

# A Weaker Regularity Condition for the Multidimensional $\nu$ -Moment Problem

Bin Zhu and Mattia Zorzi

**Abstract**—We consider the problem of finding a  $d$ -dimensional spectral density through a moment problem which is characterized by an integer parameter  $\nu$ . Previous results showed that there exists an approximate solution under the regularity condition  $\nu \geq d/2 + 1$ . To realize the process corresponding to such a spectral density, one would take  $\nu$  as small as possible. In this paper we show that this condition can be weakened as  $\nu \geq d/2$ .

## I. INTRODUCTION

Multidimensional stationary processes (or stationary random fields) represent a fundamental tool in many applications of signal and image processing. In those applications we have to estimate the multidimensional spectral density of the process. This task can be addressed by means of a moment problem, more precisely, a *convex optimization problem subject to moment constraints*.

In the unidimensional case ( $d = 1$ ) a wide range of spectral estimation paradigms based on moment problems have been proposed, see for instance [1], [2]: in this case the moments correspond to some covariance lags of the process and in the simplest setup the optimal spectrum maximizes the entropy rate. The appealing property of these paradigms is that the optimal spectrum is rational and thus leading to a finite-dimensional linear stochastic system (called “shaping filter” in the literature of signal processing) after spectral factorization.

In the case where the moments include both covariance lags and cepstral coefficients (i.e., logarithmic moments), then it is possible to characterize only an approximate solution to the moment problem: the spectrum maximizing the entropy rate matches the covariance lags and approximately the cepstral coefficients [3]. Such a solution is obtained by considering a *regularized* version of the dual optimization problem.

These paradigms have been extended also to the multidimensional case, see e.g., [4], [5]. Although spectral factorization is not always possible in the multidimensional setting, rationality still seems to be a key ingredient toward a finite-dimensional realization theory [6], [7]. The main issue is, however, that the solution of the moment problem

This work was supported in part by the National Natural Science Foundation of China with grant 62103453, and the “Hundred Talent” Program of Sun Yat-sen University.

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

M. Zorzi is with Department of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131 Padova, Italy zorzimat@dei.unipd.it

is not necessarily a spectral density, but rather a *spectral measure* that may contain a singular part [8]. In particular, if the moments are both the covariance lags and the cepstral coefficients, the existence of an approximate rational solution is only guaranteed when the dimension is  $d \leq 2$  [8].

In order to overcome this limitation on the dimension  $d$ , we have proposed a new moment problem, hereafter called  *$\nu$ -moment problem*, in which the entropy rate has been replaced by a more general definition of entropy, called  *$\nu$ -entropy*, whose derivation comes from the  $\alpha$ -divergence [9]. The definition of cepstral coefficients has been generalized accordingly. The  $\nu$ -moment problem is characterized by the integer parameter  $\nu$ . In [10], [11] we have shown that for any  $d > 2$  there exists an approximate rational solution to the  $\nu$ -moment problem under the regularity condition  $\nu \geq d/2 + 1$ . On the other hand, if the solution admits a spectral factorization, then the estimated process can be realized through a cascade of  $\nu$  identical linear filters. Therefore, the larger  $\nu$  is, the larger the complexity is in order to realize such a process. Accordingly, from a practical perspective there are situations in which one would take  $\nu$  as small as possible.

The aim of the present paper is to show that the regularity condition can be weakened as  $\nu \geq d/2$  and thus we have more freedom in choosing the rational structure than the one characterized in our previous work [10]. Moreover, it is possible to estimate a  $d$ -dimensional process with  $\nu = \lceil d/2 \rceil$  whose realization, if admissible, is simpler than the one obtained using the theory in [10]. Such a result is achieved through a new regularization technique. Moreover, the technical proof of the existence of an approximate rational solution to the  $\nu$ -moment problem takes a rather different route from the one in [10].

## II. PROBLEM FORMULATION

Consider a  $d$ -dimensional real<sup>1</sup> stationary random field  $\{y(\mathbf{t}) : \mathbf{t} \in \mathbb{Z}^d\}$  with zero mean and a spectral density  $\Phi(\boldsymbol{\theta})$ . The latter is a nonnegative function on the  $d$ -dimensional frequency domain  $\mathbb{T}^d := (0, 2\pi]^d$  and  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d) \in \mathbb{T}^d$  is a frequency vector. In the sequel, we shall use the notation  $\Phi(e^{i\boldsymbol{\theta}})$  which is common in Complex Analysis. Here  $e^{i\boldsymbol{\theta}}$  is a shorthand for the vector  $(e^{i\theta_1}, \dots, e^{i\theta_d})$  representing a point on the  $d$ -torus which is isomorphic to  $\mathbb{T}^d$ .

The  $\nu$ -entropy of the random field  $y$  with the integer

<sup>1</sup>We choose to present the theory for real random fields for simplicity. After suitable adaptation, complex random fields can be handled as well.

parameter  $\nu > 1$  is defined as

$$\mathbb{H}_\nu(\Phi) := \frac{\nu^2}{\nu - 1} \left( \int_{\mathbb{T}^d} \Phi^{\frac{\nu-1}{\nu}} d\mu - 1 \right) \quad (1)$$

where  $d\mu = \frac{1}{(2\pi)^d} \prod_{j=1}^d d\theta_j$  is the normalized Lebesgue measure on  $\mathbb{T}^d$ . It is worth noting that  $\mathbb{H}_\nu$  is a more general definition of entropy. Indeed, for the case  $\nu = 1$  (which is understood in a suitable limit sense), we obtain the usual entropy rate  $\mathbb{H}_1(\Phi) = \int_{\mathbb{T}^d} \log \Phi d\mu$  [10]. It is worth noting that  $\mathbb{H}_\nu$  measures the discrepancy between  $\Phi$  and the constant function 1. In this paper, we suppose that a finite index set  $\Lambda \subset \mathbb{Z}^d$  is given and contains  $\mathbf{0}$  and is symmetric with respect to the origin, i.e.,  $\mathbf{k} \in \Lambda$  implies  $-\mathbf{k} \in \Lambda$ . Let  $\Lambda_0 := \Lambda \setminus \{\mathbf{0}\}$  be another index set. Then suppose further that the following data are given:

- $\mathbf{c} = \{c_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$  which is a covariance multisequence of the random field  $y$ , namely  $c_{\mathbf{k}} := \mathbb{E}[y(\mathbf{t} + \mathbf{k})y(\mathbf{t})]$  where  $\mathbb{E}$  denotes the expectation operator. It is well known that the covariances are the Fourier coefficients of the spectral density  $\Phi$ , i.e.,

$$c_{\mathbf{k}} = \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} \Phi d\mu \quad \forall \mathbf{k} \in \Lambda, \quad (2)$$

where  $\langle \mathbf{k}, \boldsymbol{\theta} \rangle = \sum_{j=1}^d k_j \theta_j$  is the standard inner product in  $\mathbb{R}^d$  and the multidimensional exponential function is understood as  $e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} = \prod_{j=1}^d e^{ik_j \theta_j}$ .

- $\mathbf{m} = \{m_{\mathbf{k}} : \mathbf{k} \in \Lambda_0\}$  which is a multisequence of generalized cepstral coefficients, called  $\nu$ -cepstral coefficients, associated to the same random field  $y$ , see [10]. A mathematical definition similar to (2) is given as

$$m_{\mathbf{k}} := \frac{\nu}{\nu - 1} \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} \Phi^{\frac{\nu-1}{\nu}} d\mu \quad \forall \mathbf{k} \in \Lambda_0. \quad (3)$$

It is a generalization of the classic logarithmic moments used in [3], [4]. Here  $m_{\mathbf{0}}$  is excluded for technical reasons.

Now we can formulate the multidimensional  $\nu$ -moment problem:

$$\max_{\Phi \geq 0} \mathbb{H}_\nu(\Phi) \quad \text{s.t.} \quad (2) \text{ and } (3), \quad (4)$$

where the data  $\mathbf{c}$  and  $\mathbf{m}$  enter as measurements on the spectrum  $\Phi$ . Notice that, the notion of  $\nu$ -cepstral coefficients (3) is consistent with the objective functional (1) employed in (4). Moreover, Problem (4) aims to select the power spectral density having the most regular shape while matching the given moments.

The  $\nu$ -moment problem (4), which is also referred to as the primal optimization problem, can be used to perform spectral estimation. Assume that a dataset generated by  $y$  has been collected. Then, it is possible to compute the sample estimates  $\hat{c}_{\mathbf{k}}$  and  $\hat{m}_{\mathbf{k}}$  of  $c_{\mathbf{k}}$  and  $m_{\mathbf{k}}$ , respectively. Then, the estimate of  $\Phi$  is the solution of (4) in which  $c_{\mathbf{k}}$  and  $m_{\mathbf{k}}$  are substituted by  $\hat{c}_{\mathbf{k}}$  and  $\hat{m}_{\mathbf{k}}$ .

In [10], we derived the dual optimization problem and showed that, if the dual problem admits an interior-point solution, the optimal spectral density (primal variable) is a rational function of the form

$$\Phi_\nu = (P/Q)^\nu \quad (5)$$

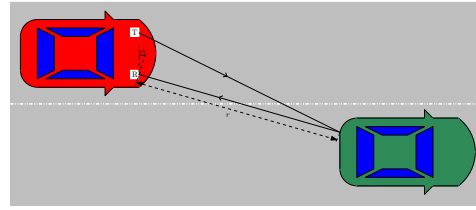


Fig. 1. Target parameters estimation problem: T is source generating the pulse train signals, R is the receiver,  $\alpha$  is the azimuth angle and  $r$  is the range.

where  $P$  and  $Q$  are positive<sup>2</sup> trigonometric polynomials associated with the Lagrange multipliers. However, it is highly nontrivial to prove that the optimal dual variable lies in the interior of the feasible set. In order to overcome this difficulty, we introduced a regularization term in the dual objective function which depends solely on  $P$ :

$$\frac{\lambda}{\nu - 1} \int_{\mathbb{T}^d} \frac{1}{P^{\nu-1}} d\mu \quad (6)$$

where  $\lambda > 0$  is a regularization parameter. Such a regularizer can of course be interpreted as a *barrier function* (for  $P$ ) since under the regularity condition

$$\nu \geq d/2 + 1, \quad (7)$$

the regularizer takes an infinite value if  $P$  has a zero on the  $d$ -torus. Hence, the optimal  $P$  is forced to be an interior point, i.e., a positive polynomial. Obviously, one can always choose  $\nu$  such that the condition (7) is met, and indeed in this case, we showed that the optimal  $Q$  is also positive so that the optimal form (5) is true for the primal problem (4). The price to pay for the regularization is that the  $\nu$ -cepstral constraints (3) are only approximately satisfied with an error that decreases as  $\lambda \rightarrow 0_+$ , i.e., the rational function (5) represents an approximate solution to (4). It is worth noting that a process with the spectrum (5) can be generated by a cascade of  $\nu$  identical multidimensional filters if  $P/Q$  admits a spectral factor corresponding to a state space realization. Clearly, the larger  $\nu$  is, the larger the complexity of the realization is. Accordingly, the key point from a practical perspective is to have the possibility to take  $\nu$  as small as possible.

Having in mind this important requirement, in this paper we want to face the following problem in which we consider a new regularization term.

**Problem 1.** Take the barrier function in the dual problem as

$$\frac{\lambda}{\nu - 1} \int_{\mathbb{T}^d} \frac{1}{P^\nu} d\mu \quad (8)$$

and show that the corresponding approximate solution of the primal problem (4) exists under the weaker condition  $\nu \geq d/2$ .

<sup>2</sup>By positive we mean that  $P > 0$  for any point on the  $d$ -torus. In contrast, if  $P \geq 0$  and  $P = 0$  for certain points, we say that  $P$  is *nonnegative*.

**Example 1.** Consider an automotive radar system installed in the red car of Fig. 1 that employs coherent linear frequency-modulated pulse trains signals (T) and uses a uniform linear array of receive antennas for the measurement (R). The target (green car in Fig. 1) is identified by the range  $r$ , the azimuth angle  $\alpha$  and the relative velocity  $v$ . The problem of estimating the target parameters can be formulated as a multidimensional spectral estimation problem with  $d = 3$  [12], [13], [14]. Under the aforementioned hypothesis on the existence of a “realizable” spectral factor, the estimated model can be realized by a cascade of at least 3 identical filters using the theory developed in [10], while only 2 using the theory that we will develop in this paper which makes more efficient the implementation of the model for simulation purposes. Indeed, the computational cost of data generation via white noise filtering is proportional to  $\nu$ . In addition, it is important to point out that the computational burden of multidimensional filtering is remarkable because it grows exponentially with respect to  $d$ .

### III. DUAL PROBLEM

Let us recall the innocuous condition  $\nu \geq 2$  which is assumed throughout this paper. In [10, Sec. 4], it has been shown that the dual function of the primal problem (4) is

$$J_{\nu}(P, Q) := \frac{1}{\nu - 1} \int_{\mathbb{T}^d} \frac{P^{\nu}}{Q^{\nu-1}} d\mu + \langle \mathbf{q}, \mathbf{c} \rangle - \langle \mathbf{p}, \mathbf{m} \rangle, \quad (9)$$

where:

- $\mathbf{q} = \{q_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$  contains the real coefficients of the nonnegative trigonometric polynomial  $Q(e^{i\theta}) := \sum_{\mathbf{k} \in \Lambda} q_{\mathbf{k}} e^{-i\langle \mathbf{k}, \theta \rangle}$  such that  $q_{-\mathbf{k}} = q_{\mathbf{k}}$ ;
- similarly,  $\mathbf{p} = \{p_{\mathbf{k}} : \mathbf{k} \in \Lambda_0\}$  contains the real coefficients of the nonnegative polynomial  $P(e^{i\theta}) = \sum_{\mathbf{k} \in \Lambda} p_{\mathbf{k}} e^{-i\langle \mathbf{k}, \theta \rangle}$  with  $p_{-\mathbf{k}} = p_{\mathbf{k}}$  and  $p_0 = 1$  fixed;
- $\langle \mathbf{q}, \mathbf{c} \rangle := \sum_{\mathbf{k} \in \Lambda} q_{\mathbf{k}} c_{\mathbf{k}}$  inner product of two multisequences indexed in  $\Lambda$ , and  $\langle \mathbf{p}, \mathbf{m} \rangle$  is understood similarly with the index set replaced by  $\Lambda_0$ .

Since it is rather difficult to prove that the dual problem admits an interior-point solution, see [10], we consider the regularized dual function

$$J_{\nu, \lambda}(P, Q) := J_{\nu}(P, Q) + \frac{\lambda}{\nu - 1} \int_{\mathbb{T}^d} \frac{1}{P^{\nu}} d\mu \quad (10a)$$

$$= \frac{1}{\nu - 1} \int_{\mathbb{T}^d} g(P(e^{i\theta}), Q(e^{i\theta})) d\mu + \langle \mathbf{q}, \mathbf{c} \rangle - \langle \mathbf{p}, \mathbf{m} \rangle \quad (10b)$$

where the bivariate function is defined as

$$g(x, y) := x^{\nu}/y^{\nu-1} + \lambda/x^{\nu} \quad (11)$$

with  $x > 0$ ,  $y > 0$  and the regularization parameter  $\lambda > 0$  is fixed. Now let us introduce the feasible sets  $\mathfrak{F}_+ := \{Q(e^{i\theta}) = \sum_{\mathbf{k} \in \Lambda} q_{\mathbf{k}} e^{-i\langle \mathbf{k}, \theta \rangle} : Q > 0 \text{ on } \mathbb{T}^d\}$ ,  $\mathfrak{F}_{+, o} := \{P(e^{i\theta}) = \sum_{\mathbf{k} \in \Lambda} p_{\mathbf{k}} e^{-i\langle \mathbf{k}, \theta \rangle} \in \mathfrak{F}_+ : p_0 = 1\}$  so that  $Q \in \mathfrak{F}_+$  and  $P \in \mathfrak{F}_{+, o}$ . The domain of definition of  $J_{\nu, \lambda}$  can be extended to the boundary of the feasible set  $\mathfrak{F}_{+, o} \times \mathfrak{F}_+$  by excluding the zero sets of  $P$  and  $Q$  from the domain of integration, which does not change the values of

the integrals since the zero sets have zero Lebesgue measure. Moreover,  $J_{\nu, \lambda}$  may take a value of  $\infty$  at some boundary points, and hence it is understood as an *extended real-valued* function.

**Lemma 1.** *The function  $g$  in (11) is strictly convex in the domain  $x > 0$ ,  $y > 0$  (the first quadrant).*

*Proof.* We shall prove the claim via the derivative test. After some straightforward computations, we arrive at

$$\nabla g(x, y) = x^{\nu-1} y^{-\nu} \begin{bmatrix} \nu y \\ (1-\nu)x \end{bmatrix} + \begin{bmatrix} -\lambda \nu x^{-\nu-1} \\ 0 \end{bmatrix}, \quad (12a)$$

$$\nabla^2 g(x, y) = \nu(\nu-1) x^{\nu-2} y^{-\nu-1} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} + \begin{bmatrix} \lambda \nu(\nu+1) x^{-\nu-2} & 0 \\ 0 & 0 \end{bmatrix}, \quad (12b)$$

where  $\nu(\nu-1) > 0$  is a positive integer. It is readily observed that every diagonal element in the Hessian of  $g$  is positive, and the first matrix in (12b) is positive semidefinite. After checking the determinant of the Hessian, we conclude that  $\nabla^2 g(x, y)$  is positive definite in the first quadrant and the strict convexity follows.  $\square$

The next proposition is a direct consequence of Lemma 1.

**Proposition 1.** *The regularized dual function  $J_{\nu, \lambda}$  is strictly convex in the closed set  $\overline{\mathfrak{F}}_{+, o} \times \overline{\mathfrak{F}}_+$ .*

*Proof.* We only need to show the strict convexity of the integral term in (10b) since the inner products are linear in  $(P, Q)$ . In the interior of the feasible set, namely  $P \in \mathfrak{F}_{+, o}$  and  $Q \in \mathfrak{F}_+$ , the strict convexity of the integral term follows from that of  $g$  (see Lemma 1), and this can be seen by a pointwise argument on the integrand (see e.g., Proposition 5.3 in [12]). The same reasoning works if  $P$  or  $Q$  is on the boundary of the respective feasible set because, once the zero sets of  $P$  and  $Q$  are excluded from the domain of integration, the function  $g(P(e^{i\theta}), Q(e^{i\theta}))$  is well defined and the proof holds verbatim.  $\square$

We conclude that the regularized dual optimization problem can be formulated as

$$\min J_{\nu, \lambda}(P, Q) \quad \text{s.t. } P \in \overline{\mathfrak{F}}_{+, o}, Q \in \overline{\mathfrak{F}}_+. \quad (13)$$

It is worth noting that the regularizer (6) in [10] is designed so that the regularized dual function always diverges on the boundary of the feasible set. Hence any boundary point is not a minimizer. The main result in [12] also exploits a similar property. In the present setting, however, the route to establish the existence of an interior-point minimizer will be completely different, and this is done in the next section.

### IV. UNIQUE SOLUTION UNDER THE CONDITION $\nu \geq d/2$

In this section we shall see how the condition  $\nu \geq d/2$  guarantees that the regularized dual problem (13) has an interior-point solution. By Proposition 1, we know that a minimizer of  $J_{\nu, \lambda}$  in  $\overline{\mathfrak{F}}_{+, o} \times \overline{\mathfrak{F}}_+$  is unique *provided that it exists*. We shall first establish such existence. Then, under

the regularity condition  $\nu \geq d/2$  which is weaker than (7), we aim to exclude the possibility that a minimum of  $J_{\nu,\lambda}$  may fall on the boundary of the feasible set, which is a weaker version of Lemma 5.8 in [10], using a Byrnes–Gusev–Lindquist-type argument that first appeared in [1] and subsequently in e.g., [4].

#### A. Existence of a minimizer in $\overline{\mathfrak{P}}_{+,o} \times \overline{\mathfrak{P}}_+$

The existence of a solution to (13) can be shown via arguments similar to the ones in [10]: such existence depends on the following feasibility assumption which holds for instance, when the covariance lags are estimated from data using the so called *standard biased covariance estimates* [15].

**Assumption 1** (Feasibility). *The given covariances  $\{c_{\mathbf{k}}\}_{\mathbf{k} \in \Lambda}$  admit an integral representation*

$$c_{\mathbf{k}} = \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \theta \rangle} \Phi_0 d\mu \quad \forall \mathbf{k} \in \Lambda, \quad (14)$$

where  $\Phi_0$  is a nonnegative function on  $\mathbb{T}^d$  and is positive on some open ball  $B_1 \subset \mathbb{T}^d$ .

In order to prove our existence result, we need the following lemmas.

**Lemma 2.** *The unregularized dual function  $J_\nu$  in (9) and the regularized version  $J_{\nu,\lambda}$  in (10) are lower-semicontinuous on  $\overline{\mathfrak{P}}_{+,o} \times \overline{\mathfrak{P}}_+$ . In particular, they are both continuous on  $\mathfrak{P}_{+,o} \times \mathfrak{P}_+$ .*

*Proof.* The proof is similar to that of Lemma 5.5 in [10]. The only difference here is that the power of  $P$  in the regularizer in (10a) is  $\nu$  instead of  $\nu - 1$  in [10].  $\square$

**Lemma 3.** *Suppose that Assumption 1 holds. If a sequence  $\{(P_j, Q_j)\}_{j \geq 1} \subset \overline{\mathfrak{P}}_{+,o} \times \overline{\mathfrak{P}}_+$  is such that  $\|(P_j, Q_j)\| \rightarrow \infty$  as  $j \rightarrow \infty$ , then  $J_{\nu,\lambda}(P_j, Q_j) \rightarrow \infty$ .*

*Proof.* See the proof of Lemma 5.6 in [10] which uses Assumption 1.  $\square$

**Proposition 2.** *Under Assumption 1, the regularized dual optimization problem (13) admits a solution.*

*Proof.* Take a sufficiently large real number  $\beta$ , so that the sublevel set of the regularized dual function  $J_{\nu,\lambda}^{-1}(-\infty, \beta] := \{(P, Q) \in \overline{\mathfrak{P}}_{+,o} \times \overline{\mathfrak{P}}_+ : J_{\nu,\lambda}(P, Q) \leq \beta\}$  is not empty. Then Lemma 2 implies that the sublevel set is closed; Lemma 3 implies that the sublevel set is bounded. Obviously the polynomial pair  $(P, Q)$ , parametrized by their coefficients, belongs to a finite-dimensional vector space. It follows that the sublevel set is compact. Given the lower-semicontinuity of the objective function  $J_{\nu,\lambda}$  (see Lemma 2), a minimizer exists in  $J_{\nu,\lambda}^{-1}(-\infty, \beta]$  by the extreme value theorem of Weierstrass.  $\square$

#### B. Non-optimality of boundary points given $\nu \geq d/2$

In this subsection we prove that the optimal solution of (13) cannot belong to the boundary of  $\overline{\mathfrak{P}}_{+,o} \times \overline{\mathfrak{P}}_+$  using arguments which conceptually differ from the ones used for

the case  $\nu \geq d/2 + 1$  in [10]. Notice first that if  $P \in \partial\overline{\mathfrak{P}}_{+,o}$  where  $\partial$  denotes the boundary of a set, then the regularization term (8) employed in (10a) takes a value of  $\infty$  under the condition  $\nu \geq d/2$ , see Proposition A.4 in [12]. The other term, namely the unregularized function  $J_\nu(P, Q)$  whose expression is given in (9), is bounded from below under Assumption 1, see the proof of [10, Lemma 5.6]. Therefore, in this case we have  $J_{\nu,\lambda}(P, Q) = \infty$  which is certainly not a minimum. Consequently, an optimal  $(P, Q)$  must have  $P \in \mathfrak{P}_{+,o}$ .

Next, we work on the case of  $(P, Q) \in \mathfrak{P}_{+,o} \times \partial\overline{\mathfrak{P}}_+$ . It is still possible that the integral term in (9) diverges so that  $J_{\nu,\lambda}(P, Q) = \infty$ , and such a point is obviously not a minimizer. Therefore, we only need to consider points  $(P, Q) \in \mathfrak{P}_{+,o} \times \partial\overline{\mathfrak{P}}_+$  such that the function value  $J_{\nu,\lambda}(P, Q)$  is finite. Construct the real-valued function

$$f(t) := J_{\nu,\lambda}(P, Q + t\mathbf{1}) \quad (15)$$

defined for  $t \geq 0$  where  $\mathbf{1}$  denotes the constant polynomial taking value one. One can show without difficulty that  $f$  “inherits” the strict convexity from  $J_{\nu,\lambda}$  (see Proposition 1). It is worth noting that the continuity of  $f(t)$  can be extended to  $t = 0_+$  which corresponds to the boundary point  $(P, Q)$ , in contrast with the general result that  $J_{\nu,\lambda}$  is only lower-semicontinuous on the boundary of the feasible set (see Lemma 2). These properties are established in the next statement.

**Lemma 4.** *The function  $f(t)$  in (15), where  $(P, Q)$  is an arbitrarily fixed point in  $\mathfrak{P}_{+,o} \times \partial\overline{\mathfrak{P}}_+$  such that  $J_{\nu,\lambda}(P, Q) < \infty$ , is strictly convex and continuous in  $[0, \infty)$ .*

*Proof.* The strict convexity of  $f$  follows directly from Proposition 1, and it remains to show the continuity. For  $t > 0$ , the argument  $(P, Q + t\mathbf{1})$  belongs to the interior  $\mathfrak{P}_{+,o} \times \mathfrak{P}_+$ , and the continuity of  $f$  follows from that of  $J_{\nu,\lambda}$  (see Lemma 2). Hence, we are only concerned with the right continuity of  $f$  at  $t = 0$ . By the lower-semicontinuity of  $J_{\nu,\lambda}$  at  $(P, Q)$  (Lemma 2 again), for any  $\varepsilon > 0$ , there exists  $\delta > 0$  such that whenever  $(P_1, Q_1) \in \overline{\mathfrak{P}}_{+,o} \times \overline{\mathfrak{P}}_+$  satisfies  $\|(P_1, Q_1) - (P, Q)\| < \delta$ , we have  $J_{\nu,\lambda}(P_1, Q_1) > J_{\nu,\lambda}(P, Q) - \varepsilon$ . We can always choose  $t$  sufficiently small to make the argument on the right-hand side of (15) sufficiently close to  $(P, Q)$ , which leads to the inequality  $f(t) > f(0) - \varepsilon$ . At the same time, by strict convexity we have  $f(t) < tf(1) + (1 - t)f(0) = f(0) + t[f(1) - f(0)]$ . The last term, namely  $t[f(1) - f(0)]$  can be made smaller than  $\varepsilon$  (in absolute value) for  $t$  sufficiently small. Therefore, we reach the inequality  $|f(t) - f(0)| < \varepsilon$  which proves the continuity at  $t = 0_+$  since  $\varepsilon > 0$  is arbitrarily chosen.  $\square$

**Proposition 3.** *Fix any point  $(P, Q) \in \mathfrak{P}_{+,o} \times \partial\overline{\mathfrak{P}}_+$  such that  $J_{\nu,\lambda}(P, Q) < \infty$ . Then the function  $f(t)$  in (15) has a derivative  $f'(t) \rightarrow -\infty$  as  $t \rightarrow 0_+$ . Therefore,  $(P, Q)$  cannot be a minimizer of  $J_{\nu,\lambda}$ .*

*Proof.* After some computations, we have for  $t > 0$  that

$$f'(t) = c_0 - \int_{\mathbb{T}^d} \left[ \frac{P}{Q + t\mathbf{1}} \right]^\nu d\mu. \quad (16)$$

Since  $P \in \mathfrak{P}_{+,o}$ , we know  $P_{\min} := \min_{\theta \in \mathbb{T}^d} P(e^{i\theta}) > 0$ . Then the following relation

$$\int_{\mathbb{T}^d} \left[ \frac{P}{Q+t\mathbf{1}} \right]^\nu d\mu \geq P_{\min}^\nu \int_{\mathbb{T}^d} \frac{1}{(Q+t\mathbf{1})^\nu} d\mu \rightarrow \infty \quad (17)$$

holds as  $t \rightarrow 0_+$  by Lebesgue's monotone convergence theorem [16, p. 21] and Proposition A.4 in [12]. This shows that  $f'(t)$  in (16) tends to  $-\infty$  as  $t$  goes to zero from the right. Taking Lemma 4 into account, we conclude that 0 is not a local minimizer of  $f(t)$ . Consequently, any boundary point  $(P, Q) \in \mathfrak{P}_{+,o} \times \partial\mathfrak{P}_+$  is not a minimizer of  $J_{\nu,\lambda}$ , because taking an arbitrarily small step along the direction  $(\mathbf{0}, \mathbf{1})$ , which points towards the interior  $\mathfrak{P}_{+,o} \times \mathfrak{P}_+$ , will result in a decrease of the objective function value.  $\square$

We summarize what we have got so far in the next theorem.

**Theorem 1.** *Under Assumption 1 and the condition  $\nu > d/2$ , the optimization problem (13) admits a unique interior-point solution  $(\hat{P}, \hat{Q}) \in \mathfrak{P}_{+,o} \times \mathfrak{P}_+$  such that*

$$\begin{aligned} c_{\mathbf{k}} &= \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} (\hat{P}/\hat{Q})^\nu d\mu \quad \forall \mathbf{k} \in \Lambda, \\ m_{\mathbf{k}} &= \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} \frac{\nu}{\nu-1} \left[ (\hat{P}/\hat{Q})^{\nu-1} - \lambda/\hat{P}^{\nu+1} \right] d\mu \quad \forall \mathbf{k} \in \Lambda_0. \end{aligned} \quad (18a)$$

$$(18b)$$

*In plain words, the spectral density  $\hat{\Phi}_\nu = (\hat{P}/\hat{Q})^\nu$  achieves covariance matching and approximate  $\nu$ -cepstral matching with errors*

$$\varepsilon_{\mathbf{k}} = \frac{\lambda\nu}{\nu-1} \int_{\mathbb{T}^d} e^{i\langle \mathbf{k}, \boldsymbol{\theta} \rangle} \frac{1}{\hat{P}^{\nu+1}} d\mu, \quad \mathbf{k} \in \Lambda_0.$$

*Proof.* The existence of a solution is guaranteed by Proposition 2 and the uniqueness by Proposition 1. Moreover, given the reasoning at the beginning of this subsection and Proposition 3, the optimal  $(\hat{P}, \hat{Q})$  must be an interior point, i.e., both polynomials are positive. As a consequence, it must satisfy the stationary-point equation  $\nabla J_{\nu,\lambda}(P, Q) = \mathbf{0}$ , which is equivalent to the conditions in (18). Indeed, this point can be seen by setting the first differential of the regularized dual function

$$\begin{aligned} \delta J_{\nu,\lambda}(P, Q; \delta P, \delta Q) &= \langle \mathbf{c}, \delta \mathbf{q} \rangle - \int_{\mathbb{T}^d} \delta Q (P/Q)^\nu d\mu \\ &\quad - \langle \mathbf{m}, \delta \mathbf{p} \rangle + \frac{\nu}{\nu-1} \int_{\mathbb{T}^d} \delta P [(P/Q)^{\nu-1} - \lambda/P^{\nu+1}] d\mu \end{aligned} \quad (19)$$

equal to zero for any direction  $(\delta P, \delta Q)$ .  $\square$

*Remark 1.* It is important to observe that the continuous dependence of the solution  $(\hat{P}, \hat{Q})$  to (13) on the covariance and  $\nu$ -cepstral data  $(\mathbf{c}, \mathbf{m})$  can be established similarly to [10, Sec. 6], so that the optimization problem (13) is in fact well-posed in the sense of Hadamard.

## V. NUMERICAL SIMULATIONS

In this section we present some numerical experiments in which the problem (4) is used to reconstruct the spectrum of a 3-d stationary random field  $y(t_1, t_2, t_3)$  starting from a finite set of its covariance lags and  $\nu$ -cepstral coefficients. The aim is to show that it is possible to compute numerically an approximate solution to (4) for a setup in which the theory developed in [10] does not guarantee the existence of a solution, while the theory developed here does.

We assume that the underlying process  $y$  can be described as the output of a cascade linear shaping filter  $\tilde{W}(z_1, z_2, z_3) = W^\nu(z_1, z_2, z_3)$  with a white noise input  $e(t_1, t_2, t_3)$ ,  $\nu = 2$ , and  $d = 3$ . We shall also assume that the transfer function  $W$  has the structure consistent with our optimal form (5) for the spectrum. More precisely, we consider the class of rational models such that the numerator and denominator polynomials both have degree one:

$$\tilde{W}(\mathbf{z}) = \left[ \frac{b(\mathbf{z})}{a(\mathbf{z})} \right]^\nu = \left[ \frac{b_0 - b_1 z_1^{-1} - b_2 z_2^{-1} - b_3 z_3^{-1}}{a_0 - a_1 z_1^{-1} - a_2 z_2^{-1} - a_3 z_3^{-1}} \right]^\nu \quad (20)$$

where  $\mathbf{z}$  represents  $(z_1, z_2, z_3)$  for short. Obviously, the polynomials  $a(\mathbf{z})$  and  $b(\mathbf{z})$  are described by the respective vectors  $\mathbf{a} = [a_0, \dots, a_3]$  and  $\mathbf{b} = [b_0, \dots, b_3]$  of coefficients. If the white noise input  $e$  has unit variance, then the spectral density of the output process  $y$  is  $(P/Q)^\nu$  where  $P(e^{i\theta}) = |b(e^{i\theta})|^2$  and  $Q(e^{i\theta}) = |a(e^{i\theta})|^2$  are two nonnegative polynomials in  $\overline{\mathfrak{P}}_+$ . Throughout this section, we set the parameter  $\nu = 2$  to meet the condition  $\nu \geq d/2 = 3/2$ . In such scenario it is worth stressing that the theory developed in [10] does not work. Indeed that theory requires  $\nu \geq 3$  to guarantee the existence of an approximate solution to the primal problem (4).

In what follows we consider two models of the form (20). We take two sets of *real* parameters  $(\mathbf{a}_j, \mathbf{b}_j)$  with  $j = 1, 2$  such that  $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a} = [1, 0.3, 0.3, 0.3]$ ,  $\mathbf{b}_1 = [1, -0.2, -0.3, -0.4]$ ,  $\mathbf{b}_2 = [1, -0.2, -0.3, -0.5]$ , and  $\mathbf{b}_j = \tilde{\mathbf{b}}_j / \|\tilde{\mathbf{b}}_j\|$ . The last operation of normalization gives  $\|\mathbf{b}\|^2 = 1$  which is equivalent to  $p_0 = 1$  for the numerator polynomial  $P$ . The first model, hereafter called “zeroless model”, corresponds to the process with spectrum  $\Phi_1 = (P_1/Q)^2$  where the polynomials  $Q = |a|^2$  and  $P_1 = |b_1|^2$  are positive on  $\mathbb{T}^3$ ; the second model, hereafter called “model with a spectral zero”, corresponds to the process with spectrum  $\Phi_2 = (P_2/Q)^2$  where  $P_2 = |b_2|^2$  (and also the spectrum) has a zero at the frequency vector  $(\pi, \pi, \pi)$ . The index set  $\Lambda$  is identified as  $\Lambda = \Lambda_+ - \Lambda_+$  with  $\Lambda_+ := \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ . Here the set difference is understood as  $A - B := \{x - y : x \in A, y \in B\}$ . Since  $q_{-\mathbf{k}} = q_{\mathbf{k}}$  and  $p_{-\mathbf{k}} = p_{\mathbf{k}}$ , the total number of variables is 13.

By means of the discrete formulation described in Section 7 of [10], we consider the following procedure to test the ability to reconstruct the spectra through the solution of (4) for the previous two models:

- 1) Fix  $N = 20$  and discretize  $\mathbb{T}^3$  into  $N^3$  regular grid points by gridding the interval  $[0, 2\pi]$  into  $N$  equidistant

points in each dimension.

- 2) Compute the “true” covariances  $\{c_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$  and the  $\nu$ -cepstral coefficients  $\{m_{\mathbf{k}} : \mathbf{k} \in \Lambda_0\}$  of  $\Phi = |\tilde{W}|^2$  via (2) and (3) where  $d\mu$  is replaced by a discrete measure with equal mass  $1/N^3$  on the grid points.
- 3) Solve the discrete version of the regularized dual problem (13) using  $(\mathbf{c}, \mathbf{m})$  computed above and  $\lambda > 0$  chosen sufficiently small.
- 4) Let  $(\hat{\mathbf{p}}, \hat{\mathbf{q}})$  be the optimal solution to (13) and  $(\mathbf{p}, \mathbf{q})$  be the polynomial coefficients corresponding to the true spectrum  $\Phi = |\tilde{W}|^2$ . Finally, evaluate the reconstruction error  $\|(\hat{\mathbf{p}}, \hat{\mathbf{q}}) - (\mathbf{p}, \mathbf{q})\|$ .

In Step 3, the optimization problem is solved using Newton’s method. Some computational details can be found in [10], and suitable modifications on the gradient and Hessian of the objective function are needed for the current problem (13).

The left panel of Fig. 2 shows the reconstruction errors for the two models with different values of the regularization parameter  $\lambda = 10^{-n}$ ,  $n = 0, 2, 4, 6, 8, 10$ . It is readily observed in both cases that the errors decrease monotonically as  $\lambda \rightarrow 0$ . In view of Remark 1, therefore, if we have a dataset generated by a  $d$ -dimensional process and consider the approximate solution to the problem (4) with  $\nu = \lceil d/2 \rceil$  where  $\mathbf{c}$  and  $\mathbf{m}$  are replaced by their sample estimators (computed from the data), then the resulting spectral estimator is characterized by a small estimation error provided that  $\lambda$  is chosen sufficiently small and the underlying process has a spectrum of the form (5) with  $\nu = \lceil d/2 \rceil$  which agrees with our parameter specification.

In the right panel of Fig. 2, we compare the spectral density of the model with a spectral zero and the reconstructed spectra for values of  $\lambda = 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}$  (corresponding to the orange line in the left panel) along a cross section  $[\cdot, 11, 11]$  of the regular grid for  $\mathbb{T}^3$ , where the true spectral zero is located at the index<sup>3</sup>  $[11, 11, 11]$ . Notice that the other two cross sections  $[11, \cdot, 11]$  and  $[11, 11, \cdot]$  of the spectral densities are not shown because they are visually similar to this figure. We conclude that as  $\lambda$  goes to zero, the true nonnegative spectrum is well approximated by a positive spectral density with smaller and smaller errors, which is consistent with the left panel.

## VI. CONCLUSIONS

We have shown the existence of a rational approximate solution to the  $\nu$ -moment problem under the condition  $\nu \geq d/2$  which is weaker than the one in [10]. Future research should regard the spectral factorization theory in the multidimensional case. It is well known that a positive trigonometric polynomial of several variables in general can only be factored into a *sum of squares* [17] which creates difficulty in the multidimensional stochastic realization theory. Hence it will be interesting to investigate when a usual factorization (into one square) will hold in the multidimensional setting.

<sup>3</sup>In Matlab, the array indices start with 1. The notation  $[\cdot, 11, 11]$  means the collection of indices  $[j, 11, 11]$  with  $j = 1, \dots, 20$ . The other two similar symbols are understood accordingly.

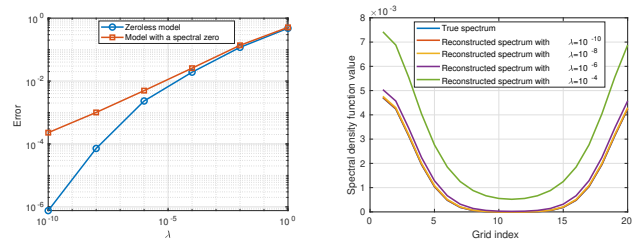


Fig. 2. *Left:* Error of spectrum reconstruction versus the regularization parameter  $\lambda = 10^{-n}$ ,  $n = 0, 2, 4, 6, 8, 10$ , where both axes are in the logarithmic scale. *Right:* The true spectrum with a zero and the reconstructed spectra with choices of  $\lambda = 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}$  at the cross section  $[\cdot, 11, 11]$ , i.e.,  $\hat{\Phi}_{\nu, \lambda}(e^{i\theta})$  with  $\theta = 2\pi \times [(k-1)/20, 11/20, 11/20]$  for grid indices  $k = 1, \dots, 20$ . Note that the reconstructed spectra with  $\lambda = 10^{-10}$  and  $10^{-8}$  (red and yellow lines) and the true spectrum (blue line) almost overlap each other.

## REFERENCES

- [1] C. I. Byrnes, S. V. Gusev, and A. Lindquist, “A convex optimization approach to the rational covariance extension problem,” *SIAM Journal on Control and Optimization*, vol. 37, no. 1, pp. 211–229, 1998.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] —, “Multidimensional rational covariance extension with approximate covariance matching,” *SIAM Journal on Control and Optimization*, vol. 56, no. 2, pp. 913–944, 2018.
- [6] J. S. Geronimo and M. J. Lai, “Factorization of multivariate positive Laurent polynomials,” *Journal of Approximation Theory*, vol. 139, no. 1–2, pp. 327–345, 2006.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] B. Zhu and M. Zorzi, “A well-posed multidimensional rational covariance and generalized cepstral extension problem,” *SIAM J. Control Optimization*, vol. 61, no. 3, pp. 1532–1556, 2023.
- [11] —, “On the statistical consistency of a generalized cepstral estimator,” Submitted to IEEE Transactions on Automatic Control. arXiv preprint: 2301.06784, 2022.
- [12] B. Zhu, A. Ferrante, J. Karlsson, and M. Zorzi, “ $M^2$ -spectral estimation: A flexible approach ensuring rational solutions,” *SIAM Journal on Control and Optimization*, vol. 59, no. 4, pp. 2977–2996, 2021.
- [13] F. Engels, “Target shape estimation using an automotive radar,” in *Smart Mobile In-Vehicle Systems: Next Generation Advancements*. Springer Science+Business Media, 2014, ch. 16, pp. 271–290.
- [14] 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.
- [15] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.
- [16] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill Book Company, 1987.
- [17] M. Bakonyi and H. J. Woerdeman, *Matrix Completions, Moments, and Sums of Hermitian Squares*. Princeton University Press, 2011.