A New Family of High-Resolution Multivariate Spectral Estimators
Mattia Zorzi

Abstract—In this paper, we extend the Beta divergence family to multivariate power spectral densities. Similarly to the scalar case, we show that it smoothly connects the multivariate Kullback-Leibler divergence with the multivariate Itakura-Saito distance. We successively study a spectrum approximation problem, based on the Beta divergence family, which is related to a multivariate extension of the THREE spectral estimation technique. It is then possible to characterize a family of solutions to the problem. An upper bound on the complexity of these solutions will also be provided. Finally, we will show that the most suitable solution of this family depends on the specific features required from the estimation problem.

Index Terms—Generalized covariance extension problem, Spectrum approximation problem, Structured covariance estimation problem, Beta divergence, Convex optimization

I. INTRODUCTION

The recent development of THREE-like approaches to multivariate spectral estimation has triggered a renewed interest for multivariate distance measures (or simply divergence indexes) among (power) spectral densities, [1]. In the THREE approach, the output covariance of a bank of filters is used to extract information on the input spectral density. More precisely, the family of spectral densities matching the output covariance matrix is considered and a spectrum approximation problem, which “chooses” an estimate of the input spectral density in this family, is then employed. The choice criterion is based on finding the spectral density which minimizes a divergence index with respect to an a priori spectral density. Note that, the problem of parameterizing the family of feasible spectral densities may be viewed as a generalized covariance extension problem [2], [4], [5], [6], [7]. The key feature for these estimators concerns the high resolution achievable in prescribed frequency bands, in particular with short data records. Significant applications to these methods can be found in $H_\infty$ robust control [8], [9], biomedical engineering [10], and modeling and identification [11], [12], [13].

The most delicate issue for this theory deals with the choice of the divergence index. In fact, the corresponding solution to the spectrum approximation problem (that heavily depends on the divergence index) must be computable and possibly with bounded McMillan degree. Accordingly, it is important to have many different indexes available in such a way to choose the most appropriate index in relation to the specific application. The THREE estimator, introduced by Byrnes, Georgiou and Lindquist in [14], has been extended to the multichannel case by suggesting different multivariate divergence indexes, [15], [16], [17]. In particular, Georgiou introduced a multivariate version of the Kullback-Leibler divergence, [15], which has been frequently used within information theory, and a multivariate extension of the Itakura-Saito distance has been recently presented by Ferrante et al., [17]. The latter metric has an interpretation in terms of relative entropy rate among processes. Finally, it is worth noting that the output covariance is not available in a THREE-like spectral estimation method. Indeed, we need to estimate it by using a collection of sample data generated by feeding the filters bank with the signal whose spectral density is to be estimated. Moreover, the family of spectral densities matching the estimated output covariance must be non-empty. This covariance estimation task is accomplished by solving a structured covariance estimation problem, [18], [19]. Therefore, a THREE-like spectral estimation procedure consists in solving a structured covariance estimation problem and then a spectrum approximation problem.

The main results of this paper are three. Firstly, we extend to the multivariate case the Beta divergence family (introduced for the scalar case in [20]) which smoothly connects the Kullback-Leibler divergence with the Itakura-Saito distance. It is worth mentioning that the Beta divergence family for scalar spectral densities has been widely used in many applications: Robust principal component analysis and clustering [21], robust independent component analysis [22], and robust nonnegative matrix and tensor factorization [23], [24].

Secondly, we consider a spectrum approximation problem which employs the multivariate Beta divergence family. It turns out that it is possible to characterize a family of solutions to the problem with bounded McMillan degree. Moreover, the limit of the family coincides to the solution obtained by using the Kullback-Leibler divergence.

Finally, we tackle the related structured covariance estimation problem which can be viewed as the static version of the previous spectrum approximation problem. Also in this case, a Beta matrix divergence family for covariance matrices, leading to a family of solutions to the structured covariance estimation problem, may be introduced.

The paper is outlined as follows. Section II introduces THREE-like spectral estimation methods. Section III presents the new extension to the multivariate case of the Beta divergence family. In Section IV the corresponding spectrum approximation problem is introduced. More precisely, we derive the solution thanks to the means of the convex optimiza-
II. THREE-LIKE SPECTRAL ESTIMATION

Let us consider an unknown zero mean, m-dimensional, \( \mathbb{R}^m \)-valued, purely non-deterministic, full-rank, stationary process \( y = \{ y_k : k \in \mathbb{Z} \} \) with spectral density \( \Omega(e^{j\theta}) \) defined on the unit circle \( T \). Assume that the a priori information on \( \Omega \) is given by a prior spectral density \( \Psi \in \mathcal{S}_+^m(T) \). Here, \( \mathcal{S}_+^m(T) \) denotes the family of \( C_+^{m \times m} \)-valued spectral density functions on \( T \) which are bounded and coercive, i.e. \( \Psi \in \mathcal{S}_+^m(T) \) if there exist two constants \( \mu_1 \geq \mu_2 > 0 \) such that \( \mu_2 I \leq \Psi(e^{j\theta}) \leq \mu_1 I \) on \( T \). Then, a finite-length data \( y_1 \ldots y_N \) generated by \( y \) is observed. We want to find an estimate \( \hat{\Sigma} \in \mathcal{S}_+^m(T) \) of \( \Omega \) by using \( \Psi \) and \( y_1 \ldots y_N \).

This spectral estimation task is accomplished by employing a THREE-like approach which hinges on the following four elements:

1. A prior spectral density \( \Psi \in \mathcal{S}_+^m(T) \);
2. A rational filter to process the data
   \[
   G(z) = (zI - A)^{-1}B, \tag{1}
   \]
   where \( A \in \mathbb{R}^{n \times n} \) is a stability matrix, \( B \in \mathbb{R}^{n \times m} \) is full rank with \( n > m \), and \( (A, B) \) is a reachable pair;
3. An estimate \( \hat{\Sigma} \), based on the data \( y_1 \ldots y_N \), of the steady state covariance \( \Sigma = \Sigma^T > 0 \) of the state \( x_k \) of the filter
   \[
   x_{k+1} = Ax_k + B y_k;
   \]
4. A divergence index \( S \) between two spectral densities. According to the THREE-like approach, an estimate \( \hat{\Sigma} \in \mathcal{S}_+^m(T) \) of \( \Omega \) is given by solving the problem\(^1\)

\[
\minimize_{\hat{\Sigma}} S(\Psi) \text{ over the set } \left\{ \hat{\Sigma} \in \mathcal{S}_+^m(T) \mid \int G \hat{\Sigma} G^* = \hat{\Sigma} \right\}. \tag{2}
\]

Note that \( \Psi \) is generally not consistent with \( \hat{\Sigma} \), i.e. \( \int G \hat{\Sigma} G^* \neq \hat{\Sigma} \). Hence, we have a spectrum approximation problem. The parametrization of all spectral densities satisfying constraint in (2) may be viewed as a generalized moment problem. For instance, the covariance extension problem may be recovered by setting

\[
G(z) = \begin{bmatrix} z^{-n} I_m & \ldots & z^{-1} I_m \end{bmatrix}^T. \tag{3}
\]

In this case, the state covariance has a block Toeplitz structure:

\[
\Sigma = \begin{bmatrix}
\Sigma_0 & \Sigma_1 & \Sigma_2 & \ldots & \Sigma_{n-1} \\
\Sigma_1^T & \Sigma_0 & \Sigma_1 & \ldots & \Sigma_{n-2} \\
\Sigma_2^T & \ldots & \Sigma_0 & \ldots & \ldots \\
\vdots & \ddots & \ddots & \ldots & \ddots \\
\Sigma_{n-1}^T & \ldots & \ldots & \ldots & \Sigma_0
\end{bmatrix}, \quad \Sigma_l = \mathbb{E}[y_k y_{k+l}^T].
\]

A. Feasibility of the problem

The first issue arising with the previous spectrum approximation problem concerns its feasibility, i.e. the existence of \( \Phi \in \mathcal{S}_+^m(T) \) satisfying the constraint in (2) for a given \( \hat{\Sigma} \). To deal with this issue, we first introduce some notation: \( Q_n \subseteq \mathbb{R}^{n \times n} \) denotes the \( n(n+1)/2 \)-dimensional real vector space of \( n \)-dimensional symmetric matrices and \( Q_{n,+} \) denotes the corresponding cone of positive definite matrices. We denote as \( \mathcal{V}(\mathcal{S}_+^m) \) the linear space generated by \( \mathcal{S}_+^m(T) \). Finally, we introduce the linear operator

\[
\Gamma : \mathcal{V}(\mathcal{S}_+^m) \to Q_n, \quad \Phi \mapsto \int G \Phi G^*. \tag{4}
\]

An equivalent condition, \( [18] \), is that the kernel of the linear operator

\[
V : Q_n \to Q_n, \quad Q \mapsto \Pi^\perp_G (Q - A \Sigma C) \Pi^\perp_G \tag{5}
\]

contains \( P \), namely \( V(P) = 0 \). Here, \( \Pi^\perp_G := I - B(B^T B)^{-1} B^T \). It turns out that the spectrum approximation problem is feasible if and only if \( \Sigma \in \text{Range } \Gamma \cap Q_{n,+} \). \( \mathbb{R}^{n \times n} \). Let \( x_1 \ldots x_N \) be the output data generated by feeding the filters bank with the finite-length data \( y_1 \ldots y_N \). An estimate of \( \Sigma \) is therefore given by the sample covariance matrix \( \hat{\Sigma}_C := \frac{1}{N} \sum_{k=1}^N x_k x_k^T \) which is normally positive definite. It may not, however, belong to \( \text{Range } \Gamma \). Accordingly, we need to compute a new estimate \( \hat{\Sigma} \in \text{Range } \Gamma \) which is positive definite and “close” to the estimate \( \hat{\Sigma}_C \). Hence, we have to solve a structured covariance estimation problem which lead us to consider the following optimization task.

Problem 1: Given \( \hat{\Sigma}_C > 0 \),

\[
\minimize_{P \in Q_{n,+}} D(P|\hat{\Sigma}_C) \text{ over the set } \{ P \in Q_{n,+} \mid V(P) = 0 \}.
\]

Here, \( D \) is a suitable divergence index among (positive definite) covariance matrices. Furthermore, by choosing \( D \) convex with respect to \( P \), Problem 1 can be efficiently solved by means of convex optimization. For instance, in \( [18] \) the

\[^1\] Here and throughout the paper, integration, when not otherwise specified, is on the unit circle with respect to the normalized Lebesgue measure. Moreover, a star denotes transposition plus conjugation.
information divergence among two Gaussian densities with covariance $P$ and $Q$, respectively, has been considered:

$$D_l(P||Q) := \frac{1}{2} \text{tr} [\log(Q) - \log(P) + PQ^{-1} - I].$$  

(6)

Another approach characterizes $\Sigma$ in terms of the filter parameters and the sequence of the covariance lags of $y$. Once we have $\Sigma$ in such a way that the spectrum approximation problem is feasible, we can replace $G$ with $\hat{G} = \Sigma^{-\frac{1}{2}} G$ and $(A, B)$ with $(\hat{A} = \Sigma^{-\frac{1}{2}} A \Sigma^{-\frac{1}{2}}, \hat{B} = \Sigma^{-\frac{1}{2}} B)$. Thus, the constraint may be rewritten as $\int \hat{G} \Phi G^* = I$. Accordingly, from now on we assume that the spectrum approximation problem in (2) is feasible and we consider the following equivalent formulation.

Problem 2: Given $\Psi \in S_m^+(T)$ and $G(z) = (zI - A)^{-1} B$ such that $I \in \text{Range } \Gamma$,

$$\text{minimize } S(\Phi||\Psi) \text{ over the set } \left\{ \Phi \in S_m^+(T) \mid \int G \Phi G^* = I \right\}.$$  

(7)

B. Choice of the divergence index

A divergence index among spectral densities in $S_m^+(T)$ must satisfy the following basic property for all $\Phi, \Psi \in S_m^+(T)$:

$$S(\Phi||\Psi) \geq 0 \quad \text{and only if } \Phi = \Psi.$$  

(8)

Moreover, the corresponding Problem 2 should lead to a computable solution, by typically solving the dual optimization problem. In [15], a Kullback-Leibler divergence for multivariate spectral densities with the same trace of the zeroth-moment has been introduced

$$S_{KL}(\Phi||\Psi) := \int \text{tr} [\Phi \log(\Phi) - \log(\Psi)]$$  

(9)

where $\log(.)$, whose definition will be given in Section III-B, is the matrix logarithm. This divergence is inspired by the Umegaki-von Neumann’s relative entropy [27] of statistical quantum mechanics. Moreover, may be readily extended to the general case, see [28] for the scalar case,

$$S_{KL}(\Phi||\Psi) = \int \text{tr} [\Phi \log(\Phi) - \log(\Psi)] - \Phi + \Psi$$  

(10)

and $S_{KL}(\Phi||\Psi) = S_{KL}(\Psi||\Phi)$ when $\int \text{tr} \Phi = \int \text{tr} \Psi$. However, the resulting solution to the spectrum approximation problem is generically non-rational. On the contrary, by considering the multivariate extension of the Itakura-Saito distance

$$S_{IS}(\Phi||\Psi) = \int \text{tr} [\log(\Psi) - \log(\Phi) + \Phi^{-1} - I],$$

the solution is rational when $\Psi$ is rational, [17]. We will show in the following section that the divergence indexes (10) and (11) belongs to the same multivariate Beta divergence family. Moreover, this family leads, under a suitable choice of $\Psi$, to a family of solutions to the spectrum approximation problem.

Observe that, it is also possible to rewrite Problem 2 by considering $S_{KL}(\Phi||\Psi)$. The resulting solution is, however, only computable when $\Psi$ is a scalar process [14], or $\Psi = I$. [5], [30], [15]. Finally, we mention that there exists another multivariate distance, called Hellinger distance, which gives a rational solution to Problem [2] [16].

III. BETA DIVERGENCE FAMILY FOR SPECTRAL DENSITIES

In this section we extend the notion of Beta divergence (family) for scalar spectral densities, firstly introduced in [20] and [22], to the multivariate case. All the proofs of the propositions stated in this section are placed in Appendix B.

A. Scalar case

We recall the definition of the scalar Beta divergence by adopting the same notation employed in [28]. First of all, we need to introduce the following function

$$\log_c : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$$

$$(x, y) \rightarrow \begin{cases} \frac{1}{1-c} (\frac{x}{y})^{1-c} - 1, & c \in \mathbb{R} \setminus \{1\} \\ \log(x) - \log(y), & c = 1 \end{cases}$$

which is referred to as generalized logarithm discrepancy function throughout the paper. Notice that $\log_c$ is a continuous function of real variable $c$ and $\log_c(x, y) = 0$ if and only if $x = y$. The (asymmetric) Beta divergence between two scalar spectral densities $\Phi, \Psi \in S_1^+(T)$ is defined by

$$S_\beta(\Phi||\Psi) := -\frac{1}{\beta} \int (\Phi^\beta \log(\Phi^\beta, \Phi^\beta) + \Phi^\beta - \Psi^\beta)$$

$$= \int \left( \frac{1}{\beta - 1} (\Phi^\beta - \Psi^\beta) - 1 \right)$$

where the parameter $\beta$ is a real number. For $\beta = 0$ and $\beta = 1$, it is defined by continuity in the following way

$$\lim_{\beta \rightarrow 0} S_\beta(\Phi||\Psi) = S_{IS}(\Phi||\Psi) \quad \text{and} \quad \lim_{\beta \rightarrow 1} S_\beta(\Phi||\Psi) = S_{KL}(\Phi||\Psi),$$

where $S_{IS}$ and $S_{KL}$ are the scalar versions of (11) and (10), respectively. Moreover, the Beta divergence is a continuous function of real variable $\beta$ in the whole range including singularities. Thus, it smoothly connects the Itakura-Saito distance with the Kullback-Leibler divergence. Since property (8) is satisfied, $S_\beta$ is a divergence index. Finally, $S_\beta$ is always strictly convex in the first argument, but is often not in the second argument.

B. Multivariate case

Likewise to the scalar case, we start by introducing the generalized multivariate logarithm discrepancy. To this aim, recall that the exponentiation of a positive definite matrix $X$ to an arbitrary real number $c$, is defined as $X^c := U \text{diag}(d_1, \ldots, d_m) U^T$ where $X := U \text{diag}(d_1, \ldots, d_m) U^T$ is the usual spectral decomposition with $U$ orthogonal, i.e. $UU^T = I$, and $\text{diag}(d_1, \ldots, d_m) > 0$ diagonal matrix. We

\footnote{It is also possible to take the exponentiation of positive semidefinite matrices when $c \neq 0$.}
are now ready to extend the definition of generalized logarithm discrepancy to the multivariate case

\[
\log_c : \mathbb{Q}_{m,+} \times \mathbb{Q}_{m,+} \to \mathbb{R}^{m \times m},
\]

\[
(X, Y) \mapsto \left\{ \begin{array}{ll}
\frac{1}{c} \log(X) - \log(Y), & c \in \mathbb{R} \setminus \{1\} \\
1, & c = 1
\end{array} \right.
\]

(11)

where \( \log(X) = U \text{diag}(\log(d_1), \ldots, \log(d_m)) U^T \) is the matrix logarithm of \( X \).

**Proposition 3.1:** The generalized multivariate logarithm discrepancy is a continuous function of real variable \( c \) in the whole range. Moreover, \( \log_c(X, Y) = 0 \) if and only if \( X = Y \). The exponentiation of a spectral density \( \Phi(e^{j\vartheta}) \in \mathbb{S}^m_+(\mathbb{T}) \) to an arbitrary real number \( c \) is pointwise defined by using the previous spectral decomposition:

\[
\Phi(e^{j\vartheta})^c = U(e^{j\vartheta}) \text{diag}(d_1(e^{j\vartheta})^c, \ldots, d_m(e^{j\vartheta})^c) U(e^{j\vartheta})^T
\]

where \( \Phi(e^{j\vartheta}) = U(e^{j\vartheta}) \text{diag}(d_1(e^{j\vartheta}), \ldots, d_m(e^{j\vartheta})) U(e^{j\vartheta})^T \) with \( U(e^{j\vartheta}) \in \mathbb{L}^{m \times m}_+ (\mathbb{T}) \) such that \( U(e^{j\vartheta}) U(e^{j\vartheta})^T = I \). Observe that \( \Phi^c \) belongs to \( \mathbb{S}^m_+(\mathbb{T}) \). We are now ready to introduce the multivariate (asymmetric) Beta divergence among \( \Phi, \Psi \in \mathbb{S}^m_+(\mathbb{T}) \):

\[
S_\beta(\Phi \| \Psi) := \frac{1}{\beta} \int \left[ \psi^\beta \log \frac{\psi}{\Phi} (\psi^\beta - \Phi^\beta) + \Phi^\beta - \Psi^\beta \right]
\]

\[
= \int \left[ \frac{1}{\beta - 1} (\Phi^\beta - \Phi \Psi^\beta) - \frac{1}{\beta} (\Phi^\beta - \Psi^\beta) \right]
\]

(13)

where \( \beta \in \mathbb{R} \setminus \{0, 1\} \). Similarly to the scalar case, we can extend by continuity the definition of Beta divergence for \( \beta = 0 \) and \( \beta = 1 \).

**Proposition 3.2:** The following limits hold:

\[
\lim_{\beta \to 0} S_\beta(\Phi \| \Psi) = S_{IS}(\Phi \| \Psi)
\]

\[
\lim_{\beta \to 1} S_\beta(\Phi \| \Psi) = S_{KL}(\Phi \| \Psi).
\]

IV. SPECTRUM APPROXIMATION PROBLEM

Since the Beta divergence is well-defined for \( \beta \in \mathbb{R} \), we choose \( \beta = -\frac{1}{2} + 1 \) with \( \nu \in \mathbb{N}_+ \) and we define \( S_\nu(\Phi \| \Psi) := S_\beta(\Phi \| \Psi) \) with \( \beta = -\frac{1}{2} + 1 \). Moreover, here and in the remainder of the paper we assume that \( \Psi(z) \) is a rational matrix function. The aim of this section and Section V is to prove the following statement.

**Theorem 4.1:** Given \( \Psi \in \mathbb{S}^m_+(\mathbb{T}) \) such that \( \Psi(z) \) is rational, and \( G(z) \) such that \( I \in \text{Range } \Gamma \), the problem

\[
\min \{ \Phi \in \mathbb{S}^m_+(\mathbb{T}) \mid \int G \Phi G^* = I \}
\]

always admits a unique solution when \( \nu \in \mathbb{N}_+ \). Moreover, such a solution is rational with McMillan degree less than or equal to \( \nu(\deg(\Psi^\frac{1}{2}) + 2\nu) \).

Since (14) is a constrained convex optimization problem, we consider the corresponding Lagrange functional

\[
\mathcal{L}_\nu(\Phi, \Lambda) = S_\nu(\Phi \| \Psi) + \frac{\nu}{1 - \nu} \int \text{tr}[\Lambda \Psi^{-\frac{1}{2}} - \Phi^{-\frac{1}{2}}] + \int G \Phi G^* - I, \lambda
\]

\[
= \int \text{tr}[\nu(\Phi^{-\frac{1}{2}} - \Psi^{-\frac{1}{2}}) + \frac{\nu}{1 - \nu} \Phi^{-\frac{1}{2}} + G^* \Lambda G \Phi] - \text{tr}[\Lambda]
\]

where we exploited the fact that the term \( \int \text{tr}[\Psi^{-\frac{1}{2}}] \) plays no role in the optimization problem. Note that, the Lagrange multiplier \( \Lambda \in \mathbb{Q}_m \) can be uniquely decomposed as \( \Lambda = \Lambda^G + \Lambda^L \) where \( \Lambda^G \in \text{Range } \Gamma \), \( \Lambda^L \in \text{Range } \Gamma^\perp \). Since \( \Lambda^L \) is such that \( G^*(e^{j\vartheta}) \Lambda^L G(e^{j\vartheta}) \equiv 0 \) and \( \text{tr}[\Lambda^L] = \langle \Lambda^L, I \rangle = 0 \) (see [31], Section III), it does not affect the Lagrangian, i.e., \( \mathcal{L}_\nu(\Phi, \Lambda) = \mathcal{L}_\nu(\Phi, \Lambda^G) \). Accordingly we can impose from now on that \( \Lambda \in \text{Range } \Gamma \).

Consider now the unconstrained minimization problem

\[
\min \{ \mathcal{L}_\nu(\Phi, \Lambda) \mid \Phi \in \mathbb{S}^m_+(\mathbb{T}) \}.
\]

Since \( \mathcal{L}_\nu(\cdot, \Lambda) \) is strictly convex over \( \mathbb{S}^m_+(\mathbb{T}) \), its unique minimum point \( \Phi_\nu \) is given by annihilating its first variation in each direction \( \delta \Phi \in L^m_{\text{infty}}(\mathbb{T}) \):

\[
\delta \mathcal{L}_\nu(\Phi, \Lambda; \delta \Phi) = \int \text{tr}[\nu (\Psi^{-\frac{1}{2}} - \Phi^{-\frac{1}{2}}) + G^* \Lambda G \Phi] \delta \Phi
\]

(15)

where we exploited (18). Note that, \( \nu(\Psi^{-\frac{1}{2}} - \Phi^{-\frac{1}{2}}) + G^* \Lambda G \in L^m_{\text{infty}}(\mathbb{T}) \). Thus, (15) is zero \( \forall \delta \Phi \in L^m_{\text{infty}}(\mathbb{T}) \) if and only if

\[
\Phi^{-\frac{1}{2}} = \Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda G \Phi.
\]

Since \( \Phi^{-\frac{1}{2}} \in \mathbb{S}^m_+(\mathbb{T}) \), the set of the admissible Lagrange multipliers is

\[
\mathcal{L}_\nu := \{ \Lambda \in \mathbb{Q}_m \mid \Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda G > 0 \text{ on } \mathbb{T} \}.
\]

Therefore, the natural set for \( \Lambda \) is

\[
\mathcal{L}^\nu = \mathcal{L}_\nu \cap \text{Range } \Gamma.
\]

\( ^3 \mathbb{N}_+ \) denotes the set of the positive natural numbers.
In conclusion, the unique minimum point of the Lagrange functional has the form
\[
\Phi_\nu(\Lambda) := \left( \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G \right)^{-\nu}.
\] (16)

**Proposition 4.1:** If \( \Phi_\nu \) is a minimizer of Problem \( \nu \) then it is a rational function with McMillan degree less than or equal to \( \nu (\deg[\Psi^1] + 2n) \). Moreover, the following facts hold:

1) If \( \Psi \) is constant then, among all the spectral density \( \Phi_\nu \) with \( \nu \in \mathbb{N}_+ \), the spectral density with the smallest upper bound on the McMillan degree corresponds to the Itakura-Saito distance

2) As \( \nu \to +\infty \), \( \Phi_\nu \) tends to the spectral density corresponding to the Kullback-Leibler divergence.

**Proof:** Since both \( \Psi^\lambda \) and \( G \) are rational matrix functions and \( \nu \) is an integer, then also \( \Phi_\nu \) is rational. Moreover, in view of (16), \( \deg[\Phi_\nu] \leq \nu (\deg[\Psi^1] + 2n) \), where \( n \) is the McMillan degree of \( G(z) \).

1) Since \( \Psi \) is constant, we get \( \deg[\Phi_\nu] \leq \nu 2n \) with \( \nu \in \mathbb{N}_+ \). Thus, the spectral density with the smallest upper bound on the McMillan degree is attained for \( \nu = 1 \), i.e. \( \beta = 0 \), which is the optimal form related to \( S_{\| \Phi \|}\Psi \). Note that, \( \Phi_1(\Lambda) = (\Psi^{-1} + G^* \Lambda G)^{-1} \), which is the same optimal form found in (17) for the multivariate Itakura-Saito distance.

2) Firstly, it is possible to show that the optimal form obtained by using the Kullback-Leibler divergence is \( \Phi_{KL}(\Lambda) = e^{\log(\Psi)} G^* \Lambda G \), which is a straightforward generalization of the optimal form for \( S_{KL} \) presented in (15). We want to show that \( \Phi_\nu \to \Phi_{KL} \) as \( \nu \to +\infty \). Let us consider the function \( F(\Lambda) := \log(\Psi^{-\lambda} + \lambda G^* A G) \) with \( \lambda \in \mathbb{R} \) such that \( \Psi^{-\lambda} + \lambda G^* \Lambda G > 0 \) on \( \mathbb{R} \). Its first order Taylor expansion with respect to \( \lambda = 0 \) is \( \Psi^{-\lambda} + \lambda G^* \Lambda G - I \). Accordingly,

\[
\lim_{\nu \to +\infty} \nu \log \left( \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G \right) = \lim_{\nu \to +\infty} \Psi^{-\frac{1}{\nu}} - I + G^* \Lambda G - \log(\Psi) + G^* \Lambda G
\]
where we exploited (30) and the previous Taylor expansion. Finally,

\[
\lim_{\nu \to +\infty} \Phi_\nu(\Lambda) = \lim_{\nu \to +\infty} e^{\log \left( \left( \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G \right)^{-\nu} \right)} = e^{\log(\Psi) - G^* \Lambda G} = \Phi_{KL}(\Lambda).
\]

In this section we showed that \( \Phi_\nu(\Lambda) \) is the unique minimum point of \( L_\nu(\cdot, \Lambda) \), namely
\[
L_\nu(\Phi_\nu(\Lambda), \Lambda) < L_\nu(\Phi, \Lambda), \ \forall \Phi \in S_{\nu}^\nu(\mathbb{T})
\]
s.t. \( \Phi \neq \Phi_\nu(\Lambda), \ \Lambda \in \mathcal{L}_\nu^\nu \). (17)

Hence, if we produce \( \Lambda^0 \in \mathcal{L}_\nu^\nu \) such that \( \Phi_\nu(\Lambda^0) \) is satisfying the constraint in (7), inequality (17) implies \( S_{\nu}(\Phi_\nu(\Lambda^0)) \leq S_{\nu}(\Phi \| \Psi), \ \forall \Phi \in S_{\nu}^\nu(\mathbb{T}) \) s.t. \( G^* \Phi G = I \) and equality holds if and only if \( \Phi = \Phi_\nu(\Lambda^0) \), namely such a \( \Phi_\nu(\Lambda^0) \) is the unique solution to Problem \( \nu \) with \( S_{\nu} \). The following step consists in showing the existence of such a \( \Lambda^0 \) by using the duality theory.

**V. DUAL PROBLEM**

Here, we deal with the case \( \nu \in \mathbb{N}_+ \setminus \{1\} \), since the existence of the solution to the dual problem for \( \nu = 1 \) was already showed in (17). The dual problem consists in maximizing the functional
\[
\inf_{\Phi} L_\nu(\Phi, \Lambda) = L_\nu(\Phi_\nu(\Lambda), \Lambda) = \frac{\nu}{1 - \nu} \int \left( \left( \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G \right)^{-\nu} \right) - \text{tr}[\Lambda]
\]

where we recall that \( \Psi^{-1} \) (and thus also \( \Psi^{-\frac{1}{\nu}} \)) is by assumption a rational matrix function. Hence, it is equivalent to minimize the following functional hereafter referred to as dual functional:
\[
J_\nu(\Lambda) = \frac{\nu}{1 - \nu} \int \left( \left( \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G \right)^{-\nu} \right) + \text{tr}[\Lambda].
\]

**Theorem 5.1:** The dual functional \( J_\nu \) belongs to \( C^\infty(\mathcal{L}_\nu^\nu) \) and it is strictly convex over \( \mathcal{L}_\nu^\nu \).

**Proof:** In order to prove the statement, we need the following first variation of the map \( X \mapsto \text{tr}[X^c] \) (further details may be found in Appendix A):
\[
\delta(\text{tr}[X^c]; \delta X) = c \text{tr}[X^{c-1} \delta X].
\] (18)

The first variation of \( J_\nu(\Lambda) \) in direction \( \delta \Lambda_1 \in \mathbb{R}_n \) is
\[
\delta J_\nu(\Lambda; \delta \Lambda_1) = -\int \left( \left( \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G \right)^{-\nu} G^* \delta \Lambda_1 G \right) + \text{tr}[\delta \Lambda_1]
\]
\[
= -\int \left( (W(I + G^* \Lambda G_1)^{-1} W)^* G^* \delta \Lambda_1 G \right) + \text{tr}[\delta \Lambda_1]
\]
\[
= -\int \left( (W(I + G^* \Lambda G_1)^{-1} W)^* G^* \delta \Lambda_1 G \right) + \text{tr}[\delta \Lambda_1].
\] (19)

The linear form \( \nabla J_\nu(\Lambda; \cdot) := \delta J_\nu(\Lambda; \cdot) \) is the gradient of \( J_\nu \) at \( \Lambda \). In order to prove that \( J_\nu(\Lambda) \in C^1(\mathcal{L}_\nu^\nu) \) we have to show that \( \delta(\lambda_1; \delta \Lambda_1) \), for any fixed \( \delta \Lambda_1 \), is continuous in \( \Lambda \). To this aim, consider a sequence \( M_n \in \text{Range} \Gamma \) such that \( M_n \to 0 \) and define \( Q_n(z) = W(z)(I + G_1(z)^* N G_1(z))^{-1} W(z)^* \) with \( N \in \mathbb{R}_n \). By Lemma 5.2 in (31) and since \( W \) is bounded
on $\mathbb{T}$, $Q_{\Lambda^{\alpha} M_\nu}$ converges uniformly to $Q_{\Lambda}$, Thus, applying
the bounded convergence theorem, we obtain
$$
\lim_{n \to \infty} \int GQ'_{\Lambda^{\alpha} M_\nu} G^* = \int GQ'_{\Lambda} G^*.
$$
Accordingly, $\delta(J_\nu(\Lambda); \delta \Lambda)$ is continuous, i.e. $J_\nu$ belongs to $C^1(L^\nu_\nu)$. In order to compute the second variation, notice that
$$
Q_\Lambda = \left( \Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda G \right)^{-1}
$$
and its first variation in direction $\delta \Lambda \in \mathcal{Q}_n$ is
$$
\delta Q_\Lambda ; \delta \Lambda = -\frac{1}{\nu} Q_\Lambda G^* \delta \Lambda GGQ_\Lambda.
$$
Furthermore, consider the operator $\mathcal{I} : A \mapsto A'$. By applying the chain rule, we get
$$
\delta(\mathcal{I}(A); \delta A) = \sum_{i=1}^\nu A'^{-1} \delta A A'^{-1}.
$$
Since
$$
\delta J_\nu(\Lambda; \delta \Lambda_1) = -\int \text{tr} \bigl[ Q_{\Lambda^1} G^* \delta \Lambda_1 G + \text{tr}[\delta \Lambda_1] \bigr],
$$
the second variation of $J_\nu$ in direction $\delta \Lambda_1, \delta \Lambda_2 \in \mathcal{Q}_n$ is
$$
\delta^2 J_\nu(\Lambda; \delta \Lambda_1, \delta \Lambda_2) = -\sum_{i=1}^\nu \int \text{tr} \bigl[ Q_{\Lambda^i}^{-1} \delta Q_\Lambda \delta \Lambda_2 Q_{\Lambda^i}^{-1} G^* \delta \Lambda_1 G \bigr]
= -\frac{1}{\nu} \sum_{i=1}^\nu \int \text{tr} \bigl[ Q_{\Lambda^i} G^* \delta \Lambda_2 Q_{\Lambda^i}^{-1} G^* \delta \Lambda_1 G \bigr].
$$
The bilinear form $H_{\nu, \Lambda}(\cdot, \cdot) = \delta^2 J_\nu(\Lambda; \cdot, \cdot)$ is the Hessian of $J_\nu$ at $\Lambda$. The continuity of $\delta^2 J_\nu$ can be established by using the previous argumentation. In similar way, we can show that $J_\nu$ has continuous directional derivatives of any order, i.e. $J_\nu \in C^k(L^\nu_\nu)$ for any $k$. Finally, it remains to be shown that $J_\nu$ is strictly convex on the open set $L^\nu_\nu$. Since $J_\nu \in C^\infty(L^\nu_\nu)$, it is sufficient to show that $H_{\nu, \Lambda}(\delta \Lambda, \delta \Lambda) \geq 0$ for each $\delta \Lambda \in \text{Range} \Gamma$ and equality holds if and only if $\delta \Lambda = 0$. Since $\nu > 0$ and the trace of integrands in (20) is positive semidefinite when $\delta \Lambda_1 = \delta \Lambda_2$, we have $H_{\nu, \Lambda}(\delta \Lambda, \delta \Lambda) \geq 0$. If $H_{\nu, \Lambda}(\delta \Lambda, \delta \Lambda) = 0$, then $G^* \delta \Lambda G \equiv 0$ namely $\delta \Lambda \in \text{Range} \Gamma$ (see [31] Section III). Since $\delta \Lambda \in \text{Range} \Gamma$, it follows that $\delta \Lambda = 0$. In conclusion, the Hessian is positive definite and the dual functional is strictly convex on $L^\nu_\nu$. \\
In view of Theorem 3.1, the dual problem
$$
\min_{\Lambda} \left\{ J_\nu(\Lambda) \mid \Lambda \in L^\nu_\nu \right\}
$$
admits at most one solution $\Lambda^\circ$. Since $L^\nu_\nu$ is an open set, such a $\Lambda^\circ$ (if it does exist) annihilates the first directional derivative (19) for each $\delta \Lambda \in \mathcal{Q}_n$,
$$
\left\langle I - \int G \left( \Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda^\circ G \right)^{-\nu} G^*, \delta \Lambda \right\rangle = 0 \quad \forall \delta \Lambda \in \mathcal{Q}_n
or, equivalently,
$$
I = \int G \left( \Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda^\circ G \right)^{-\nu} G^* = \int G \Phi_\nu(\Lambda^\circ) G^*.
$$
This means that $\Phi_\nu(\Lambda^\circ) \in S^N_\nu(\mathbb{T})$ satisfies the constraint in (7) and $\Phi_\nu(\Lambda^\circ)$ is therefore the unique solution to Problem 2.

The next step concerns the existence issue for the dual problem. Although the existence question is quite delicate, since set $L^\nu_\nu$ is open and unbounded, we will show that a $\Lambda^\circ$ minimizing $J_\nu$ over $L^\nu_\nu$ does exist.

**Theorem 5.2:** Let $\nu \in \mathbb{N}_+ \setminus \{1\}$, then the dual functional $J_\nu$ has a unique minimum point in $L^\nu_\nu$.

**Proof:** Since the solution of the dual problem (if it does exist) is unique, we only need to show that $J_\nu$ takes a minimum value on $L^\nu_\nu$. First of all, note that $J_\nu$ is continuous on $L^\nu_\nu$, see Theorem 5.1. Secondly, we show that $\text{tr}[\Lambda]$ is bounded from below on $L^\nu_\nu$. Since Problem 2 is feasible, there exists $\Phi_1 \in S^N_\nu(\mathbb{T})$ such that $\int G \Phi_1 G^* = I$. Thus,
$$
\text{tr}[\Lambda] = \text{tr} \left[ \int G \Phi_1 G^* \Lambda \right] = \text{tr} \left[ \int G^* \Lambda G \Phi_1 \right].
$$
Defining $\alpha = -\nu \text{tr} \Psi^{-\frac{1}{2}} \Phi_1$, we obtain
$$
\text{tr}[\Lambda] = \nu \text{tr} \left[ \int \left( \Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda G \right) \Phi_1 \right] + \alpha.
$$
Since $\Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda G$ is positive definite on $\mathbb{T}$ for $\Lambda \in L^\nu_\nu$, there exists a right spectral factor $\Delta$ such that $\Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda G = \Delta^* \Delta$. Moreover, $\Phi_1$ is a coercive spectrum, namely there exists a constant $\mu > 0$ such that $\Phi_1(e^{i\theta}) \geq \mu I$, $\forall e^{i\theta} \in \mathbb{T}$. Starting from the fact that the trace and the integral are monotonic functions, we get
$$
\text{tr}[\Lambda] = \nu \text{tr} \left[ \int \Delta \Phi_1 \Delta^* \right] + \alpha \geq \nu \mu \text{tr} \left[ \int \Delta \Delta^* \right] + \alpha
= \nu \mu \text{tr} \left[ \Psi^{-\frac{1}{2}} + \frac{1}{\nu} G^* \Lambda G \right] + \alpha > \alpha
$$
where we have used the fact that $\text{tr} \Psi^{-\frac{1}{2}} > 0$ when $\Lambda \in L^\nu_\nu$. Finally, notice that $J_\nu(0) = -\nu \text{tr} \left[ \Psi^{-\frac{1}{2}} \right]$. Accordingly, we can restrict the search of a minimum point to the set $\left\{ \Lambda \in L^\nu_\nu \mid J_\nu(\Lambda) \leq J_\nu(0) \right\}$. Now we show that this set is compact. Accordingly, the existence of the solution to the dual problem follows from the Weierstrass’ Theorem. To prove the compactness of the set, it is sufficient to show that:
$$
1) \lim_{\Lambda \to \partial L^\nu_\nu} J_\nu(\Lambda) = +\infty;
2) \lim_{\|\Lambda\| \to \infty} J_\nu(\Lambda) = +\infty.
$$
1) Firstly, recall that $\Psi(z)^{-\frac{1}{2}}$ is rational by assumption, thus
$$
R_\Lambda(z) := \Psi(z)^{-\frac{1}{2}} + \frac{1}{\nu} G(z)^* \Lambda G(z)
$$
is a rational matrix function. Let $\lambda_{\Lambda, i} (z), i = 1 \ldots m$, denote the eigenvalues of $R_\Lambda(z)$. In view of (12), the eigenvalues of $R_\Lambda(z)^1-\nu$ are $\lambda_{\Lambda, i}(z)^{1-\nu}$. Moreover, $\text{tr}[R_\Lambda(z)^1-\nu] = \sum_{i=1}^m \lambda_{\Lambda, i}(z)^{1-\nu}$ is a rational function because $R_\Lambda(z)$ is a rational matrix function and $\nu - 1 \in \mathbb{N}_+$. Observe that $L^\nu_\nu$ is the set of $\Lambda \in \text{Range} \Gamma$ such that $\lambda_{\Lambda, i}(e^{i\theta}) \geq 0$ on $\mathbb{T}$ and there exists $\theta$ and $i$ such that $\lambda_{\Lambda, i}(e^{i\theta}) = 0$. Thus for $\Lambda \to \partial L^\nu_\nu, \lambda_{\Lambda, i}(z)^{1-\nu}$ has a pole tending to $e^{i\theta} \in \mathbb{T}$. Accordingly, $\lim_{\|\Lambda\| \to \infty} R_\Lambda(\Lambda)^{1-\nu} \to \infty$ as $\Lambda \to \partial L^\nu_\nu$. In view of (21), we conclude that $J_\nu(\Lambda) =$
Moreover, let $\Lambda^0 = \frac{\Lambda}{\|\Lambda\|}$. Since $L^\nu_0$ is convex and $0 \leq L^\nu_0$, if $\Lambda \in L^\nu_0$ then $\xi \Lambda \in L^\nu_0 \quad \forall \xi \in [0,1]$. Therefore $\Lambda^0 \in L^\nu_0$ for $k$ sufficiently large. Let $\eta := \lim \inf \|\Lambda^0_k\|$. In view of (21),
\[
\begin{align*}
\text{tr}[\Lambda^0_k] &= \frac{1}{\|\Lambda_k\|} \text{tr}[\Lambda_k] > \frac{1}{\|\Lambda_k\|} \alpha \to 0,
\end{align*}
\]
for $k \to \infty$, so $\eta > 0$. Thus, there exists a subsequence of $\{\Lambda^0_k\}$ such that the limit of its trace is equal to $\eta$. Moreover, this subsequence remains on the surface of the unit ball $\partial B = \{\Lambda = \Lambda^T \mid \|\Lambda\| = 1\}$ which is compact. Accordingly, it has a subsequence $\{\Lambda^0_{k_i}\}$ converging in $\partial B$. Let $\Lambda^\infty \in \partial B$ be its limit, thus $\lim \text{tr}[\Lambda^0_{k_i}] = \text{tr}[\Lambda^\infty] = \eta$. We now prove that $\Lambda^\infty \in L^\nu_0$. First of all, note that $\Lambda^\infty$ is the limit of a sequence in the finite dimensional linear space $\Gamma$, hence $\Lambda^\infty \in \Gamma$. It remains to be shown that $\Psi^{-\frac{1}{2}} + \frac{1}{\nu} \lambda^+ \Lambda^\infty G$ is positive definite on $\Gamma$. Consider the subsequences $\{\Lambda^0_k\} \subset L^\nu_0$: We have that $\Psi^{-\frac{1}{2}} + \frac{1}{\nu} \lambda^+ \Lambda^0_k G > 0$ on $\Gamma$ so that $\frac{1}{\nu} \lambda^+ \Psi^{-\frac{1}{2}} + \frac{1}{\nu} \lambda^+ \Lambda^0_k G$ is also positive definite on $\Gamma$ for each $i$. Taking the limit for $i \to \infty$, we get $G \lambda^+ \Lambda^\infty G$ is positive semidefinite on $\Gamma$ so that $\Psi^{-\frac{1}{2}} + \frac{1}{\nu} \lambda^+ \Lambda^\infty G > 0$ on $\Gamma$. Hence, $\Lambda^\infty \in L^\nu_0$. Since Problem 2 is feasible, there exists $\Phi \in S^m_\nu(\Gamma)$ such that $\Gamma = f(G \Phi \Gamma)$, accordingly
\[
\begin{align*}
\eta = \text{tr}[\Lambda^\infty] &= \text{tr} \int G \Phi_i G^* \Lambda^\infty = \text{tr} \int \Phi_i \frac{\nu}{\nu} G \Lambda^\infty G \Phi_i.
\end{align*}
\]
Moreover, $\Lambda^\infty G^*G$ is not identically equal to zero. In fact, if $G \lambda^+ \Lambda^\infty G \equiv 0$, then $\Lambda^\infty \in [\text{Range} \Gamma]^{-1}$ and $\Lambda^\infty \neq 0$ since it belongs to the surface of the unit ball. This is a contradiction because $\Lambda^\infty \in \Gamma$. Thus, $G \lambda^+ \Lambda^\infty G$ is not identically zero and $\eta > 0$. Finally, we have
\[
\begin{align*}
\lim_{k \to \infty} J_\nu(\Lambda^0_k) &= \lim_{k \to \infty} \frac{\nu}{1 - \nu} \text{tr} \left[ \left( \Psi^{-\frac{1}{2}} + \frac{1}{\nu} \lambda^+ \Lambda^0_k G \right)^{1-\nu} \right] + \text{tr}[\Lambda_k] > \lim_{k \to \infty} \|\Lambda_k\| \text{tr}[\Lambda^0_k] = \eta \lim_{k \to \infty} \|\Lambda_k\| = \infty.
\end{align*}
\]
\[\text{Remark 5.1: For the case } \nu \in \mathbb{Z} \text{ such that } \nu < 0, \text{ still holds. Moreover, if } \{\nu\} \text{ is a minimizer of Problem 2 then it is a rational matrix function with } \deg[\Phi_\nu] < \nu (\deg[\Psi^\frac{1}{2}] + 2m). \text{ However, the dual problem may not have solution: The minimum point for } J_\nu(\Lambda) \text{ may lie on } \partial L^\nu, \text{ since } J_\nu \text{ takes finite values on the boundary of } L^\nu.\]

\[\text{VI. COMPUTATION OF } \Lambda^\nu \]

We showed that the dual problem always admits a unique solution $\Lambda^\nu$ on $L^\nu$ for $\nu \in \mathbb{N}_+$. In order to find $\Lambda^\nu$, we use the following matrical Newton algorithm with backtracking stage proposed in [31]:
\begin{enumerate}
\item Set $\Lambda_0 = I \in L^\nu$;
\item At each iteration, compute the Newton step $\Delta_\nu$, by solving the linear equation $H_{\nu,\Lambda}(\Delta_\nu, \cdot) = -\nabla J_{\nu,\Lambda}(\cdot)$ where, once fixed $\Lambda$, $\nabla J_{\nu,\Lambda}(\cdot)$ and $H_{\nu,\Lambda}(\cdot, \cdot)$ must be understood as a linear and bilinear form of $[19]$ and $[20]$, respectively;
\item Set $t^0 = 1$ and let $t^{k+1} = t^k / 2$ until both of the following conditions hold:
\begin{align*}
\Delta_\nu + t^k \Delta_\nu &\in L^\nu \quad (22) \\
J_\nu(\Delta_\nu + t^k \Delta_\nu) &< J_\nu(\Lambda_k) + \alpha t^k \langle \nabla J_{\nu,\Lambda}(\Delta_\nu) \rangle (23)
\end{align*}
with $0 < \alpha < 1 / 2$;
\item Set $\Lambda_{k+1} = \Lambda_k + t^k \Delta_\nu$;
\item Repeat steps 2, 3 and 4 until $\|\nabla J_{\nu,\Lambda}(\cdot)\| < \varepsilon$ where $\varepsilon$ is a tolerance threshold. Then set $\Lambda^\nu = \Lambda_k$.
\end{enumerate}

The computation of the search direction $\Delta_\nu$ is the most delicate part of the procedure. The corresponding linear equation reduces to
\[
\begin{align*}
\frac{1}{\nu} \sum_{i=1}^\nu \int GQ_{\Lambda} \lambda^+ \Lambda^\nu GQ_{\Lambda}^{\nu+1-1} G = \int GQ_{\Lambda} \lambda^+ \Lambda^\nu G - I \quad (24)
\end{align*}
\]
where $Q_{\Lambda} = W(I + G^* \lambda^+ \gamma_{\Lambda})^{-1} W^*$. By similar arguments used in [16] Proposition 8.1, it is possible to prove that there exists a unique solution $\Delta_\nu \in \text{Range} \Gamma$ to (24). Accordingly, we can easily compute $\Delta_\nu$ in this way:
\begin{enumerate}
\item Compute
\[
\begin{align*}
Y &= \int GQ_{\Lambda} \lambda^+ \Lambda^\nu G - I;
\end{align*}
\]
\item Compute a basis $\{\Sigma_1 \ldots \Sigma_M\}$ for Range $\Gamma$ from [4] and for each $\Sigma_k$, $k = 1 \ldots M$, compute
\[
\begin{align*}
Y_k &= \frac{1}{\nu} \sum_{i=1}^\nu \int GQ_{\Lambda} \lambda^+ \Sigma_k GQ_{\Lambda}^{\nu+1-1} G^*; \\
\end{align*}
\]
\item Find $\{\alpha_k\}$ such that $Y = \sum_k \alpha_k Y_k$. Then set $\Delta_\nu = \sum_k \alpha_k \Sigma_k$.
\end{enumerate}

Concerning the evaluation of the integrals in (25), (25) and (26), a sensible and efficient method based on spectral factorization techniques may be employed. For further details, including the checking of condition (22), we refer to Section VI in [31].

\[\text{Finally, it is possible to prove that:}\]
\begin{enumerate}
\item $J_\nu(\cdot) \in C^\infty(L^\nu)$ strongly convex on the sublevel set $K = \{\Lambda \in L^\nu \mid J_\nu(\Lambda) \leq J_\nu(\Lambda_0)\}$;
\item The Hessian is Lipschitz continuous in $K$.
\end{enumerate}

\[\text{The proof follows the ones in [31] Section VIII and [17] Section VI-C] faithfully. These properties allow us to conclude that the proposed Newton algorithm globally converges, [32] Chapter 9]. In particular the rate of convergence is quadratic during the last stage. In this way, the solution to Problem 2 may be efficiently computed.}\]

\[\text{VII. SIMULATIONS RESULTS - PART I}\]

In order to test the features of the family of solutions $\Phi_\nu$ with $\nu \in \mathbb{N}_+$, we take into account the following comparison procedure:
\begin{enumerate}
\item Choose a zero mean stationary process $y = \{y_k; k \in \mathbb{Z}\}$ with spectral density $\Omega \in S^m_\nu(\Gamma)$;
\end{enumerate}
2) Design a filters bank \(G(z)\) as in (1);
3) Choose a prior spectral density \(\Psi \in \mathcal{S}^2_+ (\mathbb{T})\) such that \(\tilde{\psi}^\pm\) is rational;
4) Set \(\Sigma = \Sigma \in \Gamma \cap Q_{n,+}\), i.e. the corresponding spectrum approximation problem is feasible;
5) Solve Problem 2 (with \(S_n\)) by means of the proposed algorithm with the chosen \(\Psi\) and \(\Sigma^{-\frac{1}{2}}G(z)\) as filters bank.

In the above comparison procedure we assume to know \(\Sigma\). In this way, we avoid the approximation errors introduced by the estimation of \(\Sigma\) from the finite-length data \(y_1 \ldots y_N\). As noticed in Section II \(\Psi\) incorporates the a priori information on \(y\). More specifically, \(\Psi\) is designed by using some given partial information on \(y\) (e.g. its zeroth moment), or given laws (e.g. physical laws if \(y\) describes a physical phenomenon) which describe its theoretical features. When no a priori information is available, we choose \(\Psi = I\) which represents the spectral density of the most unpredictable random process. Concerning the design of the filter, its role consists in providing the interpolation conditions for the solution to the spectrum approximation problem. More specifically, a higher resolution can be attained by selecting poles in the proximity of the unit circle, with arguments in the range of frequency of interest.

A. Scalar case

We start by considering Example described in [31, Section VIII-B] (the unique difference is that we assume to know \(\Sigma\) and \(\int \Omega\)). Consider the following ARMA process:

\[
y(t) = 0.5y(t - 1) - 0.42y(t - 2) + 0.602y(t - 3) - 0.0425y(t - 4) + 0.1192y(t - 5) + e(t) + 1.1e(t - 1) + 0.08e(t - 2) - 0.15e(t - 3)
\]

where \(e\) is a zero-mean Gaussian white noise with unit variance. In Figure 1 the spectral density \(\Omega \in \mathcal{S}^2_+ (\mathbb{T})\) of the ARMA process is depicted (gray line). \(\Psi\) is equal to \(\int \Omega\) and \(G(z)\) is structured according to the covariance extension setting (3) with 6 covariance lags (i.e. \(n = 6\)). In Figure 1 the different solutions obtained by fixing \(\nu = 1\), dashed line, \(\nu = 2\), solid line, and \(\nu = 3\), thick line, are shown. The solution obtained by minimizing the multivariate Itakura-Saito distance (\(\nu = 1\)) is characterized by peaks which are taller than these in \(\Omega\). On the contrary, the peaks are reduced by increasing \(\nu\). Finally, the solutions with \(\nu = 2\) and \(\nu = 3\) are closer to \(\Omega\) than the one with \(\nu = 1\).

As second example we consider the scalar bandpass random process with spectral density \(\Omega\) depicted in Figure 2 (gray curve). Here, the cutoff frequencies are \(\vartheta_1 = 0.89\) and \(\vartheta_2 = 2.46\). Moreover, \(\Omega(e^{j\vartheta}) \geq 2 \cdot 10^{-3}\) in the stopband, accordingly \(\Omega \in \mathcal{S}^2_+ (\mathbb{T})\). Matrix \(B\) is a column of ones. Matrix \(A\) is chosen as a block-diagonal matrix with one eigenvalue equal to zero and eight eigenvalues equispaced on the circle of radius 0.8

\[
\pm 0.8, \ 0.8 e^{\pm j \frac{\pi}{8}}, \ 0.8 e^{\pm j \frac{2 \pi}{8}}, \ 0.8 e^{\pm j \frac{4 \pi}{8}}.
\]

B. Multivariate case

We consider a bivariate bandpass random process with spectral density \(\Omega\) plotted in Figure 3 (gray curve). Here,

\[
\Psi(z) = (W\Psi(z)W\Psi(z^{-1}))^6,
\]

\[
W\Psi(z) = \frac{5}{6} \left(\frac{z + 0.6}{z - 0.4e^{2j\pi/3}}(z - 0.4e^{-j\pi/3})\right).
\]

In this way \(\tilde{\psi}^\pm\) with \(\nu = 1, \nu = 2, \text{and} \nu = 3\) are rational. Figure 2 also shows \(\Psi\) and the obtained solutions. The one with \(\nu = 1\) turns out inadequate. The solutions with \(\nu = 2\) and \(\nu = 3\) are, instead, similar and closer to \(\Omega\).
The found solutions, however, exhibit peaks which are reduced out the features of the family of solutions. In the above result, consider two scalar spectral densities $\Psi, \Phi \in \mathbb{S}_+^1(\mathbb{T})$. Let $\psi = \Psi(e^{j\bar{\nu}})$, $\phi = \Phi(e^{j\bar{\nu}})$, where $\bar{\nu} \in [0, 2\pi)$ is fixed, and consider the following function

$$s_\nu(\phi, \psi) = \begin{cases} 
\log \psi - \log \phi + \phi \psi^{-1} - 1, & \nu = 1 \\
-\nu(\frac{\nu+1}{\nu} - \phi \psi^{-\frac{1}{\nu}}) - \frac{\nu}{\nu-1}(\frac{\nu+1}{\nu} - \psi^{-\frac{1}{\nu}}), & 1 < \nu < \infty \\
\phi(\log \phi - \log \psi) - \phi + \psi, & \nu = \infty
\end{cases}$$

Informally stated, $s_\nu$ represents the (infinitesimal) contribution at $\bar{\nu}$ to $S_\nu(\Phi \parallel \Psi)$. Note that $s_\nu(\phi, \psi) = 0$ for each $\nu$. Since $\psi$ is given in Problem 2, we assume that $\psi$ is a fixed parameter and we consider

$$s'_\nu(\phi, \psi) := \frac{ds_\nu(x, \psi)}{dx} \bigg|_{x=\phi}$$

$$= \begin{cases} 
\nu(\psi^{-\frac{1}{\nu}} - \phi^{-\frac{1}{\nu}}), & 1 \leq \nu < \infty \\
\log \phi - \log \psi, & \nu = \infty
\end{cases}$$

which represents the instantaneous rate of change of $s_\nu(\cdot, \psi)$ at point $\phi$. The first Taylor expansion of $s'_\nu(\cdot, \psi)$ with respect to $\phi = \psi$ is the straight line

$$\begin{cases} 
\psi^{-1+\frac{1}{\nu}}(\phi - \psi), & 1 \leq \nu < \infty \\
\psi^{-1}(\phi - \psi), & \nu = \infty
\end{cases}$$

having a slope equal to $\psi^{-1+\frac{1}{\nu}}$ when $1 \leq \nu < \infty$ and $\psi^{-1}$ for $\nu = \infty$. Once $\nu$ is fixed, the slope decreases as $\psi$ increases, and it is close to zero for $\psi$ sufficiently large. Thus the critical cases, i.e. when $s_\nu(\phi, \psi)$ is not able to discriminate $\phi$ from $\psi$ sufficiently well, happen when $\psi$ is large, because $s_\nu(\cdot, \psi)$ is almost flat in a neighborhood of $\psi$. On the other hand, if $\psi$ is greater than one then the slope increases as $\nu$ increases, i.e. $s_\nu(\phi, \psi)$ improves the ability to discriminate $\phi$ from $\psi$ by increasing $\nu$. Accordingly, a sufficiently large value of $\nu$ avoids solutions $\Phi$ which are very different from $\Psi$ in narrow ranges of frequencies. This explains the presence of relevant peaks only for $\nu = 1$. The same conclusion can be obtained by considering multivariate spectral densities. Concerning the complexity upper bound of the found solutions, in view of Proposition 4.1 it is easy to check that the upper bound on the McMillan degree of $\Phi_\psi$ increases as $\nu$ increases. For instance, in the first example of Section VII-A we have $\deg[\Phi_\psi]_n \leq 12$, $\deg[\Phi_\psi]_2 \leq 24$, $\deg[\Phi_\psi]_4 \leq 36$. Thus, the solution with $\nu = 1$ guarantees a simple model for the process $y$. Finally, we require that $\Psi$ is rational, accordingly the solution with $\nu = 1$ is the most appropriate to incorporate rational priors.

VIII. STRUCTURED COVARIANCE ESTIMATION PROBLEM

As mentioned in Section II-A we only have a prior $\Psi$ and a finite-length data $y_1 \ldots y_N$ in the THREE-like spectral estimation procedure. Moreover, $\Phi_\psi$ represents a family of estimates of $\Omega$ and we showed how to compute it starting from $\Psi$ and $\hat{\Sigma} \in \text{Range } \Gamma \cap Q_{n,+}$. Accordingly, it remains to find $\hat{\Sigma}$ from $y_1 \ldots y_N$. To deal with this issue, we consider Problem 2 which can be viewed as the static version of Problem 1. Indeed, in both problems minimization of a divergence index, with respect to the first argument, is performed on the intersection among a vector space and an open cone. In this section, we briefly show that it is also possible to find a family of solutions to the structured covariance estimation problem.

The Beta matrix divergence (family) among two covariance matrices $P, Q \in Q_{n,+}$ with $\beta \in \mathbb{R} \setminus \{0, 1\}$ is defined as

$$D_\beta(P \parallel Q) := \text{tr} \left[ \frac{1}{\beta-1} (P^\beta - P Q^\beta) - \frac{1}{\beta} (P^\beta - Q^\beta) \right].$$

In fact, $D_\beta(P \parallel Q)$ is the Beta divergence $S_\beta(\Phi \parallel \Psi)$ among the two constant spectral densities $\Phi(\omega^\beta) \equiv P$ and $\Psi(\omega^\beta) \equiv Q$. Since $D_\beta$ is a special case of $S_\beta$, it is strictly convex with respect to the first argument. Moreover, it is a continuous function of real variable $\beta \in \mathbb{R}$ with

$$\lim_{\beta \to 0} D_\beta(P \parallel Q) = D_0(P \parallel Q),$$
$$\lim_{\beta \to 1} D_\beta(P \parallel Q) = D_{\text{KL}}(P \parallel Q).$$

Fig. 1. Approximation of an ARMA (6, 4) spectral density.
Theorem. The uniqueness of $\Delta$ be extended to $P$ $\Delta$ employed. Therefore, once we computed the multipliers is $\nu$ Lagrange multiplier. Consider now Problem 1 with $\nu$ $L$ $(\nu)$. Since $P$, the existence and uniqueness of the solution to the problem for $\nu = 1$ has been showed. Moreover, the form of the optimal solution is $P_\beta(\Delta) = \left( \frac{1}{\nu} \right)^{-1} + V^*(\Delta)^{-1}$, where $V^*(\Delta) = \Pi_{C}^0 \Pi_{B}^0 - A^2 \Pi_{B}^0 \Pi_{B}^0 A$ is the adjoint operator of the linear map $V$ defined in (5) and $\Delta \in Q_n$ is the Lagrange multiplier. Consider now Problem 1 with $\nu$. The corresponding Lagrange functional is $L_\nu(P, \Delta) : = D_\nu(P\|\Sigma_C) + \nu \frac{1}{1 - \nu} \left( \frac{1}{\nu} \right)^{-1} + V^*(\Delta)^{-1}$, where $L_\nu(P, \Delta + \Delta) = L_\nu(P, \Delta)$ $\forall \Delta \in \ker(V^*)$, which can be assumed that $\Delta \in [\ker(V^*)]$. Moreover, $L_\nu(\cdot, \Delta)$ is strictly convex over $Q_{n+}$. Thus, the unique minimum point of $L_\nu(\cdot, \Delta)$, which is given by annihilating the first directional derivative of $L_\nu(\cdot, \Delta)$, is $P_\nu(\Delta) = \left( \frac{1}{\nu} \right)^{-1} + V^*(\Delta)^{-1}$.

Since $P_\nu(\Delta) \in Q_{n+}$, the set of the admissible Lagrange multipliers is $L_\nu := \{ \Delta \in Q_n | \frac{1}{\nu} \left( \frac{1}{\nu} \right)^{-1} + V^*(\Delta)^{-1} > 0 \} \cap [\ker(V^*)]$, which is an open and bounded set (the proof is similar to the one of Proposition 5.1 in [18]). Then, the dual problem is $\Delta^\circ = \arg\min_{\Delta \in L_\nu} J_\nu(\Delta)$ where

$$J_\nu(\Delta) := -\inf_{P} L_\nu(P, \Delta) = \nu \frac{1}{1 - \nu} \left( \frac{1}{\nu} \right)^{-1} + V^*(\Delta)^{-1}. \quad (27)$$

Note that $J_\nu(0) = \frac{1}{\nu} \left( \frac{1}{\nu} \right)^{-1}$. Accordingly, we can restrict the search of a minimum point to the set $L^* := \{ \Delta \in L_\nu | J_\nu(\Delta) \leq J_\nu(0) \} \subset L_\nu$, which is bounded. Following the same lines in [18], it is possible to prove that $J_\nu \in C^\infty(L_\nu)$ is strictly convex on $L_\nu$ and $\lim_{\Delta \to \partial L_\nu^*} J_\nu(\Delta) = +\infty$ (the limit diverges because the exponent in (27) is negative). Thus, $L^*$ is a compact set (i.e. closed and bounded) and $J_\nu$ admits a minimum point $\Delta^\circ$ over $L^*$ by the Weierstrass’ Theorem. The uniqueness of $\Delta^\circ$ follows from the fact that $J_\nu$ is strictly convex over $L_\nu$. Also in this case, a globally convergent matricial Newton algorithm for finding $\Delta^\circ$ may be employed. Therefore, once we computed $\Delta^\circ$ the solution to Problem 1 is given by $P_\nu(\Delta^\circ)$. Finally, the same analysis may be extended to $D_{KL}$. In this case, $P_{KL}(\Delta) = e^{\log(\Sigma_C)^{-1} + V^*(\Delta)}$.

To sum up, a family of solutions $P_\nu \in \text{Range } \Gamma \cap Q_{n+}$ to the structured covariance estimation problem has been found. In this way, we have a complete tool to compute the family of estimates $\Phi_\nu$ of $\Omega$ starting from a prior $\Psi$ and a finite-length data $y_1 \cdots y_N$. We compute $P_\nu$ from $y_1 \cdots y_N$ and we then find $\Phi_\nu$ starting from $P_\nu$ and $\Psi$.

IX. SIMULATION RESULTS - PART II

We consider the bivariate bandpass random process $y$ of Section VII-B and we take into account the following THREE-like spectral estimation procedure:

1) We start from a finite sequence $y_1 \cdots y_N$ extracted from a realization of the process $y$.
2) Fix $G(z)$ as in Section VII-B.
3) Choose a prior spectral density $\Psi \in \mathbb{S}_+^\infty(T)$ such that $\Psi^\perp$ is rational;
4) Feed the filters bank with the data sequence $y_1 \cdots y_N$, collect the output data $x_1 \ldots x_N$ and compute $\hat{\Sigma}_C = \frac{N}{\nu} \sum_{k=1}^{N} x_k x_k^\perp$.
5) Compute $P_\nu \in \text{Range } \Gamma \cap Q_{n+}$ by solving Problem 1 (with $D_{KL}$), then set $\hat{\Sigma} = P_\nu$;
6) Compute $\Phi_\nu$ by solving Problem 2 (with $S_\nu$) by means of the proposed algorithm with the chosen $\Psi$ and $\hat{\Sigma}^\perp G(z)$ as filters bank.

As noticed in Section VII, $\Psi$ represents the a priori information on $y$. Accordingly, $\Phi_\nu$ is a spectral density (with bounded McMillan degree) which is consistent with the interpolation constraint in (2) and is as close as possible to the a priori information, encoded in $\Psi$, according to the divergence index $S_\nu$.

In the following example, one can consider the prior in Figure 3. If no a priori information is given, we set $\Psi = I$. However, one can get information on $y$ by choosing $\Psi$ as the constant spectral density equal to the variance of the given data sequence. In this way the corresponding estimate will possess at least the zeroth moment similar to the estimated one by the given data. In what follows, the latter has been considered. In Figure 4 the obtained estimates with $N = 50$ (i.e. we have considered a short-length data) are depicted. For the extracted sequence, the estimators for $\nu = 2$ and $\nu = 3$ appear to perform better than the one for $\nu = 1$. More precisely, the peaks of the estimates are reduced by increasing $\nu$. In fact, as shown in Section VII, large values of $\nu$ penalize solutions which are very different from $\Psi$ in narrow ranges of frequencies. In this case $\Psi$ is constant, thus solutions with $\nu$ large will be more “flat” than the one with $\nu = 1$.

In the light of the results found in Section VII and here, we can outline the application scenarios for the presented family of estimators. The estimator with $\nu = 1$ is preferable when the a priori model for $y$ is rational (i.e. $\Psi$ rational) and a simple model for $y$ is required. On the contrary, estimators with $\nu$ large are preferable when the model for $y$ must be similar to the a priori model also in narrow ranges of frequencies and it must exhibit a “rich” dynamic. Increasing $\nu$, the previous features become more remarked. The limit case is $\nu \to \infty$ and the corresponding model is generically non-rational.
X. Conclusions

A multivariate Beta divergence family connecting the Itakura-Saito distance with the Kullback-Leibler divergence has been introduced. The corresponding solutions to the spectrum approximation problem are rational when the parametrization in Theorem 4.1 of the parameter of the process.

Fig. 4. Estimation of the spectral density of a bivariate bandpass random process.

ACKNOWLEDGMENTS

The author wish to thank Prof. A. Ferrante and Prof. M. Pavon for useful discussions about THREE-like spectral estimation approaches.

APPENDIX

A. On the exponentiation of positive definite matrices

We collect some technical result concerning the exponentiation of positive definite matrices to an arbitrary real number. We start by introducing the differential of the matrix exponential and the matrix logarithm (see [15]).

Proposition A.1: Given $Y \in \mathbb{Q}_n$, the differential of $Y \mapsto e^Y$ in the direction $\Delta \in \mathbb{Q}_n$ is given by the linear map

$$M_Y : \Delta \mapsto \int_0^1 e^{(1-\tau)Y} \Delta e^{\tau Y} d\tau.$$

Proposition A.2: Given $Y \in \mathbb{Q}_{n^+}$, the differential of $Y \mapsto \log(Y)$ in the direction $\Delta \in \mathbb{Q}_n$ is given by the linear map

$$N_Y : \Delta \mapsto \int_0^\infty (Y + tI)^{-1} \Delta (Y + tI)^{-1} dt.$$

Let us consider now a positive definite matrix $X \in \mathbb{Q}_{n^+}$ and a real number $c$. The exponentiation of $X$ to $c$ may be rewritten in the following way

$$X^c = e^{c \log X}.$$

Accordingly, by applying the chain rule, the differential of $X \mapsto X^c$ in the direction $\Delta \in \mathbb{Q}_n$ is given by

$$M_{c \log X}(c \delta X(\Delta)) = c \int_0^1 X^{c(1-\tau)} \int_0^\infty (X + tI)^{-1} \times \Delta (X + tI)^{-1} dt X^{c\tau} d\tau.$$

We summarize this result below.

Proposition A.3: The differential of $X \mapsto X^c$ in direction $\Delta \in \mathbb{Q}_n$ is given by the linear map

$$O_{X,c} : \Delta \mapsto c \int_0^1 X^{c(1-\tau)} \int_0^\infty (X + tI)^{-1} \times \Delta (X + tI)^{-1} dt X^{c\tau} d\tau.$$

Corollary A.1: The first variation of $X \mapsto \text{tr}[X^c]$ in direction $\Delta \in \mathbb{Q}_n$ is

$$\delta([\text{tr}[X^c]]; \Delta) = c \text{tr}(X^{c-1} \Delta).$$

Proof: Since $X^{c\tau}$ and $(X + tI)$ commute, we get

$$\delta([\text{tr}[X^c]]; \Delta) = \text{tr}(O_{X,c}(\Delta)) = c \text{tr} \left[ X^c \int_0^\infty (X + tI)^{-2} dt \Delta \right] = c \text{tr} \left[ X^cX^{-\Delta} \right] = c \text{tr} \left[ X^{c-1} \Delta \right].$$

B. Proofs of Section III

Proof of Proposition 3.1. By definition $X^{1-c}$ and $Y^{c-1}$ are continuous function of real variable $c$. Thus, the function $\log_c(X,Y)$ of real variable $c$ is continuous in $\mathbb{R} \setminus \{1\}$. It remains to prove that $\log_c$ is continuous in $c = 1$. This is equivalent to show that $\lim_{c \to 1} \log_c(X,Y) = \log(X) - \log(Y)$. Let $X = U \text{diag}(d_1, \ldots, d_m) U^T$, then

$$\frac{1}{1-c}(X^{1-c} - I) = U \text{diag} \left( \frac{d_1^{1-c} - 1}{1-c}, \ldots, \frac{d_m^{1-c} - 1}{1-c} \right) U^T.$$

Taking the limit for $c \to 1$, we get

$$\lim_{c \to 1} \frac{1}{1-c}(X^{1-c} - I) = U \text{diag} \left( \lim_{c \to 1} \frac{d_1^{1-c} - 1}{1-c}, \ldots, \lim_{c \to 1} \frac{d_m^{1-c} - 1}{1-c} \right) U^T.$$

Accordingly,

$$\lim_{c \to 1} \log_c(X,Y) = \lim_{c \to 1} \left( \frac{1}{1-c}(X^{1-c} - I) - \frac{1}{1-c}(Y^{1-c} - I) \right) Y^{c-1} = \lim_{c \to 1} \left( \frac{1}{1-c}(X^{1-c} - I) - \lim_{c \to 1} \frac{1}{1-c}(Y^{1-c} - I) \right) = \log(X) - \log(Y).$$
which proves that $\log_c$ is continuous in $c = 1$. Concerning the last statement, it is straightforward that $X = Y$ implies $\log_c(X, Y) = 0$. On the contrary, $\log_c(X, Y) = 0$, with $c \neq 1$, implies $X^{1-c}Y^{c-1} = I$ which is equivalent to $X^{1-c} = Y^{1-c}$, since $X, Y \in \mathcal{Q}_{m,+}$. Thus, $X = Y$. We get the same conclusion for $c = 1$ by using similar arguments.

Proof of Proposition 3.2 Since $\Phi$ and $\Psi$ belong to $\mathcal{S}_\mathbb{T}(\mathbb{T})$, i.e. $\Phi$ and $\Psi$ are coercive and bounded, it is possible to show by standard arguments that the integrand function of (13) uniformly converges on $T$ for $\beta \to 0$ and $\beta \to 1$. Hence, it is allowed to pass the limits, for $\beta \to 0$ and $\beta \to 1$, under the integral sign. Taking into account the first limit, we get

$$
\lim_{\beta \to 0} \mathcal{S}_\beta(\Phi \| \Psi) = \lim_{\beta \to 0} \int \left[ \frac{1}{\beta - 1} (\Phi^\beta - \Phi \Psi^{\beta - 1}) - \frac{1}{\beta} (\Phi^\beta - \Psi^\beta) \right]
= \int \left[ -I + \Phi \Psi^{-1} - \frac{1}{\beta} (\Phi^\beta - I) \right]
= \int \left[ -I + \Phi \Psi^{-1} - \log(\Phi) + \log(\Psi) \right]
= S_{\text{KL}}(\Phi \| \Psi)
$$

where we exploited (30). For the second limit, we obtain

$$
\lim_{\beta \to 1} \mathcal{S}_\beta(\Phi \| \Psi) = \lim_{\beta \to 1} \left( \frac{1}{\beta} \int \left[ \Phi^\beta \log_\beta \left( \Psi^\beta, \Phi^\beta \right) + \Phi^\beta - \Psi^\beta \right] \right)
= -\int \left[ \Phi \lim_{\beta \to 1} \left( \Phi^\beta \log_\beta \left( \Psi^\beta, \Phi^\beta \right) + \Phi - \Psi \right) \right]
= -\int \left[ \Phi \lim_{\beta \to 1} \log_2 \left( \Psi^\beta, \Phi^\beta \right) + \Phi - \Psi \right]
= \int \left[ \Phi \log(\Phi) - \log(\Psi) \right] + \Psi - \Phi
= S_{\text{KL}}(\Phi \| \Psi)
$$

where we exploited (31).

Proof of Proposition 3.7 The proof will be divided in the following three cases: $0 < \beta < 1$, $\beta = 1$ and $\beta = 0$.

Case $0 < \beta < 1$: Point 1. The first variation of $\mathcal{S}_\beta(\Phi \| \Psi)$, with respect to $\Phi$, in direction $\delta \Phi \in L_{\infty}^{m \times m}(\mathbb{T})$ is

$$
\delta(\mathcal{S}_\beta(\Phi \| \Psi); \delta \Phi) = \frac{1}{\beta - 1} \int_0^{2\pi} \text{tr} \left[ (\Phi^{\beta-1} - \Psi^{\beta-1}) \delta \Phi \right] \frac{d\theta}{2\pi}
$$

where we exploited (29). The second variation in direction $\delta \Phi$ is

$$
\delta^2(\mathcal{S}_\beta(\Phi \| \Psi); \delta \Phi) = \int_0^{2\pi} \text{tr} \left[ \left( \frac{1}{\beta - 1} \int_0^{2\pi} \Phi^{\beta-1} \frac{d\theta}{2\pi} \right) \left( \Phi^{\beta-1} - \Psi^{\beta-1} \right) \delta \Phi \right] \frac{d\theta}{2\pi}
$$

Case $\beta = 1$: Firstly, given $X \in \mathcal{Q}_{m,+}$ and $\delta X \in \mathcal{Q}_{m}$ we have

$$
\delta(\text{KL}(X \| \Phi); \delta X) = \text{tr} \left[ X^{-1} \delta X \right]
$$

it is sufficient to apply Proposition A.2 in Appendix A. For $\beta = 1$ we get the Kullback-Leibler divergence in (10). Taking into account (34), its first and second variations with respect to $\Phi$ in direction $\delta \Phi \in L_{\infty}^{m \times m}(\mathbb{T})$ are, respectively,

$$
\delta(\mathcal{S}_{\text{KL}}(\Phi \| \Psi); \delta \Phi) = \int \left[ \log(\Phi) - \log(\Psi) \right] \frac{d\Phi}{d\Psi}
$$

Since the second variation is non negative and equal to zero if and only if $\delta \Phi = 0$, $\mathcal{S}_{\text{KL}}(\cdot \| \cdot)$ is strictly convex over $\mathcal{S}_\mathbb{T}(\mathbb{T})$ and the (unique) minimum point is given by annihilating the first variation which leads to condition $\log(\Phi) = \log(\Psi)$. Thus, $\Phi = \Psi$ and $\mathcal{S}_{\text{KL}}(\Phi \| \Psi) = 0$.

Case $\beta = 0$: In this case we have the Itakura-Saito distance. Using similar arguments used for the case $\beta = 1$, we get the statement.
REFERENCES

[1] X. Jiang, L. Ning, and T. Georgiou, “Distances and Riemannian metrics for multivariate spectral densities,” IEEE Trans. Autom. Control, vol. 57, no. 7, pp. 1723–1735, Jul. 2012.

[2] T. Georgiou, “Realization of power spectra from partial covariance sequences,” IEEE Trans. Acoust., Speech Signal Processing, vol. 35, no. 4, pp. 438–449, Apr. 1987.

[3] C. Byrnes, A. Lindquist, S. Gusev, and A. Matveev, “A complete parameterization of all positive rational extensions of a covariance sequence,” IEEE Trans. Autom. Control, vol. 40, no. 11, pp. 1841–1857, Nov. 1995.

[4] C. Byrnes, T. Georgiou, and A. Lindquist, “A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint,” IEEE Trans. Autom. Control, vol. 46, no. 5, pp. 822–839, May 2001.

[5] T. Georgiou, “Spectral analysis based on the state covariance: The maximum entropy spectrum and linear fractional parametrization,” IEEE Trans. Autom. Control, vol. 47, no. 11, pp. 1811–1823, Nov. 2002.

[6] C. Byrnes and A. Lindquist, “On the partial stochastic realization problem,” IEEE Trans. Autom. Control, vol. 42, no. 8, pp. 1049–1070, Aug. 1997.

[7] C. Byrnes, S. Gusev, and A. Lindquist, “A convex optimization approach to the rational covariance extension problem,” SIAM J. Control and Optimization, vol. 37, no. 1, pp. 211–229, Oct. 1998.

[8] C. Byrnes, T. Georgiou, A. Lindquist, and A. Megretski, “Generalized interpolation in H-infinity with a complexity constraint,” Trans. American Math. Society, vol. 358, no. 3, pp. 965–987, Dec. 2004.

[9] T. Georgiou and A. Lindquist, “Remarks on control design with degree constraint,” IEEE Trans. Autom. Control, vol. 51, no. 7, pp. 1150–1156, Jul. 2006.

[10] A. Nasiri Amini, E. Ebbini, and T. Georgiou, “Noninvasive estimation of tissue temperature via high-resolution spectral analysis techniques,” IEEE Trans. Biomed. Eng., vol. 52, no. 2, pp. 221–228, Feb. 2005.

[11] C. Byrnes, P. Enqvist, and A. Lindquist, “Identifiability and well-posedness of shaping-filter parameterizations: A global analysis approach,” SIAM J. Control and Optimization, vol. 41, no. 1, pp. 23–39, Mar. 2002.

[12] T. Georgiou and A. Lindquist, “A convex optimization approach to ARMA modeling,” IEEE Trans. Autom. Control, vol. 53, no. 5, pp. 1108–1119, Jun. 2008.

[13] J. Karlsson and T. Georgiou, “Uncertainty bounds for spectral estimation,” IEEE Trans. Autom. Control, vol. 58, no. 7, pp. 1659–1673, Jul. 2013.

[14] C. Byrnes, T. Georgiou, and A. Lindquist, “A new approach to spectral estimation: A tunable high-resolution spectral estimator,” IEEE Trans. Signal Processing, vol. 48, no. 11, pp. 3189–3205, Nov. 2000.

[15] T. Georgiou, “Relative entropy and the multivariable multidimensional moment problem,” IEEE Trans. Inform. Theory, vol. 52, no. 3, pp. 1052–1066, Mar. 2006.

[16] A. Ferrante, M. Pavon, and F. Ramponi, “Hellinger versus Kullback-Leibler multivariable spectrum approximation,” IEEE Trans. Autom. Control, vol. 53, no. 4, pp. 954–967, May 2008.

[17] A. Ferrante, C. Masiero, and M. Pavon, “Time and spectral domain relative entropy: A new approach to multivariate spectral estimation,” IEEE Trans. Autom. Control, vol. 57, no. 10, pp. 2561–2575, Oct. 2012.

[18] A. Ferrante, M. Pavon, and M. Zorzi, “A maximum entropy enhancement for a family of high-resolution spectral estimators,” IEEE Trans. Autom. Control, vol. 57, no. 2, pp. 318–329, Feb. 2012.

[19] M. Zorzi and A. Ferrante, “On the estimation of structured covariance matrices,” Automatica, vol. 48, no. 9, pp. 2145–2151, Sep. 2012.

[20] A. Basu, I. Harris, N. Hjort, and M. Jones, “Robust and efficient estimation by minimising a density power divergence,” Biometrika, vol. 85, no. 3, pp. 549–559, Sep. 1998.

[21] M. Mollah, N. Sultana, M. Minami, and S. Eguchi, “Robust extraction of local structures by the minimum of Beta-divergence method,” Neural Networks, vol. 23, no. 2, pp. 226–238, Mar. 2010.

[22] M. Minami and S. Eguchi, “Robust blind source separation by Beta divergence,” Neural Computation, vol. 14, no. 8, pp. 1859–1886, Aug. 2002.

[23] A. Cichocki, R. Zdunek, A. Phan, and S. Amari, Nonnegative Matrix and Tensor Factorizations. Chichester, U.K.: John Wiley & Sons, 2009.

[24] A. Cichocki, S. Cruces, and S. Amari, “Generalized Alpha-Beta divergences and their application to robust nonnegative matrix factorization,” Entropy, vol. 13, no. 1, pp. 134–170, Jan. 2011.

[25] T. Georgiou, “The structure of state covariances and its relation to the power spectrum of the input,” IEEE Trans. Autom. Control, vol. 47, no. 7, pp. 1056–1066, Jul. 2002.

[26] T. Cover and J. Thomas, Information Theory. New York: Wiley, 1991.

[27] M. Nielsen and I. Chuang, Quantum Computation and Quantum Information. Cambridge, U.K.: Cambridge University Press, 2004.

[28] A. Cichocki and S. Amari, “Families of Alpha- Beta- and Gamma-divergences: Flexible and robust measures of similarities,” Entropy, vol. 12, no. 6, pp. 1532–1568, Jun. 2010.

[29] T. Georgiou and A. Lindquist, “Kullback-Leibler approximation of spectral density functions,” IEEE Trans. Inform. Theory, vol. 49, no. 11, pp. 2910–2917, Nov. 2003.

[30] A. Blomsqist, A. Lindquist, and R. Nagamune, “Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: An optimization approach,” IEEE Trans. Autom. Control, vol. 48, no. 12, pp. 2172–2190, Dec. 2003.

[31] F. Ramponi, A. Ferrante, and M. Pavon, “A globally convergent matricial algorithm for multivariate spectral estimation,” IEEE Trans. Autom. Control, vol. 54, no. 10, pp. 2376–2388, Oct. 2009.

[32] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge University Press, 2004.