

STATIONARY GRAPH PROCESSES: PARAMETRIC POWER SPECTRAL ESTIMATION

Santiago Segarra[†], Antonio G. Marques^{*}, Geert Leus[‡], and Alejandro Ribeiro[§]

[†]IDSS, Massachusetts Institute of Technology, Cambridge, MA, USA

^{*}Dept. of TSC, King Juan Carlos University, Madrid, Spain

[‡]Dept. of EEMCS, Delft University of Technology, Delft, Netherlands

[§]Dept. of ESE, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

Advancing a holistic theory of networks and network processes requires the extension of existing results in the processing of time-varying signals to signals supported on graphs. This paper focuses on the definition of stationarity and power spectral density for random graph signals, generalizes the concepts of autoregressive and moving average random processes to the graph domain, and investigates their parametric spectral estimation. Theoretical and algorithmic results are complemented with numerical tests on synthetic and real-world graphs.

Index Terms— Graph signal processing, Stationarity, Power spectral density, Parametric estimation, ARMA graph processes.

1. INTRODUCTION

To cope with challenges posed by network science and big data, current results in modeling, analysis and processing of time-varying signals must be extended to deal with signals supported on irregular domains represented by graphs [1]. This approach has been successfully applied to a number of fundamental problems such as frequency analysis [1, 2], filtering [3–5], or sampling and reconstruction of graph signals [6–10], to name a few. The problems investigated in this paper are the generalization of the notions of stationary processes [11, 12] and power spectral density (PSD) to the graph domain [13], along with schemes to estimate the parameters describing the processes as well as their associated PSDs. To be more concrete, we first define stationary graph processes as those that can be modeled as the output of a linear shift-invariant graph filter applied to a white input and discuss the implications of such a definition in terms of the generalization of the power spectral density. Since stationary graph processes are unequivocally characterized by their generating filter, different types of processes (each associated with a type of filter) are considered. These include autoregressive (AR), moving average (MA) and ARMA processes. Such models are then leveraged to develop parametric methods to estimate the filters' parameters, which also yield a means to estimate their PSD.

Preliminary results generalizing the definition of stationarity to graph signals for Laplacian shifts were reported in [13, 14]. A straightforward generalization is not trivial because the shift operation in the graph domain is more involved. It changes the energy of the shifted signal (unless normalized [13]), and its effect in the frequency domain is more difficult to analyze. Our contribution is to consider general normal shifts and draw a parallel between the time and graph domains. While our previous work dealt with *non-parametric* PSD estimation methods [15], here the focus is on the

definition of *parametric* models for random graph processes along with efficient methods for their estimation.

1.1. Graph signals and filters

Let¹ $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be a directed graph or network with a set of N nodes $\mathcal{N}$ and directed edges $\mathcal{E}$ such that $(i, j) \in \mathcal{E}$ implies that node i is connected to node j . We associate with $\mathcal{G}$ the graph-shift operator $\mathbf{S}$, defined as an $N \times N$ matrix whose entry $S_{ji} \neq 0$ only if $i = j$ or if $(i, j) \in \mathcal{E}$ [3]. The sparsity pattern of the matrix $\mathbf{S}$ captures the local structure of $\mathcal{G}$. Frequent choices for $\mathbf{S}$ are the adjacency matrix of the graph and its Laplacian [1, 3]. The intuition behind $\mathbf{S}$ is to represent a linear transformation that can be computed locally at the nodes of the graph. More rigorously, if the set $\mathcal{N}_l(i)$ stands for the nodes within the l -hop neighborhood of node i and the signal $\mathbf{y}$ is defined as $\mathbf{y} = \mathbf{S}\mathbf{x}$, then node i can compute y_i provided that it has access to the values of x_j at $j \in \mathcal{N}_l(i)$. We assume henceforth that $\mathbf{S}$ is *normal*, so that it can be decomposed as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ with $\mathbf{V}$ being unitary and $\mathbf{\Lambda}$ diagonal.

Graph signals: We do not focus on $\mathcal{G}$, but rather on graph signals defined on the set of nodes $\mathcal{N}$. Formally, each of these signals can be represented as a vector $\mathbf{x} = [x_1, \dots, x_N]^T \in \mathbb{R}^N$ where the i -th element represents the value of the signal at node i or, alternatively, as a function $f: \mathcal{N} \rightarrow \mathbb{R}$, defined on the vertices of the graph. Given a graph signal $\mathbf{x}$, we refer to $\tilde{\mathbf{x}} := \mathbf{V}^H \mathbf{x}$ as the frequency representation of $\mathbf{x}$, with $\mathbf{V}^H$ being the graph Fourier transform (GFT) [2].

Graph filters: A graph filter is a linear graph-signal operator $\mathbf{H}: \mathbb{R}^N \rightarrow \mathbb{R}^N$ of the form $\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l$, where $\mathbf{h} = [h_0, \dots, h_{L-1}]^T$ is a vector of $L \leq N$ scalar coefficients. Graph filters are then polynomials of degree $L - 1$ in the graph-shift operator $\mathbf{S}$ [3], which due to the local structure of the shift can be implemented locally [4, 5]. It is easy to see that graph filters are invariant to applications of the shift in the sense that if $\mathbf{y} = \mathbf{H}\mathbf{x}$, it must hold that $\mathbf{S}\mathbf{y} = \mathbf{H}(\mathbf{S}\mathbf{x})$. Using the factorization $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ the filter $\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l$ can be rewritten as $\mathbf{H} = \mathbf{V}(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l) \mathbf{V}^H := \mathbf{V} \text{diag}(\tilde{\mathbf{h}}) \mathbf{V}^H$. The $N \times 1$ vector $\tilde{\mathbf{h}}$ is termed the frequency response of the filter. To relate $\tilde{\mathbf{h}}$ with the filter coefficients $\mathbf{h}$ let $\lambda_k = [\mathbf{\Lambda}]_{kk}$ be the k th eigenvalue of $\mathbf{S}$ and define the $N \times N$ Vandermonde matrix Ψ with entries $\Psi_{kl} = \lambda_k^{l-1}$. Further define Ψ_L as a tall matrix containing the first L columns of Ψ to write $\tilde{\mathbf{h}} = \Psi_L \mathbf{h}$ and conclude that the filter can be alternatively written as $\mathbf{H} = \mathbf{V} \text{diag}(\tilde{\mathbf{h}}) \mathbf{V}^H = \mathbf{V} \text{diag}(\Psi_L \mathbf{h}) \mathbf{V}^H$. This

¹Notation: Entries of a vector $\mathbf{x}$ are written as x_i and entries of a matrix $\mathbf{X}$ as X_{ij} ; $\mathbf{X}^*$, $\mathbf{X}^T$, and $\mathbf{X}^H$ denote conjugate, transpose, and conjugate transpose, respectively. For a square matrix $\mathbf{X}$, $\text{diag}(\mathbf{X})$ returns a vector with the diagonal elements of $\mathbf{X}$; for a vector $\mathbf{x}$, $\text{diag}(\mathbf{x})$ denotes a diagonal matrix with diagonal elements $\text{diag}[\text{diag}(\mathbf{x})] = \mathbf{x}$; and $\mathbf{x} \circ \mathbf{y}$ is the elementwise product of $\mathbf{x}$ and $\mathbf{y}$. We use $\mathbf{0}$ and $\mathbf{1}$ to denote the all-zero and all-one vectors.

Work in this paper is supported by USA NSF CCF-1217963 and Spanish MINECO grants TEC2013-41604-R and TEC2016-75361-R.

expression implies that if $\mathbf{y}$ is defined as $\mathbf{y} = \mathbf{H}\mathbf{x}$, its frequency representation $\tilde{\mathbf{y}} = \mathbf{V}^H \mathbf{y}$ satisfies

$$\tilde{\mathbf{y}} = \text{diag}(\Psi_L \mathbf{h}) \mathbf{V}^H \mathbf{x} = \text{diag}(\tilde{\mathbf{h}}) \tilde{\mathbf{x}} = \tilde{\mathbf{h}} \circ \tilde{\mathbf{x}}, \quad (1)$$

demonstrating that filters are orthogonal frequency operators.

2. STATIONARITY AND PSD OF GRAPH PROCESSES

For motivation purposes, let us first analyze the covariance matrix of a graph random process generated by applying a linear graph filter to a white input. Mathematically, consider a graph $\mathcal{G}$ with associated shift operator $\mathbf{S}$ and suppose that $\mathbf{w}$ is a random graph process with mean $\bar{\mathbf{w}} = \mathbb{E}[\mathbf{w}] = \mathbf{0}$ and covariance $\mathbf{C}_w = \mathbb{E}[\mathbf{w}\mathbf{w}^H] = \mathbf{I}$. Suppose now that we have a random process $\mathbf{x}$ whose realizations are generated by applying the graph filter $\mathbf{H} = \sum_{l=0}^{N-1} h_l \mathbf{S}^l$ to a realization of $\mathbf{w}$. Then it holds that the mean of that process is $\bar{\mathbf{x}} = \mathbb{E}[\mathbf{x}] = \mathbb{E}[\mathbf{H}\mathbf{w}] = \mathbf{H}\bar{\mathbf{w}} = \mathbf{0}$ and its covariance $\mathbf{C}_x = \mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{H}\mathbb{E}[\mathbf{w}\mathbf{w}^H]\mathbf{H}^H$ is

$$\mathbf{C}_x = \mathbf{H}\mathbf{H}^H = \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^H \mathbf{V}\text{diag}(\tilde{\mathbf{h}}^*)\mathbf{V}^H = \mathbf{V}\text{diag}(|\tilde{\mathbf{h}}|^2)\mathbf{V}^H. \quad (2)$$

The expression in (2) not only reveals that the color of $\mathbf{x}$ is given by the filter $\mathbf{H}$ but, equally important, that the *eigenvectors* of the covariance matrix $\mathbf{C}_x$ and those of the shift $\mathbf{S}$ are the same.

2.1. Definition of stationarity

Three conditions under which a graph process can be considered (weakly) stationary are given next. Under the assumption that the eigenvalues of $\mathbf{S}$ are distinct, it can be shown that these three conditions are equivalent – proofs are omitted due to space limitations, but they can be found in [16].

Definition 1 Given a normal graph-shift operator $\mathbf{S}$, the zero-mean random process $\mathbf{x}$ is said to be weakly stationary in $\mathbf{S}$ if it satisfies any of the following conditions: a) The process $\mathbf{x}$ can be modeled as the output of a graph filter $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$ applied to an uncorrelated input $\mathbf{w}$ with $\bar{\mathbf{w}} = \mathbf{0}$ and $\mathbf{C}_w = \mathbf{I}$; b) Matrices $\mathbf{C}_x$ and $\mathbf{S}$ are simultaneously diagonalizable; or c) The cross-correlation of the shifted versions of the process satisfies $\mathbb{E}[(\mathbf{S}^a \mathbf{x})_i ((\mathbf{S}^H)^b \mathbf{x})_j^H] = \mathbb{E}[(\mathbf{S}^{(a-c)} \mathbf{x})_i ((\mathbf{S}^H)^{(b+c)} \mathbf{x})_j^H]$ for any positive a, b , and $c \leq a$.

It is important to notice that the condition of stationarity is defined with respect to (w.r.t.) a graph-shift operator $\mathbf{S}$, which is required to be normal. Note also that b) implies that $\mathbf{C}_x$ is diagonalized by the graph Fourier basis, and that the $\mathbf{H}$ operator in c) is not required if $\mathbf{S}$ is symmetric.

When particularizing Def. 1 to time-varying signals (this can be done by setting $\mathbf{S}$ to the adjacency matrix of a directed cycle [1]), the classical definition of stationarity is recovered, requiring $\mathbf{C}_x$ to be circulant (if $\mathbf{S}$ is set to the adjacency matrix of the directed path, then $\mathbf{C}_x$ is Toeplitz). One example of a stationary graph processes is zero-mean white noise, which is stationary in any graph shift $\mathbf{S}$. Moreover, it is also true that any random process is stationary w.r.t.: i) the graph shift $\mathbf{S} = \mathbf{C}_x$ given by its covariance matrix; and ii) the graph shift $\mathbf{S} = \mathbf{C}_x^{-1}$ given by its precision matrix.

2.2. Definition of PSD

The GFT together with Def. 1 can be readily used to generalize the concept of PSD to random graph processes.

Definition 2 The PSD of a random graph process $\mathbf{x}$ that is stationary w.r.t. $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ is the nonnegative $N \times 1$ vector $\mathbf{p}$, where

$$\mathbf{p} := \text{diag}(\mathbf{V}^H \mathbf{C}_x \mathbf{V}). \quad (3)$$

Since $\mathbf{C}_x$ is diagonalized by $\mathbf{V}$, the definition in (3) corresponds to the eigenvalues of the positive-semidefinite matrix $\mathbf{C}_x$, which are always nonnegative. Thus, it holds that $\mathbf{C}_x = \mathbf{V}\text{diag}(\mathbf{p})\mathbf{V}^H$. Note also that if $\mathbf{x}$ is interpreted as the output of a graph filter applied to a white input with unitary variance, it also holds that $\mathbf{p} = |\Psi \mathbf{h}|^2 = |\tilde{\mathbf{h}}|^2$, where $\mathbf{h}$ are the filter coefficients and $|\cdot|^2$ is applied entry-wise.

Property 1 Let $\mathbf{y}$ be the random process modeling the output of a linear graph filter $\mathbf{H} = \sum_{l=0}^{N-1} h_l \mathbf{S}^l$ applied to an input $\mathbf{x}$ that is stationary in $\mathbf{S}$ with covariance $\mathbf{C}_x$ and PSD $\mathbf{p}_x$. Then it holds that process $\mathbf{y}$: a) Is stationary in $\mathbf{S}$ with covariance $\mathbf{C}_y = \mathbf{H}\mathbf{C}_x\mathbf{H}^H$; and b) Has a PSD $\mathbf{p}_y$ given by $\mathbf{p}_y = |\Psi \mathbf{h}|^2 \circ \mathbf{p}_x = |\tilde{\mathbf{h}}|^2 \circ \mathbf{p}_x$.

Property 2 Let $\mathbf{x}$ be a random stationary process in $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ and $\tilde{\mathbf{x}} = \mathbf{V}^H \mathbf{x}$ its frequency representation. Then, it holds that $\mathbf{C}_{\tilde{\mathbf{x}}} := \mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H] - \mathbb{E}[\tilde{\mathbf{x}}]\mathbb{E}[\tilde{\mathbf{x}}]^H = \text{diag}(\mathbf{p})$, so that $\tilde{x}_k$ and $\tilde{x}_{k'}$ are uncorrelated for $k \neq k'$.

Prop. 1 is the counterpart of the spectral convolution theorem for graph processes, Prop. 2 provides motivation for the analysis and modeling of stationary graph processes in the *spectral* domain. It also shows that if a process $\mathbf{x}$ is stationary in the shift $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$, then the GFT $\mathbf{V}^H$ provides the Karhunen-Loève expansion of the process. This is as it should be because $\mathbf{C}_x$ is diagonalized by $\mathbf{V}$.

3. PARAMETRIC PSD ESTIMATION

We address PSD estimation assuming that the graph process $\mathbf{x}$ is the *filtered* version of a white input, with the filter $\mathbf{H}$ being well approximated by a *parametric* model. As per Def. 1.a, this is always possible if the degree $L-1$ of $\mathbf{H}$ is $N-1$. The goal here is to devise filter representations of an order much smaller than N . Mimicking time processes, we devise MA, AR, and ARMA models [5]. To run the estimation algorithms we assume that a set $\{\mathbf{x}_r\}_{r=1}^R$ of realizations of the random process $\mathbf{x}$ is available and use those to form the *sample covariance* $\hat{\mathbf{C}}_x^{\text{sp}} = \frac{1}{R} \sum_{r=1}^R \mathbf{x}_r \mathbf{x}_r^T$ and the *average periodogram* $\hat{\mathbf{p}}_x^{\text{pg}} = \frac{1}{R} \sum_{r=1}^R |\mathbf{V}^H \mathbf{x}_r|^2$.

3.1. MA graph processes

Consider a vector of coefficients $\boldsymbol{\beta} = [\beta_0, \dots, \beta_{L-1}]^T$ and assume that $\mathbf{x}$ is stationary in $\mathbf{S}$ and generated by the FIR filter $\mathbf{H}(\boldsymbol{\beta}) = \sum_{l=0}^{L-1} \beta_l \mathbf{S}^l$. The degree of the filter is less than $N-1$ although we want in practice to have $L \ll N$. If the process is indeed generated as the response of $\mathbf{H}(\boldsymbol{\beta})$ to a white input, the covariance matrix of the process can be written as

$$\mathbf{C}_x(\boldsymbol{\beta}) = \mathbf{H}(\boldsymbol{\beta})\mathbf{H}^H(\boldsymbol{\beta}) = \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} (\beta_l \mathbf{S}^l)(\beta_{l'} \mathbf{S}^{l'})^H. \quad (4)$$

The PSD corresponding to the covariance in (4) is the magnitude squared of the frequency representation of the filter. To see this formally, notice that it follows from the definition in (3) that $\mathbf{p}(\boldsymbol{\beta}) = \text{diag}(\mathbf{V}^H \mathbf{C}_x(\boldsymbol{\beta}) \mathbf{V})$. Writing the covariance matrix as $\mathbf{C}_x(\boldsymbol{\beta}) = \mathbf{H}(\boldsymbol{\beta})\mathbf{H}^H(\boldsymbol{\beta})$ and the frequency representation of the filter as $\tilde{\mathbf{h}}(\boldsymbol{\beta}) = \text{diag}(\mathbf{V}\mathbf{H}(\boldsymbol{\beta})\mathbf{V}^H)$, it follows readily that $\mathbf{p}(\boldsymbol{\beta}) = |\tilde{\mathbf{h}}(\boldsymbol{\beta})|^2$. For the purposes of this section it is convenient to write the latter explicitly in terms of $\boldsymbol{\beta}$ as [cf. (1)]

$$\mathbf{p}(\boldsymbol{\beta}) = |\tilde{\mathbf{h}}(\boldsymbol{\beta})|^2 = |\Psi_L \boldsymbol{\beta}|^2. \quad (5)$$

The covariance and PSD expressions in (4) and (5) are graph counterparts of MA time processes generated by FIR filters.

The estimation of the coefficients $\boldsymbol{\beta}$ can be addressed in either the graph or graph frequency domains. In the graph domain we compute the sample covariance $\hat{\mathbf{C}}_x^{\text{sp}}$ and introduce a distortion function

$D_C(\hat{\mathbf{C}}_x^{\text{sp}}, \mathbf{C}_x(\beta))$ to measure the similarity of $\hat{\mathbf{C}}_x^{\text{sp}}$ and $\mathbf{C}_x(\beta)$. The coefficients β are then selected as the ones with minimal distortion

$$\hat{\beta} = \operatorname{argmin}_{\beta} D_C(\hat{\mathbf{C}}_x^{\text{sp}}, \mathbf{C}_x(\beta)). \quad (6)$$

The expression for $\mathbf{C}_x(\beta)$ in (4) is a quadratic function of β that is generally indefinite. Hence, (6) will be not convex in general.

To perform estimation in the frequency domain we first compute the periodogram $\hat{\mathbf{p}}_x^{\text{pg}}$. We then introduce a distortion measure $D_P(\hat{\mathbf{p}}_x^{\text{pg}}, |\Psi_L \beta|^2)$ to compare $\hat{\mathbf{p}}_x^{\text{pg}}$ with the PSD $|\Psi_L \beta|^2$ and select the coefficients β that solve the following optimization

$$\hat{\beta} := \operatorname{argmin}_{\beta} D_P(\hat{\mathbf{p}}_x^{\text{pg}}, |\Psi_L \beta|^2). \quad (7)$$

Since the quadratic form $|\Psi_L \beta|^2$ in (7) is also indefinite, the problem in (7) is not necessarily convex. In the particular case when the distortion $D_P(\hat{\mathbf{p}}_x^{\text{pg}}, |\Psi_L \beta|^2) = \|\hat{\mathbf{p}}_x^{\text{pg}} - |\Psi_L \beta|^2\|_2^2$ is the Euclidean 2-norm, efficient (phase-retrieval) solvers with probabilistic guarantees are available [17, 18]. Tractable formulations of (6) and (7) when $\mathbf{S}$ is symmetric, or, when $\mathbf{S}$ is positive semidefinite and the filter coefficients $\mathbf{h}$ are nonnegative are discussed below.

Symmetric shifts. If $\mathbf{S}$ is symmetric, the expression for $\mathbf{C}_x$ in (4) can be simplified to a polynomial of degree $2(L-1)$ in $\mathbf{S}$

$$\mathbf{C}_x(\beta) = \sum_{l=0}^{L-1} \beta_l \beta_{l'} \mathbf{S}^{l+l'} := \sum_{l=0}^{2(L-1)} \gamma_l \mathbf{S}^l := \mathbf{C}_x(\gamma). \quad (8)$$

In the second equality in (8) we have defined the coefficients $\gamma_l := \sum_{l'+l''=l} \beta_{l'} \beta_{l''}$ summing all the coefficient crossproducts that multiply $\mathbf{S}^l$ and introduced $\mathbf{C}_x(\gamma)$ to denote the covariance matrix written in terms of the γ coefficients. We propose now a relaxation of (6) in which $\mathbf{C}_x(\gamma)$ is used in lieu of $\mathbf{C}_x(\beta)$ to yield

$$\hat{\gamma} = \operatorname{argmin}_{\gamma} D_C(\hat{\mathbf{C}}_x^{\text{sp}}, \mathbf{C}_x(\gamma)). \quad (9)$$

If we add the constraints $\gamma_l = \sum_{l'+l''=l} \beta_{l'} \beta_{l''}$, (9) is equivalent to (6). By dropping them, we end up with a tractable relaxation because (9) is convex for all convex distortion metrics $D_C(\hat{\mathbf{C}}_x^{\text{sp}}, \mathbf{C}_x(\gamma))$. A tractable relaxation of (7) can be derived analogously.

Nonnegative filter coefficients. When the shift $\mathbf{S}$ is positive semidefinite, the elements of the matrix Ψ are all nonnegative. If we further restrict the coefficients β to be nonnegative, all the elements in the product $\Psi_L \beta$ are also nonnegative. This means that in (7) we can replace the comparison $D_P(\hat{\mathbf{p}}_x^{\text{pg}}, |\Psi_L \beta|^2)$ by $D_P(\sqrt{\hat{\mathbf{p}}_x^{\text{pg}}}, \Psi_L \beta)$. We can then replace (7) by

$$\hat{\beta} := \operatorname{argmin}_{\beta \geq 0} D_P(\sqrt{\hat{\mathbf{p}}_x^{\text{pg}}}, \Psi_L \beta). \quad (10)$$

The optimization problem in (10) is convex, therefore tractable, for all convex distortion metrics $D_P(\sqrt{\hat{\mathbf{p}}_x^{\text{pg}}}, \Psi_L \beta)$. Do notice that the objective costs in (10) and (7) are *not* equivalent and that (10) requires positive semidefinite shifts –such as the Laplacian– and restricts coefficients to satisfy $\beta \geq 0$. A tractable restriction of (6) can be derived analogously.

3.2. AR graph processes

For some processes it is more convenient to use a parametric model that generates an infinite impulse response through an AR filter. As a simple example consider a heat diffusion process of the form $\mathbf{x}_t = \alpha \mathbf{S} \mathbf{x}_{t-1} + \alpha_0 \mathbf{w}$, which is completely characterized by the diffusion rate α and the scaling coefficient α_0 . As $t \rightarrow \infty$, this process can be represented as $\mathbf{x}_{\infty} = \mathbf{H} \mathbf{w}$ with the filter $\mathbf{H} = \alpha_0 \sum_{l=0}^{\infty} \alpha^l \mathbf{S}^l$. If the series is summable, the filter can be rewritten as $\mathbf{H} = \alpha_0 (\mathbf{I} -$

$\alpha \mathbf{S})^{-1}$ from where we can conclude that its frequency response is $\tilde{\mathbf{h}} = \operatorname{diag}(\mathbf{V}^H \mathbf{H} \mathbf{V}) = \alpha_0 \operatorname{diag}(\mathbf{I} - \alpha \mathbf{\Lambda}^{-1})$. This demonstrates that $\mathbf{H}$ can be viewed as a single pole AR filter – see also [5].

Suppose now that $\mathbf{x}$ is a random graph process whose realizations are generated by applying $\mathbf{H} = \alpha_0 (\mathbf{I} - \alpha \mathbf{S})^{-1}$ to a white input $\mathbf{w}$. Then, it readily holds that its covariance $\mathbf{C}_x$ is [cf. (2)]

$$\mathbf{C}_x(\alpha_0, \alpha) = \mathbf{H} \mathbf{H}^H = \alpha_0^2 (\mathbf{I} - \alpha \mathbf{S})^{-1} (\mathbf{I} - \alpha \mathbf{S})^{-H}, \quad (11)$$

which implies that the PSD of $\mathbf{x}$ is

$$\mathbf{p}(\alpha_0, \alpha) = \operatorname{diag} [\alpha_0^2 |\mathbf{I} - \alpha \mathbf{\Lambda}|^{-2}], \quad (12)$$

confirming the fact that the expression for the PSD of $\mathbf{x}$ is similar to that of a first-order AR time-varying process. We can now proceed to estimate the PSD utilizing the AR parametric models in (11) and (12) as we did in Section 3.1 for MA models. Substituting $\mathbf{C}_x(\alpha_0, \alpha)$ for $\mathbf{C}_x(\beta)$ in (6) yields a graph domain formulation, and substituting $\mathbf{p}(\alpha_0, \alpha)$ for $|\Psi_L \beta|^2$ in (7) yields a graph frequency domain formulation. Since only two parameters must be estimated the corresponding optimization problems are tractable.

Higher order AR processes: If the filter $\mathbf{H} = \alpha_0 (\mathbf{I} - \alpha \mathbf{S})^{-1}$ is the equivalent of an AR process of order one, an AR process of order M can be written as $\mathbf{H} = \alpha_0 \prod_{m=1}^M (\mathbf{I} - \alpha_m \mathbf{S})^{-1}$ for some set of diffusion rates $\alpha = [\alpha_0, \dots, \alpha_M]^T$. The frequency response $\tilde{\mathbf{h}} = \operatorname{diag}(\mathbf{V}^H \mathbf{H} \mathbf{V})$ is given by $\tilde{\mathbf{h}} = \alpha_0 \operatorname{diag}[\prod_{m=1}^M (\mathbf{I} - \alpha_m \mathbf{\Lambda})^{-1}]$.

If we define the graph process $\mathbf{x} = \mathbf{H} \mathbf{w}$ with $\mathbf{w}$ white and of unitary energy, the covariance matrix $\mathbf{C}_x$ can be written as

$$\mathbf{C}_x(\alpha) = \alpha_0^2 \prod_{m=1}^M (\mathbf{I} - \alpha_m \mathbf{S})^{-1} (\mathbf{I} - \alpha_m \mathbf{S})^{-H}. \quad (13)$$

The process $\mathbf{x}$ is stationary in $\mathbf{S}$ and its PSD is

$$\mathbf{p}(\alpha) = \alpha_0^2 \operatorname{diag} [\prod_{m=1}^M |\mathbf{I} - \alpha_m \mathbf{\Lambda}|^{-2}]. \quad (14)$$

As before, we substitute $\mathbf{C}_x(\alpha)$ for $\mathbf{C}_x(\beta)$ in (6) to obtain a graph domain formulation and substitute $\mathbf{p}(\alpha)$ for $|\Psi_L \beta|^2$ in (7) to obtain a graph frequency domain formulation. For large degree M the problems can become intractable. Design of Yule-Walker schemes [12, Sec. 3.4] for graph signals may help, which is left as future work.

3.3. ARMA graph processes

The techniques in Secs. 3.1 and 3.2 can be combined to form ARMA models for PSD estimation. However, as is also done for time signals, we formulate ARMA filters directly in the frequency domain as a ratio of polynomials in the graph eigenvalues. We then define coefficients $\mathbf{a} := [a_1, \dots, a_M]^T$ and $\mathbf{b} := [b_0, \dots, b_{L-1}]^T$ and postulate filters with frequency response

$$\tilde{\mathbf{h}} = \operatorname{diag} \left[\left(\sum_{l=0}^{L-1} b_l \mathbf{\Lambda}^l \right) \left(1 - \sum_{m=1}^M a_m \mathbf{\Lambda}^m \right)^{-1} \right]. \quad (15)$$

To find the counterpart of (15) in the graph domain define the matrices $\mathbf{B} := \sum_{l=0}^{L-1} b_l \mathbf{S}^l$ and $\mathbf{A} := \sum_{m=1}^M a_m \mathbf{S}^m$. It then follows readily that the filter whose frequency response is in (15) is $\mathbf{H} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} = \mathbf{B} (\mathbf{I} - \mathbf{A})^{-1}$. These expressions confirm that we can interpret the filter as the sequential application of finite and infinite response filters.

If we now define the graph process $\mathbf{x} = \mathbf{H} \mathbf{w}$, its covariance matrix follows readily as

$$\mathbf{C}_x(\mathbf{a}, \mathbf{b}) = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} \mathbf{B}^H (\mathbf{I} - \mathbf{A})^{-H}. \quad (16)$$

Since $\mathbf{C}_x(\mathbf{a}, \mathbf{b})$ is diagonalized by the Fourier basis $\mathbf{V}$, the process $\mathbf{x}$ is stationary with PSD [cf. (15)]

$$\mathbf{p}(\mathbf{a}, \mathbf{b}) = \operatorname{diag} \left[\left| \sum_{l=0}^{L-1} b_l \mathbf{\Lambda}^l \right|^2 \left| 1 - \sum_{m=1}^M a_m \mathbf{\Lambda}^m \right|^{-2} \right]. \quad (17)$$

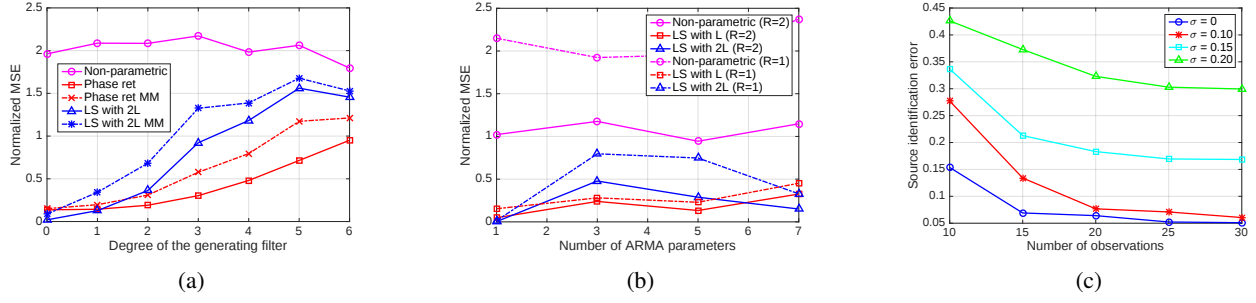


Fig. 1: Error performance for different PSD estimation schemes and networks.

As in the AR and MA models, we can identify the model coefficients by minimizing the covariance distortion $D_C(\hat{\mathbf{C}}_x^{\text{sp}}, \mathbf{C}_x(\mathbf{a}, \mathbf{b}))$ or the PSD distortion $D_P(\hat{\mathbf{p}}_x^{\text{pg}}, \mathbf{p}(\mathbf{a}, \mathbf{b}))$ [cf. (5) and (6)]. These optimization problems are computationally difficult.

Alternative estimation schemes can be obtained by reordering (16) into $(\mathbf{I} - \mathbf{A})\mathbf{C}_x(\mathbf{I} - \mathbf{A})^H = \mathbf{B}\mathbf{B}^H$ and minimizing either the graph domain distortion $D_C((\mathbf{I} - \mathbf{A})\hat{\mathbf{C}}_x^{\text{sp}}(\mathbf{I} - \mathbf{A})^H, \mathbf{B}\mathbf{B}^H)$ or the graph frequency domain distortion $D_P(|1 - \sum_{m=1}^M a_m \Lambda^m|^2 \hat{\mathbf{p}}_x^{\text{pg}}, \text{diag}(|\sum_{l=0}^{L-1} b_l \Lambda^l|^2))$. While these formulations can still be intractable, they have the same structure as their counterparts in Sec. 3.1. Hence, the tractable relaxation that we discussed for symmetric shifts and the tractable restriction to nonnegative filter coefficients for positive semidefinite shifts can be then used here for ARMA (and AR) processes as well.

Remark 1 The methods proposed for PSD estimation can be used for covariance estimation as well. This only requires substituting the optimal filter parameters into (4), (13) or (16).

4. NUMERICAL EXPERIMENTS

Two test cases (TC) are considered. The results shown in Fig. 1 are averages across 100 realizations of the particular experiment.

TC1. Synthetic graphs: We start with MA estimation. Consider the Laplacian of an Erdős-Rényi (ER) graph with $N = 100$ nodes and edge probability $p = 0.2$ [19], and processes generated by filtering white Gaussian noise with an FIR filter of length L whose coefficients β are selected randomly. The performance of an average periodogram $\hat{\mathbf{p}}_x^{\text{pg}}$ is contrasted with that of two parametric approaches: i) an algorithm that estimates the L values in β by minimizing (7) via phase-retrieval [18]; and ii) a least squares (LS) algorithm that estimates the $2L - 1$ values in γ by minimizing (8). The results are shown in Fig. 1a (solid lines). It can be observed that both parametric methods outperform $\hat{\mathbf{p}}_x^{\text{pg}}$ since they leverage the FIR structure of the generating filter. The gap is larger for smaller values of the degree, since in these cases a few parameters are sufficient to completely characterize the generating filter. We also test our schemes for a model mismatch (MM) scenario where the MA schemes assume that the order of the process is $L + 2$ instead of L (dashed lines in Fig. 1a). The results show that, although the MM degrades the performance, the parametric estimates are still superior to $\hat{\mathbf{p}}_x^{\text{pg}}$.

The second experiment considers ARMA processes with L poles and L zeros. The coefficients are drawn randomly from a uniform distribution with support $[0, 1]$ and the shift is selected as in the previous experiment. We compare $\hat{\mathbf{p}}_x^{\text{pg}}$ with two schemes: i) an LS algorithm that estimates $2L$ coefficients, i.e., the counterpart of (8) for an ARMA process; and ii) an LS algorithm that estimates L nonnegative coefficients, i.e., the counterpart of (10). The latter is computationally tractable because both the eigenvalues of $\mathbf{S}$ and the coefficients of the filters are nonnegative. The algorithms are

tested with both one and two signal realizations available. Fig. 1b shows that the parametric methods attain smaller Mean Squared Error (MSE) compared to $\hat{\mathbf{p}}_x^{\text{pg}}$. Moreover, while increasing the number of observations reduces the MSE for all tested schemes, the reduction is more pronounced for nonparametric schemes. This is a manifestation of the fact that parametric approaches tend to be more robust to noisy or imperfect observations.

TC2. Real-world graphs: Consider the social network of Zachary's karate club [20] represented by a graph $\mathcal{G}$ with 34 nodes or members of the club and 78 undirected edges symbolizing friendships among members. Denoting by $\mathbf{L}$ the Laplacian of $\mathcal{G}$, we define the graph shift operator $\mathbf{S} = \mathbf{I} - \alpha \mathbf{L}$ with $\alpha = 1/\lambda_{\max}(\mathbf{L})$, modeling the diffusion of opinions between the members of the club. A signal $\mathbf{x}$ can be regarded as a unidimensional opinion of each club member regarding a specific topic, and each application of $\mathbf{S}$ can be seen as an opinion update. We assume that an opinion profile $\mathbf{x}$ is generated by the diffusion through the network of an initially sparse (rumor) signal $\mathbf{w}$. More precisely, we model $\mathbf{w}$ as a white process such that $w_i = 1$ with probability 0.05, $w_i = -1$ with probability 0.05, and $w_i = 0$ otherwise. We are given a set $\{\mathbf{x}_r\}_{r=1}^R$ of opinion profiles generated from different sources $\{\mathbf{w}_r\}_{r=1}^R$ diffused through a filter of unknown nonnegative coefficients β . Our goal is to identify the sources of the different opinions, i.e., the nonzero entries of $\mathbf{w}_r$ for every r . Our approach proceeds in two phases. First, we use $\{\mathbf{x}_r\}_{r=1}^R$ to identify the parameters β of the generating filter. We do this by solving (10) via LS. Second, given the set of coefficients β , we have that $\mathbf{x}_r = \sum_{l=0}^{L-1} \beta_l \mathbf{S}^l \mathbf{w}_r$. Thus, we estimate the sources $\mathbf{w}_r$ by solving a ℓ_1 -regularized LS problem to promote sparsity in the input. In Fig. 1c (blue) we show the proportion of sources misidentified as a function of the number of observations R . As R increases, the estimates of the parameters β become more reliable, thus leading to a higher success rate. Finally, we consider cases where the observations are noisy. Formally, we define noisy observations $\hat{\mathbf{x}}_r$ by perturbing the original ones $\hat{\mathbf{x}}_r = \mathbf{x}_r + \sigma \mathbf{z} \circ \mathbf{x}_r$ where σ denotes the magnitude of the perturbation and $\mathbf{z}$ is a vector with elements drawn from a standard normal distribution. As expected, higher levels of noise have detrimental effects on the recovery of sources. Nevertheless, for moderate noise levels ($\sigma = 0.1$) a performance comparable to the noiseless case can be achieved when observing 20 signals or more. These findings are consistent for other types of graphs.

5. CONCLUSION

After generalizing the notions of stationarity, power spectral density and ARMA processes to random graph signals, different methods for parametric estimation were proposed. While the general estimation problem is nonconvex, several reformulations were discussed and particular cases where the problem is tractable were identified.

6. REFERENCES

- [1] D. Shuman, S. 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 Process. Mag.*, vol. 30, no. 3, pp. 83–98, Mar. 2013.
- [2] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, June 2014.
- [3] —, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [4] S. Segarra, A. G. Marques, and A. Ribeiro, "Distributed linear network operators using graph filters," *arXiv preprint arXiv:1510.03947*, 2015.
- [5] A. Loukas, A. Simonetto, and G. Leus, "Distributed autoregressive moving average graph filters," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1931–1935, Nov. 2015.
- [6] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2014, pp. 3864–3868.
- [7] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.
- [8] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 1832 – 1843, Apr. 2016.
- [9] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 4363 – 4378, Aug. 2016.
- [10] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *arXiv preprint arXiv:1605.07174*, 2016.
- [11] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 2009.
- [12] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.
- [13] B. Girault, "Stationary graph signals using an isometric graph translation," in *European Signal Process. Conf. (EUSIPCO)*, 2015, pp. 1516–1520.
- [14] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *arXiv preprint arXiv:1601.02522*, 2016.
- [15] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Stationary graph processes: Nonparametric power spectral density estimation," in *IEEE Sensor Array and Multichannel Signal Process. Wrksp.*, Rio de Janeiro, Brazil, July 10-13 2016.
- [16] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *arXiv preprint arXiv:1603.04667*, 2016.
- [17] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.
- [18] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.
- [19] B. Bollobás, *Random Graphs*. Springer, 1998.
- [20] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, 1977.