USING OPTIMAL MASS TRANSPORT FOR TRACKING AND INTERPOLATION OF TOEPLITZ COVARIANCE MATRICES

Filip Elvander, Andreas Jakobsson*, and Johan Karlsson†*

*Div. of Mathematical Statistics, Lund University, Sweden
†Dept. of Mathematics, KTH Royal Institute of Technology, Sweden
emails: {filipelv, aj}@maths.lth.se, johan.karlsson@math.kth.se

ABSTRACT

In this work, we propose a novel method for interpolation and extrapolation of Toeplitz structured covariance matrices. By considering a spectral representation of Toeplitz matrices, we use an optimal mass transport problem in the spectral domain in order to define a notion of distance between such matrices. The obtained optimal transport plan naturally induces a way of interpolating, as well as extrapolating, Toeplitz matrices. The constructed covariance matrix interpolants and extrapolants preserve the Toeplitz structure, as well as the positive semi-definiteness and the zeroth covariance of the original matrices. We demonstrate the proposed method's ability to model locally linear shifts of spectral power for slowly varying stochastic processes, illustrating the achievable performance using a simple tracking problem.

***Index Terms*—** Covariance interpolation, Optimal mass transport, Toeplitz matrices, Spectral estimation

1. INTRODUCTION

Statistical modeling is a key tool for estimation and identification and is used in most areas of signal processing. An intrinsic component of such models are covariance estimates, which is extensively used in application areas such as spectral estimation, radar, and sonar [1, 2], wireless channel estimation, medical imaging, and identification of systems and network structures [3, 4]. Although being a classical subject (see, e.g., [5]) covariance estimation has recently received considerable attention. Such contributions include works on finding robust covariance estimates with respect to outliers, as well as methods suitable for handling different distribution assumptions, including families of non-Gaussian distributions [6–10]. Another active area is covariance estimation with an inherent geometry that gives rise to a structured covariance matrix. Such structures could arise from stationarity assumptions of the underlying object [11–15] or be due to assumptions in, e.g., the underlying network structures in graphical models [16, 17]. In this work, we will focus on Toeplitz structures, which naturally arise when modelling stationary signals and processes.

Although many methods rely on stationarity for modeling signals, such assumptions are typically not valid over longer time horizons. Therefore, tools for interpolation and morphing of covariance matrices are important for modeling and fusing of information. Several such tools for interpolating covariances have recently been proposed in the literature, for example methods based on g-convexity [9], optimal mass transport [18], and information geometry [19]. An alternative approach for such interpolation is to relax, or "lift", the covariances and instead consider interpolation between the lifted objects. For example, in [20] (see also [21]), interpolation between covariance matrices is induced by the optimal mass transport geodesics between the Gaussian density functions with the corresponding covariances.

In this work, we propose a new lifting approach, where the lifting is made from the covariance domain to the frequency domain, motivated by the spectral representation of positive semi-definite Toeplitz matrices. We combine this approach with the frequency domain metric based on optimal mass transport, proposed in [22], in order to define pairwise distances between Toeplitz matrices. This is done by considering the minimum distance in the optimal mass transport sense between the sets of power spectra consistent with each of the Toeplitz matrices. We show that the proposed distance measure gives rise to a natural way of interpolating and extrapolating Toeplitz matrices. The method preserves the Toeplitz structure, the positive semi-definiteness, as well as the zeroth moment of the interpolating/extrapolating matrices.

Notation

Let $\mathbb{M}_+^n$ denote the set of positive semi-definite $n \times n$ Hermitian matrices, and let $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and the Hermitian transpose, respectively. Also, let $\overline{(\cdot)}$ denote the complex conjugate, and let $\mathbb{E}(\cdot)$ denote the expectation operator. Furthermore, let $\mathbb{T} = (-\pi, \pi]$ and let $\mathcal{M}_+(\mathbb{T})$ denote the set of generalized integrable non-negative functions on the set $\mathbb{T}$, e.g., such functions that may contain singular parts, such as Dirac delta functions.

This work was supported in part by the Swedish Research Council.

2. BACKGROUND

2.1. Stochastic processes and spectral representations

Consider a complex-valued discrete-time stochastic process (signal) $y(t)$, $t \in \mathbb{Z}$, that is zero mean and wide sense stationary (WSS), i.e., $\mathbb{E}(y(t)) = 0$ and the covariance $r_k = \mathbb{E}(y(t)\overline{y(t-k)})$ is independent of t. The power spectrum, denoted by Φ, represents the frequency content of the process $y(t)$, and is the non-negative function[1] on $\mathbb{T}$ whose Fourier coefficients coincide with the covariances, i.e.,

$$r_k = \frac{1}{2\pi}\int_{\mathbb{T}} \Phi(\theta)e^{-ik\theta}d\theta \tag{1}$$

for $k \in \mathbb{Z}$ (see, e.g., [23, Chapter 2]). Since the process $y(t)$ is WSS, the $n \times n$ covariance matrix of the signal

$$\mathbf{R} = \begin{pmatrix} r_0 & r_{-1} & r_{-2} & \cdots & r_{-n+1} \\ r_1 & r_0 & r_{-1} & \cdots & r_{-n+2} \\ r_2 & r_1 & r_0 & \cdots & r_{-n+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n-1} & r_{n-2} & r_{n-3} & \cdots & r_0 \end{pmatrix} \tag{2}$$

is a Toeplitz matrix. Typically in spectral estimation, one considers the inverse problem of recovering the power spectrum Φ from a given set of covariances r_k for $k \in \mathbb{Z}$, with $|k| \leq n-1$. A power spectrum is consistent with such a sequence if (1) holds for $|k| \leq n-1$. Expressed in matrices, a spectrum is consistent with such a partial covariance sequence if $\Gamma(\Phi) = \mathbf{R}$, where $\Gamma : \mathcal{M}_+(\mathbb{T}) \to \mathbb{M}_+^n$ is the linear operator

$$\Gamma(\Phi) \triangleq \frac{1}{2\pi}\int_{\mathbb{T}} \mathbf{a}(\theta)\Phi(\theta)\mathbf{a}(\theta)^H d\theta \tag{3}$$

and

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 & e^{i\theta} & \ldots & e^{i(n-1)\theta} \end{bmatrix}^T / \sqrt{n} \tag{4}$$

is the Fourier vector. Note that $\Gamma(\Phi)$ is a Toeplitz matrix since $\mathbf{a}(\theta)\mathbf{a}(\theta)^H$ is Toeplitz for any θ. For any positive semi-definite Toeplitz matrix $\mathbf{R}$, there always exists a consistent power spectrum; in fact, if $\mathbf{R}$ is positive definite, there is an infinite family of consistent power spectra [24]. We will in the main section utilize such spectral representations in order to define distances between Toeplitz matrices, with the distance being defined in terms of the optimal mass transport cost between consistent power spectra.

2.2. Optimal mass transport

The Monge-Kantorovich transportation problem is the problem of finding an optimal transport plan between two given mass distributions [25, 26]. Here, the cost of moving a unit mass is defined on the underlying space, and the optimal transport plan is the one with minimal total cost. The resulting total cost associated with the transport plan is then used as a measure of similarity, or distance, between the two mass distributions. These ideas have been used for defining metrics between power spectra [22], as well as for tracking stochastic processes with smoothly varying spectral content and spectral morphing for speech signals [27]. Recently, the notion has also been utilized in fundamental frequency estimation algorithms as a means of clustering [28]. As in [22], we define the distance between two spectra Φ_0 and Φ_1 as

$$T(\Phi_0, \Phi_1) \triangleq \min_{M \in \mathcal{M}_+(\mathbb{T}^2)} \int_{\mathbb{T}^2} c(\theta,\varphi)M(\theta,\varphi)d\theta d\varphi \tag{5a}$$

$$\text{subject to} \quad \Phi_0(\theta) = \int_{\mathbb{T}} M(\theta,\varphi)d\varphi \tag{5b}$$

$$\Phi_1(\varphi) = \int_{\mathbb{T}} M(\theta,\varphi)d\theta \tag{5c}$$

where $\mathbb{T}^2 = \mathbb{T}\times\mathbb{T}$ denotes the 2-D frequency space. Here, the cost function, $c(\theta,\varphi)$, denotes the transportation cost of moving one unit of mass between the frequencies θ and φ. The transport plan, $M(\theta,\varphi)$, specifies the amount of mass moved from frequency θ to frequency φ. The objective in (5a) is the total cost incurred by the transport plan M and the constraints (5b) and (5c) ensure that M is a valid transport plan from Φ_0 to Φ_1. It may be noted that the distance measure $T(\Phi_0, \Phi_1)$ is only defined for spectra of the same mass. However, this may be generalized in order to handle mass distributions with different total mass by including a cost for adding and subtracting mass (see [22] for details).

3. A NOTION OF DISTANCE FOR TOEPLITZ MATRICES

Given a pair of positive semi-definite Toeplitz matrices $\mathbf{R}_0$ and $\mathbf{R}_1$, there is, as noted above, a family of power spectra consistent with each of them. Accordingly, we define the distance between $\mathbf{R}_0$ and $\mathbf{R}_1$ as the minimum transportation cost between the corresponding spectral families, i.e.,

$$\begin{aligned} d(\mathbf{R}_0, \mathbf{R}_1) \triangleq \min_{\Phi_j \in \mathcal{M}_+(\mathbb{T})} \quad & T(\Phi_0, \Phi_1) \\ \text{subject to} \quad & \Gamma(\Phi_j) = \mathbf{R}_j \text{ for } j = 0, 1 \end{aligned} \tag{6}$$

using the cost function $c(\theta,\varphi) = |e^{i\theta} - e^{i\varphi}|^2$. Using (5), the expression (6) may be formulated as the convex optimization problem

$$\begin{aligned} \min_{M \in \mathcal{M}_+(\mathbb{T}^2)} \quad & \int_{\mathbb{T}^2} c(\theta,\varphi)M(\theta,\varphi)d\theta d\varphi \\ \text{subject to} \quad & \Gamma\left(\int_{\mathbb{T}} M(\theta,\varphi)d\varphi\right) = \mathbf{R}_0 \\ & \Gamma\left(\int_{\mathbb{T}} M(\theta,\varphi)d\theta\right) = \mathbf{R}_1. \end{aligned} \tag{7}$$

As we will see next, this construction also implies a natural way of interpolating and extrapolating Toeplitz matrices.

[1] A power spectrum is in general a non-negative bounded measure and may contain, e.g., spectral lines.

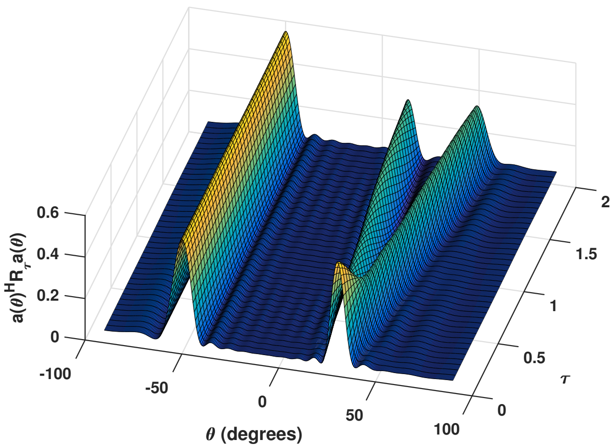

Fig. 1. Interpolated spatial spectrum estimated as $\mathbf{a}(\theta)^H\mathbf{R}_\tau\mathbf{a}(\theta)$, where $\mathbf{R}_\tau$ is obtained by solving (6).

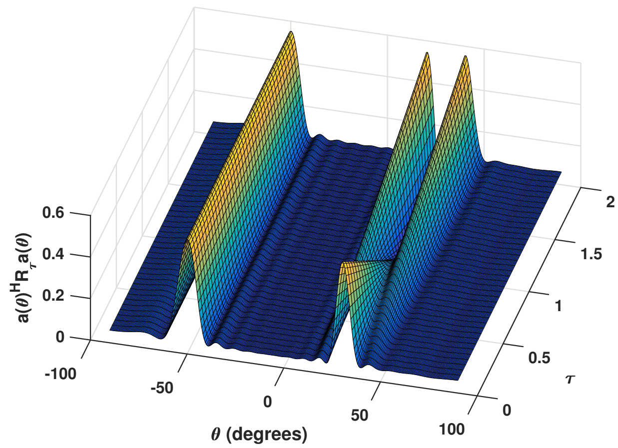

Fig. 2. Interpolated spatial spectrum estimated as $\mathbf{a}(\theta)^H\mathbf{R}_\tau^{\mathrm{conv}}\mathbf{a}(\theta)$, where $\mathbf{R}_\tau^{\mathrm{conv}}$ is the linear combination of $\mathbf{R}_0$ and $\mathbf{R}_1$.

3.1. Interpolation, extrapolation, and tracking

Based on a transport plan, M, one may define intermediate spectra by linearly shifting the frequency locations of the mass as dictated by M. That is, any mass transfered from ϕ to $\phi+\varphi$ is at $\tau\in[0,1]$ located at $\phi+\tau\varphi$, and thus the intermediate spectrum is given by

$$\begin{aligned}\Phi_\tau^M(\theta) &= \int_{\mathbb{T}^2}\delta_\theta(\{\phi+\tau\varphi\}_{\mathrm{mod}\,\mathbb{T}})M(\phi,\phi+\varphi)d\phi d\varphi \\ &= \int_{\mathbb{T}} M(\theta-\tau\varphi,\theta+(1-\tau)\varphi)d\varphi. \end{aligned} \tag{8}$$

Here, δ_θ is the Dirac delta function localized at θ and the integrands are extended periodically with period 2π outside the domain of integration. Also, we denote by $\{x\}_{\mathrm{mod}\,\mathbb{T}}$ the value in $\mathbb{T}$ that is congruent with x modulo 2π. This construction allows one to define a corresponding interpolating covariance matrix $\mathbf{R}_\tau$ according to

$$\begin{aligned}&\mathbf{R}_\tau = \Gamma(\Phi_\tau^M) \qquad (9)\\ &= \int_{\mathbb{T}}\mathbf{a}(\theta)\left(\int_{\mathbb{T}} M(\theta-\tau\varphi,\theta+(1-\tau)\varphi)d\varphi\right)\mathbf{a}(\theta)^H d\theta \\ &= \int_{\mathbb{T}^2}\mathbf{a}(\{\phi+\tau\varphi\}_{\mathrm{mod}\,\mathbb{T}})M(\phi,\phi+\varphi)\mathbf{a}(\{\phi+\tau\varphi\}_{\mathrm{mod}\,\mathbb{T}})^H d\phi d\varphi\end{aligned}$$

for $\tau\in[0,1]$. From this, it may be noted that (9) naturally lends itself to extrapolation, i.e., directly allows for choosing τ outside the interval $[0,1]$. The following proposition follows directly from the definition of $\mathbf{R}_\tau$.

Proposition 1. *Let $\mathbf{R}_0,\mathbf{R}_1\in\mathbb{M}_+^n$ be Toeplitz matrices with same zeroth covariance. For any $\tau\in\mathbb{R}$, let $\mathbf{R}_\tau$ be given by* (9) *where M is the minimizing transport plan in* (7)*. Then the following properties hold:*

a) $\mathbf{R}_\tau$ *is a Toeplitz matrix*

b) $\mathbf{R}_\tau$ *is positive semi-definite*

c) $\mathbf{R}_\tau$ *has same zeroth covariance as* $\mathbf{R}_0$ *and* $\mathbf{R}_1$.

Due to these properties, the proposed method offers a way of interpolating the covariances of, e.g., slowly varying time series as the interpolant $\mathbf{R}_\tau$ allows for modeling linear changes in the spectrum of the process. The proposed method can also be readily used for spectral tracking by fitting a covariance path $\mathbf{R}_\tau$ to a sequence of covariance matrix estimates, $\hat{\mathbf{R}}_{\tau_j}$, for $j=1,\ldots,J$. In order to formulate this, we let $\mathcal{I}_\tau(M)\triangleq\Gamma(\Phi_\tau^M)$ denote the linear operator in (9) that maps the transport plan to a covariance matrix on the path. Thus, the tracking problem can be expressed as the convex optimization problem

$$\underset{M\in\mathcal{M}_+(\mathbb{T}^2)}{\text{minimize}}\int_{\mathbb{T}^2}c(\theta,\varphi)M(\theta,\varphi)d\theta d\varphi+\lambda\sum_{j=1}^{J}\left\|\mathcal{I}_{\tau_j}(M)-\hat{\mathbf{R}}_{\tau_j}\right\|_F^2 \tag{10}$$

where $\lambda>0$ is a user-specified regularization parameter. Thus, the added regularization term will penalize deviations of the induced interpolant $\mathbf{R}_\tau=\mathcal{I}_\tau(M)$ from the estimated covariances, as measured by the squared Frobenius norm.

3.2. Comparison with other methods

The properties in Proposition 1 distinguish the proposed interpolant $\mathbf{R}_\tau$ from other proposed matrix geodesics. As an example, consider the basic method of forming interpolants using convex combinations of $\mathbf{R}_0$ and $\mathbf{R}_1$, i.e., $\mathbf{R}_\tau^{\mathrm{conv}}=\tau\mathbf{R}_0+(1-\tau)\mathbf{R}_1$, for $\tau\in[0,1]$. This preserves the Toeplitz structure, as well as the zeroth covariance and the positive semi-definiteness. However, from a spectral representation point of view, the convex combination gives rise to fade-in fade-out effects, i.e., only spectral modes directly related to $\mathbf{R}_0$ and $\mathbf{R}_1$ can be represented, and there can be no shift in the location of these modes (see the example in Section 4.1 and Figure 2). Other more sophisticated options include, e.g.,

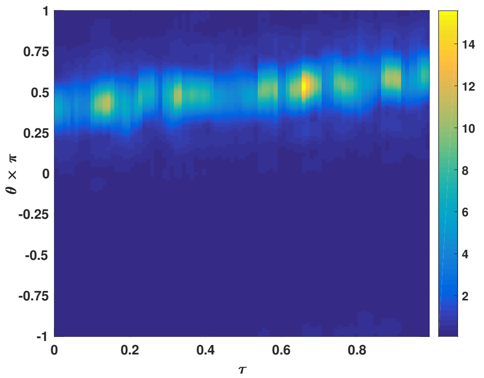

Fig. 3. Spectrum estimated as $\mathbf{a}(\theta)^H\hat{\mathbf{R}}\mathbf{a}(\theta)$, where $\hat{\mathbf{R}}$ is estimated as the sample covariance matrix based on 100 samples in each window.

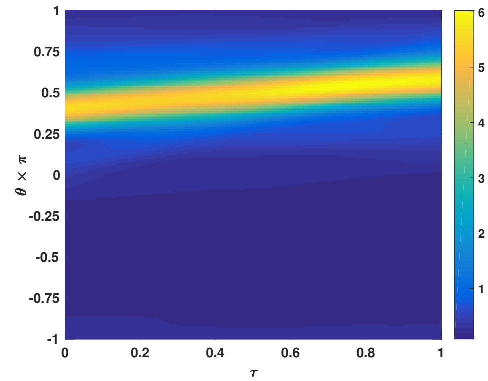

Fig. 4. Spectrum estimated as $\mathbf{a}(\theta)^H\mathbf{R}_\tau\mathbf{a}(\theta)$, where $\mathbf{R}_\tau$ is obtained by solving (10), fitted to a sequence of five covariance estimates.

the geodesic with respect to g-convexity [9]

$$\tilde{\mathbf{R}}_\tau = \mathbf{R}_0^{1/2}\left(\mathbf{R}_0^{-1/2}\mathbf{R}_1\mathbf{R}_0^{-1/2}\right)^\tau \mathbf{R}_0^{1/2},$$

and the geodesic in [20, 21], which builds on optimal mass transport of Gaussian distributions

$$\breve{\mathbf{R}}_\tau = \left((1-\tau)\mathbf{R}_0^{1/2}+\tau\mathbf{R}_1^{1/2}\mathbf{U}\right)\left((1-\tau)\mathbf{R}_0^{1/2}+\tau\mathbf{R}_1^{1/2}\mathbf{U}\right)^H$$

where $\mathbf{U} = \mathbf{R}_1^{-1/2}\mathbf{R}_0^{-1/2}(\mathbf{R}_0^{1/2}\mathbf{R}_1\mathbf{R}_0^{1/2})^{1/2}$. Although both of these geodesics preserve positive semi-definiteness, they neither preserve the Toeplitz structure nor the zeroth covariance. As noted above, the three properties in Proposition 1 hold for all $\tau \in \mathbb{R}$ for the proposed approach, and thus directly allows for extrapolation using (9). In contrast, it may be noted that for the linear combination $\mathbf{R}_\tau^{\text{conv}}$ there are no guarantees that the resulting matrix is positive semi-definite for $\tau \notin [0, 1]$. Also, note that neither of the alternative geodesics, $\tilde{\mathbf{R}}_\tau$ and $\breve{\mathbf{R}}_\tau$, naturally generalize to extrapolation.

4. NUMERICAL EXAMPLES

4.1. Interpolation and extrapolation for DOA

We begin by illustrating the performance of the proposed method on an interpolation example for direction-of-arrival (DOA) estimation. Consider a uniform linear array (ULA) with 15 sensors with half-wavelength sensor spacing and a scenario where two covariance matrices

$$\mathbf{R}_0 = \frac{1}{2}\sum_{\ell=1}^{2}\mathbf{a}(\theta_\ell^{(0)})\mathbf{a}(\theta_\ell^{(0)})^H + \sigma^2\mathbf{I}$$

$$\mathbf{R}_1 = \frac{1}{2}\mathbf{a}(\theta_1^{(1)})\mathbf{a}(\theta_1^{(1)})^H + \frac{1}{4}\sum_{\ell=2}^{3}\mathbf{a}(\theta_\ell^{(1)})\mathbf{a}(\theta_\ell^{(1)})^H + \sigma^2\mathbf{I}$$

are available[2]. Here $\theta_1^{(0)} = \theta_1^{(1)} = -50°$, $\theta_2^{(0)} = 30°$, $\theta_2^{(1)} = 20°$, and $\theta_3^{(1)} = 40°$, and $\sigma^2 = 0.05$. Such a scenario may be interpreted as a target at $\theta_2^{(0)}$ splitting up into two targets at $\theta_2^{(1)}$ and $\theta_3^{(1)}$ as time progresses, whereas the target at $\theta_1^{(0)}$ stays put. We use the proposed method to find covariances $\mathbf{R}_\tau$, for $\tau \in [0, 2]$, and compare these to the basic interpolant $\mathbf{R}_\tau^{\text{conv}}$ based on convex and linear combinations. Let the spectral estimates obtained using these interpolants be $\Psi_\tau(\theta)$ and $\Psi_\tau^{\text{conv}}(\theta)$, respectively, defined as

$$\begin{aligned}\Psi_\tau(\theta) &= \mathbf{a}(\theta)^H\mathbf{R}_\tau\mathbf{a}(\theta) \\ \Psi_\tau^{\text{conv}}(\theta) &= \mathbf{a}(\theta)^H\mathbf{R}_\tau^{\text{conv}}\mathbf{a}(\theta).\end{aligned} \tag{11}$$

The spectral estimates are shown in Figures 1 and 2. As can be seen, the obtained covariance interpolant $\mathbf{R}_\tau$ results in spectra Ψ_τ that model linear displacement of the targets. Note also that the extrapolated covariances $\mathbf{R}_\tau$ for $\tau \in (1, 2]$ imply that the targets continue linearly with respect to the look-angle θ, as may be expected. For the basic interpolant, $\mathbf{R}_\tau^{\text{conv}}$, the spectral estimates Ψ_τ^{conv} display undesirable behavior; clear fade-in fade-out effects are visible, and non-negativity is violated for some of the extrapolated spectra.

4.2. Tracking of an AR-process

Next, we illustrate the approach in (10) for the tracking of signals with slowly varying spectral content. To this end, consider a complex autoregressive (AR) process with one complex, time-varying pole. The pole is placed at a constant radius of 0.9, and moves from the frequency 0.4π to 0.6π. Spectral estimates based directly on covariance matrix estimates $\hat{\mathbf{R}}$ are shown in Figure 3. These covariance matrix estimates are obtained as the outer product estimate, based on 100 samples each, where the overlap between each estimate is 80 samples. As can be seen, the spectral estimates are very noisy and vary greatly in power. Using five of these covariance matrix estimates $\hat{\mathbf{R}}$, evenly spaced throughout the signal, we solve (10) in order to obtain an estimated covariance path, $\mathbf{R}_\tau$. The resulting spectra, estimated using (11), are shown in Figure 4. As can be seen, the path resulting from the proposed method allows for a smooth tracking of the shift in spectral content.

[2] Note that θ in this example denotes spatial frequency. For simplicity, we retain the notation $\mathbf{a}(\theta)$ also for this case.

5. REFERENCES

[1] J. P. Burg, "Maximum Entropy Spectral Analysis," in *Proceedings of the* 37^{th} *Annual International SEG Meeting*, Oklahoma City, OK., 1967.

[2] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, N.J., 2005.

[3] A. P. Dempster, "Covariance Selection," *Biometrics*, vol. 28, no. 1, pp. 157–175, 1972.

[4] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," *The Annals of Statistics*, vol. 40, no. 4, pp. 1935–1967, 2012.

[5] A. K. Kshirsagar, *Multivariate Analysis*, Dekker, 1972.

[6] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, July 2012.

[7] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex Elliptically Symmetric Distributions: Survey, New Results and Applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov 2012.

[8] E. Ollila and D. E. Tyler, "Regularized m -Estimators of Scatter Matrix," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 6059–6070, Nov 2014.

[9] A. Wiesel, "Geodesic Convexity and Covariance Estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 23, pp. 6182–6189, Dec 2012.

[10] Y. Sun, P. Babu, and D. P. Palomar, "Robust Estimation of Structured Covariance Matrix for Heavy-Tailed Elliptical Distributions," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3576–3590, July 2016.

[11] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of Structured Covariance Matrices," *Proc. IEEE*, vol. 70, no. 9, pp. 963–974, 1982.

[12] C. L. Byrnes, T. T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: a tunable high-resolution spectral estimator," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3189–3205, Nov 2000.

[13] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal*, vol. 47, no. 5, pp. 1314–1323, May 1999.

[14] T. T. Georgiou, "Structured covariances and related approximation questions," in *Directions in Mathematical Systems Theory and Optimization*. 2003, pp. 135–140, Springer.

[15] J. Karlsson and P. Enqvist, "Input-to-state covariances for spectral analysis: The biased estimate," in *Proc. of Int. Symp. Mathematical Theory of Networks and Systems*, 2012.

[16] E. Avventi, A. G. Lindquist, and B. Wahlberg, "ARMA Identification of Graphical Models," *IEEE Trans. Autom. Control*, vol. 58, no. 5, pp. 1167–1178, May 2013.

[17] M. Zorzi and R. Sepulchre, "AR identification of latent-variable graphical models," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2327–2340, Oct 2016.

[18] K. Yamamoto, Y. Chen, L. Ning, T. T. Georgiou, and A. Tannenbaum, "Regularization and Interpolation of Positive Matrices," *arXiv preprint arXiv:1611.07945*, 2016.

[19] F. Barbaresco, "Information geometry of covariance matrix: Cartan-Siegel homogeneous bounded domains, Mostow/Berger fibration and Frechet media," in *Matrix Information Geometry*. 2013, pp. 199–255, Springer.

[20] M. Knott and C. S. Smith, "On the optimal mapping of distributions," *Journal of Optimization Theory and Applications*, vol. 43, no. 1, pp. 39–49, 1984.

[21] L. Ning, X. Jiang, and T. T. Georgiou, "On the Geometry of Covariance Matrices," *IEEE Signal Process. L*, vol. 20, no. 8, pp. 787–790, August 2013.

[22] T. T. Georgiou, J. Karlsson, and M. S. Takyar, "Metrics for power spectra: an axiomatic approach," *IEEE Trans. Signal*, vol. 57, no. 3, pp. 859–867, March 2009.

[23] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall, Englewood Cliffs, N.J., 1993.

[24] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, University of California Press, Los Angeles, 1958.

[25] C. Villani, *Optimal transport: old and new*, Springer Science & Business Media, 2008.

[26] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal Mass Transport: Signal processing and machine-learning applications," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 43–59, July 2017.

[27] X. Jiang, Z. Q. Luo, and T. T. Georgiou, "Geometric Methods for Spectral Analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1064–1074, March 2012.

[28] F. Elvander, S. I. Adalbjörnsson, J. Karlsson, and A. Jakobsson, "Using Optimal Transport for Estimating Inharmonic Pitch Signals," in *42nd IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, New Orleans, LA, USA, March 5-9 2017, pp. 331–335.