Defining Graph Signal Distances Using an Optimal Mass Transport Framework

Maria Juhlin
Mathematical Statistics
Lund University, Sweden
juhlin@maths.lth.se

Filip Elvander
Mathematical Statistics
Lund University, Sweden
filipelv@maths.lth.se

Andreas Jakobsson
Mathematical Statistics
Lund University, Sweden
aj@maths.lth.se

***Abstract*—In this work, we propose a novel measure of distance for quantifying dissimilarities between signals observed on a graph. Building on a recently introduced optimal mass transport framework, the distance measure is formed using the second-order statistics of the graph signals, allowing for comparison of graph processes without direct access to the signals themselves, while explicitly taking the dynamics of the underlying graph into account. The behavior of the proposed distance notion is illustrated in a graph signal classification scenario, indicating attractive modeling properties, as compared to the standard Euclidean metric.**

***Index Terms*—Graph Signal Processing, Optimal mass transport, Graph signal similarity**

I. Introduction

In recent years, graph signal processing has attracted much interest in the signal processing community due to the topic's diverse area of application, ranging from the processing of brain signals, transportation networks, social media, and sensor networks (see, e.g., [1]–[4]). Generally, one is interested in processing signals that are defined on an underlying graph or network, where often the signal resides on the vertices of the graph, with the edges modeling the dependencies between the different signal values [2]. Extending the signal processing framework to graph signals enables signal modeling not only of inter-temporal dependencies for the individual node signals but also includes the modeling of the joint dependencies between the node signals, enabling the modeling of complex dynamic networks [5]. As in conventional signal processing, identification and estimation problems are often at the heart of ones interest, and to address such problems, a notion of signal similarity and distance for graph signals has to be introduced. This has led to the development of several different similarity measures, with notable examples including graph kernel based approaches [6], the Earth mover's distance for comparing images [7], and the conventional Euclidean distance. Of these, the graph kernel approach builds on identifying and comparing the underlying graph structure, and is thus not concerned with the graph signals directly [6]. Which similarity formulation that is most suitable to employ typically depends on the application at hand, but to fully utilize the information supplied by the graph formulation, it is desirable to use a measure that reflects the underlying graph topology. In some applications, it is not feasible to compute measures of similarity defined for observations of the signal directly, e.g., due to restrictions on the storage or transmission capacity of the system, or due to a sensitive nature of the measured data. Earlier Earth mover's distance measures, such as the one introduced in [7], compares the actual graph signals, necessitating access to these. An alternative is then to instead employ covariance estimation techniques, already commonly used in applications such as spectral estimation, radar, sonar and medical imaging [8]. Then, instead of storing the entire graph signal, only the signal covariance matrix needs to be stored or transmitted. In order to comply with such restrictions, we in this work propose a graph signal distance notion relying only on the availability of the second-order statistics of the signals, as well as knowledge of the underlying graph. In [9], a problem reminiscent of this, specifically, covariance interpolation, was explored and a framework for defining distances between covariance matrices in terms of their spectral representations was presented. As illustrated in [9], the commonly used Euclidean metric does not, in general, respect the underlying topology for spectral distances, and typically imply fade-in fade-out effects not desirable from a modeling perspective. As an alternative, distances induced by optimal mass transport (OMT), in some specific settings known as Earth mover's distance (see, e.g., [10] for an introduction and overview of the field), was considered in [9], and showed to exhibit several desirable modeling properties. Earth mover's distance has also recently been considered in the context of graph signals in [11], where it was shown to accurately model heat diffusion on various graphs. Extending upon the work in [9], this work proposes a notion of distances for graph signals such that the underlying topology of the graph, and the graph signal dynamics, are fully taken into account. In contrast to the distances introduced in [6], [7], the presented measure operates only on the second-order representation of the data, thereby allowing for a formulation that does not require access to the actual measurements. To illustrate the performance of the proposed measure, as compared to the conventional Euclidean distance, we examine the problem of classifying a measured spectrum using a set of mismatched spectra.

This work was supported in part by the Crafoord foundation (Grant no. 20180641), eSSENCE (Grant no 2017-4:2), and the Swedish Research Council (Grant no. 2015-04148).

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

II. Signal model

Consider a graph consisting of N nodes, and let $\mathbf{x}_t \in \mathbb{C}^N$ denote the signal over the graph at time t. Reminiscent of [12], [13], we will herein assume that the dynamics of $\mathbf{x}_t$ may be well modeled using the system

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{N\times N}$, $\mathbf{B} \in \mathbb{C}^{N\times P}$, and with

$$\mathbf{u}_t = \begin{bmatrix} u_t^{(1)} & \dots & u_t^{(P)} \end{bmatrix}^T \in \mathbb{C}^P$$

denoting an external input affecting the graph signal. Using such a model, the matrix $\mathbf{A}$ encodes the internal structure of the graph, i.e., the matrix details the signal's evolution when not affected by an external input. In applications, $\mathbf{A}$ may be inferred from, e.g., the sensor network topology or the coupling structure. The external input signal may be used to model, e.g., electromagnetic, acoustic, or seismic waves impinging on a sensor network, with $\mathbf{B}$ reflecting, e.g., the considered source locations. To simplify the exposition below, let the column structure of $\mathbf{B}$ be formed as

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_P \end{bmatrix}.$$

In this work, we will consider input signals $\mathbf{u}_t$ where each component $u_t^{(k)}$ may be well modelled as a wide-sense stationary (WSS) stochastic process, i.e., the covariance function $r_k(\tau)$, defined as $\mathbb{E}(u_t^{(k)}\overline{u}_{t-\tau}^{(k)})$, where $\mathbb{E}(\cdot)$ and $\overline{u}$ denote the statistical expectation and the complex conjugate, respectively, is assumed to only depend on the lag τ and not the actual time t. As noted, some applications prohibit a direct processing of the measured signal $\mathbf{x}_t$, sampled at times $t = 1, 2, \dots, T$, directly, due to, e.g., security or computational considerations. In order to allow for measuring distances between such graph signals, for instance in order to classify an observed signal in such applications, we will here only exploit second-order moments, i.e., the covariance matrix of the signal $\mathbf{x}_t$, in forming the distance measure. Similar to [9], the signals' spectral representations are here used to define the distance between the covariance matrices corresponding to the processes on the graph. Specifically, these spectral representations are directly related to the exogenous input signals, as detailed in the following proposition[1] (see also [14], [15]).

Proposition 1. *Let the matrix $\mathbf{A}$ be stable, and let the components of the input signal $\mathbf{u}_t$ be mutually independent WSS stochastic processes with power spectral densities $\Phi_1, \dots, \Phi_P \in \mathcal{M}_+(\mathbb{T})$, with $\mathcal{M}_+(\mathbb{T})$ denoting non-negative measures on $\mathbb{T}$, and $\mathbb{T} = (-\pi, \pi]$. Then, the graph signal $\mathbf{x}_t$ is a WSS process, and the covariance matrix $\mathbf{R} \triangleq \mathbb{E}\left(\mathbf{x}_t\mathbf{x}_t^H\right)$ is the unique matrix satisfying*

$$\mathbf{R} - \mathbf{A}\mathbf{R}\mathbf{A}^H = \sum_{k=1}^{P} \Gamma_k\left(\Phi_k\right) \tag{2}$$

[1]The proofs for the propositions are given in the appendix.

where $(\cdot)^H$ denotes conjugate transpose, and $\Gamma_k : \mathcal{M} \to \mathbb{C}^{N\times N}$ is detailed as

$$\begin{aligned}\Gamma_k(\Phi) = &\int_{-\pi}^{\pi} \Phi(\theta) \sum_{\tau=0}^{\infty} e^{-i\theta\tau}\mathbf{A}^\tau\mathbf{b}_k\mathbf{b}_k^H \frac{d\theta}{2\pi} \\ &+ \int_{-\pi}^{\pi} \Phi(\theta)\mathbf{b}_k\mathbf{b}_k^H \sum_{\tau=0}^{\infty} e^{i\theta\tau}(\mathbf{A}^\tau)^H \frac{d\theta}{2\pi} \\ &- \mathbf{b}_k\mathbf{b}_k^H \int_{-\pi}^{\pi} \Phi(\theta)\frac{d\theta}{2\pi},\end{aligned} \tag{3}$$

and where $\mathbf{b}_k$, for $k = 1, \dots, P$, are the columns of $\mathbf{B}$.

Proposition 1 provides a direct link between the power spectra of the input signals and the graph signal covariance matrix, thereby allowing for analysis of the behavior of the graph signal, as detailed by its second-order statistics, in terms of the structure of the driving exogenous process. Here, we say that a set $\{\Phi_k \mid k = 1, \dots, P\}$ of power spectra corresponding to the input signal $\mathbf{u}_t$ are *consistent* with an observed covariance matrix $\mathbf{R}$ if equation (2) holds. It may be noted that the application of the operators Γ_k in (3) involve computing a power series in the matrix $\mathbf{A}$. This may at first glance pose a problem, unless $\mathbf{A}$ is nil-potent. However, under the quite general assumption of diagonalizability, the following result may be utilized.

Proposition 2. *In addition to the assumptions of Proposition 1, let $\mathbf{A}$ be diagonalizable according to $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$, where $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_N)$. Then, the operators Γ_k may be expressed as*

$$\begin{aligned}\Gamma_k(\Phi) = &\left(\int_{-\pi}^{\pi} \Phi(\theta)\mathbf{\Psi}(\theta)\frac{d\theta}{2\pi}\right)\mathbf{b}_k\mathbf{b}_k^H + \\ &+ \mathbf{b}_k\mathbf{b}_k^H\left(\int_{-\pi}^{\pi} \Phi(\theta)\mathbf{\Psi}(\theta)\frac{d\theta}{2\pi}\right)^H \\ &- \mathbf{b}_k\mathbf{b}_k^H \int_{-\pi}^{\pi} \Phi(\theta)\frac{d\theta}{2\pi},\end{aligned} \tag{4}$$

for $k = 1, \dots, P$, where the matrix valued function $\mathbf{\Psi} : \mathbb{T} \to \mathbb{C}^{N\times N}$ is detailed as

$$\mathbf{\Psi}(\theta) = \mathbf{S}\operatorname{diag}\left(\frac{1}{1-e^{-i\theta}\lambda_1}, \dots, \frac{1}{1-e^{-i\theta}\lambda_N}\right)\mathbf{S}^{-1}.$$

It should be noted that the representation of Γ_k in (4) more easily lends itself to a practical implementation.

III. Distances between graph processes

Consider a scenario in which two processes, $\{\mathbf{x}_t\}_t$ and $\{\tilde{\mathbf{x}}_t\}_t$, belonging to the same graph description $(\mathbf{A}, \mathbf{B})$, are observed. Let the corresponding, potentially different, covariance matrices be denoted $\mathbf{R}$ and $\tilde{\mathbf{R}}$, respectively. Based on this pair of observed covariance matrices, we want to have a measure of distance between $\mathbf{R}$ and $\tilde{\mathbf{R}}$ in order to quantify dissimilarities between the processes $\{\mathbf{x}_t\}_t$ and $\{\tilde{\mathbf{x}}_t\}_t$. A potential candidate distance is the Euclidean metric on $\mathbb{C}^{N\times N}$, i.e., the standard Frobenius norm. However, this distance does not take the underlying graph structure into account, nor

does it provide a direct link to the underlying generative mechanism of the process, i.e., the influence of the input signal. In order to take the graph structure into account and allow for the influence of the input, we herein propose to utilize the framework recently proposed in [9] for defining distances between covariance matrices by relating these to an underlying spectral domain. These ideas have recently also been used in the context of robust direction-of-arrival estimation localization [16], [17]. The work-horse of these methods is the Monge-Kantorovich problem of optimal mass transport (see, e.g., [10]). This problem is concerned with finding the optimal coupling between two mass distributions [10], and has recently been used to induce metric structure on the space of power spectra [18], convex clustering [19], regularizing inverse problems [20], and modeling and control of stochastic systems [21]. The modeling potential of OMT has also gathered increasing attention in a plethora of signal processing and machine learning applications (see, e.g., [22] and the references therein).

Building on the framework introduced in [9], we here propose to define the distance between $\mathbf{R}$ and $\tilde{\mathbf{R}}$, and thereby $\{\mathbf{x}_t\}_t$ and $\{\tilde{\mathbf{x}}_t\}_t$, by considering the transport distances between the power spectra of the components of the input signal $\mathbf{u}_t$. Specifically, letting $c : \mathbb{T}^2 \to \mathbb{R}_+$, where $\mathbb{T}^2 = \mathbb{T} \times \mathbb{T}$, be a cost function, we define the distance between $\mathbf{R}$ and $\tilde{\mathbf{R}}$ as the minimum objective value of

$$\begin{aligned} &\underset{M_k \in \mathcal{M}_+(\mathbb{T}^2)}{\text{minimize}} \sum_{k=1}^{P} \int_{\mathbb{T}^2} M_k(\theta,\varphi) c(\theta,\varphi) d\theta d\varphi \\ &\text{subject to } \mathbf{R} - \mathbf{A}\mathbf{R}\mathbf{A}^H = \sum_{k=1}^{P} \Gamma_k \left(\int_{\mathbb{T}} M_k(\cdot,\varphi) d\varphi \right), \quad (5) \\ &\qquad \tilde{\mathbf{R}} - \mathbf{A}\tilde{\mathbf{R}}\mathbf{A}^H = \sum_{k=1}^{P} \Gamma_k \left(\int_{\mathbb{T}} M_k(\theta,\cdot) d\theta \right). \end{aligned}$$

Here, M_k denotes the transport plan (see, e.g., [10]) between two spectra corresponding to the input signal component $u^{(k)}$. That is, the margins of M_k, i.e., $\int_{\mathbb{T}} M_k(\cdot,\varphi)d\varphi$ and $\int_{\mathbb{T}} M_k(\theta,\cdot)d\theta$ are power spectra of $u^{(k)}$. The two constraint equations ensure that these margins are consistent with the observed covariance matrices $\mathbf{R}$ and $\tilde{\mathbf{R}}$, i.e., $\left\{\int_{\mathbb{T}} M_k(\cdot,\varphi)d\varphi \mid k = 1,\dots,P\right\}$ and $\left\{\int_{\mathbb{T}} M_k(\theta,\cdot)d\theta \mid k = 1,\dots,P\right\}$ are sets of spectra consistent with $\mathbf{R}$ and $\tilde{\mathbf{R}}$, respectively. Thus, the minimum objective value corresponds to the most cost efficient way of morphing power spectra of the input signal components, such that these spectra are consistent with the measurements. By definition, the distance measure defined via the problem in (5) directly takes the graph structure and underlying dynamic into account. That is, perturbations of the spectra of the input signal are explicitly modeled, allowing for interpretability of the obtained distance (see also [9] and [18] for a discussion of the modeling properties). It may be noted that the problem in (5) is convex, and a direct discretization of the problem yields a linear program that may be directly implemented using off-the-shelf solvers. In the numerical examples of this work, we use CVX [23] to form the implementations. However, it should be noted that computationally efficient solution algorithms may be constructed along the lines in [24]. For the choice of the cost function c, this may be chosen as to reflect modeling considerations. Throughout the numerical section, we will use the squared distance on the unit torus, i.e., $c(\theta,\varphi) = \left|e^{i\theta} - e^{i\varphi}\right|^2$.

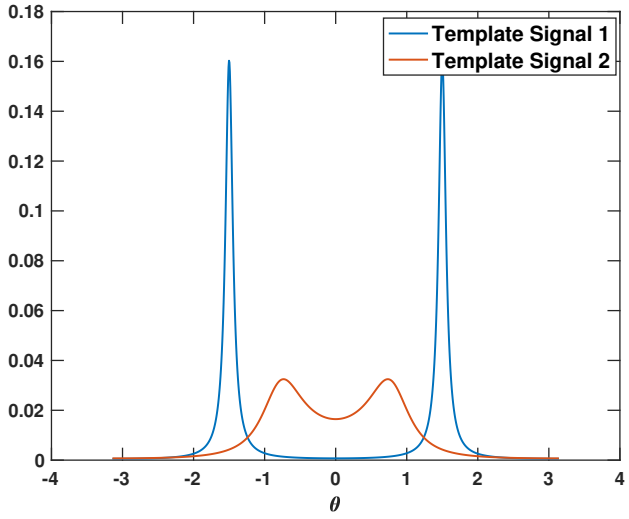

Fig. 1. Spectra of the two template signals.

IV. Numerical examples

To illustrate the use of the proposed framework, we consider the classification of a signal that is somewhat mismatched to its assumed template, a situation commonly occurring in various applications. Such a situation would, for example, occur in several forms of spectroscopy, where the signal frequencies will depend on the temperature of the observed sample, implying that even minor deviations in temperature between the reference measurement and the sample of interest will result in an undesired spectral mismatch (see, e.g., [25]). Another common situation would be when the template spectra have been formed under ideal conditions, for example in a laboratory, whereas the measurement of interest is made in a less ideal situation, possibly using cheaper sensors. To mimic such a situation, we here consider two template graph signals, and a test signal that is to be matched. The template signals are illustrated in Figure 1. Here, both the templates and the measured signals are generated using (1), with each template being formed using a randomly generated symmetric and stable $\mathbf{A}$ matrix, and a randomly generated $\mathbf{B}$. The input signals for the first and second templates were generated as AR(2)-processes, with complex conjugated poles at $0.9395e^{1.5i}$ and $0.7002e^{0.7984i}$, respectively. For the measured signal, which is a mismatched version of the second template, the input was instead generated using an AR(2)-process $0.7002e^{\theta i}$, where the angle, θ, varies along a 100 point grid on $(0,\pi]$. Figure 2 illustrates how the distance measures change for different frequencies between the measured signal and the

two template signals using both the OMT framework and the Euclidean distance. In the figure, it may be seen that when the frequency of the measured signal aligns with the frequency with one of the templates the distance is minimal using both the Euclidean distance and the OMT framework. However, what is interesting to note is that the OMT distance is, as one may expect, small in the region around the true frequency of the template signal, whereas the Euclidean distance makes a sharp dip at just the correct frequency, only to quickly increase again for small frequency offsets. This behavior indicates that the Euclidean distance is not a good similarity measure, as it will be very sensitive to small perturbations, whereas the OMT on the other hand is more robust. As this example illustrates, the OMT measure is able to better represent the distance between the graph signals than the Euclidean measure allows. In these simulations, we have implemented both the proposed OMT distance, given in (5), and the Euclidean distance using CVX [23].

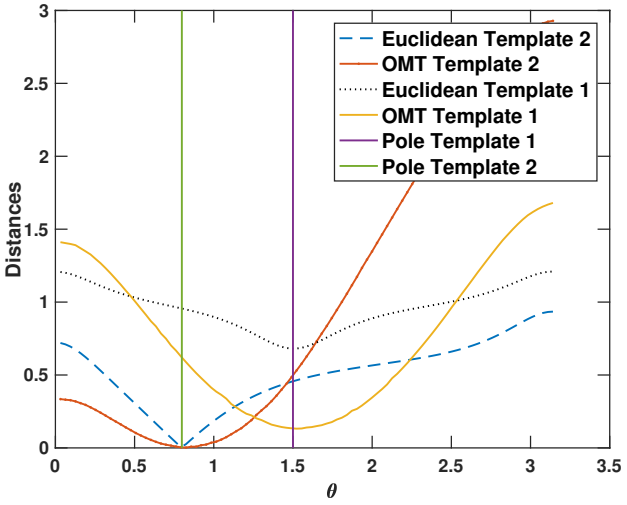

Fig. 2. OMT distance as well as Euclidean distance between the measured signal and the template spectra for different frequencies of the measured signal. The poles of the templates are marked by horizontal lines. with the green line and purple line corresponding to the poles of templates 1 and 2, respectively.

V. Conclusion

In this paper, we have utilized the framework of optimal mass transport in order to formulate a measure of distance, or dissimilarity, between signals defined on a graph. Relying only on the availability of the second-order moments of the signal, the proposed distance is applicable also to scenarios in which storing or transmission of the original graph signal is not feasible. Specifically, the proposed distance is defined in terms of the power spectral densities of external input signals affecting the signal observed on the graph. The considered numerical examples indicate promising robustness and interpretability of the proposed distance, as compared to the Euclidean metric.

Appendix

A. Proof of Proposition 1

As the system is linear and time-invariant and $\mathbf{A}$ is stable, it holds that $\{\mathbf{x}_t\}_t$ is a WSS process. Then,

$$\begin{aligned}\mathbf{R} &= \mathbb{E}\left[(\mathbf{A}\mathbf{x}_{t-1}+\mathbf{B}\mathbf{u}_t)(\mathbf{A}\mathbf{x}_{t-1}+\mathbf{B}\mathbf{u}_t)^H\right]\\ &= \mathbf{A}\mathbf{R}\mathbf{A}^H + \mathbf{A}\mathbb{E}\left[\mathbf{x}_{t-1}\mathbf{u}_t^H\right]\mathbf{B}^H\\ &\quad + \left(\mathbf{A}\mathbb{E}\left[\mathbf{x}_{t-1}\mathbf{u}_t^H\right]\mathbf{B}^H\right)^H + \mathbf{B}\mathbb{E}\left[\mathbf{u}_t\mathbf{u}_t^H\right]\mathbf{B}^H.\end{aligned}$$

Expanding the second term yields

$$\begin{aligned}\mathbf{A}\mathbb{E}\left[\mathbf{x}_{t-1}\mathbf{u}_t^H\right]\mathbf{B}^H &= \sum_{\tau=1}^{\infty}\mathbf{A}^\tau\mathbf{B}\mathbb{E}\left[\mathbf{u}_{t-\tau}\mathbf{u}_t^H\right]\mathbf{B}^H\\ &= \sum_{\tau=1}^{\infty}\mathbf{A}^\tau\mathbf{B}\mathbf{R}_u(\tau)\mathbf{B}^H\\ &= \sum_{\tau=1}^{\infty}\mathbf{A}^\tau\left(\sum_{k=1}^{P}r_k(\tau)\mathbf{b}_k\mathbf{b}_k^H\right)\\ &= \sum_{k=1}^{P}\left(\sum_{\tau=1}^{\infty}r_k(\tau)\mathbf{A}^\tau\right)\mathbf{b}_k\mathbf{b}_k^H,\end{aligned}$$

where we have used the assumption of independence between the components of the input signal, i.e.,

$$\mathbf{R}_u(\tau) \triangleq \mathbb{E}\left(\mathbf{u}_t\mathbf{u}_{t-\tau}^H\right) = \operatorname{diag}\begin{pmatrix}r_1(\tau) & \dots & r_P(\tau)\end{pmatrix}.$$

As $\mathbf{A}$ is assumed stable, the involved series are convergent. Finally, as each component $u_t^{(k)}$ is WSS, there exists non-negative measures, i.e., power spectra, Φ_k, such that (see, e.g., [26])

$$r_k(\tau) = \int_{-\pi}^{\pi} e^{-i\theta\tau}\Phi_k(\theta)\frac{d\theta}{2\pi},$$

allowing us to write

$$\begin{aligned}\sum_{\tau=1}^{\infty}r_k(\tau)\mathbf{A}^\tau &= \sum_{\tau=1}^{\infty}\int_{-\pi}^{\pi}\Phi_k(\theta)e^{-i\theta\tau}\frac{d\theta}{2\pi}\mathbf{A}^\tau\\ &= \int_{-\pi}^{\pi}\Phi_k(\theta)\sum_{\tau=1}^{\infty}e^{-i\theta\tau}\mathbf{A}^\tau\frac{d\theta}{2\pi}.\end{aligned}$$

The uniqueness of $\mathbf{R}$ follows directly from the stability of $\mathbf{A}$, concluding the proof. □

B. Proof of Proposition 2

By assumption, $|\lambda_n| < 1$ for $n = 1, \dots, N$. Then,

$$\begin{aligned}\sum_{\tau=0}^{\infty}e^{-i\theta\tau}\mathbf{A}^\tau &= \mathbf{S}\sum_{\tau=0}^{\infty}e^{-i\theta\tau}\mathbf{\Lambda}^\tau\mathbf{S}^{-1}\\ &= \mathbf{S}\sum_{\tau=0}^{\infty}\operatorname{diag}\left(e^{-i\theta\tau}\lambda_1^\tau,\dots,e^{-i\theta\tau}\lambda_N^\tau\right)\mathbf{S}^{-1}\\ &= \mathbf{S}\operatorname{diag}\left(\sum_{\tau=0}^{\infty}e^{-i\theta\tau}\lambda_1^\tau,\dots,\sum_{\tau=0}^{\infty}e^{-i\theta\tau}\lambda_N^\tau\right)\mathbf{S}^{-1}\\ &= \mathbf{S}\operatorname{diag}\left(\frac{1}{1-e^{-i\theta}\lambda_1},\dots,\frac{1}{1-e^{-i\theta}\lambda_N}\right)\mathbf{S}^{-1}.\end{aligned}$$

The statement of the proposition then follows directly. □

REFERENCES

[1] W. Huang, T. A. W. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van. De Ville, "A Graph Signal Processing Perspective on Functional Brain Imaging," *arXiv preprint arXiv:1710.01135v3*, 2018.

[2] D. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The Emerging Field of Signal Processing on Graphs: Extending High Dimensional Data Analysis to Networks and Other Irregular Domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[3] P. M. Djuric and C. Richard, *Cooperative and Graph Signal Processing, Principles and Applications*, Academic Press, 2018, pp. 261-282.

[4] I. Jablonski, "Graph Signal Processing in Applications to Sensor Networks, Smart Grids, and Smart Cities," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7659 – 7666, 2017.

[5] R. Pastor-Satorras, C. Castellano, P.V. Mieghem, and A. Vespignani, "Epidemic Processes in Complex Networks," *Reviews of Modern Physics*, vol. 87, no. 3, pp. 925–979, 2015.

[6] G. Nikolentzos, P. Meladianos, F. Rousseau, Y. Stavrakas, and M. Vazirgiannis, "Shortest-path Graph Kernels for Document Similarity," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 1890–1900.

[7] Y. Rubner, C. Tomasi, and L. Guibas, "The Earth Mover's Distance as a Metric for Image Retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[8] H. Krim and M. Viberg, "Two Decades of Array Signal Processing Research," *IEEE Signal Process. Mag.*, pp. 67–94, July 1996.

[9] F. Elvander, A. Jakobsson, and J. Karlsson, "Interpolation and Extrapolation of Toeplitz Matrices via Optimal Mass Transport," *IEEE Trans. Signal. Process*, vol. 66, no. 20, pp. 5285 – 5298, Oct. 2018.

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

[11] E. Simou and P. Frossard, "Graph Signal Representation with Wasserstein Barycenters," in *Proc. 44th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Brighton, UK, May 13-17 2019, pp. 5386–5390.

[12] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal Transport Over a Linear Dynamical System," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2137–2152, May 2017.

[13] Y. Chen and J. Karlsson, "State Tracking of Linear Ensembles via Optimal Mass Transport," *IEEE Control Systems Letters*, vol. 2, no. 2, pp. 260–265, Apr 2018.

[23] Inc. CVX Research, "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta," `http://cvxr.com/cvx`, Sept. 2012.

[14] T. 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, 2002.

[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] F. Elvander, I. Haasler, A. Jakobsson, and J. Karlsson, "Tracking and Sensor Fusion in Direction of Arrival Estimation Using Optimal Mass Transport," in *26th European Signal Processing Conference*, Rome, Italy, Sep. 3-7 2018, pp. 1617–1621.

[17] F. Elvander, I. Haasler, A. Jakobsson, and J. Karlsson, "Non-Coherent Sensor Fusion via Entropy Regularized Optimal Mass Transport," in *Proc. 44th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Brighton, UK, May 13-17 2019, pp. 4415–4419.

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

[19] 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.

[20] J. Adler, A. Ringh, O. Öktem, and J. Karlsson, "Learning to solve inverse problems using Wasserstein loss," *arXiv preprint arXiv:1710.10898*, 2017.

[21] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal Steering of a Linear Stochastic System to a Final Probability Distribution - Part III," *IEEE Trans. Autom. Control*, vol. 63, no. 9, pp. 3112–3118, Sep. 2018.

[22] 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.

[24] F. Elvander, I. Haasler, A. Jakobsson, and J. Karlsson, "Multi-Marginal Optimal Mass Transport with Partial Information," *arXiv:1905.03847*, 2019.

[25] A. Jakobsson, M. Mossberg, M. Rowe, and J. A. S. Smith, "Exploiting Temperature Dependency in the Detection of NQR Signals," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1610–1616, May 2006.

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