadband blind signal separation," *Sig. Proc.*, **134**:76–86, May 2017.

[4] S. Weiss, S. Redif, T. Cooper, C. Liu, P. Baxter, and J. McWhirter, "Paraunitary oversampled filter bank design for channel coding," *EURASIP J. Adv. Sig. Proc.*, **2006**:1–10, 2006.

[5] S. Redif, J. McWhirter, and S. Weiss, "Design of FIR paraunitary filter banks for subband coding using a polynomial eigenvalue decomposition," *IEEE Trans. SP*, **59**(11):5253–5264, November 2011.

[6] C.H. Ta and S. Weiss, "A jointly optimal precoder and block decision feedback equaliser design with low redundancy," in *15th EUSIPCO*, Poznan, Poland, pp. 489–492, Sept. 2007.

[7] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs: Prentice Hall, 1993.

[8] J.G. McWhirter, P.D. Baxter, T. Cooper, S. Redif, and J. Foster, "An EVD algorithm for para-Hermitian polynomial matrices," *IEEE Trans. SP*, **55**(5):2158–2169, May 2007.

[9] P. Vaidyanathan, "Theory of optimal orthonormal subband coders," *IEEE Trans. SP*, **46**(6):1528–1543, June 1998.

[10] Z. Wang, J.G. McWhirter, J. Corr, and S. Weiss, "Multiple shift second order sequential best rotation algorithm for polynomial matrix EVD," in *23rd EUSIPCO*, Nice, France, pp. 844–848, Sept. 2015,

[11] V.W. Neo and P.A. Naylor, "Second order sequential best rotation algorithm with householder reduction for polynomial matrix eigenvalue decomposition," in *ICASSP*, Brighton, UK, pp. 8043–8047, May 2019.

[12] S. Redif, S. Weiss, and J. McWhirter, "Sequential matrix diagonalization algorithms for polynomial EVD of parahermitian matrices," *IEEE Trans. SP*, **63**(1):81–89, Jan. 2015.

[13] J. Corr, K. Thompson, S. Weiss, J. McWhirter, S. Redif, and I. Proudler, "Multiple shift maximum element sequential matrix diagonalisation for parahermitian matrices," in *IEEE SSP*, Gold Coast, Australia, pp. 312–315, June 2014.

[14] J.G. McWhirter and Z. Wang, "A novel insight to the SBR2 algorithm for diagonalising para-hermitian matrices," in *11th IMA Maths in Sig. Proc*, Birmingham, UK, Dec. 2016.

[15] S. Weiss, J. Pestana, and I.K. Proudler, "On the existence and uniqueness of the eigenvalue decomposition of a parahermitian matrix," *IEEE Trans. SP*, **66**(10):2659–2672, May 2018.

[16] S. Weiss, J. Pestana, I.K. Proudler, and F.K. Coutts, "Corrections to "on the existence and uniqueness of the eigenvalue decomposition of a parahermitian matrix"," *IEEE Trans. SP*, **66**(23):6325–6327, Dec. 2018.

[17] L. Dieci and T. Eirola, "On smooth decompositions of matrices," *SIAM J. Matrix Analysis & Applications*, **20**(3):800–819, 1999.

[18] E. S. Van Vleck, *Numerical algebra, matrix theory, differential-algebraic equations and control theory: Festschrift in honor of Volker Mehrmann*. Springer, 2015, ch. Continuous Matrix Factorizations, pp. 299–318.

[19] B. De Moor and S. Boyd, "Analytic properties of singular values and vectors," KU Leuven, Tech. Rep., 1989.

[20] A. Bunse-Gerstner, R. Byers, V. Mehrmann, and N. K. Nicols, "Numerical computation of an analytic singular value decomposition of a matrix valued function," *Numer. Math*, **60**(1):1–40, 1991.

[21] K. Wright, "Differential equations for the analytic singular value decomposition of a matrix," *Numer. Math.*, **63**(1):283–295, Dec. 1992.

[22] D. Janovská, V. Janovský, and K. Tanabe, "An algorithm for computing the analytic singular value decomposition," *Int. J. Math. & Comp. Sciences*, **2**(11):765–770, 2008.

[23] Y. Nakatsukasa, V. Noferini, and N. Trefethen, "Computing the analytic SVD," Chebfun – numerical computing with functions, 2016.

[24] D. Cescato and H. Bölcskei, "QR decomposition of Laurent polynomial matrices sampled on the unit circle," *IEEE Trans. IT*, **56**(9):4754–4761, Sept. 2010.

[25] ——, "Algorithms for interpolation-based QR decomposition in MIMO-OFDM systems," *IEEE Trans. SP*, **59**(4):1719–1733, Apr. 2011.

[26] M. Tohidian, H. Amindavar, and A.M. Reza, "A DFT-based approximate eigenvalue and singular value decomposition of polynomial matrices," *EURASIP J. Adv. Sig. Proc.*, **2013**(1):1–16, 2013.

[27] S. Weiss, I.K. Proudler, F.K. Coutts, and J. Pestana, "Iterative approximation of analytic eigenvalues of a parahermitian matrix EVD," in *IEEE ICASSP*, Brighton, UK, May 2019.

[28] S. Weiss and M.D. Macleod, "Maximally smooth Dirichlet interpolation from complete and incomplete sample points on the unit circle," in *IEEE ICASSP*, Brighton, UK, May 2019.

[29] S. Weiss, J. Selva, M.D, Macleod, "Measuring smoothness of trigonometric interpolation through incomplete sample points," in *EUSIPCO*, 2020.

[30] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[31] F. Rellich, "Störungstheorie der Spektralzerlegung. I. Mitteilung. Analytische Störung der isolierten Punkteigenwerte eines beschränkten Operators," *Mathematische Annalen*, vol. 113, pp. DC–DCXIX, 1937.

[32] L.V. Ahlfors, *Complex analysis: An introduction to the theory of analytic functions of one complex variable*. NY: McGraw-Hill, 1953.

[33] S. Mandelbrojt, "Analytic functions and classes of infinitely differentiable functions," Rice University Pamphlet, V29, Jan. 1942.

[34] A.V. Oppenheim, R.W. Schafer, and J.R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Pearson, 1999.

[35] B. Girod, R. Rabenstein, and A. Stenger, *Signals and Systems*. Chichester: J. Wiley & Sons, 2001.

[36] P.G. Dirichlet, "Sur la convergence des séries trigonométriques qui servent à preséenter une fonction arbitraire entre des limites données," *Journal für reine und angewandte Mathematik*, **4**(2):157–169, 1829.

[37] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1983.

[38] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*, 2nd ed. Springer, 2017.

[39] A. Zygmund, *Trigonometric Series*, 2nd ed., Cambridge University Press, 1959.

[40] D. Burdick and F.D. Lesley, "Some uniqueness theorems for analytic functions," *The American Mathematical Monthly*, **82**(2):152–155, 1975.

[41] I. Bronshtein and K. Semendyayew, *Handbook of Mathematics*. Heidelberg: Springer, 2015.

[42] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, **61**(3):268–278, Mar. 1973.

[43] M. Ghandi, M.M.J. Yekta, and F. Marvasti, "Some nonlinear/adaptive methods for fast recovery of the missing samples of signals," *Signal Processing*, **88**(3):624 – 638, 2008.

[44] J. Selva, "FFT interpolation from nonuniform samples lying in a regular grid," *IEEE Trans. SP*, **63**(11):2826–2834, June 2015.

[45] S. Weiss, I.K. Proudler, and M.D. Macleod, "Measuring smoothness of real-valued functions defined by sample points on the unit circle," in *SSPD*, Brighton, UK, May 2019.

[46] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, Maryland: John Hopkins University Press, 1996.

[47] H.L. Van Trees, *Detection, Estimation and Modulation Theory: Optimum Array Processing*. New York: Wiley, 2002.

[48] F. Coutts, K. Thompson, S. Weiss, and I.K. Proudler, "A comparison of iterative and DFT-based polynomial matrix eigenvalue decompositions," in *IEEE CAMSAP*, Curacao, Dec. 2017.

[49] J. Foster, J.G. McWhirter, and J. Chambers, "Limiting the order of polynomial matrices within the SBR2 algorithm," in *IMA Maths in Sig. Proc.*, Cirencester, UK, Dec. 2006.

[50] C.H. Ta and S. Weiss, "Shortening the order of paraunitary matrices in SBR2 algorithm," in *6th Int. Conf. ICSP*, Singapore, pp. 1–5, Dec. 2007.

[51] J. Corr, K. Thompson, S. Weiss, I.K. Proudler, and J.G. McWhirter, "Row-shift corrected truncation of paraunitary matrices for PEVD algorithms," in *23rd EUSIPCO*, Nice, France, pp. 849–853, Aug. 2015,

[52] ——, "Shortening of paraunitary matrices obtained by polynomial eigenvalue decomposition algorithms," in *SSPD*, Edinburgh, Sep. 2015.

[53] S. Weiss, I.K Proudler, F.K Coutts, and J. Deeks, "Extraction of analytic eigenvectors from a parahermitian matrix," in *SSPD*, Sept. 2020.

[54] M. Alrmah, S. Weiss, and S. Lambotharan, "An extension of the MUSIC algorithm to broadband scenarios using polynomial eigenvalue decomposition," in *EUSIPCO*, Barcelona, pp. 629–633, Aug. 2011.

[55] V. Neo, C. Evers, and P. Naylor, "Speech enhancement using polynomial eigenvalue decomposition," in *IEEE WASPAA*, New Paltz, NY, 2019.

[56] C.H. Ta and S. Weiss, "A design of precoding and equalisation for broadband MIMO systems," in *41st Asilomar Conf. SSC*, Pacific Grove, CA, pp. 1616–1620, Nov. 2007.

[57] N. Moret, A. Tonello, and S. Weiss, "MIMO precoding for filter bank modulation systems based on PSVD," in *IEEE VTC-Spring*, May 2011.

[58] X. Mestre and D. Gregoratti, "A parallel processing approach to filter-bank multicarrier MIMO transmission under strong frequency selectivity," in *IEEE ICASSP*, pp. 8078–8082, May 2014.

[59] A.I. Pérez-Neira, M. Caus, R. Zakaria, D.L. Ruyet, E. Kofidis, M. Haardt, X. Mestre, and Y. Cheng, "MIMO signal processing in offset-QAM based filter bank multicarrier systems," *IEEE Trans. SP*, **64**(21):5733–5762, Nov. 2016.

[60] C. Delaosa, F.K. Coutts, J. Pestana, and S. Weiss, "Impact of space-time covariance estimation errors on a parahermitian matrix EVD," in *IEEE SAM*, pp. 1–5, July 2018.

APPENDIX

A. Proof of Theorem 2: Laurent polynomial approximation

Proof. We consider one eigenvalue $\lambda(z)$ in $\boldsymbol{\Lambda}(z)$. Because of its analyticity, it can be represented as

$$\lambda(z) = \sum_{n=-\infty}^{\infty} c_n z^{-n} \,,$$

with some coefficients $c_n \in \mathbb{C}$. For its approximation $\hat{\lambda}^{(N)}(z)$, we employ a Laurent polynomial

$$\hat{\lambda}^{(N)}(z) = \sum_{n=-N/2}^{N/2} \hat{c}_n z^{-n}$$

of even order N. Evaluating the least squares approximation error on the unit circle, $z = \mathrm{e}^{\mathrm{j}\Omega}$, yields

$$\begin{aligned}\xi &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |\lambda(\mathrm{e}^{\mathrm{j}\Omega}) - \hat{\lambda}^{(N)}(\mathrm{e}^{\mathrm{j}\Omega})|^2 \mathrm{d}\Omega \\ &= \sum_{n=-N/2}^{N/2} |c_n - \hat{c}_n|^2 + 2 \sum_{n=N/2+1}^{\infty} |c_n|^2 \,,\end{aligned}$$

where we have exploited $\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{j}\Omega n}\mathrm{d}\Omega = \delta(n)$, $\forall n \in \mathbb{Z}$, and the parahermitian property of $\lambda(z)$, s.t. $c_{-n} = c_n^*$. Hence,

$$\min \xi \qquad \longleftrightarrow \qquad \hat{c}_n = c_n \;\; \forall\, |n| \leq \tfrac{N}{2} \;,$$

i.e., $\hat{\lambda}^{(N)}(z)$ is indeed a truncation of $\lambda(z)$. ■

B. Proof of Lemma 1: Switching points of a discontinuous function

Proof: Assume that $\Phi_m(\mathrm{e}^{\mathrm{j}\Omega})$ comprises two analytic segments $\lambda_1(\mathrm{e}^{\mathrm{j}\Omega})$ and $\lambda_2(\mathrm{e}^{\mathrm{j}\Omega})$, with a switching point or discontinuity at Ω_1, where w.l.o.g. $\Omega_1 \neq 0$ i.e.,

$$\Phi_m(\mathrm{e}^{\mathrm{j}\Omega}) = \begin{cases} \lambda_1(\mathrm{e}^{\mathrm{j}\Omega}), & 0 \leq \Omega < \Omega_1 \;, \\ \lambda_2(\mathrm{e}^{\mathrm{j}\Omega}), & \Omega_1 \leq \Omega < 2\pi \;. \end{cases}$$

Due to its 2π periodicity, the two segments also meet at $\Omega = 0$. To be analytic and therefore infinitely differentiable at $\Omega = 0$, we require

$$\lim_{\Omega \to 0+} \frac{\mathrm{d}^p}{\mathrm{d}\Omega^p}\lambda_1(\mathrm{e}^{\mathrm{j}\Omega}) = \lim_{\Omega \to 0-} \frac{\mathrm{d}^p}{\mathrm{d}\Omega^p}\lambda_2(\mathrm{e}^{\mathrm{j}\Omega}) \qquad \forall p \in \mathbb{N} \;. \quad (51)$$

Since analytic functions are unique [40], this can only be satisfied if $\lambda_1(\mathrm{e}^{\mathrm{j}\Omega}) = \lambda_2(\mathrm{e}^{\mathrm{j}\Omega})$, i.e. there cannot be any switching point at Ω_1 either. If (51) is violated for any differentiation order p, then a second switching point has occurred at $\Omega = 0$. ■

Stephan Weiss received a Dipl.-Ing. degree from the University of Erlangen-Nürnberg, Erlangen, Germany, in 1995, and a Ph.D. degree from the University of Strathclyde, Glasgow, Scotland, in 1998, both in electronic and electrical engineering.

He is professor and head of the Centre for Signal and Image Processing (CeSIP) at the University of Strathclyde, Glasgow, following previous academic appointments at both the Universities of Strathclyde and Southampton. His research interests lie in adaptive, multirate, and array signal processing with applications in acoustics, communications, audio, and biomedical signal processing, where he has published more than 300 technical papers. For his work in biomedical signal processing, he was a co-recipient of the 2001 research award of the German society on hearing aids. In 2011 he was a co-recipient of the VTC-Spring best paper award in the MIMO systems track.

Dr Weiss is a member of EURASIP and a senior member of the IEEE. He was the technical co-chair for EUSIPCO 2009 and general chair of IEEE ISPLC 2014, both organised in Glasgow, and special session co-chair for ICASSP 2019.

Ian K. Proudler graduated from Oxford University in 1978 having read Physics. He spent two years doing R&D work in the electronics industry before obtaining a Ph.D. in Digital Signal Processing from Cambridge University in 1984.

He is currently a Visiting Professor of Signal Processing at the University of Strahclyde. From 1986 until 2011 he worked in the Defence sector looking into various adaptive digital signal processing issues such as: numerical stability and efficient computation; antenna algorithm for HF communications; signal separation for ESM purposes; magnetic detection for maritime surveillance; and GPS anti-jam systems. He has published some 100+ research papers, contributed to four textbooks and holds a patent on an adaptive filtering architecture.

He was awarded the John Benjamin Memorial Prize, in 1992 and 2001, and the IEE J.J. Thomson Medal, in 2002, for his work on signal processing algorithms. He was an Honorary Editor for IEE Proceedings: Radar, Sonar and Navigation for ten years. He has been on the organising committee of several international conferences.

Fraser K. Coutts graduated with an MEng (distinction) in Electronic & Electrical Engineering from the University of Strathclyde in 2015, and received his PhD in 2019 from the same institution.

Dr Coutts is a research associate at the Institute for Digital Communications (IDCOM) at the University of Edinburgh, where he is working on secondary inference from compacted data. He previously was a PhD researcher within the Centre for Signal & Image Processing (CeSIP) at the University of Strathclyde, holding a prestigious Caledonian Scholarship from the Carnegie Trust. He has collaborated widely, and his interests span linear algebraic methods for signal processing, hyperspectral imaging, Doppler radar, as well as general algorithm development and implementations. He has published more than 20 peer-reviewed conference and journal papers in these areas, and received a best paper award at IEEE SAM'18.