On the Statistical Consistency of a Generalized Cepstral Estimator

Bin Zhu and Mattia Zorzi

***Abstract*—We consider the problem to estimate the generalized cepstral coefficients of a stationary stochastic process or stationary multidimensional random field. It turns out that a naive version of the periodogram-based estimator for the generalized cepstral coefficients is not consistent. We propose a consistent estimator for those coefficients. Moreover, we show that the latter can be used in order to build a consistent estimator for a particular class of cascade linear stochastic systems.**

***Index Terms*—Generalized cepstral coefficients, periodogram, consistent estimator, system identification.**

I. Introduction

Given a stationary time series $y_0, y_1, \dots, y_{N-1}$, the estimation of some second-order statistics of the series, e.g., the covariances $c_k := \mathbb{E}\,[y_{t+k}\, y_t^*]$ where $\mathbb{E}$ denotes mathematical expectation and $(\cdot)^*$ means complex conjugation, has been a basic problem in the fields of signal processing, identification, and systems theory [1]–[4]. In particular, the covariance estimates are commonly used in practical applications where high resolution spectral estimators are needed. More specifically, we would like to mention the line of research on *rational covariance extension*, see for instance [5]–[16].

Let $\mathbb{T} := [0, 2\pi)$ be the domain for the angular frequency. It follows from the classical theory of stationary processes that the covariances of y_t are the Fourier coefficients of the *power spectral density* $\Phi(\theta)$ which is a *nonnegative* function on $\mathbb{T}$, that is

$$c_k = \int_{\mathbb{T}} e^{ik\theta}\Phi(\theta)\frac{\mathrm{d}\theta}{2\pi}, \quad k \in \mathbb{Z}. \tag{1}$$

In view of the above relation, the covariances c_k can be viewed as *moments* (a term from Mechanics) of the power spectrum Φ where the complex exponentials $\{e^{ik\theta}\}_{k\in\mathbb{Z}}$ play the role of *basis functions*. Another important instance of moments is the cepstral coefficients

$$m_k := \int_{\mathbb{T}} e^{ik\theta}\log\Phi(\theta)\frac{\mathrm{d}\theta}{2\pi}, \quad k \in \mathbb{Z}, \tag{2}$$

assuming that the integration gives a finite number, e.g., when $\log\Phi \in L^1(\mathbb{T})$. These m_k's find applications notably in speech processing, see e.g., [17]–[19]. In addition, cepstral coefficients represent a fundamental tool in covariance extension problems for the simultaneous estimation of poles and zeros in rational spectral densities, see [8], [20].

Inspired by the integral representations (1) and (2), a natural choice of the estimators $\hat{c}_k$ and $\hat{m}_k$ is to replace the true spectrum Φ with the *periodogram* $\hat{\Phi}$ (a simple spectral estimator) and to discretize the integral into a finite summation. An important question is whether these estimators are statistically consistent, that is: does $\hat{c}_k$ (and $\hat{m}_k$) converge in a certain stochastic sense to the true value c_k (and m_k, respectively) as the length N of the available time series goes to infinity? Such a consistency property is useful in e.g., evaluating the consistency of those moment-based spectral estimators when the model class is properly chosen, see e.g., [21], [22].

For the covariance estimator $\hat{c}_k$, an affirmative answer to the consistency question can be found in [23, Subsec. 5.3.3] under some Gaussian assumptions, which is somewhat surprising because the periodogram $\hat{\Phi}$ alone is *not* a consistent estimator of the true spectrum Φ since its variance (at each frequency θ) does not converge to zero as $N \to \infty$, see e.g., [23, p. 425]. Consistency of $\hat{c}_k$ comes out as a result of the averaging operation in the discrete version of (1) involving $\hat{\Phi}$. The consistency question for the cepstral estimator $\hat{m}_k$ is significantly harder than that for $\hat{c}_k$, because the latter estimator is linear in $\hat{\Phi}$ while the former one involves the logarithm (which is obviously nonlinear). The answer is still affirmative and can be found in [17] under the assumption that the spectral components Y_ℓ (to be defined later in Section III) are independent Gaussian random variables. That result has also been extended to the case in which these components are correlated [18].

In the case of multidimensional random fields, however, the classical cepstral coefficients (2) do not always guarantee that the covariance extension problem admits a solution, see [12], [24]. In order to overcome this difficulty, in the recent papers [21], [25], [26] of the authors, the *generalized cepstral coefficients* [27], which are the moments of a power function of the spectrum, have been used to set up a rational covariance extension problem which admits as the solution a rational multidimensional spectral density with zeros. In this context, the consistency of the covariance estimator has been established in [23, Sec. 9.7] for the two-dimensional case, and the result is extendable to general d-dimensional random fields in view of [28]. To our surprise, the periodogram-based estimator for the generalized cepstral coefficients appeared to be inconsistent when we were carrying out simulation studies. However, it turns out in the later development of the current paper that the generalized cepstral estimator using $\hat{\Phi}$ is indeed *consistent up to a constant multiplicative factor* under the same Gaussian assumption for the usual cepstral estimator in

B. Zhu is with the School of Intelligent Systems Engineering, Sun Yat-sen University, Gongchang Road 66, 518107 Shenzhen, China (email: zhub26@mail.sysu.edu.cn).

M. Zorzi is with the Department of Information Engineering, University of Padova, Via Giovanni Gradenigo, 6b, 35131 Padova, Italy (email: zorzimat@dei.unipd.it).

This work was supported in part by the National Natural Science Foundation of China under the grant number 62103453 and the "Hundred-Talent Program" of Sun Yat-sen University. Corresponding author B. Zhu. Tel. +86 14748797525. Fax +86(20) 39336557.

[17]. In other words, the unwindowed periodogram suffices to produce a consistent estimator of the generalized cepstral coefficients and all one needs to do is to suitably rescale the computed quantities. To the best of our knowledge, such a consistency question has never been addressed before in the literature. Moreover, the consistency result is partially extended to the general case of correlated spectral components, and generalization to multidimensional random fields is also briefly discussed. Finally, we show that our result can be used to build a consistent estimator for a class of *cascade* linear stochastic systems. As explained in e.g., [29], [30], cascade systems are very common in process industry, signal processing, and other engineering applications.

This paper is organized as follows. Section II reviews the DFT and sets up the notation. Section III formulates the problem of generalized cepstral estimation and its statistical consistency. Section IV shows the consistency of the generalized cepstral estimator under the assumption of independent Gaussian spectral components. Section V deals with the more general case of correlated spectral components. Section VI extends the consistency result of the generalized cepstral estimation to multidimensional random fields. Section VII describes an application of the generalized cepstral estimator to cascade system identification, for which some simulation results are presented in Section VIII. Finally, Section IX draws the conclusions.

II. Review of the Discrete Fourier Transform

The aim of this short section is to clarify the notation of the DFT that is used throughout the paper.

The standard DFT (in the signal processing literature) of a finite complex-valued sequence $y_0, \dots, y_{N-1}$ of length N is defined as

$$Y_\ell = \sum_{t=0}^{N-1} y_t \, e^{-it\frac{2\pi}{N}\ell}, \quad \ell = 0, \dots, N-1. \tag{3}$$

The inverse DFT is known as

$$y_t = \frac{1}{N} \sum_{\ell=0}^{N-1} Y_\ell \, e^{it\frac{2\pi}{N}\ell}, \quad t = 0, \dots, N-1. \tag{4}$$

The above two expression can be rewritten in matrix forms as

$$\mathbf{Y} = \mathbf{F}\mathbf{y} \quad \text{and} \quad \mathbf{y} = \frac{1}{N}\mathbf{F}^*\mathbf{Y}, \tag{5}$$

in which $\mathbf{Y} = [Y_0, \dots, Y_{N-1}]^\top$, $\mathbf{y} = [y_0, \dots, y_{N-1}]^\top$, and

$$\mathbf{F} = \begin{bmatrix} \omega_N^{0\cdot 0} & \omega_N^{0\cdot 1} & \cdots & \omega_N^{0\cdot(N-1)} \\ \omega_N^{1\cdot 0} & \omega_N^{1\cdot 1} & \cdots & \omega_N^{1\cdot(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_N^{(N-1)\cdot 0} & \omega_N^{(N-1)\cdot 1} & \cdots & \omega_N^{(N-1)\cdot(N-1)} \end{bmatrix} \tag{6}$$

is the DFT matrix where $\omega_N = e^{-i\frac{2\pi}{N}}$ is a primitive N-th root of unity. It follows immediately that $\mathbf{F}^{-1} = \frac{1}{N}\mathbf{F}^*$. Notice also that $\mathbf{F}$ is a complex symmetric matrix so that $\mathbf{F}^* = \overline{\mathbf{F}}$ (elementwise complex conjugate).

III. Problem Statement

In order to introduce the generalized cepstral estimation problem, we need to first review some basic notions of the periodogram and covariance estimation.

Consider a zero-mean second-order stationary complex-valued discrete-time random signal y_t (so that $t \in \mathbb{Z}$). Suppose that we are given a finite number of random samples $y_0, y_1, \dots, y_{N-1}$. Taking the (unilateral) finite Fourier transform, we obtain

$$Y(\theta) = \sum_{t=0}^{N-1} y_t \, e^{-it\theta}, \tag{7}$$

a random variable that depends on the angular frequency $\theta \in \mathbb{T} = [0, 2\pi)$. Notice that the frequency interval $[0, 2\pi)$ is understood as the quotient group $\mathbb{R}/2\pi\mathbb{Z}$, so that the function $Y(\theta)$ is 2π-periodic on $\mathbb{R}$. In the spectral theory of stationary processes, $\mathbb{T}$ is often identified as the unit circle on the complex plane, and one writes $Y(e^{i\theta})$ instead [2]. Here we keep the simpler notation $Y(\theta)$. The (unwindowed) periodogram, which is a first estimate of the power spectrum $\Phi(\theta)$ of the random process y_t, is defined as

$$\hat{\Phi}(\theta) := \frac{1}{N}|Y(\theta)|^2. \tag{8}$$

In addition, the periodogram $\hat{\Phi}$ admits a *correlogram* interpretation as the *bilateral* finite Fourier transform of the standard biased covariance estimates [1], that is,

$$\hat{\Phi}(\theta) = \sum_{k=-N+1}^{N-1} \hat{c}_k e^{-ik\theta}, \tag{9}$$

where

$$\hat{c}_k = \frac{1}{N} \sum_{t=0}^{N-1-k} y_{t+k} \, y_t^*, \quad k = 0, 1, \dots, N-1 \tag{10}$$

and $\hat{c}_{-k} = \hat{c}_k^*$. Here $(\cdot)^*$ means taking complex conjugate when applied to a complex number.

In practice, the periodogram $\hat{\Phi}$ is evaluated on a regular grid of $\mathbb{T}$ of size N equal to the length of the time series y_t. In other words, the frequency interval $[0, 2\pi)$ is partitioned into N subintervals of equal length $2\pi/N$, and the grid points are collected in the set

$$\mathbb{T}_N := \left\{ \frac{2\pi}{N}\ell \,:\, \ell = 0, 1, \dots, N-1 \right\}. \tag{11}$$

Let us write $Y_\ell = Y(2\pi\ell/N)$, and similarly $\hat{\Phi}_\ell = \hat{\Phi}(2\pi\ell/N)$ for simplicity. The random variables $\{Y_\ell\}_{\ell=0}^{N-1}$ are termed *spectral components* of the process y_t [17]. The covariance estimates can then be computed from the discrete spectrum $\{\hat{\Phi}_\ell\}_{\ell=0}^{N-1}$ via the inverse DFT (see also Sec. II)

$$\hat{c}_k \approx \frac{1}{N} \sum_{\ell=0}^{N-1} e^{ik\frac{2\pi}{N}\ell} \hat{\Phi}_\ell, \tag{12}$$

which is seen as an approximation to the Fourier integral

$$\hat{c}_k = \int_{\mathbb{T}} e^{ik\theta} \hat{\Phi}(\theta) \frac{\mathrm{d}\theta}{2\pi}. \tag{13}$$

Remark 1. Alternatively, we can evaluate the periodogram on a denser grid of size $K \geq 2N-1$. The motivation for doing so is that such a choice of the grid size is sufficient to guarantee that the covariance estimates can be *exactly* recovered from the discrete spectrum $\{\hat{\Phi}_\ell\}_{\ell=0}^{K-1}$ via the formula

$$\hat{c}_k = \frac{1}{K}\sum_{\ell=0}^{K-1} e^{ik\frac{2\pi}{K}\ell}\hat{\Phi}(2\pi\ell/K), \tag{14}$$

in contrast to the usual Fourier integral (13). Computation via (14) is more efficient than the direct time average (10) thanks to the FFT routines. However, we also want to point out that as N (hence also K) tends to infinity, the difference between (12) and (14) will become negligible.

In this paper, we deal with the problem of estimating from samples of a stationary process its generalized cepstral coefficients. The latter object is obtained via replacing the logarithm in the definition of the classical cepstral coefficients (2) with the so-called *generalized logarithmic function*:

$$s_\alpha(x) = \begin{cases} (x^\alpha - 1)/\alpha, & 0 < |\alpha| \leq 1, \\ \log x, & \alpha = 0. \end{cases} \tag{15}$$

The generality comes from the fact that for a fixed $x > 0$, the limit $(x^\alpha - 1)/\alpha \to \log x$ holds as the parameter $\alpha \to 0$. The generalized logarithm leads to the following definition where we are mostly interested in positive α's.

Definition 1 ([27])**.** Given a power spectral density $\Phi(\theta)$, a real number $\alpha \in (0, 1]$, the generalized cepstral coefficients are defined as

$$m_{\alpha,k} = \begin{cases} \frac{1}{\alpha}\int_{\mathbb{T}} e^{ik\theta}\Phi(\theta)^\alpha \frac{\mathrm{d}\theta}{2\pi} & \text{if } k \neq 0, \\ \frac{1}{\alpha}\left[\int_{\mathbb{T}} \Phi(\theta)^\alpha \frac{\mathrm{d}\theta}{2\pi} - 1\right] & \text{if } k = 0. \end{cases} \tag{16}$$

Notice that the constant one in the numerator of $(x^\alpha - 1)/\alpha$ disappears in the case of $k \neq 0$ due to the integration against the complex exponential function $e^{ik\theta}$. Moreover, we are not so interested in $\alpha = 1$ since in that case, $m_{\alpha,k}$ reduces to the covariance c_k in (1) (with a difference of one for $k = 0$). Thus in the remaining part of this paper, we only consider the case $\alpha \in (0, 1)$. With a slight abuse of the notation, we shall omit α in the subscript and simply write m_k instead.

An estimator of the generalized cepstral coefficients using the periodogram, inspired by (12), is given as

$$\hat{m}_k = \begin{cases} \frac{1}{\alpha N}\sum_{\ell=0}^{N-1} e^{ik\frac{2\pi}{N}\ell}\hat{\Phi}_\ell^\alpha & \text{if } k \neq 0, \\ \frac{1}{\alpha}\left(\frac{1}{N}\sum_{\ell=0}^{N-1}\hat{\Phi}_\ell^\alpha - 1\right) & \text{if } k = 0. \end{cases} \tag{17}$$

We end this section by formally stating the consistency problem for the generalized cepstral estimator above.

Problem 1. Given a sample path $y_0, \ldots, y_{N-1}$ of a zero-mean stationary complex random process y_t, understand whether the estimator (17) converges to the true value (16) in some stochastic sense as the length of the sample path N tends to infinity.

IV. Consistency in the Case of Independent Spectral Components

First, let us introduce a normalized version of the spectral components:

$$\bar{Y}_\ell := \frac{1}{\sqrt{N}} Y_\ell \quad \text{so that} \quad \hat{\Phi}_\ell = |\bar{Y}_\ell|^2. \tag{18}$$

Two assumptions for our signal y_t are stated next.

Assumption 1. The normalized spectral components $\{\bar{Y}_\ell\}_{\ell=0}^{N-1}$ of the signal $\{y_t\}_{t=0}^{N-1}$ are zero-mean complex (circular) Gaussian random variables such that $\operatorname{var}\bar{Y}_\ell = \mathbb{E}\,\hat{\Phi}_\ell = \lambda_\ell > 0$. More precisely, the real and imaginary parts of each Y_ℓ are independent real Gaussian random variables having equal variance, namely $\lambda_\ell/2$, see e.g., [1, p. 361].

Assumption 2. The normalized spectral components $\{\bar{Y}_\ell\}_{\ell=0}^{N-1}$ are independent.

Observe that the above two assumptions hold in the following two cases.

Case 1. The process y_t is i.i.d. Gaussian with a variance $\sigma^2 > 0$, e.g., a Gaussian white noise. In this case, the random vector $\bar{\mathbf{Y}} = \frac{1}{\sqrt{N}}\mathbf{F}\mathbf{y}$ (see (5) for the notation) is Gaussian with a covariance matrix

$$\mathbb{E}\bar{\mathbf{Y}}\bar{\mathbf{Y}}^* = \frac{1}{N}\mathbf{F}(\mathbb{E}\mathbf{y}\mathbf{y}^*)\mathbf{F}^* = \sigma^2\mathbf{I}, \tag{19}$$

where $\mathbf{F}$ is the DFT matrix in (6).

Case 2. The process y_t is Gaussian and N-periodic, i.e., $y_t = y_{t+N}$ almost surely for all $t \in \mathbb{Z}$. An immediate consequence is the cyclic symmetry $c_{N-k} = c_{-k} = c_k^*$ for the covariances. It follows that the covariance matrix of the random vector $\mathbf{y}$ has a *circulant* (more than being Toeplitz) structure, namely

$$\mathbf{C} := \mathbb{E}\mathbf{y}\mathbf{y}^* = \begin{bmatrix} c_0 & c_{N-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & c_{N-1} & & c_2 \\ \vdots & c_1 & c_0 & \ddots & \vdots \\ c_{N-2} & & \ddots & \ddots & c_{N-1} \\ c_{N-1} & c_{N-2} & \cdots & c_1 & c_0 \end{bmatrix}. \tag{20}$$

In plain words, each column of $\mathbf{C}$ is the cyclic shift of the first column (the same for rows). Since $\mathbf{C}$ is also a covariance matrix, we shall in addition require it to be positive definite[1]. It is well known that any circulant matrix can be diagonalized by the DFT matrix. More precisely, define a unitary matrix $\mathbf{U} := \frac{1}{\sqrt{N}}\mathbf{F}$ and let $\mathbf{c} = [c_0, \ldots, c_{N-1}]^\top$ be the first column of $\mathbf{C}$. Then the cirulant matrix $\mathbf{C}$ admits a spectral decomposition $\mathbf{C} = \mathbf{U}^*\Psi\mathbf{U}$ where the columns of $\mathbf{U}^*$ are the normalized eigenvectors and the diagonal matrix $\Psi = \operatorname{diag}(\mathbf{F}\mathbf{c})$ contains the positive eigenvalues [31], see also [32]. Therefore, the covariance matrix of the vector $\bar{\mathbf{Y}}$ of spectral components is

$$\mathbb{E}\bar{\mathbf{Y}}\bar{\mathbf{Y}}^* = \frac{1}{N}\mathbf{F}\mathbf{C}\mathbf{F}^* = \Psi. \tag{21}$$

The main result of this section is summarized below.

[1]The Hermitian structure of $\mathbf{C}$ comes from the relation $c_{N-k} = c_k^*$.

Theorem 1. *Under Assumptions 1 and 2, the estimator* (17) *of the generalized cepstral coefficients is consistent up to a constant multiplicative factor. More precisely, we have for* $k \neq 0$*,*

$$C\hat{m}_k = \frac{C}{\alpha N}\sum_{\ell=0}^{N-1} e^{ik\frac{2\pi}{N}\ell}\hat{\Phi}_\ell^\alpha \xrightarrow{m.s.} m_k \tag{22}$$

and

$$C\hat{m}_0 + \frac{C-1}{\alpha} = \frac{C}{\alpha N}\sum_{\ell=0}^{N-1}\hat{\Phi}_\ell^\alpha - \frac{1}{\alpha} \xrightarrow{m.s.} m_0 \tag{23}$$

as $N \to \infty$*, where the constant* $C = 1/\Gamma(\alpha+1)$ *and the convergence is understood in the mean-square sense. Here* $\Gamma(\cdot)$ *is the gamma function.*

The proof of this theorem is built on the following lemma concerning mean square convergence of random variables [33].

Lemma 1. *Let* X_n *be a sequence of complex-valued random variables such that* $\mathbb{E}X_n \to C$ *as* $n \to \infty$ *where* C *is a (complex) constant. Then we have* $X_n \xrightarrow{m.s.} C$ *if and only if* $\operatorname{var} X_n \to 0$ *as* $n \to \infty$.

The proof of the lemma is a simple exercise which reduces to definition checking. Therefore, the claim of Theorem 1 can be established if we can show in the case of $k \neq 0$ that $\mathbb{E}(C\hat{m}_k) \to m_k$ for the given constant C and that $\operatorname{var}\hat{m}_k \to 0$ as the number of samples N tends to infinity, and similarly for the case $k = 0$. The rest of this section is devoted to the computation of the expectation and the variance of $\hat{m}_k$.

A. Mean of $\hat{\Phi}_\ell^\alpha$ and $\hat{m}_k$

It follows from Assumption 2 that $\hat{\Phi}_\ell = |\bar{Y}_\ell|^2 = (\mathrm{Re}\bar{Y}_\ell)^2 + (\mathrm{Im}\bar{Y}_\ell)^2$ is χ^2-distributed with a degree of freedom 2 after suitable scaling. In fact, the probability density function of $\hat{\Phi}_\ell$ is

$$f(x) = \frac{1}{\lambda_\ell}e^{-x/\lambda_\ell}, \quad x \geq 0, \tag{24}$$

which is just an exponential distribution. The following computation is straightforward:

$$\mathbb{E}(\hat{\Phi}_\ell^\alpha) = \frac{1}{\lambda_\ell}\int_0^\infty x^\alpha e^{-x/\lambda_\ell}\mathrm{d}x = \lambda_\ell^\alpha\Gamma(\alpha+1), \tag{25}$$

where the second equality comes from [34, Eq. 3.381.4].

We are now ready to compute the mean of $\hat{m}_k$. For $k \neq 0$, we have

$$\begin{aligned}\mathbb{E}\,\hat{m}_k &= \mathbb{E}\left(\frac{1}{\alpha N}\sum_{\ell=0}^{N-1} e^{ik\frac{2\pi}{N}\ell}\hat{\Phi}_\ell^\alpha\right)\\ &= \frac{1}{\alpha N}\sum_{\ell=0}^{N-1} e^{ik\frac{2\pi}{N}\ell}\lambda_\ell^\alpha\Gamma(\alpha+1) &\text{(26a)}\\ &\to \frac{\Gamma(\alpha+1)}{\alpha}\int_{\mathbb{T}} e^{ik\theta}\Phi(\theta)^\alpha\frac{\mathrm{d}\theta}{2\pi} \quad \text{as} \quad N\to\infty. &\text{(26b)}\end{aligned}$$

To see the latter limit, recall first that under a mild condition for the decay of the true covariances c_k of the process y_t, $\mathbb{E}\,\hat{\Phi}_\ell = \lambda_\ell \to \Phi(2\pi\ell/N)$ as $N \to \infty$ for each $\ell \in \{0,\dots,N-1\}$ where Φ is the true spectrum [1, p. 7]. Therefore, the term $\frac{1}{N}\sum_{\ell=0}^{N-1} e^{ik\frac{2\pi}{N}\ell}\lambda_\ell^\alpha$ in (26a) converges to the integral in (26b), where the convergence is understood from the normalized Riemann sum to the integral.

Similarly for $k = 0$, we have

$$\begin{aligned}\mathbb{E}\,\hat{m}_0 + \frac{1}{\alpha} &= \mathbb{E}\left(\frac{1}{\alpha N}\sum_{\ell=0}^{N-1}\hat{\Phi}_\ell^\alpha\right)\\ &\to \frac{\Gamma(\alpha+1)}{\alpha}\int_{\mathbb{T}}\Phi(\theta)^\alpha\frac{\mathrm{d}\theta}{2\pi} \quad \text{as} \quad N\to\infty,\end{aligned} \tag{27}$$

which is equivalent to $\mathbb{E}[C\hat{m}_0 + (C-1)/\alpha] \to m_0$ for the random variable in (23) and the constant C given in Theorem 1.

B. Variance of $\hat{\Phi}_\ell^\alpha$ and $\hat{m}_k$

The second moment of $\hat{\Phi}_\ell^\alpha$ is just

$$\mathbb{E}[(\hat{\Phi}_\ell^\alpha)^2] = \mathbb{E}(\hat{\Phi}_\ell^{2\alpha}) = \lambda_\ell^{2\alpha}\Gamma(2\alpha+1), \tag{28}$$

where the second equality follows immediately from (25) since they have the same functional form. Consequently, we have

$$\begin{aligned}\operatorname{var}(\hat{\Phi}_\ell^\alpha) &= \mathbb{E}(\hat{\Phi}_\ell^{2\alpha}) - [\mathbb{E}(\hat{\Phi}_\ell^\alpha)]^2\\ &= \lambda_\ell^{2\alpha}\left\{\Gamma(2\alpha+1) - [\Gamma(\alpha+1)]^2\right\}.\end{aligned} \tag{29}$$

Example 1. In the papers [21], [25], we took some special values of the parameter α, i.e.,

$$\alpha = 1 - \frac{1}{\nu}, \quad \nu \in \mathbb{N}_+, \quad \nu \geq 2, \tag{30}$$

where $\mathbb{N}_+ = \{1, 2, \dots\}$ denotes the set of positive integers. For instance, take $\nu = 2$ so $\alpha = 1/2$, and we have

$$\begin{aligned}\mathbb{E}(\hat{\Phi}_\ell^\alpha) &= \sqrt{\lambda_\ell}\,\Gamma(3/2) = \frac{\sqrt{\pi\lambda_\ell}}{2},\\ \mathbb{E}(\hat{\Phi}_\ell^{2\alpha}) &= \lambda_\ell\Gamma(2) = \lambda_\ell,\end{aligned} \tag{31}$$

and $\operatorname{var}(\hat{\Phi}_\ell^\alpha) = (1-\pi/4)\lambda_\ell$.

Now we can continue to compute the variance of $\hat{m}_k$, where the independence between the spectral components in Assumption 2 plays an important role. For $k \neq 0$, we have

$$\begin{aligned}\operatorname{var}(\hat{m}_k) &= \operatorname{var}\left(\frac{1}{\alpha N}\sum_{\ell=0}^{N-1} e^{ik\frac{2\pi}{N}\ell}\hat{\Phi}_\ell^\alpha\right)\\ &= \frac{1}{\alpha^2N^2}\sum_{\ell=0}^{N-1}\operatorname{var}(\hat{\Phi}_\ell^\alpha) &\text{(32a)}\\ &= \frac{C_1}{\alpha^2N^2}\sum_{\ell=0}^{N-1}\lambda_\ell^{2\alpha} \to 0 \quad \text{as} \quad N\to\infty, &\text{(32b)}\end{aligned}$$

where, in (32a) we have used the independence assumption between the spectral components $\{\bar{Y}_\ell\}_{\ell=0}^{N-1}$, and in (32b) the constant $C_1 = \Gamma(2\alpha+1) - [\Gamma(\alpha+1)]^2$ is determined from (29). To see the last limit, recognize

$$\frac{1}{N}\sum_{\ell=0}^{N-1}\lambda_\ell^{2\alpha} \to \int_0^{2\pi}\Phi(\theta)^{2\alpha}\frac{\mathrm{d}\theta}{2\pi} \tag{33}$$

as the convergence of the (normalized) Riemann sum to the integral. Hence the Riemann sum is bounded, and its product with $C_1/(\alpha^2N)$ tends also to zero as $N \to \infty$.

For $k=0$, we have

$$\operatorname{var}(\hat{m}_0) = \operatorname{var}\left(\frac{1}{\alpha N}\sum_{\ell=0}^{N-1}\hat{\Phi}_\ell^\alpha\right) \tag{34}$$

which leads to the same result as (32a). Therefore, we also have $\operatorname{var}(\hat{m}_0)\to 0$ as $N\to\infty$.

Remark 2. It is implicitly assumed that the underlying true spectral density function Φ is sufficiently "nice" in order for such convergences as (26b) and (33) to make sense. For example, when the true covariance sequence c_k belongs to the Wiener class, i.e., $\sum_{k=-\infty}^{\infty}|c_k|<\infty$, the corresponding spectral density Φ is in fact continuous (a consequence of Lebesgue's dominated convergence theorem).

Remark 3. In view of the formulas (1), (26b), and (33), the true spectral density Φ of the underlying process y has to satisfy the integrability condition: $\Phi\in L^1(\mathbb{T})\cap L^\alpha(\mathbb{T})\cap L^{2\alpha}(\mathbb{T})$. Since the set $\mathbb{T}$ has a finite Lebesgue measure, we have for $0<p<q\le\infty$ that $L^q(\mathbb{T})\subset L^p(\mathbb{T})$ [35]. Recall that the parameter α takes value in the interval $(0,1)$. Therefore, the former intersection reduces to $L^1(\mathbb{T})$ if $0<\alpha\le 1/2$, and to $L^{2\alpha}(\mathbb{T})$ if $1/2<\alpha<1$. For both cases, $\Phi\in L^2(\mathbb{T})$ is a sufficient condition.

V. The Case of Correlated Spectral Components

Independence[2] between spectral components as dictated by Assumption 2 holds in an asymptotic sense [23, Theorem 6.2.3, p. 426]. With a finite number of measurements (samples) however, two spectral components $\bar{Y}_{\ell_1}$ and $\bar{Y}_{\ell_2}$ with $\ell_1\neq\ell_2$ are in general correlated. Indeed, a direct computation using the unnormalized spectral components (7) leads to the expression

$$\operatorname{cov}(Y_{\ell_1},Y_{\ell_2}) = \frac{1}{1-e^{-i(\theta_1-\theta_2)}}\left[\sum_{k=0}^{N-1}c_k a_k - \sum_{k=-N+1}^{-1}c_k a_k\right] \tag{35}$$

where $\theta_j=2\pi\ell_j/N$ for $j=1,2$, c_k is the true covariance of the process y_t in (1), and $a_k=e^{-i\theta_1 k}-e^{-i\theta_2 k}$.

The consistency problem for the classical cepstral estimation in the case of correlated spectral components has been investigated in [18]. In what follows we review some well known facts that will be used later. Assume that $\bar{Y}_{\ell_1}$ and $\bar{Y}_{\ell_2}$ are zero-mean jointly circular Gaussian random variables, that is, we drop the independence part in Assumption 2. Let

$$R_{\ell_1\ell_2} = \begin{bmatrix} \lambda_{\ell_1} & r_{\ell_1\ell_2} \\ r_{\ell_1\ell_2}^* & \lambda_{\ell_2}\end{bmatrix} \tag{36}$$

be the covariance matrix of the random vector $[\,\bar{Y}_{\ell_1}\ \bar{Y}_{\ell_2}\,]^\top$. According to [18, Eq. 16], we have for $\alpha\in(0,1)$ that

$$\begin{aligned}\mathbb{E}[\hat{\Phi}_{\ell_1}^\alpha\hat{\Phi}_{\ell_2}^\alpha] &= \mathbb{E}[\,|\bar{Y}_{\ell_1}|^{2\alpha}|\bar{Y}_{\ell_2}|^{2\alpha}\,]\\ &= \lambda_{\ell_1}^\alpha\lambda_{\ell_2}^\alpha[\Gamma(\alpha+1)]^2\,{}_2F_1(-\alpha,-\alpha;1;\rho_{\ell_1\ell_2}^2)\end{aligned} \tag{37}$$

where,

$$\rho_{\ell_1\ell_2}^2 := |r_{\ell_1\ell_2}|^2/(\lambda_{\ell_1}\lambda_{\ell_2})$$

[2]Since we always work under Assumption 1 of Gaussianity, the terms "independent" and "uncorrelated" can be interchanged.

is the modulus squared of the Pearson correlation coefficient between $\bar{Y}_{\ell_1}$ and $\bar{Y}_{\ell_2}$, and

$${}_2F_1(a,b;c;z) = \sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n}\frac{1}{n!}z^n \tag{38}$$

is the hypergeometric series [36, Ch. 15] with a convergence region $|z|<1$ (when c is not a negative integer). Here

$$(a)_n = \begin{cases}1, & n=0\\ a(a+1)\cdot\ldots\cdot(a+n-1), & n>0\end{cases}$$

is the *Pochhammer symbol* for the *rising factorial*. In addition, the series (38) also converges on the unit circle $|z|=1$ if $\operatorname{Re}(c-a-b)>0$ which holds in the case of (37).

Next, we show that in the case with correlated spectral components, the statistical consistency of the estimator (17) using the unwindowed periodogram *cannot* be guaranteed in general, although (26) and (27) still hold because they have been derived without using the independence assumption. We shall only work on $\hat{m}_k$ with $k\neq 0$ since the conclusion for the case of $k=0$ follows similarly.

It is not difficult to see that

$$\begin{aligned}&\operatorname{var}(\hat{m}_k) = \mathbb{E}\left[|\hat{m}_k-\mathbb{E}\,\hat{m}_k|^2\right]\\ =&\frac{1}{\alpha^2N^2}\sum_{\ell_1=0}^{N-1}\sum_{\ell_2=0}^{N-1}e^{ik\frac{2\pi}{N}(\ell_1-\ell_2)}\left[\mathbb{E}(\hat{\Phi}_{\ell_1}^\alpha\hat{\Phi}_{\ell_2}^\alpha)-\mathbb{E}(\hat{\Phi}_{\ell_1}^\alpha)\mathbb{E}(\hat{\Phi}_{\ell_2}^\alpha)\right]\\ =&\frac{1}{\alpha^2N^2}\sum_{\ell_1=0}^{N-1}\sum_{\ell_2=0}^{N-1}e^{ik\frac{2\pi}{N}(\ell_1-\ell_2)}\lambda_{\ell_1}^\alpha\lambda_{\ell_2}^\alpha[\Gamma(\alpha+1)]^2\\ &\qquad\times\left[{}_2F_1(-\alpha,-\alpha;1;\rho_{\ell_1\ell_2}^2)-1\right]\end{aligned} \tag{39}$$

where in the last equality we exploited (25) and (37). The computation can be continued as follows:

$$\begin{aligned}&\operatorname{var}(\hat{m}_k) = |\operatorname{var}(\hat{m}_k)|\\ \le&\frac{[\Gamma(\alpha+1)]^2}{\alpha^2N^2}\sum_{\ell_1=0}^{N-1}\sum_{\ell_2=0}^{N-1}\lambda_{\ell_1}^\alpha\lambda_{\ell_2}^\alpha\left[{}_2F_1(-\alpha,-\alpha;1;\rho_{\ell_1\ell_2}^2)-1\right]\end{aligned} \tag{40a}$$

$$\le\frac{[\Gamma(\alpha+1)]^2}{\alpha^2N^2}\sum_{\ell_1=0}^{N-1}\sum_{\ell_2=0}^{N-1}\lambda_{\ell_1}^\alpha\lambda_{\ell_2}^\alpha\left[{}_2F_1(-\alpha,-\alpha;1;1)-1\right] \tag{40b}$$

$$=\frac{C_1}{\alpha^2N^2}\sum_{\ell_1=0}^{N-1}\lambda_{\ell_1}^\alpha\sum_{\ell_2=0}^{N-1}\lambda_{\ell_2}^\alpha\to\frac{C_1}{\alpha^2}\left(\int_{\mathbb{T}}\Phi(\theta)^\alpha\frac{\mathrm{d}\theta}{2\pi}\right)^2=\text{const.} \tag{40c}$$

as $N\to\infty$ where,

- (40a) holds because using the series expansion (38), we see that the quantity

$${}_2F_1(-\alpha,-\alpha;1;\rho_{\ell_1\ell_2}^2)-1=\sum_{n=1}^{\infty}\left[\frac{(-\alpha)_n}{n!}\right]^2\rho_{\ell_1\ell_2}^{2n}$$

  is positive;
- (40b) follows from the relation $0\le\rho_{\ell_1\ell_2}^2\le 1$ for the correlation coefficient;

- (40c) results from an application of Gauss's summation theorem [36, Eq. 15.1.20]:

$$ {}_2F_1(-\alpha,-\alpha;1;1) = \frac{\Gamma(2\alpha+1)}{[\Gamma(\alpha+1)]^2} \tag{41} $$

which holds since we have restricted ourselves to the case $0 < \alpha < 1$, and C_1 is the same constant as the one in (32b).

As revealed by (40c), the upper bound for $\hat{m}_k$ is a constant that does not improve as N increases. Although such a bound may not be tight, consistency of the generalized cepstral estimator (17) cannot be directly derived in this case. There are two ways to resolve this issue. One way is to appeal to a *windowed periodogram* $\check{\Phi}$ instead of the unwindowed version $\hat{\Phi}$, as hinted at the end of [18, Sec. IV]. More precisely, one first introduces a suitable window function $w : \mathbb{Z} \to \mathbb{R}$, see [23, Subsec. 6.2.3], and then modifies the correlogram (9) as

$$ \check{\Phi}(\theta) = \sum_{k=-N+1}^{N-1} w(k)\hat{c}_k e^{-ik\theta}. \tag{42} $$

A modified estimator $\check{m}_k$ results if one replaces $\hat{\Phi}$ in (17) with the above $\check{\Phi}$. Such a practice is intuitively reasonable since under quite general conditions, $\check{\Phi}$ is a consistent estimator of the true spectrum Φ [23, Subsec. 6.2.4], and one may expect that the conclusion of Theorem 1 holds for $\check{m}_k$. A rigorous proof seems hard (if possible) as for example in the computation of (25), now $\check{\Phi}_\ell$ is a linear combination of $\hat{\Phi}_\ell$ and has a *generalized chi-squared* distribution [37]. Unfortunately, the latter distribution in general does not admit a closed-form expression for the probability density function.

The other way is to make an additional assumption about the underlying random process y_t such as the following one.

Assumption 3. The variances of the normalized spectral components $\{\bar{Y}_\ell\}_{\ell=0}^{N-1}$ are asymptotically bounded, that is, the quantity

$$ \gamma_N := \max_{0\le\ell\le N-1} \lambda_\ell \tag{43} $$

satisfies $\lim_{N\to\infty}\gamma_N < \infty$. Moreover, the growth of the correlation coefficients is under control in the sense that

$$ \sum_{\ell_2=0}^{N-1} \rho^2_{\ell_1\ell_2} \le f(N), \quad \forall \ell_1 \in \{0,1,\dots,N-1\} \tag{44} $$

where $f(N)$ is a function such that

$$ \lim_{N\to\infty} f(N)/N = 0. \tag{45} $$

In plain words, condition (44) requires that the correlation between the spectral components is mild. We then have the next result.

Proposition 1. *Let Assumption 1 hold. Assume further that the true covariances c_k of the signal y_t satisfy the decay condition [1, Eq. (1.3.11), p. 7], and that the true spectral density Φ of y_t is square integrable on $\mathbb{T}$. Then under the additional Assumption 3, the generalized cepstral estimator* (17) *is mean-square consistent up to a constant multiplicative factor, that is,* (22) *and* (23) *hold.*

Proof. Apparently, the calculations in (39) and (40) are all valid. We shall modify the inequalities after (40a) and show that the power of Assumption 3 can lead to the result $\mathrm{var}(\hat{m}_k) \to 0$ as $N \to \infty$.

Given the fact $0 \le \rho^2_{\ell_1\ell_2} \le 1$ and the condition (44), we have that

$$ \sum_{\ell_2=0}^{N-1} \rho^{2n}_{\ell_1\ell_2} \le f(N), \quad \forall \ell_1 \in \{0,\dots,N-1\} \ \text{ and } \ \forall n \ge 1. $$

Next we expand (40a) using the series (38):

$$ \begin{aligned} \mathrm{var}(\hat{m}_k) &\le \frac{[\Gamma(\alpha+1)]^2}{\alpha^2 N^2} \sum_{\ell_1=0}^{N-1}\sum_{\ell_2=0}^{N-1} \lambda^\alpha_{\ell_1}\lambda^\alpha_{\ell_2} \sum_{n=1}^{\infty}\left[\frac{(-\alpha)_n}{n!}\right]^2 \rho^{2n}_{\ell_1\ell_2} \\ &\le \frac{[\Gamma(\alpha+1)]^2\gamma_N^{2\alpha}}{\alpha^2 N^2} \sum_{n=1}^{\infty}\left[\frac{(-\alpha)_n}{n!}\right]^2 \sum_{\ell_1=0}^{N-1}\sum_{\ell_2=0}^{N-1} \rho^{2n}_{\ell_1\ell_2} \qquad (46a) \\ &\le \frac{[\Gamma(\alpha+1)]^2 M^{2\alpha}}{\alpha^2}\left[{}_2F_1(-\alpha,-\alpha;1;1)-1\right]\frac{f(N)}{N} \qquad (46b) \\ &= \text{const.} \times \frac{f(N)}{N} \to 0 \ \text{ by } \ (45), \end{aligned} $$

where,

- (46a) follows from (43) and interchanging the order of summation,
- and in (46b) the constant M is an upper bound for γ_N (which exists by Assumption 3) and we have utilized (44).

□

It is worth noting that the consistency property established in Theorem 1 and the proposition above does *not* require that the process y_t has a rational spectrum.

Example 2. Although the condition in (44) is not easy to check in practice, it seems to be a reasonable requirement for stationary stochastic processes generated by a stable rational shaping filter driven by white noise. Indeed, consider the ARMA (autoregressive moving-average) process $y_t = W(z)e_t$ where e_t is a normalized white Gaussian noise, the shaping filter is defined as

$$ W(z) = \frac{z^2 - z + 0.8}{z^2 - 1.6z + 0.81} = \sum_{t=0}^{\infty} w_t z^{-t}, $$

and w_t denotes the impulse response of the shaping filter. Then, it is not difficult to see that

$$ \begin{aligned} &\mathbb{E}[\bar{Y}_{\ell_1}\bar{Y}^*_{\ell_2}] \\ &\quad = \frac{1}{N}\sum_{t_1,t_2=0}^{N-1}\sum_{s=0}^{\min\{t_1,t_2\}} e^{-i\frac{2\pi}{N}(\ell_1 t_1 - \ell_2 t_2)}(w_{t_1-s}w_{t_2-s}). \end{aligned} $$

Based on the computation with this formula, Fig. 1 shows the value of the sum $\sum_{\ell_2=0}^{N-1}\rho^2_{\ell_1\ell_2}$ as a function of N with $\ell_1 = 3N/4$ (i.e., it corresponds to the angular frequency $\theta_1 = 3\pi/2$). A similar behavior has been found for different values of ℓ_1. This experiment suggests that the sum of the modulus squared correlation coefficients does not grow as N

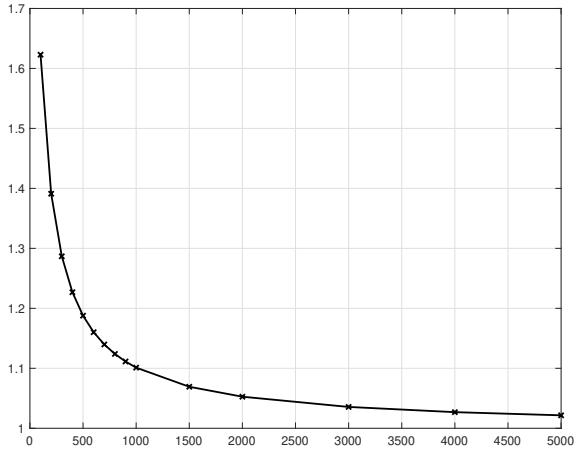

Figure 1: The partial sum of the correlation coefficients in (44) with $\ell_1 = 3N/4$ as a function of N. The crosses represent the values of the partial sum that have been computed.

approaches infinity and thus conditions (44) and (45) are satisfied. However, whether this is true or not for general ARMA processes seems difficult to prove via direct calculations.

Remark 4. It is worth pointing out that it seems hard to construct a counterexample in which Assumption 3 does not hold. The reason is the following. It is known, at least in the unidimensional case, that the values of the unwindowed periodogram $\hat{\Phi}$ at any two fixed neighboring frequencies θ_1 and θ_2 are *asymptotically uncorrelated*, see e.g., [23, Theorem. 6.2.3, p. 426] for the precise statement. Hence, $\hat{\Phi}_\ell$ and $\hat{\Phi}_k$ will become essentially uncorrelated if the sample size N is very large. After raising up to the α-th power, the same should be true for $\hat{\Phi}_\ell^\alpha$ and $\hat{\Phi}_k^\alpha$, and the computation in (32) should go through, although we do not have a proof for this point (and it seems rather nontrivial). In other words, our intuition is that Assumption 3 should hold even though the spectral components of the signal are in general correlated when the sample size N is relatively small making difficult the construction of a counterexample.

VI. Extension to Multidimensional Random Fields

We want to point out that the results obtained in Secs. IV and V transit smoothly to the multidimensional case. In order to see this, let $y_{\mathbf{t}}$ be a zero-mean stationary d-dimensional complex-valued random field where $\mathbf{t} = (t_1, \dots, t_d) \in \mathbb{Z}^d$. Given a vector $\mathbf{N} = (N_1, \dots, N_d) \in \mathbb{N}_+^d$, define a finite index set

$$\mathbb{Z}_{\mathbf{N}}^d := \{(t_1, \dots, t_d) \,:\, 0 \le t_j \le N_j - 1,\ j = 1, \dots, d\}, \tag{47}$$

which is just the integer grid of a d-dimensional box with a cardinality $|\mathbf{N}| := \prod_{j=1}^d N_j$. Suppose that we have obtained samples of the field $\{y_{\mathbf{t}} : \mathbf{t} \in \mathbb{Z}_{\mathbf{N}}^d\}$. The spectral domain now becomes $\mathbb{T}^d$ containing frequency vectors $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$. For the computation of the DFT, the multidimensional counterpart of (11) for the discretization of $\mathbb{T}^d$ is given as

$$\mathbb{T}_{\mathbf{N}}^d := \left\{ \left(\frac{2\pi}{N_1}\ell_1, \dots, \frac{2\pi}{N_d}\ell_d \right) \in \mathbb{T}^d : \boldsymbol{\ell} \in \mathbb{Z}_{\mathbf{N}}^d \right\}. \tag{48}$$

Obviously, the map $\boldsymbol{\ell} \mapsto \boldsymbol{\theta}_{\boldsymbol{\ell}} := (2\pi\ell_1/N_1, \dots, 2\pi\ell_d/N_d)$ is a bijection from $\mathbb{Z}_{\mathbf{N}}^d$ to $\mathbb{T}_{\mathbf{N}}^d$. The spectral components are given by the multidimensional DFT

$$Y_{\boldsymbol{\ell}} := Y(\boldsymbol{\theta}_{\boldsymbol{\ell}}) = \sum_{t \in \mathbb{Z}_{\mathbf{N}}^d} y_{\mathbf{t}}\, e^{-i\langle \mathbf{t}, \boldsymbol{\theta}_{\boldsymbol{\ell}} \rangle}, \quad \boldsymbol{\theta}_{\boldsymbol{\ell}} \in \mathbb{T}_{\mathbf{N}}^d \tag{49}$$

whose normalized version is $\bar{Y}_{\boldsymbol{\ell}} := \frac{1}{\sqrt{|\mathbf{N}|}} Y_{\boldsymbol{\ell}}$. The periodogram is still given by $\hat{\Phi}_{\boldsymbol{\ell}} = |\bar{Y}_{\boldsymbol{\ell}}|^2$. The multidimensional version of Theorem 1 is summarized below.

Proposition 2. *Assume that the normalized spectral components $\{\bar{Y}_{\boldsymbol{\ell}} : \boldsymbol{\ell} \in \mathbb{Z}_{\mathbf{N}}^d\}$ of the random field $\{y_{\mathbf{t}} : \mathbf{t} \in \mathbb{Z}_{\mathbf{N}}^d\}$ are independent zero-mean complex (circular) Gaussian random variables such that $\operatorname{var} \bar{Y}_{\boldsymbol{\ell}} = \mathbb{E}\,\hat{\Phi}_{\boldsymbol{\ell}} = \lambda_{\boldsymbol{\ell}} > 0$. Assume in addition that the underlying true spectral density $\Phi(\boldsymbol{\theta})$ of the field $y_{\mathbf{t}}$ satisfies some regularity properties as discussed in Remarks 2 and 3. Then, the estimator*

$$\hat{m}_{\mathbf{k}} = \begin{cases} \frac{1}{\alpha|\mathbf{N}|} \sum_{\boldsymbol{\ell} \in \mathbb{Z}_{\mathbf{N}}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta}_{\boldsymbol{\ell}} \rangle} \hat{\Phi}_{\boldsymbol{\ell}}^\alpha & \text{if } \mathbf{k} \neq \mathbf{0}, \\ \frac{1}{\alpha} \left(\frac{1}{|\mathbf{N}|} \sum_{\boldsymbol{\ell} \in \mathbb{Z}_{\mathbf{N}}^d} \hat{\Phi}_{\boldsymbol{\ell}}^\alpha - 1 \right) & \text{if } \mathbf{k} = \mathbf{0} \end{cases} \tag{50}$$

of the generalized cepstral coefficients

$$m_{\mathbf{k}} = \begin{cases} \frac{1}{\alpha} \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} \Phi(\boldsymbol{\theta})^\alpha \mathrm{d}\mu(\boldsymbol{\theta}) & \text{if } \mathbf{k} \neq \mathbf{0}, \\ \frac{1}{\alpha} \left[\int_{\mathbb{T}^d} \Phi(\boldsymbol{\theta})^\alpha \mathrm{d}\mu(\boldsymbol{\theta}) - 1 \right] & \text{if } \mathbf{k} = \mathbf{0} \end{cases} \tag{51}$$

is consistent up to a constant multiplicative factor. Here $\mathrm{d}\mu(\boldsymbol{\theta}) = \frac{1}{(2\pi)^d} \prod_{j=1}^d \mathrm{d}\theta_j$ is the normalized Lebesgue measure on $\mathbb{T}^d$. More precisely, we have for $\mathbf{k} \neq \mathbf{0}$ the convergence

$$C\hat{m}_{\mathbf{k}} \xrightarrow{m.s.} m_{\mathbf{k}}, \quad \text{and} \quad C\hat{m}_{\mathbf{0}} + (C-1)/\alpha \xrightarrow{m.s.} m_{\mathbf{0}} \tag{52}$$

as $\min(\mathbf{N}) \to \infty$, where the constant $C = 1/\Gamma(\alpha + 1)$.

Notice that the proof of the above proposition is almost identical to that in the unidimensional case (hence omitted), as the computation of the quantities such as (25) and (29) is left unchanged. The only difference in the multidimensional case is that we need to use multiple indices for the summation and write multivariable integrals. For example, (33) should be rewritten as

$$\frac{1}{|\mathbf{N}|} \sum_{\boldsymbol{\ell} \in \mathbb{Z}_{\mathbf{N}}^d} \lambda_{\boldsymbol{\ell}}^{2\alpha} \to \int_{\mathbb{T}^d} \Phi(\boldsymbol{\theta})^{2\alpha} \mathrm{d}\mu(\boldsymbol{\theta}). \tag{53}$$

As for the case with correlated spectral components, we can either use windowing techniques as explained in [23, Sec. 9.7] for the 2-d case and in [38], [39] for 3-d radar signal processing, or make an argument along the lines of Section V where conditions in Assumption 3 are now replaced by

$$\lim_{\min(\mathbf{N}) \to \infty} \gamma_{\mathbf{N}} < \infty, \quad \sum_{\boldsymbol{\ell}_2 \in \mathbb{Z}_{\mathbf{N}}^d} \rho_{\boldsymbol{\ell}_1 \boldsymbol{\ell}_2}^2 \le f(\mathbf{N}), \quad \forall \boldsymbol{\ell}_1 \in \mathbb{Z}_{\mathbf{N}}^d,$$

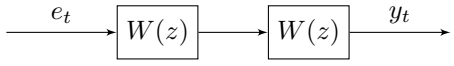

Figure 2: A cascade linear stochastic system with two identical subsystems.

where

$$\gamma_{\mathbf{N}} := \max_{\boldsymbol{\ell} \in \mathbb{Z}^d_{\mathbf{N}}} \lambda_{\boldsymbol{\ell}}, \qquad \lim_{\min(\mathbf{N}) \to \infty} f(\mathbf{N})/|\mathbf{N}| = 0.$$

Below we summarize the procedure in Matlab for computing the generalized cepstral estimates $\hat{m}_{\mathbf{k}}$ from a realization of the random field $y_{\mathbf{t}}$. The input of the procedure includes the parameter α, the index vector $\mathbf{k}$, and the finite samples $\{y_{\mathbf{t}} : \mathbf{t} \in \mathbb{Z}^d_{\mathbf{N}}\}$, while the output is the estimate $\hat{m}_{\mathbf{k}}$:

1) Evaluate the DFT multisequence $Y_{\boldsymbol{\ell}}$ in (49) using the `fftn` routine;
2) Take the modulus squared of $Y_{\boldsymbol{\ell}}$ and divide by $|\mathbf{N}|$ to obtain the periodogram $\hat{\Phi}_{\boldsymbol{\ell}}$;
3) Compute the estimate $\hat{m}_{\mathbf{k}}$ via (50) where the `ifftn` routine can be used on $\hat{\Phi}^{\alpha}_{\boldsymbol{\ell}}$, cf. e.g., [39];
4) Correct the estimate with the constant $C = 1/\Gamma(\alpha + 1)$ according to (52).

VII. Cascade System Identification

In this section, we show that our generalized cepstral estimator can be integrated into a spectral estimation procedure for the identification of cascade (possibly multidimensional) linear stochastic systems with *identical* subsystems, see Fig. 2 where $W(z)$ denotes the transfer function of each subsystem, y_t is the output, and e_t is the white noise input. In practice, of course the system need not be driven by a white noise but rather has an input u_t which can be modeled as a stationary stochastic process. The output of the cascade system then becomes (symbolically) $y_t = [W(z)]^2 u_t$. Assume that we are able to measure both y_t and u_t. At that point one can estimate the power spectral density of the input and thus also its shaping filter $\hat{W}_u$. After that, we can filter the output to obtain $\tilde{y}_t = [\hat{W}_u(z)]^{-1} y_t$ which approximately admits the model of Fig. 2. In this sense we have *whitened* the input process. Therefore, it is not restrictive to consider a white noise input in the cascade system above.

A cascade of identical systems is commonly used to model reactors in chemical engineering: each subsystem corresponds to a continuous stirred-tank reactor (CSTR) and the aim of the cascade is to increase the residence time distribution (RTD) of the reactor [40]. Moreover, the multidimensional setting allows to model the fact that influent and effluent concentration distributions are non-uniform, see for instance [41].

We assume that the number of subsystems which we call ν, is *a priori known* so that the parameter α is specialized as $1 - \frac{1}{\nu}$ (see also Example 1 in Section IV). The latter choice of the parameter turns out to correctly recover the number of subsystems. At the heart of our system identification approach lies a spectral estimation problem presented in [21], [25] which we shall briefly recall next.

Suppose that from some underlying random field we are given a number of covariances $\{c_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$ and generalized cepstral coefficients $\{m_{\mathbf{k}} : \mathbf{k} \in \Lambda_0\}$, and from these numbers we want to reconstruct the spectrum $\Phi(\boldsymbol{\theta})$, a nonnegative function in $L^1(\mathbb{T}^d)$. Here the index set Λ can be any finite set that contains the zero vector and is symmetric with respect to the origin, namely $\mathbf{k} \in \Lambda \implies -\mathbf{k} \in \Lambda$. The other index set $\Lambda_0 := \Lambda \backslash \{\mathbf{0}\}$ so $m_{\mathbf{0}}$ is excluded for technical reasons. Inspired by the famous *Maximum Entropy* method [42], we set up the following general optimization problem:

$$\max_{\Phi \geq 0} \quad \mathbb{H}_\nu(\Phi) := \frac{\nu^2}{\nu - 1} \left(\int_{\mathbb{T}^d} \Phi(\boldsymbol{\theta})^{\frac{\nu-1}{\nu}} \mathrm{d}\mu(\boldsymbol{\theta}) - 1 \right) \tag{54a}$$

$$\text{s.t.} \quad c_{\mathbf{k}} = \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} \Phi(\boldsymbol{\theta}) \mathrm{d}\mu(\boldsymbol{\theta}) \ \ \forall \mathbf{k} \in \Lambda, \tag{54b}$$

$$m_{\mathbf{k}} = \frac{\nu}{\nu - 1} \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} \Phi(\boldsymbol{\theta})^{\frac{\nu-1}{\nu}} \mathrm{d}\mu(\boldsymbol{\theta}) \ \ \forall \mathbf{k} \in \Lambda_0, \tag{54c}$$

where the objective functional $\mathbb{H}_\nu(\Phi)$ is a generalized version of the entropy of Φ derived from the α-divergence [43], [44]. The above formulation can be viewed as a considerable generalization of the simultaneous covariance-cepstral extension theory in [20], [45] as their problem can be recovered from ours by letting $\nu = 1$ (understood in a suitable limit sense, see [21, Sec. 3]). The optimization problem in (54) is infinite-dimensional, and it is usually more convenient to work with its finite-dimensional dual problem. Following the analysis in [21], a suitable form of *regularization* is needed in order to promote a unique positive solution to the dual problem, so that the resulting primal solution Φ is a strictly positive spectral density which is desired in many practical applications. More specifically, the regularized dual optimization problem is formulated as:

$$\begin{aligned} &\min_{\mathbf{p},\, \mathbf{q}} \ J_{\nu,\lambda}(\mathbf{p}, \mathbf{q}) := J_\nu(\mathbf{p}, \mathbf{q}) + \frac{\lambda}{\nu - 1} \int_{\mathbb{T}^d} \frac{1}{P(\boldsymbol{\theta})^{\nu-1}} \mathrm{d}\mu(\boldsymbol{\theta}) \\ &\text{s.t. } P(\boldsymbol{\theta}) \geq 0, \quad Q(\boldsymbol{\theta}) \geq 0, \quad \forall \boldsymbol{\theta} \in \mathbb{T}^d \end{aligned} \tag{55}$$

where, the dual variables (Lagrange multipliers) $\{q_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$ and $\{p_{\mathbf{k}} : \mathbf{k} \in \Lambda_0\}$ correspond to the covariance and cepstral constraints (54b) and (54c), respectively, $P(\boldsymbol{\theta}) := 1 + \sum_{\mathbf{k} \in \Lambda_0} p_{\mathbf{k}} e^{-i\langle \mathbf{k}, \boldsymbol{\theta} \rangle}$ and $Q(\boldsymbol{\theta}) := \sum_{\mathbf{k} \in \Lambda} q_{\mathbf{k}} e^{-i\langle \mathbf{k}, \boldsymbol{\theta} \rangle}$ are trigonometric polynomials, and $\lambda > 0$ is the regularization parameter. The first term in the dual objective function $J_{\nu,\lambda}(\mathbf{p}, \mathbf{q})$ is the unregularized dual function

$$J_\nu(\mathbf{p}, \mathbf{q}) := \frac{1}{\nu - 1} \int_{\mathbb{T}^d} \frac{P(\boldsymbol{\theta})^\nu}{Q(\boldsymbol{\theta})^{\nu-1}} \mathrm{d}\mu(\boldsymbol{\theta}) + \langle \mathbf{q}, \mathbf{c} \rangle - \langle \mathbf{p}, \mathbf{m} \rangle, \tag{56}$$

where the real-valued inner product $\langle \mathbf{q}, \mathbf{c} \rangle := \sum_{\mathbf{k} \in \Lambda} q_{\mathbf{k}} c^*_{\mathbf{k}}$.

Under a suitable feasibility assumption for the covariance (multi)sequence $\{c_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$ (see [21, Assumption 5.1]), if the integer parameter ν, the number of subsystems in the cascade, satisfies the condition $\nu \geq d/2 + 1$, then the regularized dual problem (55) admits a unique solution $(\hat{\mathbf{p}}, \hat{\mathbf{q}})$ such that the corresponding polynomials $\hat{P}$ and $\hat{Q}$ are both strictly positive. Moreover, the positive rational function

$$\hat{\Phi}_\nu(\boldsymbol{\theta}) = [\hat{P}(\boldsymbol{\theta}) / \hat{Q}(\boldsymbol{\theta})]^\nu \tag{57}$$

matches the given covariances $\{c_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$ exactly and the generalized cepstral coefficients $\{m_{\mathbf{k}} : \mathbf{k} \in \Lambda_0\}$ approxi-

mately. Clearly we have set $p_{\mathbf{0}} = 1$ in order to avoid trivial cancellations between P and Q in the rational function Φ_ν.

Remark 5 (On the regularity condition $\nu \geq d/2 + 1$). The condition $\nu \geq d/2 + 1$ was proposed in [21, Sec. 5] and a similar condition appeared in [46] for a multidimensional and multivariate spectral estimation problem. Such a condition is crucial in guaranteeing that the dual problem (55) admits an *interior-point* solution which gives a rational spectral density as an approximate solution to the primal problem (54). This point is highly nontrivial in multidimensional problems since in general a solution to the primal problem is a *spectral measure* which contains[3] a density function and a singular part like the Dirac impulse [45]. Moreover, the singular part may not be uniquely determined which creates difficulty in engineering applications. The strength of the regularity condition $\nu \geq d/2 + 1$ is precisely to exclude this pathology and to promote well-posedness of the dual problem, see [21, Sec. 6]. In addition, as shown in the recent paper [26], the condition can be further weakened as $\nu \geq d/2$ which agrees better with unidimensional results [20] since in this way $\nu = 1$ is sufficient to handle the case $d = 1$.

From the perspective of numerical computation, it is better to *discretize* the problems (55) on a grid similar to (48) but of a different size $\mathbf{K} = (K_1, \ldots, K_d) \in \mathbb{N}_+^d$. The formulas will be similar to the ones above and are omitted here. For details see [21, Section 8] where it is also explained that the solution to the discrete problem is a reasonable approximation of the solution to the corresponding continuous problem when the grid size $\mathbf{K}$ is sufficiently large (componentwise).

To summarize, our system identification procedure can be divided into four steps:

- **(i)** Feed the system with normalized white noise $e_{\mathbf{t}}$ and collect the output random field $\{y_{\mathbf{t}} : \mathbf{t} \in \mathbb{Z}_{\mathbf{N}}^d\}$;
- **(ii)** Choose an index set Λ and estimate $\{c_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$ via (12) adapted to multidimensional random fields (see e.g., [48]) and $\{m_{\mathbf{k}} : \mathbf{k} \in \Lambda_0\}$ via the procedure outlined at the end of Sec. VI;
- **(iii)** Fix a grid size $\mathbf{K}$ and solve the discrete version of the regularized dual optimization problem (55) given $c_{\mathbf{k}}$'s and $m_{\mathbf{k}}$'s in Step (ii) and a regularization parameter $\lambda > 0$;
- **(iv)** Factor the optimal spectrum $\hat{\Phi}(\boldsymbol{\theta})$ in order to obtain a transfer function $\hat{W}(\mathbf{z})$ where $\mathbf{z} := (z_1, \ldots, z_d)$.

We must point out that there is a significant technical difficulty in the last step of the procedure concerning *spectral factorization*. Since our optimal spectrum is a *rational* function, we end up factoring positive Laurent trigonometric polynomials of several variables into *one square*, which is in general an impossible task.[4] Fortunately in the 2-d case, Theorems 1.1.1 & 1.1.3 in [50] have given a sufficient and necessary condition to check such factorability and an explicit formula to compute the factor when the factorization is possible. The nontrivial part of the condition states that a certain submatrix of the full covariance matrix should have a specific low rank.

In addition, we want to stress that the regularization employed in (55) has the power of promoting strict convexity[5] and well-posedness of the dual problem [21]. Of course, the regularization term does not appear naturally in the system identification problem. Thus in order to get a reasonable consistency result for the identified model, we should take the regularization parameter λ small as it directly controls the error in generalized cepstral matching. The next proposition expresses the idea.

Proposition 3. *Assume that the model class is correctly specified, i.e., the unknown system transfer function $W(\mathbf{z})$ is such that true spectrum of the output random field $y_{\mathbf{t}}$ (when the input $e_{\mathbf{t}}$ is a normalized white noise) falls within the class of rational functions having the form* (57)*. Then as the number of samples $\min(\mathbf{N}) \to \infty$ and the regularization parameter $\lambda \to 0$, the optimal solution to* (55) *recovers the true spectrum of $y_{\mathbf{t}}$ as well as the correct system parameters in $W(\mathbf{z})$ after spectral factorization.*

Proof. The proposition is a direct consequence of Theorem 6.1 in [21]. □

In plain words the above result says: given a d-dimensional system, which is a cascade of $\nu \geq d/2 + 1$ identical subsystems, it is possible to construct a consistent estimator of it using some second-order statistics of the output process.

VIII. Simulation Results

In this section, we present simulation results in 1-d and 2-d cascade system identification because theoretical results and algorithms for polynomial spectral factorization are available only when the dimension is such that $d \leq 2$, as highlighted in the previous section.

A. Identification of 1-D dynamical systems

Consider a 1-d linear time-invariant (LTI) system described by the transfer function $W(z)$. Furthermore, let W have the cascade structure that corresponds to our optimal spectrum (57), namely $W(z) = [W_1(z)]^\nu$ where each subsystem

$$W_1(z) = \frac{b(z)}{a(z)} = \frac{\sum_{k=0}^n b_k z^{-k}}{\sum_{k=0}^n a_k z^{-k}}. \tag{58}$$

Here the index set is $\Lambda_+ = \{0, 1, \ldots, n\}$ and we take $n = 2$ which corresponds to a second-order system for simplicity.

Let us write $a(\theta) \equiv a(e^{i\theta})$ for the value of the polynomial $a(z)$ on the unit circle. If the white noise input e_t has unit variance, then the spectral density of the output process y_t is $\Phi(\theta) = [P(\theta)/Q(\theta)]^\nu$ where $P(\theta) = |b(\theta)|^2$, $Q(\theta) = |a(\theta)|^2$. We shall take the integer $\nu = 3$ which is known by our identification procedure. Moreover, we shall parametrize the polynomials in terms of their roots, that is, $a(z) = (1 - r_{a,1}z^{-1})(1 - r_{a,2}z^{-1})$ and $b(z) = b_0(1 - r_{b,1}z^{-1})(1 - r_{b,2}z^{-1})$ with b_0 a normalization constant such

[3] See Lebesgue's decomposition theorem in e.g., [47] for the precise meaning.

[4] It has been shown that *sum-of-squares* factorization is always possible for multivariate Laurent polynomials that are strictly positive on the multi-torus [49]. However, the factors in general have degrees larger than the original Laurent polynomial.

[5] The unregularized dual function (56) is only convex but not strictly convex [21].

that $P(z) = b(z)b(z^{-1})$ has $p_0 = 1$. The true system parameters are specified as $r_{a,1}, r_{a,2} = 0.5e^{\pm i\pi/3}$ (complex conjugate roots) and $r_{b,1}, r_{b,2} = -0.8$ and 0.6 (real roots), so the system W_1 (also W) has *real* parameters and is clearly stable and minimum-phase. This in turn gives

$$\begin{aligned} \mathbf{a} &= \begin{bmatrix} a_0 & a_1 & a_2 \end{bmatrix} = \begin{bmatrix} 1 & -0.5 & 0.25 \end{bmatrix} \quad \text{and} \\ \mathbf{b} &= \begin{bmatrix} b_0 & b_1 & b_2 \end{bmatrix} = \begin{bmatrix} 0.8872 & 0.1774 & -0.4259 \end{bmatrix}. \end{aligned} \tag{59}$$

Next, we follow the system identification procedure listed in the previous section, and demonstrate how to estimate the system parameters $\mathbf{a}$ and $\mathbf{b}$ from the samples of the output process y_t. Such samples of size N are generated by feeding the normalized *real* white noise e_t to the true system (so y_t is also real). Then the covariances and generalized cepstral coefficients of the output y_t with the indices in the set $\Lambda = \{k \in \mathbb{Z} : -2 \le k \le 2\}$ are estimated using (12) and (17). Due to the symmetry, we only need to compute

$$\hat{\mathbf{c}} = \begin{bmatrix} \hat{c}_0 & \hat{c}_1 & \hat{c}_2 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{m}} = \begin{bmatrix} \hat{m}_1 & \hat{m}_2 \end{bmatrix}.$$

In particular, the constant C in Theorem 1 for correcting the generalized cepstral estimation is equal to $1/\Gamma(5/3) \approx 1.1077$ with $\alpha = (\nu - 1)/\nu = 2/3$. The estimation errors are defined as $\|[\hat{\mathbf{c}}, \hat{\mathbf{m}}] - [\mathbf{c}, \mathbf{m}]\|$ and $\|\hat{\mathbf{m}} - \mathbf{m}\|$ where the true $\mathbf{c}$ and $\mathbf{m}$ are evaluated using the true spectrum. Such errors against different sample sizes N are shown in the orange and yellow lines[6] of Fig. 3a in the double-logarithmic scale (base 10). At the same time, Fig. 3c indicates the errors when $\hat{\mathbf{m}}$ is not corrected by the constant C. To be more clear, one time series y_t of length $N_{\max} = 10^4$ is generated first and simulations are carried out using the same series truncated to different lengths N. One can see the general trend of the lines in Fig. 3a that the error reduces as the sample size grows. The unique exception at $N = 500$ can be explained as random fluctuation before the convergence of the estimators. In this particular example, the uncorrected generalized cepstral estimator begins to produce a large error (relative to that of the corrected version) after $N = 2500$, see Fig. 3c. Such an erratic behavior will be more apparent in the 2-d example presented in the next subsection.

Then the regularized dual problem (55) is solved on a discrete grid of size $K = N$ with the estimated $\hat{\mathbf{c}}$, $\hat{\mathbf{m}}$, and $\lambda = 10^{-6}$ using a gradient descent algorithm. The algorithm is initialized at $q_0 = 1$ and the rest variables equal to 0 which corresponds to constant polynomials $P(\theta) = Q(\theta) \equiv 1$. The iterations terminate when the norm of the gradient is less than 10^{-6}. Given the optimal polynomials $\hat{P}$ and $\hat{Q}$, we proceed to compute their factors $\hat{a}$ and $\hat{b}$ using standard methods for (1-d) polynomial spectral factorization, e.g., the Bauer method [51]. After a comparison with (59), we compute the error $\|[\hat{\mathbf{a}}, \hat{\mathbf{b}}] - [\mathbf{a}, \mathbf{b}]\|$ which is shown in the blue line of Fig. 3a for different values of N: it shares the same trend with the other two lines in the same figure as the sample size N increases, which is an expected result since the true system $W(z)$ belongs to the model class dictated by the solution form (57) of our optimization problem, as explained in Proposition 3. Furthermore, in Fig. 3b we plot the true spectrum $\Phi(\theta)$ on the interval $[0, \pi)$ against the estimated spectra $\hat{\Phi}(\theta)$. The convergence of the estimated spectra to the true one is clearly observed. In addition, Fig. 3d shows the estimated spectra which are computed using the uncorrected generalized cepstral estimates, and they obviously deviate from the true spectrum even when the sample size is large.

[6]The lines are drawn to facilitate observation and has no meaning here.

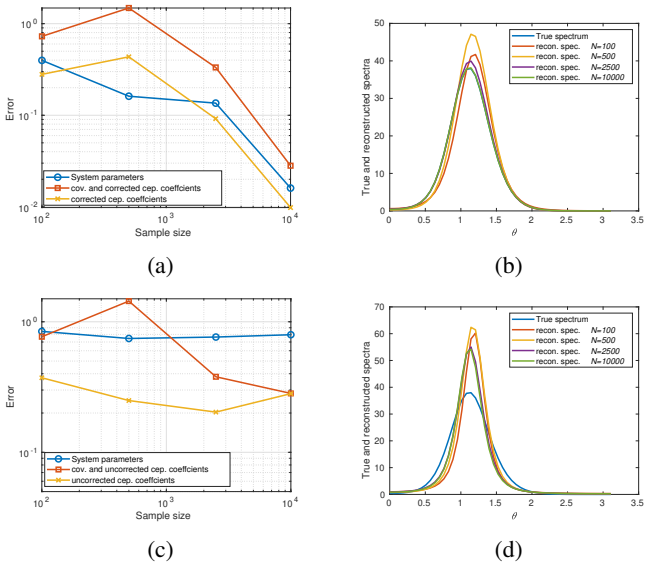

Figure 3: Simulation results in the 1-d case: *Upper and lower left.* Estimation errors of the covariances and the generalized cepstral coefficients (yellow and orange lines) and the system parameters (blue lines) when the sample size N of the random process is equal to $100, 500, 2500, 10^4$. Subfigs. 3a and 3c show the results for corrected and uncorrected estimates of the generalized cepstral coefficients, respectively. *Upper and lower right.* The true spectrum (blue lines) $\Phi(\theta)$ of y_t versus the estimated spectra $\hat{\Phi}(\theta)$ via solving the discrete version of the regularized dual optimization problem (55) with $\lambda = 10^{-6}$ and the sample sizes indicated above. Only the segment corresponding to $\theta \in [0, \pi)$ is shown due to the symmetry of the spectra of real processes. Subfigs. 3b and 3d show the results for corrected and uncorrected estimates of the generalized cepstral coefficients, respectively. Notice that in Subfigs. 3b the estimated spectrum with $N = 10^4$ virtually coincide with the true one.

Remark 6. In practice, the order of the model ($n = 2$ in the previous example) is usually unknown and can be selected from data using classic techniques for system identification, see [1, Appendix C] and [52]; for instance, we can estimate the spectrum using different orders and then score the estimated model according to the BIC criterion (which is equivalent to selecting the best order). Moreover, we have done an additional set of simulations using the same time series as above to see what happens to the spectrum estimate if the model order is mis-specified or more precisely, overestimated. Empirical evidences suggest that our method is robust against mis-specification of the model order because the spectra with overestimated orders stay close to the true spectrum as measured by the differences of function values.

B. Modeling 2-D random fields

In this subsection, we consider the problem of identifying a 2-d LTI system $W(z_1, z_2)$ from samples of a planar random process. The procedure will be similar to the previous subsection which we will describe briefly. Let $W(\mathbf{z}) = [W_1(\mathbf{z})]^\nu$ with $\nu = 2$ and the subsystem

$$W_1(\mathbf{z}) = \frac{b(\mathbf{z})}{a(\mathbf{z})} = \frac{\sum_{\mathbf{k}\in\Lambda_+} b_{\mathbf{k}}\mathbf{z}^{-\mathbf{k}}}{\sum_{\mathbf{k}\in\Lambda_+} a_{\mathbf{k}}\mathbf{z}^{-\mathbf{k}}}, \tag{60}$$

where the index set $\Lambda_+ := \{(k_1, k_2) \in \mathbb{Z}^2 : 0 \le k_1, k_2 \le 1\}$ so that W_1 (roughly) has order one in each dimension, the indeterminates (z_1, z_2) are abbreviated as $\mathbf{z}$, and $\mathbf{z}^{\mathbf{k}}$ stands for $z_1^{k_1} z_2^{k_2}$. For notational convenience, the value of the polynomial $a(\mathbf{z})$ on the unit torus is written as $a(\boldsymbol{\theta}) \equiv a(e^{i\theta_1}, e^{i\theta_2})$. For simplicity, we also impose a separable form $(1 - r_{a,1}z_1^{-1})(1 - r_{a,2}z_2^{-1})$ on the polynomials a, b, and take $|r_{a,j}| < 1$ for $j = 1, 2$. The system parameters can be assigned via

$$\begin{bmatrix} a_{0,0} & a_{0,1} & a_{1,0} & a_{1,1} \end{bmatrix} = \begin{bmatrix} 1 & -r_{a,2} & -r_{a,1} & r_{a,1}r_{a,2} \end{bmatrix}.$$

It is convenient to collect the 2-d system parameters into matrices. In our particular example, we take real parameters

$$A = \begin{bmatrix} 1 & -0.7 \\ -0.5 & 0.35 \end{bmatrix}, \quad B = \begin{bmatrix} 0.6696 & -0.5357 \\ -0.4018 & 0.3214 \end{bmatrix} \tag{61}$$

where $a_{k_1,k_2} = A(k_1+1, k_2+1)$ and similar for b_{k_1,k_2}. Notice that the constraint $p_{\mathbf{0}} = 1$ translates into a normalization condition $\|B\|_{\mathrm{F}} = 1$ where the subscript $_{\mathrm{F}}$ denotes the Frobenius norm. The true spectrum $\Phi(\boldsymbol{\theta})$ defined on the 2-d domain is shown in Fig. 4b.

Next we implement the system identification procedure in order to estimate the "system matrices" A and B from the samples of the output process $y_{\mathbf{t}}$ with a size $\mathbf{N} = (N_1, N_2)$. The covariances and generalized cepstral coefficients of $y_{\mathbf{t}}$ indexed by the set $\Lambda = \{(k_1, k_2) \in \mathbb{Z}^2 : -1 \le k_1, k_2 \le 1\}$ are estimated using the unwindowed periodogram as explained in Section VI. Due to the symmetry of $c_{\mathbf{k}}$ and $m_{\mathbf{k}}$ with respect to the origin, we only need to compute

$$\hat{\mathbf{c}} = \begin{bmatrix} \hat{c}_{0,0} & \hat{c}_{0,1} & \hat{c}_{1,-1} & \hat{c}_{1,0} & \hat{c}_{1,1} \end{bmatrix},$$
$$\hat{\mathbf{m}} = \begin{bmatrix} \hat{m}_{0,1} & \hat{m}_{1,-1} & \hat{m}_{1,0} & \hat{m}_{1,1} \end{bmatrix}$$

which are put in the lexicographic ordering. In particular, the constant of correction C in (52) is now $1/\Gamma(3/2) = 2/\sqrt{\pi} \approx 1.1284$ since $\alpha = 1 - 1/\nu = 1/2$. The estimation errors $\|[\hat{\mathbf{c}}, \hat{\mathbf{m}}] - [\mathbf{c}, \mathbf{m}]\|$ and $\|\hat{\mathbf{m}} - \mathbf{m}\|$ versus the sample size $N_1 = N_2 = N$ are plotted in the orange and purple lines, respectively, of Fig. 4a which share the same trend as that in Fig. 3a: the errors go down as the sample size increases. The yellow and green lines corresponding to the uncorrected generalized cepstral estimator, in contrast, deviate from the orange and purple lines, respectively, as early as $N = 100$ in this example. Obviously, the uncorrected estimator does not enjoy the consistency property since the error curves stop going down as N further increases. In particular, a comparison between the yellow and green lines reveals the fact that the error in the generalized cepstral estimate dominates when the sample size is large.

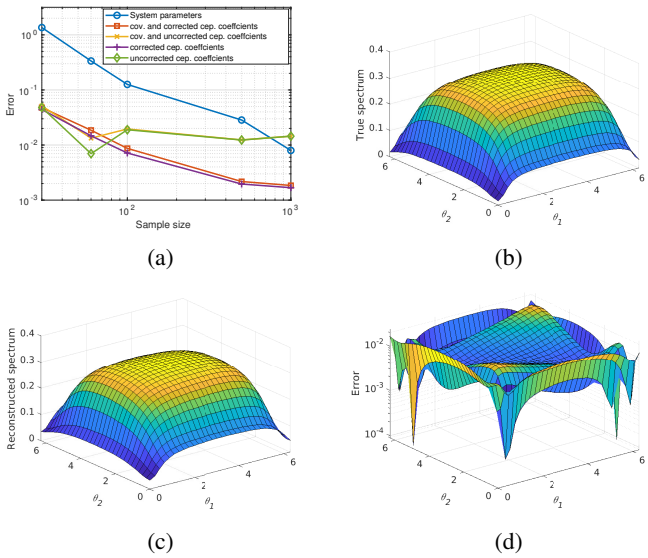

Figure 4: Simulation results in the 2-d case: *Upper left.* Estimation errors of the covariances and the generalized cepstral coefficients (the orange and purple lines with correction and the yellow and green lines without correction) and the system parameters (blue line) when the sample size $N_1 = N_2 = N$ of the random field is equal to $30, 60, 100, 500, 1000$. *Upper right.* The true spectrum $\Phi(\boldsymbol{\theta})$ of $y_{\mathbf{t}}$. *Lower left.* The estimated spectrum $\hat{\Phi}(\boldsymbol{\theta})$ via solving the discrete version of the regularized dual optimization problem (55) with $\lambda = 10^{-6}$ and the sample size $N = 100$. *Lower right.* The pointwise absolute error $|\hat{\Phi}(\boldsymbol{\theta}) - \Phi(\boldsymbol{\theta})|$ in a logarithmic scale between Subfigs. 4b and 4c.

Then we solve the regularized dual problem (55) on a discrete grid of *fixed* size $\mathbf{K} = (30, 30)$ with the estimated $\hat{\mathbf{c}}$, $\hat{\mathbf{m}}$, and $\lambda = 10^{-6}$. The optimal spectrum $\hat{\Phi}(\boldsymbol{\theta})$ returned by the solver is shown in Fig. 4c when the sample size is $N_1 = N_2 = 100$. In this case, clearly the shape of the estimated spectrum is almost visually indistinguishable from that of the true spectrum. Moreover, the pointwise absolute error $|\hat{\Phi}(\boldsymbol{\theta}) - \Phi(\boldsymbol{\theta})|$ is shown in Fig. 4d where the z-axis is in a logarithmic scale. One can see that the largest absolute error is roughly 0.02 which means that the error plot is in fact quite flat. As a complement, we also compute the cumulative relative error on the whole grid $\|\hat{\Phi} - \Phi\|_{\mathrm{F}}/\|\Phi\|_{\mathrm{F}} = 4.40\%$ which is reasonably small.

Once we have the optimal polynomials $\hat{P}$ and $\hat{Q}$, we can run the factorization algorithm in [50, Theorems 1.1.1 & 1.1.3] and compute factors $\hat{a}$ and $\hat{b}$. The numerical values of their coefficients in this particular instance with a sample size $N = 100$ are reported in the following matrices

$$\hat{A} = \begin{bmatrix} 1.0590 & -0.6896 \\ -0.4503 & 0.2854 \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} 0.7102 & -0.5282 \\ -0.3679 & 0.2688 \end{bmatrix}.$$

The error here is $\|[\hat{A}, \hat{B}] - [A, B]\|_{\mathrm{F}} = 0.1259$ which, along with other errors for different N, is depicted in the blue line of Fig. 4a. We can see the consistency of our identification procedure as reported in Proposition 3.

IX. Conclusions

In the first part of the paper we faced the problem of generalized cepstral estimation. More precisely, we have constructed an estimator based on the periodogram and showed its statistical consistency up to a multiplicative factor. We did that both for the case of uncorrelated and correlated spectral components. These findings represent a generalization of the ones presented in [17], [18] for cepstral estimation. Moreover, in the case of correlated spectral components the consistency is guaranteed under Assumption 3 and we showed through a numerical example its reasonableness (i.e., it holds empirically when the stochastic process is generated by a stable rational filter). In the second part of the paper we have considered a cascade system identification problem. More precisely, we have dealt with a special class of cascade systems (with identical subsystems) and shown the consistency of the multidimensional spectral estimator proposed in [21] when equipped with the generalized cepstral estimator developed in the first part. The consistency of the corresponding identification procedure is then a direct consequence once spectral factorization is possible.

References

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

[2] A. Lindquist and G. Picci, *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*, ser. Series in Contemporary Mathematics. Springer-Verlag Berlin Heidelberg, 2015, vol. 1.

[3] J. Li, F. Chen, Y. Wang, P. Charge, F. Ji, and H. Yu, "Spatial spectrum estimation of incoherently distributed sources based on low-rank matrix recovery," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6333–6347, 2020.

[4] J. Ramos, "A subspace algorithm for identifying 2-D separable in denominator filters," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 41, no. 1, pp. 63–67, 1994.

[5] C. I. Byrnes, S. V. Gusev, and A. Lindquist, "A convex optimization approach to the rational covariance extension problem," *SIAM Journal on Control Optimization*, vol. 37, no. 1, pp. 211–229, 1998.

[6] ——, "From finite covariance windows to modeling filters: A convex optimization approach," *SIAM Review*, vol. 43, no. 4, pp. 645–675, 2001.

[7] T. T. Georgiou, "Spectral estimation via selective harmonic amplification," *IEEE Transactions on Automatic Control*, vol. 46, no. 1, pp. 29–42, 2001.

[8] C. Byrnes, P. Enqvist, and A. Lindquist, "Cepstral coefficients, covariance lags, and pole-zero models for finite data strings," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 677–693, 2001.

[9] A. Ferrante, M. Pavon, and F. Ramponi, "Hellinger versus Kullback–Leibler multivariable spectrum approximation," *IEEE Transactions on Automatic Control*, vol. 53, no. 4, pp. 954–967, 2008.

[10] A. Ferrante, F. Ramponi, and F. Ticozzi, "On the convergence of an efficient algorithm for Kullback–Leibler approximation of spectral densities," *IEEE Transactions on Automatic Control*, vol. 56, no. 3, pp. 506–515, 2011.

[11] F. Ramponi, A. Ferrante, and M. Pavon, "A globally convergent matricial algorithm for multivariate spectral estimation," *IEEE Transactions on Automatic Control*, vol. 54, no. 10, pp. 2376–2388, 2009.

[12] A. Ringh, J. Karlsson, and A. Lindquist, "Multidimensional rational covariance extension with approximate covariance matching," *SIAM Journal on Control and Optimization*, vol. 56, no. 2, pp. 913–944, 2018.

[13] B. Zhu, "On the well-posedness of a parametric spectral estimation problem and its numerical solution," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 1089–1099, 2020.

[14] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2561–2575, 2012.

[15] M. Zorzi, "A new family of high-resolution multivariate spectral estimators," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 892–904, 2014.

[16] B. Zhu and G. Baggio, "On the existence of a solution to a spectral estimation problem *à la* Byrnes-Georgiou-Lindquist," *IEEE Transactions on Automatic Control*, vol. 64, no. 2, pp. 820–825, 2019.

[17] Y. Ephraim and M. Rahim, "On second-order statistics and linear estimation of cepstral coefficients," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 162–176, 1999.

[18] Y. Ephraim and W. J. Roberts, "On second-order statistics of log-periodogram with correlated components," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 625–628, 2005.

[19] F. Alías, J. C. Socoró, and X. Sevillano, "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds," *Applied Sciences*, vol. 6, no. 5, p. 143, 2016.

[20] P. Enqvist, "A convex optimization approach to ARMA(n, m) model design from covariance and cepstral data," *SIAM Journal on Control and Optimization*, vol. 43, no. 3, pp. 1011–1036, 2004.

[21] B. Zhu and M. Zorzi, "A well-posed multidimensional rational covariance and generalized cepstral extension problem," *SIAM Journal on Control and Optimization*, vol. 61, no. 3, pp. 1532–1556, 2023.

[22] F. Ramponi, A. Ferrante, and M. Pavon, "On the well-posedness of multivariate spectrum approximation and convergence of high-resolution spectral estimators," *Systems & Control Letters*, vol. 59, no. 3, pp. 167–172, 2010.

[23] M. B. Priestley, *Spectral Analysis and Time Series (Two-Volume Set)*, ser. Probability and Mathematical Statistics. Elsevier Academic Press, 1981, reprinted in 2004.

[24] J. Karlsson, A. Lindquist, and A. Ringh, "The multidimensional moment problem with complexity constraint," *Integral Equations and Operator Theory*, vol. 84, no. 3, pp. 395–418, 2016.

[25] B. Zhu and M. Zorzi, "A generalized multidimensional circulant rational covariance and cepstral extension problem," in *IFAC PapersOnLine: Proceedings of the 19th IFAC Symposium on System Identification (SYSID 2021)*, vol. 54, no. 7. Padova, Italy: IFAC, 2021, pp. 553–558.

[26] ——, "On a weaker regularity condition for a multidimensional spectral estimation problem," *IEEE Control Systems Letters*, vol. 7, pp. 1795–1800, 2023.

[27] K. Tokuda, T. Kobayashi, and S. Imai, "Generalized cepstral analysis of speech-unified approach to LPC and cepstral method," in *First International Conference on Spoken Language Processing*, 1990.

[28] L. Falconi, A. Ferrante, and M. Zorzi, "Mean-square consistency of the f-truncated M^2-periodogram," *Automatica*, vol. 147, 2023.

[29] B. Wahlberg, H. Hjalmarsson, and J. Mårtensson, "Variance results for identification of cascade systems," *Automatica*, vol. 45, no. 6, pp. 1443–1448, 2009.

[30] H. Sandberg, P. Hägg, and B. Wahlberg, "Approximative model reconstruction of cascade systems," *Systems & Control Letters*, vol. 69, pp. 90–97, 2014.

[31] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[32] A. Lindquist and G. Picci, "The circulant rational covariance extension problem: The complete solution," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2848–2861, 2013.

[33] A. W. Van der Vaart, *Asymptotic Statistics*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000, vol. 3.

[34] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 8th ed. Academic Press, 2014.

[35] A. Villani, "Another note on the inclusion $L^p(\mu) \subset L^q(\mu)$," *The American Mathematical Monthly*, vol. 92, no. 7, pp. 485–487, 1985.

[36] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed., ser. National Bureau of Standards Applied Mathematics Series. U.S. Government Printing Office, 1972, vol. 55, reprinted by Dover Publications in 2020 with corrections.

[37] R. B. Davies, "Algorithm AS155: The distribution of a linear combination of χ^2 random variables," *Applied Statistics*, pp. 323–333, 1980.

[38] F. Engels, P. Heidenreich, A. M. Zoubir, F. K. Jondral, and M. Wintermantel, "Advances in automotive radar: A framework on computationally efficient high-resolution frequency estimation," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 36–46, 2017.

[39] B. Zhu, A. Ferrante, J. Karlsson, and M. Zorzi, "Fusion of sensors data in automotive radar systems: A spectral estimation approach," in *58th*

IEEE Conference on Decision and Control (CDC 2019). IEEE, 2019, pp. 5088–5093.
[40] A. Renken, M. N. Kashid, and L. Kiwi-Minsker, *Microstructured Devices for Chemical Processing*. John Wiley & Sons, 2014.
[41] V. Vavilin, L. Lokshina, X. Flotats, and I. Angelidaki, "Anaerobic digestion of solid material: Multidimensional modeling of continuous-flow reactor with non-uniform influent concentration distributions," *Biotechnology and Bioengineering*, vol. 97, no. 2, pp. 354–366, 2007.
[42] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Department of Geophysics, Stanford University, 1975.
[43] M. Zorzi, "Rational approximations of spectral densities based on the Alpha divergence," *Mathematics of Control, Signals, and Systems*, vol. 26, no. 2, pp. 259–278, 2014.
[44] ——, "An interpretation of the dual problem of the THREE-like approaches," *Automatica*, vol. 62, pp. 87–92, 2015.
[45] A. Ringh, J. Karlsson, and A. Lindquist, "Multidimensional rational covariance extension with applications to spectral estimation and image compression," *SIAM Journal on Control and Optimization*, vol. 54, no. 4, pp. 1950–1982, 2016.
[46] B. Zhu, A. Ferrante, J. Karlsson, and M. Zorzi, "M^2-spectral estimation: A flexible approach ensuring rational solutions," *SIAM J. Control Optimization*, vol. 59, no. 4, pp. 2977–2996, 2021.
[47] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill Book Company, 1987.
[48] B. Zhu, A. Ferrante, J. Karlsson, and M. Zorzi, "M^2-spectral estimation: A relative entropy approach," *Automatica*, vol. 125, 2021.
[49] M. A. Dritschel, "On factorization of trigonometric polynomials," *Integral Equations and Operator Theory*, vol. 49, no. 1, pp. 11–42, 2004.
[50] J. S. Geronimo and H. J. Woerdeman, "Positive extensions, Fejér-Riesz factorization and autoregressive filters in two variables," *Annals of Mathematics*, vol. 160, no. 3, pp. 839–906, 2004.
[51] A. H. Sayed and T. Kailath, "A survey of spectral factorization methods," *Numerical Linear Algebra with Applications*, vol. 8, no. 6-7, pp. 467–496, 2001.
[52] L. Ljung, "System identification: Theory for the user." New Jersey: Prentice Hall, 1999.

Bin Zhu was born in Changshu, Jiangsu, China in 1991. He received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China in 2012 and the M.Eng. degree from Shanghai Jiao Tong University, Shanghai, China in 2015, both in control science and engineering. In 2019, he obtained a Ph.D. degree in information engineering from University of Padova, Padova, Italy, and he had a one-year postdoc position in the same university. Since December 2019, he has been working at the School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China, as an assistant professor.

His current research interest includes spectral analysis, frequency estimation, and sparsity-promoting techniques for signal processing and machine learning.

Mattia Zorzi received the M.S. degree in Automation Engineering and the Ph.D. degree in Information Engineering from the University of Padova, Padova, Italy, in 2009 and 2013, respectively. He held Postdoctoral appointments with the Department of Electrical Engineering and Computer Science, University of Liege, Liege, Belgium, and with the Human Inspired Technology Research Centre, University of Paodva, Padova, Italy. He held visiting positions with the Department of Electrical and Computer Engineering, University of California, Davis, USA, and with the Department of Engineering, University of Cambridge, Cambridge, U.K., in 2011 and 2013-2014, respectively. He is currently an Associate Professor with the Department of Information Engineering, University of Padova. His current research interests include machine learning, robust estimation, identification theory.

Dr. Zorzi has been an Associate Editor of Automatica since 2021 and IEEE Control Systems Letters since 2019. He serves as an Associate Editor on the IEEE Control System Society Conference Editorial Board since 2017 and the EUCA Conference Editorial Board since 2020. He is an IEEE Senior member and a member of the IFAC Technical Committee on Modelling, Identification and Signal Processing.