STATIONARY GRAPH PROCESSES: NONPARAMETRIC SPECTRAL ESTIMATION

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

[†]Dept. of ESE, University of Pennsylvania, Philadelphia, PA, USA
[*]Dept. of TSC, King Juan Carlos University, Madrid, Spain
[‡]Dept. of EEMCS, Delft University of Technology, Delft, Netherlands

ABSTRACT

Stationarity is a cornerstone property that facilitates the analysis and processing of random signals in the time domain. Although time-varying signals are abundant in nature, in many practical scenarios the information of interest resides in more irregular graph domains. The contribution in this paper is twofold. First, we propose several equivalent notions of weak stationarity for random graph signals, all taking into account the structure of the graph where the random process takes place. Second, we analyze the properties of the induced power spectral density along with nonparametric approaches to estimate it, including average and window-based periodograms.

Index Terms— Graph signal processing, Weak stationarity, Periodogram, Windowing, Power spectral density.

1. INTRODUCTION

The growing interest in network science and big data applications has prompted the need to extend the existing results in analysis and processing of time-varying signals to signals supported on graphs [1,2]. This paper investigates the problem of generalizing the notion of stationary processes [3,4] to the graph domain [5], along with nonparametric schemes to estimate their power spectral density (PSD). We begin by proposing different equivalent definitions of weak stationarity for graph signals, all taking into account the structure of the graph where the random process takes place, while inheriting many of the meaningful properties of the classical definition in the time domain. A straightforward generalization is not trivial because the shift (translation) operation in the graph domain is more involved, it changes the energy of the shifted signal (unless normalized [5]), and its effect in the frequency domain is more difficult to analyze. With these considerations in mind, a random process in a graph is said to be stationary if either its correlation is invariant with respect to a constant number of applications of the graph-shift operator; or, if it can be modeled as the output of a linear graph filter applied to a white input; or, if the correlation matrix of the process is diagonalized by the graph Fourier transform (GFT). Under these definitions, notions like the PSD or results such as the spectral convolution theorem can be generalized to signals supported on graphs. After showing that stationary processes are easier to understand in the frequency domain, we propose and analyze two different methods to estimate the PSD: average and window-based periodograms. Their estimation performance is characterized and differences relative to their time-domain counterparts are highlighted.

Preliminary results generalizing the definition of stationarity to graph signals for Laplacian shifts were reported in [5,6]. Our contribution here is to draw a parallel between the fundamental properties of stationary stochastic processes in time and stationary processes in graphs. We also consider general normal shifts, identify properties hitherto unreported, establish connections with application settings such as diffusion dynamics, and formulate and evaluate different approaches for PSD estimation.[1]

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

1.1. Graph Signals and Filters

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ 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 $(i, j) \in \mathcal{E}$ [2, 7]. 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 [2, 7] and its Laplacian [1]. 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 $j \in \mathcal{N}_1(i)$. 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: The focus of this paper is not on $\mathcal{G}$, but on graph signals defined on the set of nodes $\mathcal{N}$. Formally, each of these signals can be represented as a vector $\mathbf{x} = [x_1, ..., x_N]^T \in \mathbb{R}^N$ where the i-th element represents the value of the signal at node i or, alternatively, as a function $f : \mathcal{N} \to \mathbb{R}$, 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 GFT [7].

Graph filters: A graph filter is a linear graph-signal operator $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^N$ of the form $\mathbf{H} := \sum_{l=0}^{L-1} h_l\mathbf{S}^l$, where $\mathbf{h} = [h_0, \ldots, h_{L-1}]^T$ is a vector of $L \leq N$ scalar coefficients. Graph filters are then polynomials of degree $L - 1$ in the graph-shift operator $\mathbf{S}$ [2], which due to the local structure of the shift can be implemented locally too [8, 9]. 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

[1]*Notation:* Entries of a vector $\mathbf{x}$ are written as x_i and entries of a matrix $\mathbf{X}$ as X_{ij}. If needed for clarity we may alternatively write $[\mathbf{x}]_i$ and $[\mathbf{X}]_{ij}$. We use $\mathbf{X}^*$, $\mathbf{X}^T$, and $\mathbf{X}^H$ to denote conjugate, transpose, and conjugate transpose, respectively. For a square matrix $\mathbf{X}$, we use $\mathrm{tr}[\mathbf{X}]$ for its trace and $\mathrm{diag}(\mathbf{X})$ for an operator returning a vector with the diagonal elements of $\mathbf{X}$. For a vector $\mathbf{x}$, we let $\mathrm{diag}(\mathbf{x})$ denote a diagonal matrix with diagonal elements $\mathrm{diag}[\mathrm{diag}(\mathbf{x})] = \mathbf{x}$. The notation $\mathbf{x} \circ \mathbf{y}$ denotes the elementwise product of $\mathbf{x}$ and $\mathbf{y}$. We use $\mathbf{0}$ and $\mathbf{1}$ for the all-zero and all-one vectors and $\mathbf{e}_i$ for the ith element of the canonical basis of $\mathbb{R}^N$.

as $\mathbf{H} = \mathbf{V}\big(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l\big)\mathbf{V}^H := \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^H$. The $N \times 1$ vector $\tilde{\mathbf{h}}$ is termed the frequency response of the filter. To relate the frequency response $\tilde{\mathbf{h}}$ with the filter coefficients $\mathbf{h}$ let $\lambda_k = [\mathbf{\Lambda}]_{kk}$ be the kth eigenvalue of $\mathbf{S}$ and define the $N \times N$ Vandermonde matrix $\mathbf{\Psi}$ with entries $\Psi_{kl} = \lambda_k^{l-1}$. Further define $\mathbf{\Psi}_L$ as a tall matrix containing the first L columns of $\mathbf{\Psi}$ to write $\tilde{\mathbf{h}} = \mathbf{\Psi}_L\mathbf{h}$ and conclude that the filter can be alternatively written as $\mathbf{H} = \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^H = \mathbf{V}\text{diag}\big(\mathbf{\Psi}_L\mathbf{h}\big)\mathbf{V}^H$. This 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

$$\tilde{\mathbf{y}} = \text{diag}\big(\mathbf{\Psi}_L\mathbf{h}\big)\mathbf{V}^H\mathbf{x} = \text{diag}\big(\tilde{\mathbf{h}}\big)\tilde{\mathbf{x}} = \tilde{\mathbf{h}} \circ \tilde{\mathbf{x}}, \tag{1}$$

demonstrating that the output at a given frequency depends only on the value of the input and the filter response at that given frequency.

2. WEAKLY STATIONARY RANDOM GRAPH PROCESSES

To motivate our definition of stationarity, we first analyze the covariance matrix of a random process generated when applying a linear graph filter to a white input. To that end, consider a graph $\mathcal{G}$ with associated shift operator $\mathbf{S}$ and suppose that $\mathbf{w}$ is a random process independent and identically distributed (i.i.d.) taking values on the nodes of $\mathcal{G}$. If $\mathbf{w}$ is zero-mean, its covariance matrix is simply $\mathbf{C}_w = \mathbb{E}\left[\mathbf{w}\mathbf{w}^H\right] = \sigma_w^2\mathbf{I}$.

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 $\bar{\mathbf{x}} = \mathbb{E}\left[\mathbf{x}\right] = \mathbb{E}\left[\mathbf{H}\mathbf{w}\right] = \mathbf{H}\bar{\mathbf{w}} = \mathbf{0}$ (this assumption will be revisited in Remark 1) and its covariance $\mathbf{C}_x$ is

$$\begin{aligned}\mathbf{C}_x &= \mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbb{E}[\mathbf{H}\mathbf{w}\mathbf{w}^H\mathbf{H}^H] = \mathbf{H}\mathbf{C}_w\mathbf{H}^H = \sigma_w^2\mathbf{H}\mathbf{H}^H \\ &= \sigma_w^2\mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^H\mathbf{V}\text{diag}(\tilde{\mathbf{h}}^*)\mathbf{V}^H \stackrel{\star}{=} \mathbf{V}\text{diag}(|\tilde{\mathbf{h}}|^2)\mathbf{V}^H,\end{aligned} \tag{2}$$

where we assumed that $\sigma_w = 1$ since it can be absorbed into $\mathbf{h}$. 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.

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 linear graph filter $\mathbf{H} = \sum_{l=0}^{L} h_l\mathbf{S}^l$ applied to a white input $\mathbf{w}$.
b) Matrices $\mathbf{C}_x$ and $\mathbf{S}$ are simultaneously diagonalizable.
c) The cross-correlation of the shifted versions of the process satisfies $\mathbb{E}\left[[\mathbf{S}^a\mathbf{x}]_i[(\mathbf{S}^H)^b\mathbf{x}]_j^H\right] = \mathbb{E}\left[[\mathbf{S}^{(a-c)}\mathbf{x}]_i[(\mathbf{S}^H)^{(b+c)}\mathbf{x}]_j\right]$ for any positive integer a, b, and $c \leq a$.

It is important to notice that the condition of stationarity is defined 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 ^{H} operator in c) is not required if $\mathbf{S}$ is symmetric.

Proposition 1 *Suppose that the eigenvalues of* $\mathbf{S}$ *are distinct, then the three conditions in Definition 1 are equivalent*[2].

[2] Proofs are omitted due to space limitations, but they can be found in [10].

The following corollary further motivates Definition 1 by analyzing its particularization to the time domain. To that end, let us consider the directed cycle of size N, which is the support for time-varying signals. With $((\cdot))_N$ standing for the remainder operation, the adjacency matrix of the directed graph is given by $[\mathbf{A}_{dc}]_{ij} = 1$ if either $i = ((j+1))_N$, and $[\mathbf{A}_{dc}]_{ij} = 0$ otherwise. With these definitions, the application of $\mathbf{S} = \mathbf{A}_{dc}$ to a signal $\mathbf{x}$ corresponds to the classical (circular) shift in the time domain.

Corollary 1 *When* $\mathbf{x}$ *is a random time-varying process, particularizing Definition 1 to* $\mathbf{S} = \mathbf{A}_{dc}$ *yields the following conditions*
a) $x_t = \sum_{l=0}^{N-1} h_l w_{((t-l))_N}$, *with* $\mathbf{w}$ *being white.*
b) $\mathbf{C}_x$ *is a circulant matrix.*
c) $\mathbb{E}\left[x_a x_{((-b))_N}\right] = \mathbb{E}\left[x_{((a-c))_N} x_{((-b-c))_N}\right]$.

In words, when particularized for the regular time grid, Definition 1 is equivalent to the classical definition of weakly stationary time-varying random processes. Next, we leverage Definition 1 to generalize the concept of PSD to random *graph* processes.

Definition 2 The PSD of a random 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 $\mathbf{p}$

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

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 $\mathbf{p} = |\mathbf{\Psi}\mathbf{h}|^2 = |\tilde{\mathbf{h}}|^2$, where $\mathbf{h}$ are the filter coefficients. Finally, if convenient for interpretation purposes, the definition in (3) can be modified so that the left hand side is multiplied by a $1/N$ factor.

Before studying properties of the PSD, we provide some representative examples of stationary graph processes.

Example 1: White noise. Zero-mean white noise is stationary in any graph shift $\mathbf{S}$. The PSD of white noise with covariance matrix $\mathbb{E}\left[\mathbf{w}\mathbf{w}^H\right] = \sigma^2\mathbf{I}$ is $\mathbf{p} = \sigma^2\mathbf{1}$.

Example 2: Covariance matrices. Any random process is stationary with respect to: 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. Case i) implies that when $\mathcal{G}$ is built as a correlation network the data (signals) used to construct the graph are stationary in $\mathcal{G}$. Case ii) is the counterpart when $\mathcal{G}$ is built as a partial correlation network.

Example 3: Heat diffusion processes. Suppose that a zero-mean white input $\mathbf{w}$ diffuses through a graph with Laplacian $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ to generate the signal $\mathbf{x} = \alpha_0\sum_{l=0}^{\infty}(\alpha\mathbf{L})^l\mathbf{w} = \alpha_0(\mathbf{I} - \alpha\mathbf{L})^{-1}\mathbf{w}$. The process $\mathbf{x}$ is stationary in the shift $\mathbf{S} = \mathbf{L}$ with PSD $\mathbf{p} = \text{diag}[\alpha_0^2(\mathbf{I} - \alpha\mathbf{\Lambda})^{-2}]$ because we can write the covariance matrix as $\mathbf{C}_x = \alpha_0^2(\mathbf{I} - \alpha\mathbf{L})^{-2} = \mathbf{V}\alpha_0^2(\mathbf{I} - \alpha\mathbf{\Lambda})^{-2}\mathbf{V}^H$. In fact, this is a particular case of a more general category of diffusion processes [11].

2.2. Properties

We start by analyzing the effect of a linear graph filter on the covariance matrix and the PSD of a stationary graph random process.

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 $\mathbf{C}_y = \mathbf{H}\mathbf{C}_x\mathbf{H}^H$.
b) Has a PSD $\mathbf{p}_y$ given by $\mathbf{p}_y = |\mathbf{\Psi}\mathbf{h}|^2 \circ \mathbf{p}_x = |\tilde{\mathbf{h}}|^2 \circ \mathbf{p}_x$.

Property 2 Let $\mathbf{x}$ be a random process than can be modeled as the output of a linear graph filter $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$ of degree $L-1$ applied to a white input. Then, if the distance between nodes i and j is greater than $2L-2$, it holds that $[\mathbf{C}_x]_{ij} = 0$, so that x_i and x_j are not correlated.

Property 3 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 $\mathbf{C}_{\tilde{x}} := \mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right] - \mathbb{E}[\tilde{\mathbf{x}}]\mathbb{E}[\tilde{\mathbf{x}}]^H = \text{diag}(\mathbf{p})$, so that $\tilde{x}_k$ and $\tilde{x}_{k'}$ are *uncorrelated* for $k \neq k'$.

Property 1 is the counterpart of the spectral convolution theorem for graph processes, Property 2 will be useful when designing windows, and Property 3 provides motivation for the analysis and modeling of stationary graph processes in the frequency domain which we undertake in ensuing sections. 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 $\mathbf{V}$.

Remark 1 Definition 1 assumed that the random process $\mathbf{x}$ has mean $\bar{\mathbf{x}} := \mathbb{E}[\mathbf{x}] = \mathbf{0}$. In time processes, stationarity implies that the mean must be constant. This restriction can be incorporated into our definitions by requiring $\bar{\mathbf{x}} = \bar{x}\mathbf{1}$ for some scalar $\bar{x}$. This is possible but it is not difficult to see that for a generic $\mathbf{S}$ this restriction invalidates (the desirable) Property 1. Alternatively, we can require $\bar{\mathbf{x}} = \bar{x}\mathbf{v}_k$ where $\mathbf{v}_k$ is an arbitrary eigenvector of $\mathbf{S}$. This choice maintains validity of Property 1. If $\mathbf{S}$ is either the adjacency matrix of the directed cycle or the Laplacian of any graph and we set $\mathbf{v}_k = \mathbf{1}$, the choices coincide. The selection of $\mathbf{v}_k = \mathbf{1}$ is further justified because $\mathbf{1}$ is the Laplacian eigenvector associated with smallest total variation [12] and therefore a natural choice for the notion of DC component.

3. NONPARAMETRIC PSD ESTIMATION

The goal here is to estimate the PSD of a stationary random process $\mathbf{x}$ using as input either one or a few realizations $\{\mathbf{x}_r\}_{r=1}^R$ of $\mathbf{x}$.

3.1. Periodogram and Correlogram

Since Property 3 states that $\mathbf{C}_{\tilde{x}} = \text{diag}(\mathbf{p})$ we can write the PSD as $\mathbf{p} = \mathbb{E}\left[|\mathbf{V}^H\mathbf{x}|^2\right]$. This yields a natural approach to estimate $\mathbf{p}$ with the GFT of realizations of $\mathbf{x}$. Thus, compute the GFTs $\tilde{\mathbf{x}}_r = \mathbf{V}^H\mathbf{x}_r$ of each of the samples $\mathbf{x}_r$ in the training set and estimate $\mathbf{p}$ as

$$\hat{\mathbf{p}}_{\text{pg}} := \frac{1}{R}\sum_{r=1}^R |\tilde{\mathbf{x}}_r|^2 = \frac{1}{R}\sum_{r=1}^R \left|\mathbf{V}^H\mathbf{x}_r\right|^2. \quad (4)$$

The estimator in (4) is the analogous of the *periodogram* of time signals and is referred as such from now on. Its intuitive appeal is that it writes the PSD of the process $\mathbf{x}$ as the average of the squared magnitudes of the GFTs of realizations of $\mathbf{x}$. Alternatively, one can replace $\mathbf{C}_x$ in (3) by its empirical estimate $\hat{\mathbf{C}}_x = (1/R)\sum_{r=1}^R \mathbf{x}_r\mathbf{x}_r^H$ and propose the PSD estimate

$$\hat{\mathbf{p}}_{\text{cg}} := \text{diag}\big(\mathbf{V}^H\hat{\mathbf{C}}_x\mathbf{V}\big) := \text{diag}\big[\mathbf{V}^H\big[\tfrac{1}{R}\textstyle\sum_{r=1}^R \mathbf{x}_r\mathbf{x}_r^H\big]\mathbf{V}\big]. \quad (5)$$

Although different in genesis, it can be shown from their respective expressions that the periodogram in (4) and the correlogram in (5) are identical estimates. This is consistent with the equivalence of correlograms and periodograms in time signals. Henceforth, we choose to call $\hat{\mathbf{p}}_{\text{pg}} = \hat{\mathbf{p}}_{\text{cg}}$ the periodogram estimate of $\mathbf{p}$.

To evaluate the performance of the periodogram estimator in (4) we asses its mean and variance.

Proposition 2 *Let $\mathbf{p}$ be the PSD of a process $\mathbf{x}$ that is stationary in the shift $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$. Independent samples $\{\mathbf{x}_r\}_{r=1}^R$ are drawn from the distribution of the process $\mathbf{x}$ and the periodogram $\hat{\mathbf{p}}_{\text{pg}}$ is computed as in (4). The expectation of the estimator is $\mathbb{E}[\hat{\mathbf{p}}_{\text{pg}}] = \mathbf{p}$, so that the bias $\mathbf{b}_{\text{pg}} := \mathbb{E}[\hat{\mathbf{p}}_{\text{pg}}] - \mathbf{p}$ is zero. Further define the covariance matrix of the periodogram estimator as $\mathbf{\Sigma}_{\text{pg}} := \mathbb{E}\left[(\hat{\mathbf{p}}_{\text{pg}} - \mathbf{p})(\hat{\mathbf{p}}_{\text{pg}} - \mathbf{p})^H\right]$. If the process $\mathbf{x}$ is assumed Gaussian, the covariance matrix[3] can be written as $\mathbf{\Sigma}_{\text{pg}} = (2/R)\text{diag}^2(\mathbf{p})$.*

As for time signals, the variance of the periodogram is proportional to the square of the PSD. The latter fact is more often expressed in terms of the mean squared error (MSE), which we define as $\text{MSE}(\hat{\mathbf{p}}_{\text{pg}}) := \mathbb{E}\left[\|(\hat{\mathbf{p}}_{\text{pg}} - \mathbf{p})\|_2^2\right]$ and write as

$$\text{MSE}(\hat{\mathbf{p}}_{\text{pg}}) = \|\mathbf{b}_{\text{pg}}\|_2^2 + \text{tr}[\mathbf{\Sigma}_{\text{pg}}] = (2/R)\|\mathbf{p}\|_2^2. \quad (6)$$

3.2. Windowed Average Periodogram

The Bartlett and Welch methods for PSD estimation of time signals utilize windows to, in effect, generate multiple samples of the process even if only a single realization is given [4, Sec. 2.7]. These methods reduce variances of PSD estimates but introduce some bias. The purpose of this section is to define counterparts of windowing methods for PSD estimation of graph signals.

Consider then a bank of M windows $\mathcal{W} = \{\mathbf{w}_m\}_{m=1}^M$, with $\|\mathbf{w}_m\|_2^2 = \|\mathbf{1}\|_2^2 = N$ for all m, and use each of the windows $\mathbf{w}_m$ to construct the windowed signal[4] $\mathbf{x}_m := \text{diag}(\mathbf{w}_m)\mathbf{x}$. We estimate the PSD $\mathbf{p}$ with the *windowed average periodogram*

$$\hat{\mathbf{p}}_{\mathcal{W}} := \frac{1}{M}\sum_{m=1}^M \left|\mathbf{V}^H\mathbf{x}_m\right|^2 = \frac{1}{M}\sum_{m=1}^M \left|\mathbf{V}^H\text{diag}(\mathbf{w}_m)\mathbf{x}\right|^2. \quad (7)$$

The estimator $\hat{\mathbf{p}}_{\mathcal{W}}$ is reminiscent of the periodogram in (4). The difference is that in (4) the R signals $\mathbf{x}_r$ are independent observations whereas in (7) the M signals $\mathbf{x}_m$ are all generated through multiplications with the window bank $\mathcal{W}$. This means that: (i) There is some distortion in the windowed periodogram estimate because the windowed signals $\mathbf{x}_m$ are used in lieu of $\mathbf{x}$. (ii) The different signals $\mathbf{x}_m$ are correlated with each other and the reduction in variance resulting from the averaging operation in (7) is less significant than the reduction in variance that we observe in Proposition 2.

To study these effects we define the frequency counterpart of the windowing operator $\text{diag}(\mathbf{w}_m)$ as $\tilde{\mathbf{W}}_m := \mathbf{V}^H\text{diag}(\mathbf{w}_m)\mathbf{V}$ (so that $\tilde{\mathbf{x}}_m = \tilde{\mathbf{W}}_m\tilde{\mathbf{x}}$) and the power *spectrum mixing* matrix of windows m and m' as the componentwise product

$$\tilde{\mathbf{W}}_{mm'} := \tilde{\mathbf{W}}_m \circ \tilde{\mathbf{W}}_{m'}^*. \quad (8)$$

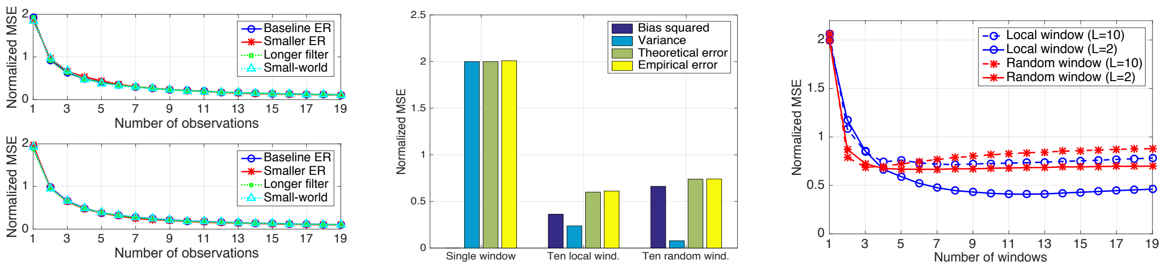
We use these matrices to give expressions for the bias and covariance of the estimator in (7) in the following proposition.

Proposition 3 *Let $\mathbf{p}$ be the PSD of a process $\mathbf{x}$ that is stationary with respect to the shift $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$. A single observation $\mathbf{x}$ is given along with the window bank $\mathcal{W} = \{\mathbf{w}_m\}_{m=1}^M$ and the windowed average periodogram $\hat{\mathbf{p}}_{\mathcal{W}}$ is computed as in (7). The expectation of the estimator $\hat{\mathbf{p}}_{\mathcal{W}}$ is*

$$\mathbb{E}[\hat{\mathbf{p}}_{\mathcal{W}}] = \frac{1}{M}\sum_{m=1}^M \tilde{\mathbf{W}}_{mm}\mathbf{p}. \quad (9)$$

[3] To help readability, the expressions for the covariance in Propositions 2 and 3 assume symmetric shifts. The expressions for normal, not necessarily symmetric, shifts are given in [10].

[4] Abusing notation, in this section $\mathbf{x}$ will also be used to denote a realization of the process $\mathbf{x}$.



(a) (b) (c)

Fig. 1: Normalized MSEs (NMSEs) averaged over 100 realizations. (a) Periodogram with Gaussian (top) and non-Gaussian (bottom) inputs. (b) Theoretical and empirical NMSE for different window strategies. (c) NMSE as a function of the number of local and random windows.

Equivalently, $\hat{\mathbf{p}}_{\mathcal{W}}$ *is biased with bias* $\mathbf{b}_{\mathcal{W}} := \mathbb{E}[\hat{\mathbf{p}}_{\mathcal{W}}] - \mathbf{p}$. *Further define the covariance matrix of the windowed periodogram as* $\boldsymbol{\Sigma}_{\mathcal{W}} := \mathbb{E}\left[(\hat{\mathbf{p}}_{\mathcal{W}} - \mathbb{E}[\hat{\mathbf{p}}_{\mathcal{W}}])(\hat{\mathbf{p}}_{\mathcal{W}} - \mathbb{E}[\hat{\mathbf{p}}_{\mathcal{W}}])^H\right]$. *If the process* $\mathbf{x}$ *is assumed Gaussian, the trace of the covariance matrix can be written as* $\mathrm{tr}[\boldsymbol{\Sigma}_{\mathcal{W}}] = \frac{2}{M^2}\sum_{m=1,m'=1}^{M} \mathrm{tr}\left[\left(\tilde{\mathbf{W}}_{mm'}\mathbf{p}\right)\left(\tilde{\mathbf{W}}_{mm'}\mathbf{p}\right)^H\right]$.

Window design. The overall MSE is given by the squared bias norm summed to the trace of the covariance matrix,

$$\mathrm{MSE}(\hat{\mathbf{p}}_{\mathcal{W}}) = \|\mathbf{b}_{\mathcal{W}}\|_2^2 + \mathrm{tr}[\boldsymbol{\Sigma}_{\mathcal{W}}]. \tag{10}$$

The expression in (10) can be used to design windows with minimum MSE. The problem is challenging because the MSE depends on the unknown PSD, the entries $\mathbf{w}_m$ may be binary, and the cost is a fourth-order polynomial. An alternative approach to design the window bank $\mathcal{W}$ is to exploit the local properties of the random process $\mathbf{x}$. As stated in Property 2, some stationary processes are expected to have correlations with local structure. It is then reasonable to expect that, e.g., designing windows as outputs of clustering algorithms could perform well since windows without overlap capture independent information that results in a reduction of the cross terms in $\mathrm{tr}[\boldsymbol{\Sigma}_{\mathcal{W}}]$ (cf. Proposition 3).

4. NUMERICAL EXPERIMENTS

We begin by evaluating the estimation performance of the average periodogram [cf. (4) and (6)] as a function of R, the number of realizations observed. Consider a baseline Erdős-Rényi (ER) graph with $N = 100$ nodes and edge probability $p = 0.05$ [13]. We define its adjacency matrix as the shift and generate signals by filtering white Gaussian noise with a filter of degree 3. In this case, the normalized MSE equals $2/R$ [cf. (6)] as can be corroborated in Fig. 1a (top). To further confirm this result, we consider three variations of the baseline setting: i) a smaller ER graph with $N = 10$ nodes and $p = 0.3$, ii) a small-world graph [14] obtained by rewiring with probability $q = 0.1$ the edges in a regular graph of the same size as the baseline ER, and iii) filtering the noise with a longer filter of degree 6. As expected, Fig. 1a (top) indicates that the normalized MSE is independent of these variations. We then repeat the above setting but for signals generated as filtered versions of *non-Gaussian* white noise drawn from a uniform distribution of unit variance. Even though the MSE expression in (6) was shown for Gaussian signals, we observe that in the tested non-Gaussian setup the evolution of the MSE with R is very similar; see Fig. 1a (bottom).

The second experiment evaluates the performance of window-based estimators. We first consider graphs generated via a stochastic block model [15] with $N = 100$ nodes and 10 communities with 10 nodes each. The edge probability within each community is $p = 0.9$, while the probability for edges across communities is $q = 0.1$. We design rectangular non-overlapping windows where the nodes are chosen following two strategies: i) $M = 10$ local windows corresponding to the 10 communities, and ii) $M = 10$ windows of equal size with randomly chosen nodes. We use the Laplacian as shift and generate the graph process using a filter with $L = 2$ coefficients. Fig. 1b shows the theoretical and empirical normalized MSE for the two designs as well as that of the periodogram (single window). We first observe that the periodogram has no bias and that the theoretical and empirical errors coincide for the three cases, validating the results in Propositions 2 and 3. Moreover, we corroborate that windowing contributes to reduce the variance of the estimator. Fig. 1b also illustrates that windows that leverage the community structure of the graph yield a better estimation performance. To gain insights on the latter observation, we now consider a small-world graph of size $N = 100$ obtained by rewiring with probability $q = 0.05$ a regular graph where each node has 10 neighbors. Both local and random windows are considered, where the local windows are obtained by cutting the dendrogram obtained when applying complete linkage clustering [16, 17] to a metric space given by the shortest path distances between nodes. The windows are tested for graph processes generated by two filters of different degrees: i) $L = 2$, modeling a localized graph process, and ii) $L = 10$. In Fig. 1c we illustrate the performance of local and random windows in these two settings as a function of M. We first observe that as M increases, the error first decreases until it reaches an optimal point after which it starts to increase. Intuitively, this indicates that at first the reduction in variance outweighs the increase in bias but, after some point, the marginal variance reduction when adding one extra window does not compensate the detrimental effect on the bias. Moreover, it can be seen that local windows outperform the random ones, especially for localized graph processes ($L = 2$). These findings are consistent for other types of graphs, although for graphs with a weaker clustered structure the benefits of local windows are less conspicuous.

5. CONCLUSION

Three equivalent ways of generalizing the notion of stationarity to graph processes were proposed. Under these definitions, the correlation of a stationary process was shown to be diagonalized by the Graph Fourier basis. This led naturally to the consideration of power spectral density for graph signals, along with methods to estimate it from a set of realizations. The focus here was on generalizing nonparametric spectral estimation methods (including average and window-based periodograms) and characterizing their estimation (bias and variance) performance.

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," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[3] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 2009.

[4] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

[5] B. Girault, "Stationary graph signals using an isometric graph translation," in *European Signal Process. Conf. (EUSIPCO)*, 2015, pp. 1516–1520.

[6] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *arXiv preprint arXiv:1601.02522*, 2016.

[7] 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.

[8] S. Segarra, A. G. Marques, and A. Ribeiro, "Distributed implementation of linear network operators using graph filters," in *Allerton Conf. on Commun. Control and Computing*, Sept 2015, pp. 1406–1413.

[9] A. Loukas, A. Simonetto, and G. Leus, "Distributed autoregressive moving average graph filters," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1931–1935, 2015.

[10] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *arXiv preprint arXiv:1603.04667*, 2016.

[11] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *arXiv preprint arXiv:1507.08364*, 2015.

[12] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2012, pp. 3921–3924.

[13] B. Bollobás, *Random Graphs*. Springer, 1998.

[14] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. Springer, 2009.

[15] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Netw.*, vol. 5, no. 2, pp. 109–137, 1983.

[16] A. Jain and R. C. Dubes, *Algorithms for Clustering Data*, ser. Prentice Hall Advanced Reference Series. Prentice Hall Inc., 1988.

[17] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Axiomatic construction of hierarchical clustering in asymmetric networks," *arXiv preprint arXiv:1301.7724*, 2014.