Stationary Graph Processes and Spectral Estimation

Antonio G. Marques, *Senior Member, IEEE*, Santiago Segarra, *Member, IEEE*, Geert Leus, *Fellow, IEEE*, and Alejandro Ribeiro, *Member, IEEE*

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. This lack of regularity hampers the generalization of the classical notion of stationarity to graph signals. This paper proposes a definition of weak stationarity for random graph signals that takes into account the structure of the graph where the random process takes place, while inheriting many of the meaningful properties of the classical time domain definition. Provided that the topology of the graph can be described by a normal matrix, stationary graph processes can be modeled as the output of a linear graph filter applied to a white input. This is shown equivalent to requiring the correlation matrix to be diagonalized by the graph Fourier transform; a fact that is leveraged to define a notion of power spectral density (PSD). Properties of the graph PSD are analyzed and a number of methods for its estimation are proposed. This includes generalizations of nonparametric approaches such as periodograms, window-based average periodograms, and filter banks, as well as parametric approaches, using moving-average, autoregressive, and ARMA processes. Graph stationarity and graph PSD estimation are investigated numerically for synthetic and real-world graph signals.

Index Terms—Graph signal processing, weak stationarity, random graph process, periodogram, windowing, power spectral density, parametric estimation.

I. INTRODUCTION

NETWORKS and graphs are used to represent pairwise relationships between elements of a set, and are objects of

A. G. Marques is with the Department of Signal Theory and Communication, King Juan Carlos University, Madrid, 28933, Spain (e-mail: antonio.garcia.marques@urjc.es).

S. Segarra is with the Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: segarra@mit.edu).

G. Leus is with the Department of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD, Delft, The Netherlands (e-mail: g.j.t.leus@tudelft.nl).

A. Ribeiro is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: aribeiro@seas.upenn.edu).

intrinsic interest. In graph signal processing (GSP), the object of study is not the network itself but a signal supported on the vertices of the graph [2], [3]. The graph is intended to represent a notion of proximity between the components of the signal and, as such, provides a structure that can be exploited in its processing. The use of a graph to process a signal is formalized by the definition of the graph shift operator (GSO), which is a matrix representation of the graph (Sec. II), and the graph Fourier transform (GFT), which projects signals in the eigenvector space of the GSO. The GSO and GFT have been proven useful in the processing of brain activity signals using brain networks, the study of epidemic spreading over social networks, and analysis of Markov random fields [4]–[6].

When the GSO is particularized to a cyclic shift, the associated GFT turns out equivalent to the classical discrete Fourier transform (DFT) [7]. Cyclic shifts and DFTs are central to the processing of time signals in general and to the analysis of (periodic) stationary stochastic processes in particular [8]. The latter is true for various reasons, one of the most important being that the DFT diagonalizes the covariance matrix of the stationary process and reduces its analysis to the study of its power spectral density (PSD) [9]. In this paper we consider analogous notions of stationary *graph* processes, their concomitant PSDs and use those to study parametric and nonparametric *graph* PSD estimation.

Stationary graph processes were first defined and analyzed in [10]. The fundamental problem identified in that paper is that GSOs do not preserve energy in general and cannot therefore be isometric [11]. This problem is addressed with the definition of an isometric graph shift that preserves the eigenvector space of the Laplacian GSO but modifies its eigenvalues [12]. A stationary graph process is then defined as one whose probability distributions are invariant with respect to multiplications with the isometric shift. It is further shown that this definition requires the covariance matrix of the signal to be diagonalized by the eigenvectors of the GSO, which are also the eigenvectors of the isometric shift. Although not formally defined as such, this conclusion implies the existence of a graph PSD with components given by the covariance eigenvalues. The requirement of having a covariance matrix diagonalizable by the eigenvectors of the Laplacian GSO is itself adopted as a definition in [13], where the requirement is shown to be equivalent to statistical invariance with respect to the translation operator introduced in [14]. The PSD is explicitly defined in [13] and a generalization of the short time Fourier transform is proposed for its estimation.

The main goal of this paper is to provide a comprehensive introduction to the spectral analysis and estimation of graph stationary processes. We start with a review of GSP concepts (Sec. II), define weakly stationary graph processes (Sec. III), introduce useful properties (Sec. III-A), and discuss some

examples (Sec. III-B). To emphasize connections with stationary processes in time we say a process is stationary with respect to a normal GSO if: (i) It can be modeled as the output of a linear shift invariant graph filter [7], [15] applied to a white input. (ii) Its covariance matrix satisfies a form of invariance with respect to applications of the GSO. (iii) Its covariance matrix is diagonalized by the GSO's eigenvector matrix. The shift invariance in definition (ii) is more involved than its time-domain counterpart, but it does preserve the locality of the given GSO and sheds light on the local structure of graph stationary processes. These three definitions are equivalent under mild conditions (Prop. 1) and, since (iii) is shared with [13] and a consequence of [10], they are also equivalent to the isometric invariance in [10] and the translation invariance in [13]. Our discussion in this section differs from [10], [13] in that we not only consider Laplacian GSOs but arbitrary normal GSOs. Although mathematically minor, this modification broadens applicability of graph stationarity, establishing connections with results in statistics and network topology inference [16]. Normal GSOs that are not Laplacians cover all symmetric shifts and a subset of non-symmetric GSOs which is small but includes the directed cycle.

Since stationary processes are easier to understand in the frequency domain, we study different methods for *spectral* (PSD) *estimation*, which, if needed, can also be used to improve the estimate of the covariance matrix itself (Sec. IV and Sec. V). We begin by looking at simple nonparametric methods for PSD estimation (Sec. IV). We first extend the periodogram and the correlogram to the graph domain, and analyze their estimation performance. We then generalize more advanced estimators such as window-based average periodograms and filter banks [9, Chs. 2 and 5]. The estimation performance is evaluated analytically and differences relative to their time-domain counterparts are highlighted. After this, we shift attention to parametric estimation (Sec. V). The focus is on estimating the PSD of autoregressive (AR), moving-average (MA), and ARMA graph processes, which are likely to arise in distributed setups driven by linear dynamics [5], [17], [18]. As in time, it turns out that the estimation of the ARMA parameters is a non-convex (quartic) problem, although for certain particular cases – including that of positive semidefinite shifts – the optimization is tractable. Numerical results are presented to confirm theoretical claims and provide insights on the implications of graph stationarity and to discuss some potential applications (Sec. VI). The latter include smoothing of face-image signals, characterization of brain networks and signals, and description of flow cytometry data.

Notation: Entries of vector $\mathbf{x}$ are $[\mathbf{x}]_i = x_i$. Entries of matrix $\mathbf{X}$ are $[\mathbf{X}]_{ij} = X_{ij}$. Conjugate, transpose, and transpose conjugate are $\mathbf{X}^*$, $\mathbf{X}^T$, and $\mathbf{X}^H$. For square matrix $\mathbf{X}$, we use $\mathrm{tr}[\mathbf{X}]$ for its trace and $\mathrm{diag}(\mathbf{X})$ for a vector with the diagonal elements of $\mathbf{X}$. For vector $\mathbf{x}$, $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with diagonal elements $\mathbf{x}$. The elementwise product of $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{x} \circ \mathbf{y}$. The all-zero and all-one vectors are $\mathbf{0}$ and $\mathbf{1}$. The ith element of the canonical basis of $\mathbb{R}^N$ is $\mathbf{e}_i$.

II. 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 if there exists a connection from node i to node j then $(i, j) \in \mathcal{E}$. We associate

with $\mathcal{G}$ the GSO $\mathbf{S}$, 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], [7]. The sparsity pattern of $\mathbf{S}$ captures the local structure of $\mathcal{G}$, but we make no specific assumptions on the values of the nonzero entries of $\mathbf{S}$. Frequent choices for $\mathbf{S}$ are the adjacency matrix [3], [7], the Laplacian [2], and their respective generalizations [19]. 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 $\mathbf{y} = \mathbf{S}^l \mathbf{x}$, then node i can compute y_i provided that it has access to the value of x_j at $j \in \mathcal{N}_l(i)$. We henceforth work with a *normal* $\mathbf{S}$ to guarantee the existence of a unitary matrix $\mathbf{V}$ and a diagonal matrix $\mathbf{\Lambda}$ such that $\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^H$.

A graph signal is 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 graph Fourier transform (GFT) [7].

We further introduce the notion of a graph filter $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^N$, defined as a linear graph signal operator of the form

$$\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l, \tag{1}$$

where $\mathbf{h} = [h_0, \ldots, h_{L-1}]^T$ is a vector of $L \leq N$ scalar coefficients. According to (1), graph filters are polynomials of degree $L - 1$ in the GSO $\mathbf{S}$ [3], which due to the local structure of the shift can be implemented locally too [15], [20]. 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 in (1) can be rewritten as $\mathbf{H} = \mathbf{V}\left(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l\right)\mathbf{V}^H$. The diagonal matrix $\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l$ is termed the frequency response of the filter and extracted in the vector $\tilde{\mathbf{h}} := \mathrm{diag}(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l)$.

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 (1) can be written as

$$\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l = \mathbf{V}\mathrm{diag}(\tilde{\mathbf{h}})\mathbf{V}^H = \mathbf{V}\mathrm{diag}(\mathbf{\Psi}_L \mathbf{h})\mathbf{V}^H. \tag{2}$$

Equation (2) 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

$$\tilde{\mathbf{y}} = \mathrm{diag}(\mathbf{\Psi}_L \mathbf{h})\mathbf{V}^H \mathbf{x} = \mathrm{diag}(\tilde{\mathbf{h}})\tilde{\mathbf{x}} = \tilde{\mathbf{h}} \circ \tilde{\mathbf{x}}, \tag{3}$$

which demonstrates that the output at a given frequency depends only on the value of the input and the filter response at that given frequency. Observe that when $L = N$ we have $\mathbf{\Psi}_N = \mathbf{\Psi}$ and that, if we are given a filter with order greater than $N - 1$, we can rewrite it as a different filter of order not larger than $N - 1$ due to the Cayley-Hamilton theorem.

III. WEAKLY STATIONARY RANDOM GRAPH PROCESSES

This section leverages notions of weak stationarity in discrete time to define stationarity in the graph domain, discusses relations with existing definitions, and presents results that will

be leveraged in the ensuing sections. The proposed definitions depend on the GSO $\mathbf{S}$, so that a process $\mathbf{x}$ can be stationary in $\mathbf{S}$ but not in $\mathbf{S}' \neq \mathbf{S}$. The shift does not need to be the adjacency or the Laplacian matrix, but it does need to be *normal* and account for the topology of the graph.

A standard zero-mean white random process $\mathbf{w}$ is one with mean $\mathbb{E}[\mathbf{w}] = \mathbf{0}$ and covariance $\mathbb{E}[\mathbf{w}\mathbf{w}^H] = \mathbf{I}$. Our first definition, states that white signals processed by a linear shift-invariant *graph* filter give rise to stationary *graph* processes.

Definition 1: Given a normal shift operator $\mathbf{S}$, a zero-mean random process $\mathbf{x}$ is said to be weakly stationary with respect to $\mathbf{S}$ if it can be written as the response of a linear shift-invariant graph filter $\mathbf{H} = \sum_{l=0}^{N-1} h_l \mathbf{S}^l$ to a zero-mean white input $\mathbf{w}$.

The definition states that the process $\mathbf{x}$ is stationary if we can write $\mathbf{x} = \mathbf{H}\mathbf{w}$ for some filter $\mathbf{H} = \sum_{l=0}^{N-1} h_l \mathbf{S}^l$ that we excite with a white input $\mathbf{w}$. If we write $\mathbf{x} = \mathbf{H}\mathbf{w}$, the covariance matrix $\mathbf{C}_x := \mathbb{E}[\mathbf{x}\mathbf{x}^H]$ of the signal $\mathbf{x}$ can be written as

$$\mathbf{C}_x = \mathbb{E}[(\mathbf{H}\mathbf{w})(\mathbf{H}\mathbf{w})^H] = \mathbf{H}\mathbb{E}[\mathbf{w}\mathbf{w}^H]\mathbf{H} = \mathbf{H}\mathbf{H}^H, \quad (4)$$

which implies that the color of the process is determined by the filter $\mathbf{H}$. Def. 1 is constructive as it describes how a stationary graph signal can be generated. Alternatively, one can define stationarity using requirements on the moments of the random graph signal in either the vertex or the frequency domain.

Definition 2: Given a normal shift operator $\mathbf{S}$, a zero-mean random process $\mathbf{x}$ is said to be weakly stationary with respect to $\mathbf{S}$ if the two equivalent properties hold

a) For any set of nonnegative integers a, b, and $c \leq b$ it holds

$$\mathbb{E}\left[(\mathbf{S}^a \mathbf{x})((\mathbf{S}^H)^b \mathbf{x})^H\right] = \mathbb{E}\left[(\mathbf{S}^{a+c}\mathbf{x})((\mathbf{S}^H)^{b-c}\mathbf{x})^H\right]. \quad (5)$$

b) Matrices $\mathbf{C}_x$ and $\mathbf{S}$ are simultaneously diagonalizable.

Before discussing its intuition, we start by showing that Defs. 2.a and 2.b are indeed equivalent. To do this, suppose that Def. 2.a holds and reorder terms in (5) to yield

$$\mathbf{S}^a \mathbf{C}_x \mathbf{S}^b = \mathbf{S}^{a+c} \mathbf{C}_x \mathbf{S}^{b-c}. \quad (6)$$

For (6) to be true, $\mathbf{S}^c$ and $\mathbf{C}_x$ must commute for all c. Since both matrices are diagonalizable, this will happen if and only if $\mathbf{S}$ and $\mathbf{C}_x$ have the same eigenvectors [21], which implies Def. 2.b. Conversely, if Def. 2.b. holds, then we can write $\mathbf{C}_x = \mathbf{V}\text{diag}(\boldsymbol{\lambda}_c)\mathbf{V}^H$. Substituting this expression into (6) the equality checks and Def. 2.a follows.

The idea under Def. 2.a is that if the total number of times that we shift our signal is constant (regardless of how many of those times we shift the signal forward $\mathbf{S}\mathbf{x}$ or backward $\mathbf{S}^H\mathbf{x}$), then the correlation has to be the same. Indeed, in the two sides of (5) the total number of times the signal has been shifted is $a + b$. This resembles what happens for stationary signals in time, where correlation depends on the total number of shifts, but not the particular time instants. Formally, when $\mathbf{S}$ is a cyclic shift, $\mathbf{S}^H$ is a shift in the opposite direction with $\mathbf{S}^H = \mathbf{S}^{-1}$. Then, if we set $a = 0$, $b = N$ and $c = l$ we recover $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbb{E}[\mathbf{S}^l \mathbf{x}(\mathbf{S}^l \mathbf{x})^H]$, which is the definition of a stationary signal in time. Intuitively, having the same number of shifts in (5) is necessary because the eigenvalues of the GSO $\mathbf{S}$ do not have unit magnitude and can change the energy of the signal. The vertex-based definition in (5) is based on the original GSO $\mathbf{S}$,

which is local and real-valued. As a result, (5) provides intuition on the relations between stationarity and locality, which can be leveraged to develop stationarity tests or estimation schemes that work with local information; see Sec. IV. This is different from the vertex-based definitions in [10] and [13], which however are more advantageous than (5) from other perspectives; see Remark 1.

While Def. 2.a captures the implications of stationarity in the vertex domain, Def. 2.b captures its implications in the graph frequency domain by requiring that the covariance $\mathbf{C}_x$ be diagonalized by the GFT matrix $\mathbf{V}$ – this was deduced as a property in [10] and proposed as a definition in [13]. When particularized to time, Def. 2.b requires $\mathbf{C}_x$ to be circulant and therefore diagonalized by the Fourier matrix. This fact is exploited in (periodic) time stationary processes to define the PSD as the eigenvalues of the circulant covariance matrix, and motivates the following definition.

Definition 3: The power spectral density (PSD) of a random process $\mathbf{x}$ that is stationary with respect to the normal graph shift $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$ is the nonnegative $N \times 1$ vector $\mathbf{p}$

$$\mathbf{p} := \text{diag}\left(\mathbf{V}^H \mathbf{C}_x \mathbf{V}\right). \quad (7)$$

Observe that since $\mathbf{C}_x$ is diagonalized by $\mathbf{V}$ (see Def. 2.b) the matrix $\mathbf{V}^H \mathbf{C}_x \mathbf{V}$ is diagonal and it follows that the PSD in (7) corresponds to the eigenvalues of the positive semidefinite correlation matrix $\mathbf{C}_x$. Thus, (7) is equivalent to

$$\mathbf{C}_x = \mathbf{V}\text{diag}(\mathbf{p})\mathbf{V}^H. \quad (8)$$

Defs. 1 and 2, which are valid for any symmetric GSO $\mathbf{S}$ as well as for nonsymmetric but normal shifts, are equivalent for time signals by construction. They are also equivalent for an important class of graphs, as shown next.

Proposition 1: If $\mathbf{S}$ is normal and its eigenvalues are all distinct, Defs. 1 and 2 are equivalent.

Proof: Since Defs. 2.b and 2.a are equivalent, we show that under the conditions in the proposition, Def. 1 is equivalent to Def. 2.b. Assume first that Def. 1 holds. Since the graph filter $\mathbf{H}$ in (4) is linear shift invariant, it is completely characterized by its frequency response $\tilde{\mathbf{h}} = \boldsymbol{\Psi}\mathbf{h}$ as stated in (2). Using this characterization we rewrite (4) as

$$\mathbf{C}_x = \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^H(\mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^H)^H = \mathbf{V}\text{diag}^2(|\tilde{\mathbf{h}}|)\mathbf{V}^H. \quad (9)$$

This means that Def. 2.b holds. Conversely, if Def. 2.b holds it means that we can write $\mathbf{C}_x = \mathbf{V}\text{diag}(\boldsymbol{\lambda}_c)\mathbf{V}^H$ for some component-wise *nonnegative* vector $\boldsymbol{\lambda}_c$. Define now vector $\sqrt{\boldsymbol{\lambda}_c}$ where the square root is applied element-wise. Then, for Def. 1 to hold, there must exist a filter with coefficients $\mathbf{h}$ satisfying $\tilde{\mathbf{h}} = \boldsymbol{\Psi}\mathbf{h} = \sqrt{\boldsymbol{\lambda}_c}$. Since $\boldsymbol{\Psi}$ is Vandermonde, the system is guaranteed to have a solution with respect to $\mathbf{h}$ provided that all modes of $\boldsymbol{\Psi}$ (the eigenvalues of $\mathbf{S}$) are distinct, as required in the proposition. ∎

For normal shifts, Def. 1 always implies Def. 2 as also shown in [13]. However, if $\mathbf{S}$ is not normal, as is the case for most directed graphs, the equivalence is lost and the question of which definition to use arises. Constructive approaches related to Def. 1 that use graph filters to model a set of observations have been successfully applied in the context of diffusions over *directed* non-normal graphs [3], [22].

Remark 1: When the shift is $\mathbf{S} = \mathbf{L}$ and the eigenvalues of $\mathbf{S}$ are all distinct, Defs. 1 and 2 are equivalent to those in [10] and [13]. The main differences between [10], [13], and our framework are in the interpretation of stationarity in the vertex domain. To define stationarity, [10] utilizes an isometric GSO $\tilde{\mathbf{S}}$ that has the same eigenvectors as $\mathbf{S}$ but whose eigenvalues are unitary complex exponentials [11]. Although the resultant operator is complex-valued and has a sparsity structure different from $\mathbf{S}$, it provides a more natural invariance. Indeed, since the resultant $\tilde{\mathbf{S}}$ preserves energy and gives rise to the same GFT, the counterpart to (5) can be simply written as $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbb{E}[\tilde{\mathbf{S}}^l \mathbf{x}(\tilde{\mathbf{S}}^l \mathbf{x})^H]$. Differently, [13] proposes as starting point Def. 2.b and analyzes the implications in the vertex domain using the so-called translation operator defined in [14]. The reason for having different possible definitions in the vertex domain is that the irregularity and finiteness of the graph support leads to non-isometric shifts. One can then modify the shift to make it isometric as in [10], define a different isometry as in [13], or define a different form of invariance as in (5). We also point out that our work differs from [13] in that we consider normal shifts instead of Laplacians and that we see Def. 1 as a definition, not a property. These are mathematically minor differences that are important in practice. For example, normal shifts that are not Laplacians include symmetric shifts with positive non-diagonal entries or the cyclic shift. Moreover, in cases where Defs. 1 and 2 differ, Def. 1 is useful because it models real-world processes (see Sec. III-B).

A. Properties of Graph Stationary Processes

Three properties that will be leveraged in the ensuing sections are listed next. Since these properties are direct generalizations of their classical time counterparts, their proofs are omitted for conciseness. For Property 3, recall that $\mathcal{N}_l(i)$ denotes the l-hop neighborhood of node i.

Property 1: Let $\mathbf{x}$ be a stationary process in $\mathbf{S}$ with covariance matrix $\mathbf{C}_x$ and PSD $\mathbf{p}_x$. Consider a filter $\mathbf{H}$ with coefficients $\mathbf{h}$ and frequency response $\tilde{\mathbf{h}}$ and define $\mathbf{y} := \mathbf{H}\mathbf{x}$ as the response of $\mathbf{H}$ to input $\mathbf{x}$. Then, the 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 given by $\mathbf{p}_y = |\tilde{\mathbf{h}}|^2 \circ \mathbf{p}_x$.

Property 2: Given a process $\mathbf{x}$ stationary in $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$ with PSD $\mathbf{p}$, define the GFT process as $\tilde{\mathbf{x}} = \mathbf{V}^H \mathbf{x}$. Then, it holds that $\tilde{\mathbf{x}}$ is uncorrelated and its covariance matrix is

$$\mathbf{C}_{\tilde{x}} := \mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right] = \mathbb{E}\left[(\mathbf{V}^H \mathbf{x})(\mathbf{V}^H \mathbf{x})^H\right] = \operatorname{diag}(\mathbf{p}). \quad (10)$$

Property 3: Let $\mathbf{x} = \mathbf{H}\mathbf{w}$ be a process written as the response of a linear graph filter $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$ of degree $L-1$ to a white input. Then, $[\mathbf{C}_x]_{ij} = 0$ for all $j \notin \mathcal{N}_{2(L-1)}(i)$.

Property 1, also identified in [13], is a statement of the spectral convolution theorem for graph signals, which can also be viewed as a generalization of Def. 1 for inputs that are not white. Property 2 (see also [10]) is fundamental to motivate 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 $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$, then the GFT $\mathbf{V}^H$ provides the Karhunen-Loève expansion of the process. The last property requires a bit more of a discussion. In time, the correlation matrix often conveys a notion of lo-

cality as the significant values of $\mathbf{C}_x$ accumulate close to the diagonal. Property 3, which follows from the locality of graph filters, puts a limit on the spatial extent of the components x_j of a graph signal that can be correlated with a given element x_i. Only those elements that are in the $2(L-1)$-hop neighborhood – i.e., elements x_j with indices $j \in \mathcal{N}_{2(L-1)}(i)$ – can be correlated with x_i. This spatial limitation of correlations can be used to design windows for spectral estimation as we explain in Sec. IV-B.

Remark 2: While Defs. 1 and 2 assume that the random process $\mathbf{x}$ has mean $\bar{\mathbf{x}} := \mathbb{E}[\mathbf{x}] = \mathbf{0}$, traditional stationary time processes are allowed to have a (non-zero) constant mean $\bar{\mathbf{x}} = \alpha \mathbf{1}$, with α being an arbitrary scalar. Our proposal here is for stationary graph processes to be required to have a first-order moment of the form $\bar{\mathbf{x}} = \alpha \mathbf{v}_k$, i.e., a scaled version of an eigenvector of $\mathbf{S}$. This choice: i) takes into account the structure of the underlying graph; ii) maintains the validity of Property 1; and iii) allows to set $\mathbf{v}_k = \mathbf{1}$ for the cases where $\mathbf{S}$ is either the adjacency matrix of the directed cycle or the Laplacian of any graph, recovering the classical definition. Note that [10], [12] require $\bar{\mathbf{x}} = \mathbf{0}$, while [13] requires $\bar{\mathbf{x}} = \alpha \mathbf{1}$.

B. Examples of Stationary Graph Processes

We close this section providing 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 $\mathbb{E}[\mathbf{w}\mathbf{w}^H] = \sigma^2 \mathbf{I}$ is $\mathbf{p} = \sigma^2 \mathbf{1}$.

Example 2: Covariance matrix graph. Any random process $\mathbf{x}$ is stationary with respect to the shift $\mathbf{S} = \mathbf{C}_x$ defined by its covariance matrix. Since in this case the eigenvalues of $\mathbf{S}$ and $\mathbf{C}_x$ are the same, it holds that the PSD is $\mathbf{p} = \operatorname{diag}(\boldsymbol{\Lambda})$. This can be exploited, for example, in the context of network topology inference. Given a set of (graph) signal observations $\{\mathbf{x}_r\}_{r=1}^R$ it is common to infer the underlying topology by building a graph $\mathcal{G}_{corr}$ whose weight links correspond to cross-correlations among the entries of the observations. In that case, the process generating those signals is stationary in the shift given by the adjacency of $\mathcal{G}_{corr}$; see [16] for details.

Example 3: Precision matrix graph. Let $\boldsymbol{\Theta}$ denote the precision matrix of $\mathbf{x}$, which is defined as the (pseudo-)inverse $\boldsymbol{\Theta} = \mathbf{C}_x^\dagger$. Per Def. 2.b, it holds then than the process $\mathbf{x}$ is stationary in $\boldsymbol{\Theta}$. The PSD in this case is $\mathbf{p} = \operatorname{diag}(\boldsymbol{\Lambda})^\dagger$. This is particularly important when $\mathbf{x}$ is a Gaussian Markov Random Field (GMRF) whose Markovian dependence is captured by the unweighted graph $\mathcal{G}_{MF}$. It is well-known [23, Ch. 19] that if $\boldsymbol{\Theta}$ is the precision matrix of a GMRF, then Θ_{ij} can be nonzero only if (i,j) is either a link of $\mathcal{G}_{MF}$, or an element in the diagonal $i = j$. Then, it holds that any GMRF is stationary with respect to the *sparse* shift $\mathbf{S} = \boldsymbol{\Theta}$, which captures the conditional independence of the elements of $\mathbf{x}$.

Example 4: Network Diffusion Processes. Many graph processes are characterized by local interactions between nodes of a (sparse) graph that can be approximated as linear [5], [17], [18]. This implies that we can track the evolution of the signal at the ith node through a local recursion of the form $x_i^{(l+1)} = x_i^{(l)} - \gamma^{(l)} \sum_j s_{ij} x_j^{(l)}$ where $x_i^{(0)}$ is the initial condition of the system, s_{ij} are elements of the GSO, and $\gamma^{(l)}$ are time varying diffusion coefficients. The diffusion halts after L itera-

tions – which can be possibly infinite – to produce the output $y_i = x_i^{(L)}$. Diffusion dynamics appear in, e.g., network control systems [24], opinion formation [18], brain networks [25], and molecular communications [26]. Utilizing the GSO, diffusion dynamics can be written as $\mathbf{x}^{(l+1)} = (\mathbf{I} - \gamma^{(l)}\mathbf{S})\mathbf{x}^{(l)}$, with the output $\mathbf{y} = \mathbf{x}^{(L)}$ being

$$\mathbf{y} = \mathbf{x}^{(L)} = \prod_{l=0}^{L}(\mathbf{I} - \gamma^{(l)}\mathbf{S})\mathbf{x}. \qquad (11)$$

The Cayley-Hamilton theorem guarantees that any matrix polynomial can be expressed as a graph filter with degree less than N. If we now think of the initial condition $\mathbf{x}$ as a process stationary in $\mathbf{S}$, it follows that $\mathbf{y}$, which is a filtered version of $\mathbf{x}$, is also stationary (cf. Property 1).

IV. NONPARAMETRIC PSD ESTIMATION

The interest in this and the following section is in estimating the PSD of a random process $\mathbf{x}$ that is stationary with respect to $\mathbf{S}$ using as input either *one* or a few realizations $\{\mathbf{x}_r\}_{r=1}^{R}$ of $\mathbf{x}$. In this section we consider nonparametric methods that do not assume a particular model for $\mathbf{x}$. We generalize to graph signals the periodogram, correlogram, windowing, and filter bank techniques that are used for PSD estimation in the time domain. Parametric methods are analyzed in Sec. V.

A. Periodogram and Correlogram

Since Property 2 implies that $\mathbf{C}_{\tilde{x}}$ is diagonal, we can rewrite (10) to conclude that the PSD can be written 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

$$\hat{\mathbf{p}}_{\text{pg}} := \frac{1}{R}\sum_{r=1}^{R}|\tilde{\mathbf{x}}_r|^2 = \frac{1}{R}\sum_{r=1}^{R}\left|\mathbf{V}^H\mathbf{x}_r\right|^2. \qquad (12)$$

The estimator in (12) is the analogous of the *periodogram* of time signals and is referred to 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 (7) by its sample-based version $\hat{\mathbf{C}}_x = (1/R)\sum_{r=1}^{R}\mathbf{x}_r\mathbf{x}_r^H$ and estimate the PSD as

$$\hat{\mathbf{p}}_{\text{cg}} := \text{diag}\left(\mathbf{V}^H\hat{\mathbf{C}}_x\mathbf{V}\right) := \text{diag}\left[\mathbf{V}^H\left[\frac{1}{R}\sum_{r=1}^{R}\mathbf{x}_r\mathbf{x}_r^H\right]\mathbf{V}\right]. \qquad (13)$$

An important observation in the correlogram definition in (13) is that the empirical covariance $\hat{\mathbf{C}}_x$ is not necessarily diagonalized by $\mathbf{V}$. However, since we know that the actual covariance $\mathbf{C}_{\tilde{x}}$ is diagonal, we retain only the diagonal elements of $\mathbf{V}^H\hat{\mathbf{C}}_x\mathbf{V}$ to estimate the PSD of $\mathbf{x}$. The expression in (13) is the analogous of the time *correlogram*. Its intuitive appeal is that it estimates the PSD with a double GFT transformation of the empirical covariance matrix.

Although different in genesis, the periodogram in (12) and the correlogram in (13) are identical estimates. To see this, consider the last equality in (13), and move matrices $\mathbf{V}^H$ and $\mathbf{V}$ into the empirical covariance sum. Observe then that the summands end up being of the form $(\mathbf{V}^H\mathbf{x}_r)(\mathbf{V}^H\mathbf{x}_r)^H$. The diagonal elements of these outer products are $|\mathbf{V}^H\mathbf{x}_r|^2$, which are the summands in

the last equality in (12). 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 (12) we assess its mean and variance. The estimator is unbiased by design and, as we shall prove next, this is easy to establish formally. To study the estimator's variance we need the additional hypothesis of the process having a Gaussian distribution. Expressions for means and variances of periodogram estimators are given in the following proposition.

Proposition 2: Let $\mathbf{p}$ be the PSD of a process $\mathbf{x}$ that is stationary with respect to the shift $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$. 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 (12). The bias $\mathbf{b}_{\text{pg}}$ of the estimator is zero,

$$\mathbf{b}_{\text{pg}} := \mathbb{E}\left[\hat{\mathbf{p}}_{\text{pg}}\right] - \mathbf{p} = \mathbf{0}. \qquad (14)$$

Further define the covariance matrix of the periodogram estimator as $\boldsymbol{\Sigma}_{\text{pg}} := \mathbb{E}\left[(\hat{\mathbf{p}}_{\text{pg}} - \mathbf{p})(\hat{\mathbf{p}}_{\text{pg}} - \mathbf{p})^H\right]$. If the process $\mathbf{x}$ is assumed Gaussian and $\mathbf{S}$ is symmetric, the covariance matrix can be written as

$$\boldsymbol{\Sigma}_{\text{pg}} := \mathbb{E}\left[(\hat{\mathbf{p}}_{\text{pg}} - \mathbf{p})(\hat{\mathbf{p}}_{\text{pg}} - \mathbf{p})^H\right] = (2/R)\text{diag}^2(\mathbf{p}). \quad (15)$$

Proof: See Appendix A. ∎

The expression in (14) states that the bias of the periodogram is $\mathbf{b}_{\text{pg}} = \mathbf{0}$, or, equivalently, that the expectation of the periodogram estimator is the PSD itself, i.e., $\mathbb{E}\left[\hat{\mathbf{p}}_{\text{pg}}\right] = \mathbf{p}$. The expression for the covariance matrix of $\boldsymbol{\Sigma}_{\text{pg}}$ holds true only when the process $\mathbf{x}$ has a Gaussian distribution. The reason for this limitation is that the determination of this covariance involves operations with the fourth order moments of the process $\mathbf{x}$. This necessity and associated limitation also arise in time signals [8, Sec. 8.2]. To help readability and intuition, the covariance $\boldsymbol{\Sigma}_{\text{pg}}$ in (15) is stated for a symmetric $\mathbf{S}$, however, in Appendix A the proof is done for a generic normal, not necessarily symmetric, $\mathbf{S}$.

The variance expression in (15) is analogous to the periodogram variances of PSDs of time domain signals in that: (i) Estimates of different frequencies are uncorrelated – because $\boldsymbol{\Sigma}_{\text{pg}}$ is diagonal. (ii) 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 [cf. (14) and (15)]

$$\text{MSE}(\hat{\mathbf{p}}_{\text{pg}}) = \|\mathbf{b}_{\text{pg}}\|_2^2 + \text{tr}[\boldsymbol{\Sigma}_{\text{pg}}] = (2/R)\|\mathbf{p}\|_2^2. \qquad (16)$$

As it happens in time signals, the MSE in (16) is large and results in estimation errors that are on the order of the magnitude of the frequency component itself. Two of the workarounds to reduce the MSE in (16) are the use of windows and filter banks. Both tradeoff bias for variance as we explain in the following sections.

B. 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 [9, Sec. 2.7]. These methods reduce variances of PSD estimates, but introduce some distortion (bias). The purpose of this

section is to define counterparts of windowing methods for PSD estimation of graph signals.

To understand the use of windows in estimating a PSD let us begin by defining windows for graph signals and understanding their effect on the graph frequency domain. We say that a signal $\mathbf{w}$ is a window if its energy is $\|\mathbf{w}\|_2^2 = \|\mathbf{1}\|_2^2 = N$. Applying the window $\mathbf{w}$ to a signal[1] $\mathbf{x}$ entails componentwise multiplication to produce the signal $\mathbf{x_w} = \mathrm{diag}(\mathbf{w})\mathbf{x}$. In the graph frequency domain we can use the definition of the GFT $\tilde{\mathbf{x}}_\mathbf{w} = \mathbf{V}^H \mathbf{x_w}$, the definition of the windowed signal $\mathbf{x_w} = \mathrm{diag}(\mathbf{w})\mathbf{x}$, and the definition of the inverse GFT to write

$$\tilde{\mathbf{x}}_\mathbf{w} = \mathbf{V}^H \mathbf{x_w} = \mathbf{V}^H \mathrm{diag}(\mathbf{w})\mathbf{x} = \mathbf{V}^H \mathrm{diag}(\mathbf{w})\mathbf{V}\tilde{\mathbf{x}} := \tilde{\mathbf{W}}\tilde{\mathbf{x}}, \tag{17}$$

where in the last equality we defined the dual of the windowing operator $\mathrm{diag}(\mathbf{w})$ in the frequency domain as the matrix $\tilde{\mathbf{W}} := \mathbf{V}^H \mathrm{diag}(\mathbf{w})\mathbf{V}$. In time signals the frequency representation of a window is its Fourier transform and the dual operator of windowing is the convolution between the spectra of the window and the signal. This decoupled explanation is lost in graph signals[2]. Nonetheless, (17) can be used to design windows with small spectral distortion. Ideal windows are such that $\tilde{\mathbf{W}} = \mathbf{I}$ which can be achieved by setting $\mathbf{w} = \mathbf{1}$, although this is unlikely to be of any use. More interestingly, (17) implies that good windows for spectral estimation must have $\tilde{\mathbf{W}} \approx \mathbf{I}$ or can allow nonzero values in columns k where the components $\tilde{x}_k$ of $\tilde{\mathbf{x}}$ are known or expected to be small.

Turning now to the problem of PSD estimation, consider the estimate $\hat{\mathbf{p}}$ obtained after computing the periodogram in (12) using a *single* realization $\mathbf{x}$, so that we have that $\hat{\mathbf{p}} = |\mathbf{V}^H \mathbf{x}|^2$. Suppose now that we window the realization $\mathbf{x}$ to produce $\mathbf{x_w} = \mathrm{diag}(\mathbf{w})\mathbf{x}$ and compute the *windowed periodogram* $\hat{\mathbf{p}}_\mathbf{w} := |\mathbf{V}^H \mathbf{x_w}|^2$. Utilizing the definition of the window's frequency representation, we can write this windowed periodogram as

$$\hat{\mathbf{p}}_\mathbf{w} := |\mathbf{V}^H \mathbf{x_w}|^2 = |\mathbf{V}^H \mathrm{diag}(\mathbf{w})\mathbf{x}|^2 = |\tilde{\mathbf{W}}\mathbf{V}^H \mathbf{x}|^2. \tag{18}$$

The expression in (18) can be used to compute the expectation of the windowed periodogram that we report in the following proposition (recall that $\circ$ denotes entrywise product).

Proposition 3: Let $\mathbf{p}$ be the PSD of a process $\mathbf{x}$ that is stationary with respect to the shift $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$. The expectation of the windowed periodogram $\hat{\mathbf{p}}_\mathbf{w}$ in (18) is,

$$\mathbb{E}[\hat{\mathbf{p}}_\mathbf{w}] = (\tilde{\mathbf{W}} \circ \tilde{\mathbf{W}}^*)\mathbf{p}. \tag{19}$$

Proof: Write $\hat{\mathbf{p}}_\mathbf{w} = |\tilde{\mathbf{W}}\mathbf{V}^H\mathbf{x}|^2 = \mathrm{diag}(\tilde{\mathbf{W}}\mathbf{V}^H\mathbf{x}\mathbf{x}^H\mathbf{V}\tilde{\mathbf{W}})$. Take the expectation and use $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{C}_x$ to write $\mathbb{E}[\hat{\mathbf{p}}_\mathbf{w}] = \mathrm{diag}(\tilde{\mathbf{W}}\mathbf{V}^H\mathbf{C}_x\mathbf{V}\tilde{\mathbf{W}}^H)$. Further observe that $\mathbf{V}^H\mathbf{C}_x\mathbf{V} = \mathrm{diag}(\mathbf{p})$ to conclude that $\hat{\mathbf{p}}_\mathbf{w} = \mathrm{diag}(\tilde{\mathbf{W}}\mathrm{diag}(\mathbf{p})\tilde{\mathbf{W}}^H)$. This latter expression is identical to (19). ∎

Prop. 3 implies that the windowed periodogram in (20) is a biased estimate of $\mathbf{p}$, with the bias being determined by the dual of the windowing operator in the frequency domain $\tilde{\mathbf{W}}$.

Multiple windows yield, in general, better estimates than single windows. Consider then a bank of M windows $\mathcal{W} = \{\mathbf{w}_m\}_{m=1}^M$ and use each of the windows $\mathbf{w}_m$ to construct the windowed signal $\mathbf{x}_m := \mathrm{diag}(\mathbf{w}_m)\mathbf{x}$. We estimate the PSD $\mathbf{p}$ with the *windowed average periodogram*

$$\hat{\mathbf{p}}_\mathcal{W} := \frac{1}{M}\sum_{m=1}^M |\mathbf{V}^H \mathbf{x}_m|^2 = \frac{1}{M}\sum_{m=1}^M |\mathbf{V}^H \mathrm{diag}(\mathbf{w}_m)\mathbf{x}|^2. \tag{20}$$

The estimator $\hat{\mathbf{p}}_\mathcal{W}$ is an average of the windowed periodograms in (18) but is also reminiscent of the periodogram in (12). The difference is that in (12) the samples $\{\mathbf{x}_r\}_{r=1}^R$ are independent observations whereas in (20) the samples $\{\mathbf{x}_m\}_{m=1}^M$ 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 (20) is less significant than the reduction of variance observed in Prop. 2.

To study these effects, given the windowing operation $\mathrm{diag}(\mathbf{w}_m)$, we obtain its dual in the frequency domain as $\tilde{\mathbf{W}}_m := \mathbf{V}^H \mathrm{diag}(\mathbf{w}_m)\mathbf{V}$ [cf. (17)], and use those to define the power *spectrum mixing* matrix of windows m and m' as the entrywise product

$$\tilde{\mathbf{W}}_{mm'} := \tilde{\mathbf{W}}_m \circ \tilde{\mathbf{W}}_{m'}^*. \tag{21}$$

We use these matrices to give expressions for the bias and covariance of the estimator in (20) in the following proposition.

Proposition 4: Let $\mathbf{p}$ be the PSD of a process $\mathbf{x}$ that is stationary with respect to the shift $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$. 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 (20). The expectation of the estimator $\hat{\mathbf{p}}_\mathcal{W}$ is

$$\mathbb{E}[\hat{\mathbf{p}}_\mathcal{W}] = \frac{1}{M}\sum_{m=1}^M \tilde{\mathbf{W}}_{mm}\mathbf{p}. \tag{22}$$

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 $\boldsymbol{\Sigma}_\mathcal{W} := \mathbb{E}[(\hat{\mathbf{p}}_\mathcal{W} - \mathbb{E}[\hat{\mathbf{p}}_\mathcal{W}])(\hat{\mathbf{p}}_\mathcal{W} - \mathbb{E}[\hat{\mathbf{p}}_\mathcal{W}])^H]$. If the process $\mathbf{x}$ is assumed Gaussian and $\mathbf{S}$ is symmetric, the trace of the covariance matrix can be written as

$$\mathrm{tr}[\boldsymbol{\Sigma}_\mathcal{W}] = \frac{2}{M^2} \sum_{m=1,m'=1}^M \mathrm{tr}\left[(\tilde{\mathbf{W}}_{mm'}\mathbf{p})(\tilde{\mathbf{W}}_{mm'}\mathbf{p})^H\right]. \tag{23}$$

Proof: See Appendix B, where the expression of $\mathrm{tr}[\boldsymbol{\Sigma}_\mathcal{W}]$ for nonsymmetric normal shifts is given too [cf. (59)]. ∎

The first claim of Prop. 4 is a generalization of the bias expression for the bias of windowed periodograms in (19). It states that the estimator $\hat{\mathbf{p}}_\mathcal{W}$ is biased with a bias determined by the average of the spectrum mixing matrices $\tilde{\mathbf{W}}_{mm} = \tilde{\mathbf{W}}_m \circ \tilde{\mathbf{W}}_m^*$. The form of these mixing matrices depends on the design of the window bank $\mathcal{W} = \{\mathbf{w}_m\}_{m=1}^M$ and on the topology of the graph $\mathcal{G}$. Observe that even if the individual spectrum mixing matrices $\tilde{\mathbf{W}}_{mm}$ are not close to identity we can still have small bias by making their weighted sum $M^{-1}\sum_m \tilde{\mathbf{W}}_{mm} \approx \mathbf{I}$.

---

[1]To keep notation simple, we use $\mathbf{x}$ to denote a realization of process $\mathbf{x}$.

[2]See [14] for a different definition of the windowing operation in the graph domain. While the definition in [14] does not amount to multiplication in the vertex domain, it exhibits a number of convenient properties.

The second claim in Prop. 4 is a characterization of the trace of the covariance matrix $\mathbf{\Sigma}_{\mathcal{W}}$. To interpret the expression in (23) it is convenient to separate the summands with $m = m'$ from the rest to write

$$
\text{tr}[\mathbf{\Sigma}_{\mathcal{W}}] = \frac{2}{M^2} \sum_{m=1}^{M} \text{tr}\left[\left(\tilde{\mathbf{W}}_{mm}\mathbf{p}\right)\left(\tilde{\mathbf{W}}_{mm}\mathbf{p}\right)^H\right]
$$
$$
+ \frac{2}{M^2} \sum_{m=1, m' \neq m}^{M} \text{tr}\left[\left(\tilde{\mathbf{W}}_{mm'}\mathbf{p}\right)\left(\tilde{\mathbf{W}}_{mm'}\mathbf{p}\right)^H\right]. \quad (24)
$$

Comparing (24) with (19), we see that the terms in the first summand are $\text{tr}[(\tilde{\mathbf{W}}_{mm}\mathbf{p})(\tilde{\mathbf{W}}_{mm}\mathbf{p})^H] = \|\mathbb{E}[\hat{\mathbf{p}}_{\mathbf{w}_m}]\|_2^2$. Given that the windows have energy $\|\mathbf{w}\|_2^2 = \|\mathbf{1}\|_2^2 = N$ we expect $\|\mathbb{E}[\hat{\mathbf{p}}_{\mathbf{w}_m}]\| \approx \|\mathbb{E}[\hat{\mathbf{p}}_{\mathbf{x}}]\|_2 = \|\mathbf{p}\|_2$. This implies that the first sum in (24) is approximately proportional to $2\|\mathbf{p}\|_2^2/M$. Thus, this first term behaves as if the different windows in the bank $\mathcal{W}$ were generating independent samples [cf. (16)].

The effect of the correlation between different windowed signals appears in the cross terms of the second sum, which can be viewed as the price to pay for the signals $\{\mathbf{x}_m\}_{m=1}^{M}$ being generated from the same realization $\mathbf{x}$ instead of actually being independent. We explain next that the price of this second sum is smaller than the gain we get in the first term.

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

$$
\text{MSE}(\hat{\mathbf{p}}_{\mathcal{W}}) = \|\mathbf{b}_{\mathcal{W}}\|_2^2 + \text{tr}[\mathbf{\Sigma}_{\mathcal{W}}]. \quad (25)
$$

The expression in (25) can be used to design (optimal) windows with minimum MSE. Do notice that the bias in (25) depends on the unknown PSD $\mathbf{p}$. This problem can be circumvented by making $\mathbf{p} = \mathbf{1}$ in the bias and trace expressions in Prop. 4. This choice implies that the PSD is assumed white a priori. If some other knowledge of the PSD is available, it can be used as an alternative prior. Irrespectively of the choice of prior, finding windows that minimize the MSE is computationally challenging.

An alternative approach to design the window bank $\mathcal{W} = \{\mathbf{w}_m\}_{m=1}^{M}$ is to exploit the local properties of the random process $\mathbf{x}$. As stated in Property 3, some stationary processes are expected to have correlations with local structure. It is then reasonable to expect that windows without overlap capture independent information that results in a reduction of the cross-terms in (24). This intuition is formalized next.

Proposition 5: Consider two windows $\mathbf{w}_m$ and $\mathbf{w}_{m'}$ and assume that the distance between any node in $\mathbf{w}_m$ and any node in $\mathbf{w}_{m'}$ is larger than $2L$ hops. If the process $\mathbf{x}$ can be modeled as the output of an L-degree filter, then it holds that $\text{tr}[(\tilde{\mathbf{W}}_{mm'}\mathbf{p})(\tilde{\mathbf{W}}_{mm'}\mathbf{p})^H] = 0$.

Proof: Let $\mathcal{N}(\mathbf{w}_m) = \{i : [\mathbf{w}_m]_i \neq 0\}$ be the set of nodes in the support of $\mathbf{w}_m$, and $\mathcal{N}(\mathbf{w}_{m'})$ the ones in the support of $\mathbf{w}_{m'}$. We know that the distance between any $i \in \mathcal{N}(\mathbf{w}_m)$ and any $i' \in \mathcal{N}(\mathbf{w}_{m'})$ is greater than $2L$. Invoking Property 3, this implies that $[\mathbf{C}_x]_{ii'} = 0$, which allows us to write $\text{diag}(\mathbf{e}_i)\mathbf{C}_x\text{diag}(\mathbf{e}_{i'}) = \mathbf{0}$. Since this is true for any pair (i, i') with $i \in \mathcal{N}(\mathbf{w}_m)$ and $i' \in \mathcal{N}(\mathbf{w}_{m'})$, we have that: *f1)* $\text{diag}(\mathbf{w}_m)\mathbf{C}_x\text{diag}(\mathbf{w}_{m'}) = \mathbf{0}$. Note now that $\tilde{\mathbf{W}}_{mm'}\mathbf{p} = \tilde{\mathbf{W}}_m \circ \tilde{\mathbf{W}}_{m'}^*\mathbf{p} = \tilde{\mathbf{W}}_m\text{diag}(\mathbf{p})\tilde{\mathbf{W}}_{m'}^*$, which using the definitions $\tilde{\mathbf{W}}_m = \mathbf{V}^H\text{diag}(\mathbf{w}_m)\mathbf{V}$ and $\mathbf{C}_x = \mathbf{V}\text{diag}(\mathbf{p})\mathbf{V}^H$ can be

written as: *f2)* $\tilde{\mathbf{W}}_{mm'}\mathbf{p} = \mathbf{V}^H\text{diag}(\mathbf{w}_m)\mathbf{C}_x\text{diag}(\mathbf{w}_{m'}^*)\mathbf{V}$. Substituting *f1)* into *f2)* yields $\tilde{\mathbf{W}}_{mm'}\mathbf{p} = \mathbf{V}^H\mathbf{0}\mathbf{V} = \mathbf{0}$. ∎

The result in Prop. 5 implies that if we use windows without overlap, the second sum in (24) is null. This means that the covariance matrix $\mathbf{\Sigma}_{\mathcal{W}}$ of the PSD estimator in (20) behaves as if the separate windowed samples were independent samples. We emphasize that Prop. 5 provides a lax bound for a rather stringent correlation model. The purpose of the result is to illustrate the reasons why local windows are expected to reduce estimation MSE. In practice, we expect local windows to reduce MSE as long as the correlation decreases with the hop distance between nodes. Windowing design can then be related to clustering (if the windows are non-overlapping) and covering (if they overlap) with the goal of keeping the diameter of the window large enough so that most of the correlation of the process is preserved. Recent results in the context of designing sampling schemes for covariance estimation of graph processes corroborate this point [27]. Sec. VI evaluates the estimation performance when estimating the PSD of a graph process for different types of windows.

C. Filter Banks

Windows reduce the MSE of periodograms by exploiting the locality of correlations. Filter banks reduce MSE by exploiting the locality of the PSD [9, Sec. 5]. To define filter bank PSD estimators for graph processes, suppose that we are given a filter bank $\mathcal{Q} := \{\mathbf{Q}_k\}_{k=1}^{N}$ with N filters (as many as frequencies). The filters $\mathbf{Q}_k$ are assumed linear shift invariant with frequency responses $\tilde{\mathbf{q}}_k$, so that we can write $\mathbf{Q}_k = \mathbf{V}\text{diag}(\tilde{\mathbf{q}}_k)\mathbf{V}^H$. We further assume that their energies are normalized $\|\tilde{\mathbf{q}}_k\|_2^2 = 1$. The kth (bandpass) filter is intended to estimate the kth component of the PSD $\mathbf{p} = [p_1, \ldots, p_N]^T$ through the energy of the output signal $\mathbf{x}_k := \mathbf{Q}_k\mathbf{x}$, i.e.,

$$
\hat{p}_{\tilde{\mathbf{q}}_k} := \|\mathbf{x}_k\|_2^2 = \|\mathbf{Q}_k\mathbf{x}\|_2^2. \quad (26)
$$

The filter bank PSD estimate is given by the concatenation of the individual estimates in (26) into the vector $\hat{\mathbf{p}}_{\mathcal{Q}} := [\hat{p}_{\tilde{\mathbf{q}}_1}, \ldots, \hat{p}_{\tilde{\mathbf{q}}_N}]^T$. We emphasize that we can think of filter banks as an alternative approach to generate multiple virtual realizations $\mathbf{x}_k$ from a single actual realization $\mathbf{x}$. In the case of windowing, realizations $\mathbf{x}_m$ correspond to different pieces of $\mathbf{x}$. In the case of filter banks, realizations $\mathbf{x}_k$ correspond to different filtered versions[3] of $\mathbf{x}$.

Using Parseval's theorem and the frequency representations $\tilde{\mathbf{q}}_k$ and $\tilde{\mathbf{x}}$ of the filter $\mathbf{Q}_k$ and the realization $\mathbf{x}$, respectively, the estimate in (26) is equivalent to

$$
\hat{p}_{\tilde{\mathbf{q}}_k} = \|\tilde{\mathbf{x}}_k\|_2^2 = \|\text{diag}(\tilde{\mathbf{q}}_k)\tilde{\mathbf{x}}\|_2^2. \quad (27)
$$

The expression in (27) guides the selection of the response $\tilde{\mathbf{q}}_k$. E.g., if the PSD values at frequencies k and k' are expected to be similar, i.e., if $p_k \approx p_{k'}$ we can make $[\tilde{\mathbf{q}}_k]_k = [\tilde{\mathbf{q}}_k]_{k'} = (1/\sqrt{2})$ so that the PSD components $[|\mathbf{V}^H\mathbf{x}|^2]_k$ and $[|\mathbf{V}^H\mathbf{x}|^2]_{k'}$ are averaged. More generically, the design of the filter $\tilde{\mathbf{q}}_k$ can be guided by the bias and variance expressions that we present in the following proposition.

---

[3]A related approach that filters $\mathbf{x}$ using a set of $N' \neq N$ graph filters whose frequency responses are designed by shifting a prespecified kernel and then obtains the PSD estimate via interpolation is presented in [13].

Proposition 6: Let $\mathbf{p} = [p_1, \ldots, p_N]^T$ be the PSD of a process $\mathbf{x}$ that is stationary with respect to the shift $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$. A single observation $\mathbf{x}$ is given along with the filter bank $\mathcal{Q} = \{\tilde{\mathbf{q}}_k\}_{k=1}^N$ and the filter bank PSD estimates $\hat{\mathbf{p}}_\mathcal{Q}$ are computed as in (27). The expectation of the kth entry of $\hat{\mathbf{p}}_\mathcal{Q}$ is

$$\mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right] := (|\tilde{\mathbf{q}}_k|^2)^T \mathbf{p}. \tag{28}$$

Equivalently, $\hat{p}_{\tilde{\mathbf{q}}_k}$ is biased with bias $b_{\tilde{\mathbf{q}}_k} := \mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right] - p_k$. Further define the variance of the kth entry of the filter bank estimate $\hat{\mathbf{p}}_\mathcal{Q}$ as $\mathrm{var}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right] := \mathbb{E}\left[(\hat{p}_{\tilde{\mathbf{q}}_k} - \mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right])^2\right]$. If the process $\mathbf{x}$ is assumed Gaussian and $\mathbf{S}$ is symmetric, the variance of $\hat{p}_{\tilde{\mathbf{q}}_k}$ is

$$\mathrm{var}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right] := \mathbb{E}\left[(\hat{p}_{\tilde{\mathbf{q}}_k} - \mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right])^2\right] = 2\left\|\mathrm{diag}\left(|\tilde{\mathbf{q}}_k|^2\right)\mathbf{p}\right\|_2^2. \tag{29}$$

Proof: See Appendix C, where the expression of $\mathrm{var}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right]$ for nonsymmetric normal shifts is given too [cf. (64)]. ∎

The variance expression in (28) shows that the estimation accuracy of a filter bank benefits from an averaging effect – recall that the filter is normalized to have unit energy $\|\tilde{\mathbf{q}}_k\|_2^2 = 1$. This averaging advantage manifests only if the bias in (28) is made small so that the overall MSE, given by $\mathrm{MSE}(\hat{\mathbf{p}}_\mathcal{Q}) = \sum_{k=1}^N b_{\tilde{\mathbf{q}}_k}^2 + \mathrm{var}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right]$, decreases. In time signals the bias is made small by exploiting the fact that the PSD of nearby frequencies are similar. In graph signals, some extra information is necessary to identify frequency components with similar PSD values. If, e.g., the process $\mathbf{x}$ is a diffusion process as the one in Example 4, the PSD components p_k and $p_{k'}$ are similar – irrespectively of α – if the eigenvalues λ_k and $\lambda_{k'}$ of the shift $\mathbf{S}$ are similar. If the eigenvalues of the Laplacian are further ordered, averaging of nearby PSD estimates can be interpreted as the use of a bandpass filter.

A generic approach to designing bandpass filters for PSD estimation is to exploit (FIR) filters, which are attractive due their ability to be implemented distributedly [15]. Formally, write $\mathbf{Q}_k = \sum_{l=1}^L q_k^l \mathbf{S}^l$, denote as $\mathbf{q}_k = [q_k^1, \ldots, q_k^L]^T$ the vector of filter coefficients, and as $\tilde{\mathbf{q}}_k = \boldsymbol{\Psi}_L \mathbf{q}_k$ its frequency response, where we recall that $\boldsymbol{\Psi}_L$ stands for the first L columns of $\boldsymbol{\Psi}$. The coefficients $\mathbf{q}_k$ could be designed upon substituting $\tilde{\mathbf{q}}_k = \boldsymbol{\Psi}_L \mathbf{q}_k$ into both (28) and (29) and minimizing the resultant MSE. This can be challenging because it involves fourth-order polynomials and requires some prior knowledge on $\mathbf{p}$. For the purpose of PSD estimation in time, a traditional approach for suboptimal FIR design is to select coefficients guaranteeing that $[\tilde{\mathbf{q}}_k]_k = 1$ while minimizing the out-of-band power [8]. Defining $\boldsymbol{\psi}_{k,L}^T$ as the kth row of $\boldsymbol{\Psi}_L$, this can be formalized as

$$\mathbf{q}_k := \underset{\mathbf{q}_L}{\mathrm{argmin}} \, \|\boldsymbol{\Psi}_L \mathbf{q}_L\|_2^2, \qquad \text{s.t.} \ \boldsymbol{\psi}_{k,L}^T \mathbf{q}_L = 1, \tag{30}$$

with the constraint forcing $[\tilde{\mathbf{q}}_k]_k = 1$ and the objective attempting to minimize the contribution to $\hat{p}_{\tilde{\mathbf{q}}_k}$ from frequencies other than k. If we make $L \geq N$ the solution to (30) is $[\tilde{\mathbf{q}}_k]_{k'} = 0$ for all $k' \neq k$. For $L < N$, the filter $\mathbf{q}_k$ has a response in which nonzero coefficients $[\tilde{\mathbf{q}}_k]_{k'}$ are clustered at frequencies k' that are similar to k – in the sense of being associated with multipliers $\lambda_{k'} \approx \lambda_k$. We further observe that (30) has the additional advantage of being solved in closed form as

$$\mathbf{q}_k = \left(\boldsymbol{\psi}_{k,L}^T \boldsymbol{\Psi}_L^H \boldsymbol{\Psi}_L \boldsymbol{\psi}_{k,L}^*\right)^{-1} \left(\boldsymbol{\Psi}_L^H \boldsymbol{\Psi}_L\right)^{-1} \boldsymbol{\psi}_{k,L}, \tag{31}$$

which does not have unit energy but can be normalized.

V. PARAMETRIC PSD ESTIMATION

We address PSD estimation assuming that the graph process $\mathbf{x}$ can be well approximated by a parametric model in which $\mathbf{x}$ is the response of a graph filter $\mathbf{H}$ to a white input. As per Def. 1, this is always possible if the filter's order is sufficiently large. The goal here is to devise filter representations of an order (much) smaller than the number of signal elements N. Mimicking time processes, we devise moving average (MA), autoregressive (AR), and ARMA models. Due to space constraints the focus here is on the modeling of graph processes and the parametric estimation of the generating filters, but not on the design of those filters. Details on this latter topic can be found in, e.g., [7], [20], [28].

A. Moving Average Graph Processes

Consider a vector of coefficients $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_{L-1}]^T$ and assume that $\mathbf{x}$ is stationary in the graph $\mathbf{S}$ 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 although we want in practice to have $L \ll N$. If the process $\mathbf{x}$ is indeed generated as the response of $\mathbf{H}(\boldsymbol{\beta})$ to a white input, the covariance of $\mathbf{x}$ can be written as

$$\mathbf{C}_x(\boldsymbol{\beta}) = \mathbf{H}(\boldsymbol{\beta})\mathbf{H}^H(\boldsymbol{\beta}) = \sum_{l=0, l'=0}^{L-1} (\beta_l \mathbf{S}^l)(\beta_{l'} \mathbf{S}^H)^{l'}. \tag{32}$$

The PSD corresponding to the covariance in (32) is the magnitude squared of the frequency representation of the filter. To see this formally, notice that it follows from the definition in (7) that $\mathbf{p}(\boldsymbol{\beta}) = \mathrm{diag}\left(\mathbf{V}^H \mathbf{C}_x(\boldsymbol{\beta})\mathbf{V}\right)$. 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 $\tilde{\mathbf{h}}(\boldsymbol{\beta}) = \mathrm{diag}(\mathbf{V}\mathbf{H}(\boldsymbol{\beta})\mathbf{V}^H)$, it follows readily that $\mathbf{p}(\boldsymbol{\beta}) = |\tilde{\mathbf{h}}(\boldsymbol{\beta})|^2$. For the purposes of this section the latter will be written explicitly in terms of $\boldsymbol{\beta}$ as [cf. (3)]

$$\mathbf{p}(\boldsymbol{\beta}) = |\tilde{\mathbf{h}}(\boldsymbol{\beta})|^2 = |\boldsymbol{\Psi}_L \boldsymbol{\beta}|^2. \tag{33}$$

The covariance and PSD expressions in (32) and (33) are graph counterparts of MA time processes generated by FIR filters – see Sec. III-B for a discussion on their practical relevance.

The estimation of the coefficients $\boldsymbol{\beta}$ can be addressed in either the graph or graph frequency domain. In the graph domain we compute the sample covariance $\hat{\mathbf{C}}_x = \mathbf{x}\mathbf{x}^H$ and introduce a distortion function $D_\mathbf{C}(\hat{\mathbf{C}}_x, \mathbf{C}_x(\boldsymbol{\beta}))$ to measure the similarity of $\hat{\mathbf{C}}_x$ and $\mathbf{C}_x(\boldsymbol{\beta})$. The filter coefficients $\boldsymbol{\beta}$ are then selected as the ones with minimal distortion,

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\mathrm{argmin}} \, D_\mathbf{C}(\hat{\mathbf{C}}_x, \mathbf{C}_x(\boldsymbol{\beta})). \tag{34}$$

The expression for $\mathbf{C}_x(\boldsymbol{\beta})$ in (32) is a quadratic function of $\boldsymbol{\beta}$ that is generally indefinite. The optimization problem in (34) will therefore be not convex in general.

To perform estimation in the frequency domain we first compute the periodogram $\hat{\mathbf{p}}_{\mathrm{pg}}$ defined in (12). We then introduce a distortion measure $D_\mathbf{p}(\hat{\mathbf{p}}_{\mathrm{pg}}, |\boldsymbol{\Psi}_L \boldsymbol{\beta}|^2)$ to compare the periodogram $\hat{\mathbf{p}}_{\mathrm{pg}}$ with the PSD $|\boldsymbol{\Psi}_L \boldsymbol{\beta}|^2$ and select the coefficients $\boldsymbol{\beta}$ that solve the following optimization

$$\hat{\boldsymbol{\beta}} := \underset{\boldsymbol{\beta}}{\mathrm{argmin}} \, D_\mathbf{p}(\hat{\mathbf{p}}_{\mathrm{pg}}, |\boldsymbol{\Psi}_L \boldsymbol{\beta}|^2). \tag{35}$$

We observe that although we use the periodogram in (35), any of the nonparametric methods of Sec. IV-A can be used instead. Since the quadratic form $|\mathbf{\Psi}_L\boldsymbol{\beta}|^2$ in (35) is also indefinite, the optimization problem in (35) is not necessarily convex. In the particular case when the distortion $D_{\mathbf{p}}(\hat{\mathbf{p}}_{\mathrm{pg}}, |\mathbf{\Psi}_L\boldsymbol{\beta}|^2) = \|\hat{\mathbf{p}}_{\mathrm{pg}} - |\mathbf{\Psi}_L\boldsymbol{\beta}|^2\|_2^2$ is the Euclidean 2-norm, efficient (phase-retrieval) solvers with probabilistic guarantees are available [29], [30]. Alternative tractable formulations of (34) and (35) when the shifts are symmetric, or, when the shifts are positive semidefinite and the filter coefficients are nonnegative are discussed below.

Symmetric shifts: If the shift $\mathbf{S}$ is symmetric, the expression for the covariance matrix in (32) can be simplified to a polynomial of degree $2(L-1)$ in $\mathbf{S}$,

$$\mathbf{C}_x(\boldsymbol{\beta}) = \sum_{l=0,l'=0}^{L-1} \beta_l\beta_{l'}\mathbf{S}^{l+l'} := \sum_{l=0}^{2(L-1)} \gamma_l\mathbf{S}^l := \mathbf{C}_x(\boldsymbol{\gamma}). \quad (36)$$

In the second equality in (36) 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 $\mathbf{C}_x(\boldsymbol{\gamma})$ to denote the covariance matrix written in terms of the $\boldsymbol{\gamma}$ coefficients. We propose now a relaxation of (34) in which $\mathbf{C}_x(\boldsymbol{\gamma})$ is used in lieu of $\mathbf{C}_x(\boldsymbol{\beta})$ to yield the optimization problem

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}}\ D_{\mathbf{C}}(\hat{\mathbf{C}}_x, \mathbf{C}_x(\boldsymbol{\gamma})). \quad (37)$$

If we add the constraints $\gamma_l = \sum_{l'+l''=l} \beta_{l'}\beta_{l''}$, the problem in (37) is equivalent to (34). By dropping these constraints we end up with a tractable relaxation because (37) is convex for all convex distortion metrics $D_{\mathbf{C}}(\hat{\mathbf{C}}_x, \mathbf{C}_x(\boldsymbol{\gamma}))$. A tractable relaxation of (35) can be derived analogously.

Nonnegative filter coefficients: When the shift $\mathbf{S}$ is positive semidefinite, the elements of the matrix $\mathbf{\Psi}$ are all nonnegative. If we further restrict the coefficients $\boldsymbol{\beta}$ to be nonnegative, all the elements in the product $\mathbf{\Psi}_L\boldsymbol{\beta}$ are also nonnegative. This means that in (35) we can replace the comparison $D_{\mathbf{p}}(\hat{\mathbf{p}}_{\mathrm{pg}}, |\mathbf{\Psi}_L\boldsymbol{\beta}|^2)$ by the comparison $D_{\mathbf{p}}(\sqrt{\hat{\mathbf{p}}_{\mathrm{pg}}}, \mathbf{\Psi}_L\boldsymbol{\beta})$. We can then replace (35) by

$$\hat{\boldsymbol{\beta}} := \underset{\boldsymbol{\beta} \geq \mathbf{0}}{\operatorname{argmin}}\ D_{\mathbf{p}}\left(\sqrt{\hat{\mathbf{p}}_{\mathrm{pg}}}, \mathbf{\Psi}_L\boldsymbol{\beta}\right). \quad (38)$$

The optimization in (38) is convex, therefore tractable, for all convex distortion metrics $D_{\mathbf{p}}(\sqrt{\hat{\mathbf{p}}_{\mathrm{pg}}}, \mathbf{\Psi}_L\boldsymbol{\beta})$. Do notice that the objective costs in (38) and (35) are *not* equivalent and that (38) requires positive semidefinite shifts – such as the Laplacian – and restricts coefficients to satisfy $\boldsymbol{\beta} \geq \mathbf{0}$. A tractable restriction of (34) can be derived analogously.

B. Autoregressive Graph Processes

For some processes it is more convenient to use a parametric model that generates an infinite impulse response through an autoregressive filter. As a simple example, consider the diffusion process driven by the graph filter $\mathbf{H} = \alpha_0 \sum_{l=0}^{\infty} \alpha^l \mathbf{S}^l$, where α represents the diffusion rate and α_0 a scaling coefficient. 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 [20].

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. (4)]

$$\mathbf{C}_x(\alpha_0, \alpha) = \mathbf{H}\mathbf{H}^H = \alpha_0^2(\mathbf{I} - \alpha\mathbf{S})^{-1}(\mathbf{I} - \alpha\mathbf{S})^{-H}, \quad (39)$$

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

$$\mathbf{p}(\alpha_0, \alpha) = \operatorname{diag}\left[\alpha_0^2|\mathbf{I} - \alpha\mathbf{\Lambda}|^{-2}\right], \quad (40)$$

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 (39) and (40) as we did in Sec. V-A for MA models. Substituting $\mathbf{C}_x(\alpha_0, \alpha)$ for $\mathbf{C}_x(\boldsymbol{\beta})$ in (34) yields a graph domain formulation and substituting $\mathbf{p}(\alpha_0, \alpha)$ for $|\mathbf{\Psi}_L\boldsymbol{\beta}|^2$ in (35) yields a graph frequency domain formulation. Since only two parameters must be estimated the corresponding optimization problems are tractable.

Since 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 $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_M]^T$. The frequency response $\tilde{\mathbf{h}} = \operatorname{diag}(\mathbf{V}^H\mathbf{H}\mathbf{V})$ of this filter is

$$\tilde{\mathbf{h}} = \alpha_0 \operatorname{diag}\left[\prod_{m=1}^{M}(\mathbf{I} - \alpha_m\mathbf{\Lambda})^{-1}\right]. \quad (41)$$

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

$$\mathbf{C}_x(\boldsymbol{\alpha}) = \alpha_0^2 \prod_{m=1}^{M}(\mathbf{I} - \alpha_m\mathbf{S})^{-1}(\mathbf{I} - \alpha_m\mathbf{S})^{-H}. \quad (42)$$

The process $\mathbf{x}$ is stationary with respect to $\mathbf{S}$, because of, e.g., Def. 1. The PSD of $\mathbf{x}$ can be written as

$$\mathbf{p}(\boldsymbol{\alpha}) = \alpha_0^2 \operatorname{diag}\left[\prod_{m=1}^{M}|\mathbf{I} - \alpha_m\mathbf{\Lambda}|^{-2}\right]. \quad (43)$$

As before, we substitute $\mathbf{C}_x(\boldsymbol{\alpha})$ for $\mathbf{C}_x(\boldsymbol{\beta})$ in (34) to obtain a graph domain formulation and substitute $\mathbf{p}(\boldsymbol{\alpha})$ for $|\mathbf{\Psi}_L\boldsymbol{\beta}|^2$ in (35) to obtain a graph frequency domain formulation. A related approach in the context of identifying the coefficients of a linear predictor using graph filters was presented in [3]. For large degree M the problems can become intractable. Yule-Walker schemes [9, Sec. 3.4] tailored to graph signals may be of help. Their derivation and analysis are left as future work.

Remark 3: To further motivate AR processes, consider the example of $\mathbf{x} = \mathbf{H}\mathbf{w}$ with $\mathbf{H}$ being a single-pole filter and $\mathbf{w}$ a white and Gaussian input, so that $\mathbf{x}$ is Gaussian too. The covariance of this first-order AR process is given by (39), with its inverse covariance (precision matrix) being simply $\mathbf{\Theta} := \mathbf{C}_x^{-1} = (\rho)^{-2}(\mathbf{I} - \alpha\mathbf{S})^H(\mathbf{I} - \alpha\mathbf{S})$. Since $\mathbf{S}$ is sparse, the precision matrix $\mathbf{\Theta}$ is sparse too. Specifically, $\Theta_{i,j} \neq 0$ only if j is in the two-hop neighborhood of i. Then, it follows that a Gaussian AR process of order one on the shift $\mathbf{S}$ is a GMRF, with the Markov blanket [23, Ch. 19] of a node i being given by $\mathcal{N}_2(i)$, i.e., the nodes that are within the two-hop neighborhood of i. As explained in Example 2, such a process is stationary

both in $\mathbf{S}$ and in $\boldsymbol{\Theta}$. The same holds true for an AR process of order M, which in this case will give rise to a GMRF whose Markov blankets are given by the $2M$-hop neighborhoods of the original graph.

C. Autoregressive Moving Average Graph Processes

The techniques in Secs. V-A and V-B 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 $\mathbf{a} := [a_1, ..., a_M]^T$ and $\mathbf{b} := [b_0, ..., b_{L-1}]^T$ and postulate filters with frequency response

$$\tilde{\mathbf{h}} = \mathrm{diag}\left[\left(\sum_{l=0}^{L-1} b_l \boldsymbol{\Lambda}^l \right) \left(1 - \sum_{m=1}^{M} a_m \boldsymbol{\Lambda}^m \right)^{-1} \right]. \quad (44)$$

To find the counterpart of (44) 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 (44) 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

$$\mathbf{C}_x(\mathbf{a}, \mathbf{b}) = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{B}^H(\mathbf{I} - \mathbf{A})^{-H}. \quad (45)$$

Since $\mathbf{C}_x(\mathbf{a}, \mathbf{b})$ is diagonalized by the GFT $\mathbf{V}$, the process $\mathbf{x}$ is stationary with PSD [cf. (44)]

$$\mathbf{p}(\mathbf{a}, \mathbf{b}) = \mathrm{diag}\left[\left| \sum_{l=0}^{L-1} b_l \boldsymbol{\Lambda}^l \right|^2 \left| 1 - \sum_{m=1}^{M} a_m \boldsymbol{\Lambda}^m \right|^{-2} \right]. \quad (46)$$

As in the AR and MA models, we can identify the model coefficients by minimizing the covariance distortion $D_{\mathbf{C}}(\hat{\mathbf{C}}_x, \mathbf{C}_x(\mathbf{a}, \mathbf{b}))$ or the PSD distortion $D_{\mathbf{p}}(\hat{\mathbf{p}}_{\mathrm{pg}}, \mathbf{p}(\mathbf{a}, \mathbf{b}))$ [cf. (33) and (34)]. These optimization problems are computationally difficult.

Alternative estimation schemes can be obtained by reordering (45) into $(\mathbf{I} - \mathbf{A})\mathbf{C}_x(\mathbf{I} - \mathbf{A})^H = \mathbf{B}\mathbf{B}^H$ and solving for either the graph domain distortion

$$(\hat{\mathbf{a}}, \hat{\mathbf{b}}) := \underset{\mathbf{a}, \mathbf{b}}{\mathrm{argmin}} \; D_{\mathbf{C}}\left((\mathbf{I} - \mathbf{A})\hat{\mathbf{C}}_x(\mathbf{I} - \mathbf{A})^H, \mathbf{B}\mathbf{B}^H \right) \quad (47)$$

or the graph frequency domain distortion

$$(\hat{\mathbf{a}}, \hat{\mathbf{b}}) := \underset{\mathbf{a}, \mathbf{b}}{\mathrm{argmin}}$$
$$D_{\mathbf{p}}\left[\left| 1 - \sum_{m=1}^{M} a_m \boldsymbol{\Lambda}^m \right|^2 \hat{\mathbf{p}}_{\mathrm{pg}}, \mathrm{diag}\left[\left| \sum_{l=0}^{L-1} b_l \boldsymbol{\Lambda}^l \right|^2 \right] \right]. \quad (48)$$

The formulations in (47) and (48) can still be intractable but we observe that the problems have the same structure as the ones considered in Sec. V-A. 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 as well.

Remark 4: The parametric methods in this section are well tailored to PSD estimation of diffusion processes – see Sec. III-B. When $L \ll N$ the dynamics in (11) are accurately

represented by a low-order FIR model. Parametric estimation with a MA model as in Sec. V-A is recommendable. Single pole AR models arise when $\gamma^{(l)} = \gamma$ for all l and $L = \infty$ as we already explained in Sec. V-B. An AR model of order M arises when M single pole models are applied sequentially. The latter implies that AR models are applicable when the diffusion coefficients $\gamma^{(l)}$ are constant during stretches of time or vary slowly with time. If we consider now M diffusion dynamics with constant coefficients $\gamma^{(l)} = \gamma_m$ for all l running in *parallel* and we produce an output as the sum of the M outcomes we obtain an ARMA system with M poles and $M - 1$ zeros [20].

Remark 5: The methods for PSD estimation that we presented in Secs. IV and V can be used for covariance estimation as well. This follows directly for some of the parametric formulations in which we estimate filter coefficients that minimize a graph domain distortion – these include (34), its relaxation in (37), and the analogous formulations in Secs. V-B and V-C. When this is not done, an estimate for $\mathbf{C}_x$ can be computed from a PSD estimate as $\hat{\mathbf{C}}_x = \mathbf{V}\mathrm{diag}(\hat{\mathbf{p}}_{\mathrm{pg}})\mathbf{V}^H$. A further step is to use the notion of stationarity and the models in this section to estimate the shift $\mathbf{S}$ itself (hence, the topology of the network) from a set of signal realizations. See [22], [31] for two recent examples along these lines.

VI. NUMERICAL EXPERIMENTS

The implementation and associated benefits of the proposed schemes are illustrated through four test cases (TCs). TC1 and TC2 rely on synthetic graphs to evaluate the performance of nonparametric and parametric PSD estimation methods. TC3 and TC4 illustrate how the concepts and tools of stationary graph processes can be leveraged in practical applications involving real-world signals and graphs. Unless otherwise stated, the results shown are averages across 100 realizations of the particular experiment.

TC1. Nonparametric methods: We first evaluate the estimation performance of the average periodogram [cf. (12) and (16)] as a function of R, the number of realizations observed. Consider a baseline Erdős-Rényi (ER) graph with $N = 100$ nodes and edge probability $p = 0.05$ [32]. 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. (16)] 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 [33] 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 (16) was shown for Gaussian signals, we observe that in the tested non-Gaussian setup the evolution of the MSE with R is the same; see Fig. 1a (bottom). This similarity can be explained by the fact that a graph filter is a linear combination of shifted signals and shifting is a linear operation. Hence, invoking the central limit theorem, the larger
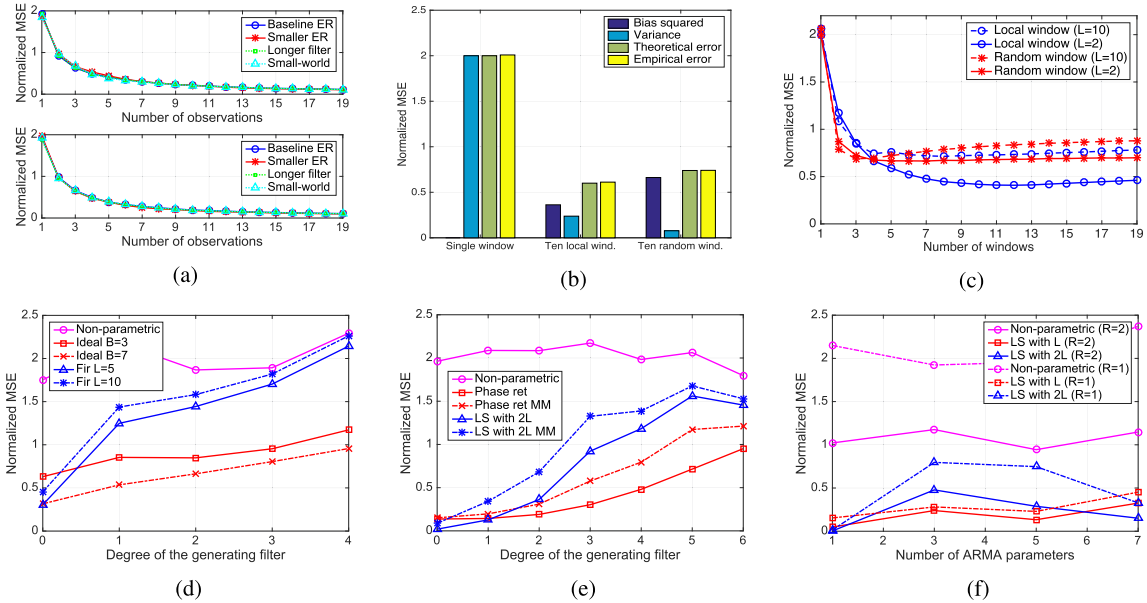
Fig. 1. Normalized MSE (NMSE) for different PSD estimation schemes. (a) Top: NMSE for the periodogram using Gaussian inputs. Bottom: NMSE for non-Gaussian inputs. (b) Theoretical and empirical NMSE for different window strategies. (c) NMSE as a function of the number of windows for local and random windows. (d) NMSE for filter banks as a function of the degree of the generating filter. (e) NMSE for MA parametric estimation as a function of the degree of the generating filter. (f) NMSE for different ARMA parametric estimators based on $R = 1$ and $R = 2$ signal realizations.

the filter degree the closer the signals at hand are from being normal.

The second experiment evaluates the performance of window-based estimators. To assess the role of locality in the window design, we consider graphs generated via a stochastic block model [34] 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 for a 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 4, respectively. 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 applying complete linkage clustering [35] to a metric space given by the shortest path distances between nodes. In order to obtain increasing number of windows M, we cut the output dendrogram at smaller resolutions [36]. The random windows are designed to have the same sizes as the local ones. The windows are tested for graph processes generated by two filters of different degrees: i) $L = 2$, so that nodes that are more than 2 hops away are not

correlated (cf. Property 3); and ii) $L = 10$, which is greater than the graph diameter, inducing correlations between every pair of nodes. 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 and then 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.

The last experiment evaluates the performance of filter-bank estimators. Two types of bandpass filters are considered. The first type designs the k-th filter as an ideal bandpass filter with unit response for the k-th frequency and the B frequencies closest to it, and zero otherwise [cf. (27)]. More precisely, being λ_k the eigenvalue associated to the k-th frequency, we consider the closest frequencies as those with eigenvalues $\lambda_{k'}$ minimizing $|\lambda_k - \lambda_{k'}|$. The second filter bank type designs the filters using the FIR approach in (31). To run the experiments we consider the adjacency matrix of an ER graph with $N = 100$ and $p = 0.05$, and generate signals by filtering white noise. Fig. 1d shows the MSE performance of both approaches as a function of the degree of the filter that generates the process as well as the nonparametric periodogram estimation. We consider two ideal bandpass filters with $B = 3$ and $B = 7$ and two FIR bandpass filters with $L = 5$ and $L = 10$. Fig. 1d indicates that filter banks contribute to reduce the MSE compared to the periodogram. Moreover, the ideal bandpass filter outperforms the FIR design and, within each type, filters with larger bandwidth ($B = 7$ and $L = 5$) tend
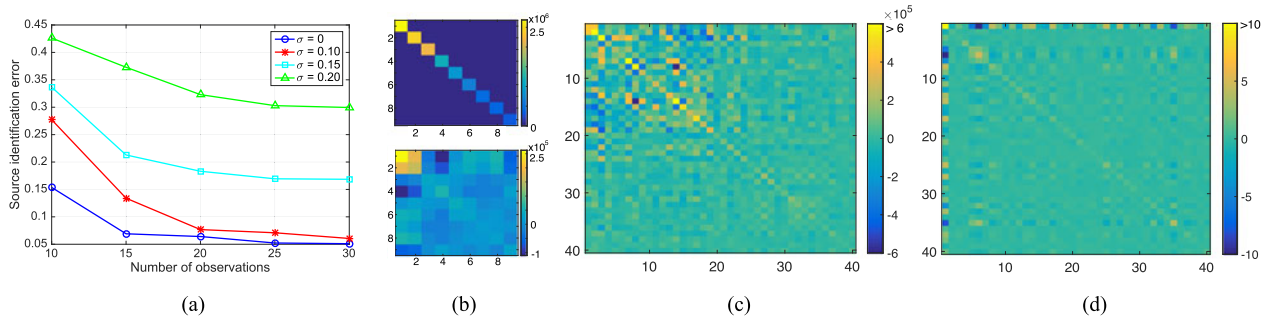
Fig. 2. (a) Error in the identification of sources in opinion formation dynamics as a function of the number of observed opinions and parametrized by the noise level σ. (b)-(c) Diagonalization of the sample covariance of the images of an individual with the basis of that same individual (b-top), a different one (b-bottom), and the ensemble sample covariance (c). (d) Diagonalization of the sample covariance of the functional brain signals in [37] with the basis corresponding to the structural brain network.

to perform better. The reason being that the periodogram-based estimation for a single observation is very noisy, thus, it benefits from the averaging effect of larger bandwidths.

TC2. Parametric methods: We first illustrate the parametric estimation of a MA process. Consider the Laplacian of an ER graph with $N = 100$ and $p = 0.2$ and processes generated by an FIR filter of length L whose coefficients $\boldsymbol{\beta}$ are selected randomly. The performance of a periodogram is contrasted with that of two parametric approaches: i) an algorithm that estimates the L values in $\boldsymbol{\beta}$ by minimizing (35) via phase-retrieval [30]; and ii) a least squares algorithm that estimates the $2L - 1$ values in $\boldsymbol{\gamma}$ by minimizing (36). The results are shown in Fig. 1e (solid lines). It can be observed that both parametric methods outperform the periodogram since they leverage the FIR structure of the generating filter. Moreover, this difference is largest for smaller values of the degree, since in these cases a few parameters are sufficient to completely characterize the filter of interest. Furthermore, 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. 1e). The results show that, although the model mismatch degrades the performance, the parametric estimates are still superior to the periodogram.

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 the periodogram estimation with two schemes: i) a least squares (LS) algorithm that estimates $2L$ coefficients, i.e., the counterpart of (36) for the problem in (48); and ii) a least squares algorithm that estimates L nonnegative coefficients, i.e., the counterpart of (38) for (48). Note that the latter is computationally tractable because both the eigenvalues of the shift and the coefficients of the filters are nonnegative. The algorithms are tested in two scenarios, with one and two signal realizations available, respectively. Fig. 1f shows that the parametric methods attain smaller MSEs compared to the periodogram. Moreover, note that 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.

TC3. Real-world graphs with synthetic signals: We now demonstrate how the tools developed in this paper can be useful in practice through a few real-world experiments. The first one

deals with source identification in opinion formation dynamics. We consider the social network of Zachary's karate club [38] represented by a graph $\mathcal{G}$ consisting of 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 GSO $\mathbf{S} = \mathbf{I} - \alpha\mathbf{L}$ 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 $\boldsymbol{\beta}$. Observe that the opinion profiles $\mathbf{x}_r$ are typically dense since the degree of the filters considered is in the order of the diameter of the graph. 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 $\boldsymbol{\beta}$ of the generating filter. We do this by solving (38) via least squares. Second, given the set of coefficients $\boldsymbol{\beta}$, 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 least squares problem to promote sparsity in the input. In Fig. 2a (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 $\boldsymbol{\beta}$ 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. Note that the elements $[\mathbf{x}_r]_i = 0$ remain unperturbed, which is equivalent to assuming that we can easily spot people that have not heard about the rumor. 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.

TC4. Real-world signals: Three experiments with real-data are presented. The first one considers grayscale images of different individuals, the second one human brain signals measuring

(a) Ground truth   (b) Noisy   (c) Gauss 2D   (d) Low-pass GF (e) Wiener GF (f) Ground truth   (g) Noisy   (h) No-glasses GF (i) Glasses GF
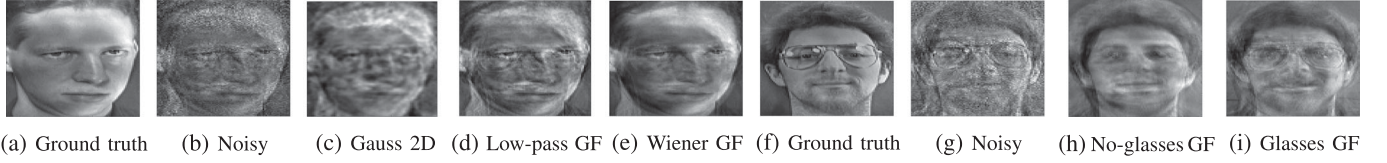
Fig. 3. The images (a-e) illustrate how, by leveraging the concept of graph stationarity, the Wiener graph filter (GF) in (e) achieves a better performance than a regular low-pass GF in (d) and both of them outperform the classical 2D Gaussian filter. The images (f-i) illustrate the difference in recovering the original image (f) from the noisy version (g) when assuming that the image is stationary in a shift build with faces without glasses (h) and with glasses (i).

activity at different regions of the brain, and the last one cellular flow cytometry data.

Face images: We consider a set of 100 grayscale images $\{\mathbf{x}_r\}_{r=1}^{100}$ corresponding to 10 face pictures of 10 different people[4]. Formally, every image is represented by a vector $\mathbf{x}_r \in \mathbb{R}^{10304}$ where the entries correspond to grayscale values of pixels, normalized to have zero mean. We consider the images $\mathbf{x}_r$ to be realizations of a graph process which, by definition, is stationary in the shift given by its covariance, here approximated by the sample covariance $\mathbf{S} = \hat{\mathbf{C}}_x = \mathbf{V}\boldsymbol{\Lambda}_c\mathbf{V}^H$. Moreover, denote by $\mathcal{I}_j$ the set of images corresponding to faces of person j and consider the sample covariance $\hat{\mathbf{C}}_x^{(j)} = \mathbf{V}_x^{(j)}\boldsymbol{\Lambda}_c^{(j)}\mathbf{V}_x^{(j)H}$ of the subset $\{\mathbf{x}_r\}_{r\in\mathcal{I}_j}$. While matrix $\hat{\mathbf{C}}_x^{(j)}$ will be perfectly diagonalized by $\mathbf{V}_x^{(j)}$, the question is whether it will be (approximately) diagonalized by $\mathbf{V}$ or $\mathbf{V}_x^{(k)}$, where $k \neq j$. After left and right multiplying the covariance $\hat{\mathbf{C}}_x^{(j)}$ with the particular GFT, the resultant off-diagonal elements account for the cross-correlation among frequency components. Hence, invoking Property 2, we know that if the process is actually stationary in the shift that generated the GFT, those off-diagonal elements must be zero. To assess this, Figs. 2b-(top), 2b-(bottom), and 2c plot blocks of the matrices $(\mathbf{V}^{(1)H}\hat{\mathbf{C}}_x^{(1)}\mathbf{V}^{(1)})$, $(\mathbf{V}^{(2)H}\hat{\mathbf{C}}_x^{(1)}\mathbf{V}^{(2)})$, and $(\mathbf{V}^H\hat{\mathbf{C}}_x^{(1)}\mathbf{V})$, respectively. The pictures reveal that while $\hat{\mathbf{C}}_x^{(1)}$ is *not* diagonalized by $\mathbf{V}_x^{(2)}$ (many of the off-diagonal elements are nonzero), it is *approximately* diagonalized by the principal component basis $\mathbf{V}$. To quantify this more rigorously, we consider the metric $\theta(\mathbf{V}, \hat{\mathbf{C}}) := \|\mathrm{diag}(\mathrm{diag}(\mathbf{V}^H\hat{\mathbf{C}}\mathbf{V}))\|_F / \|\mathbf{V}^H\hat{\mathbf{C}}\mathbf{V}\|_F$, where only the N elements of the diagonal are considered in the numerator. Then, for the example in Fig. 2c, we find that $\theta(\mathbf{V}, \hat{\mathbf{C}}_x^{(1)}) = 50\%$, confirming that the relative weight of the elements in the diagonal is high and, hence, that the process generating face images of subject j is approximately stationary in $\mathbf{S} = \hat{\mathbf{C}}_x$. To illustrate why this is useful, we describe next a couple of experiments where traditional tools for processing stationary time-signals are applied to the graph signals at hand. Our goal is to explore potential application domains for the definitions and tools introduced in this paper. In particular, we use the graph counterpart of the Wiener filter [8] to denoise face images. More precisely, given a noisy version $\mathbf{y}$ of an image of individual j, we model $\mathbf{y}$ as stationary in $\mathbf{S} = \hat{\mathbf{C}}_x$ and obtain the filtered version $\mathbf{y}^{\mathrm{Wie}}$ where $\tilde{y}_k^{\mathrm{Wie}} = p_k/(p_k + \omega_k^2)\,\tilde{y}_k$, with ω_k^2 denoting the noise power at frequency k. In the experiments we set the PSD as $\mathbf{p} = \hat{\mathbf{p}}_{\mathrm{cg}}$, i.e., the eigenvalues of the sample covariance $\hat{\mathbf{C}}_x$. For comparison purposes, we also consider a Gaussian 2D low-pass filter with unitary variance

and a 'low-pass' graph filter, where we keep unchanged the frequency components $\tilde{y}_k$ that correspond to active frequencies, i.e., all k such that $p_k > 0$, and discard the rest. The results are illustrated in Fig. 3 where inspection reveals that the Gaussian filter (which exploits neither $\mathbf{V}$ nor $\mathbf{p}$) yields the worst reconstruction performance, while the Wiener filter (which exploits both) achieves a better reconstruction than the low-pass graph filter (which exploits only $\mathbf{V}$). The second example also deals with image smoothing. In particular, we use the face with glasses in Fig. 3f and consider two different GSOs $\mathbf{S}_1$ and $\mathbf{S}_2$ corresponding to two different sample covariances that use the same number of images. The shift $\mathbf{S}_1$ is constructed based on images of people not wearing glasses whereas most of the images used in generating $\mathbf{S}_2$ contained glasses. If we implement the Wiener filter to denoise Fig. 3g using as shifts the sample covariances corresponding to $\mathbf{S}_1$ and $\mathbf{S}_2$ and as PSD their corresponding correlograms, we observe that the second filter is able to recover the glasses (Fig. 3i) while the first one fails (Fig. 3h), entailing a poorer recovery. This can be interpreted as a manifestation of the original image being closer to stationary in $\mathbf{S}_2$ than in $\mathbf{S}_1$.

Brain signals: The ensuing experiment deals with brain signals. While in the setup with face images the supporting graph was unknown and estimated as the sample covariance, in this case the graph is given. In particular, the goal is to analyze brain functional signals using as support the so-called *functional* and *structural* brain networks [4], [37]. To that end, we use the 120 brain functional signals $\{\mathbf{x}_r\}_{r=1}^{120}$ and shifts $\mathbf{S}_{\mathrm{fun}}$ and $\mathbf{S}_{\mathrm{str}}$ provided in [37]. As done before, the goal is to assess if the signals are stationary in the provided graphs. The results indicate that $\{\mathbf{x}_r\}_{r=1}^{120}$ are approximately stationary in $\mathbf{S}_{\mathrm{fun}}$. In particular, we have that $\theta(\mathbf{V}_{\mathrm{fun}}, \hat{\mathbf{C}}_x) = 99\%$, so that the contribution to the norm of the off-diagonal elements of $(\mathbf{V}_{\mathrm{fun}}^H\hat{\mathbf{C}}_x\mathbf{V}_{\mathrm{fun}})$ is negligible. More importantly and interestingly, they also indicate that $\{\mathbf{x}_r\}_{r=1}^{120}$ are approximately stationary in $\mathbf{S}_{\mathrm{str}}$, with the matrix $(\mathbf{V}_{\mathrm{str}}^H\hat{\mathbf{C}}_x\mathbf{V}_{\mathrm{str}})$ being shown in Fig. 2d. In fact, we have that $\theta(\mathbf{V}_{\mathrm{str}}, \hat{\mathbf{C}}_x) = 67\%$. This is more surprising since the construction of $\mathbf{S}_{\mathrm{str}}$ is agnostic to $\{\mathbf{x}_r\}_{r=1}^{120}$. It also points out that functional and structural networks, often viewed as separate entities, are clearly related. Such an observation can be exploited in multiple tasks. One example is to improve the algorithms to identify the underlying networks. Our results for that particular application, which build on network topology identification algorithms that view the set of available observations as stationary in the network to be identified [16], are promising and will soon be reported in a separate contribution.

Flow cytometry: A common approach to identify the graph structure associated with an observed dataset is to construct a *regularized* estimator of the precision matrix $\boldsymbol{\Theta}$; see, e.g.,

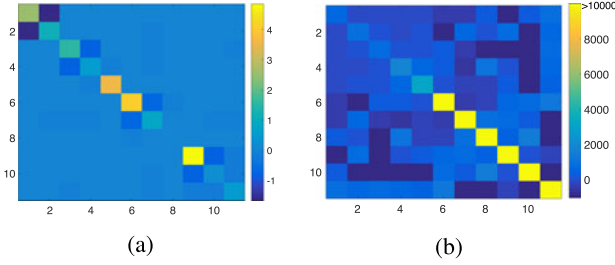[4]http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

Fig. 4. (a) Shift operator $\mathbf{S}_{10^3}$ recovered from the implementation of graphical lasso. (b) Diagonalization of $\hat{\mathbf{C}}_x$ with $\mathbf{V}_{10^3}$. The graph process is approximately stationary in $\mathbf{S}_{10^3}$.

graphical lasso [39]. Given that, by definition, a process is stationary in the graph given by its precision matrix (Example 3 in Section III-B), this same process is expected to be *approximately* stationary in the graph obtained via graphical lasso. In order to illustrate this property, we consider a flow cytometry dataset on $N = 11$ proteins and $R = 7466$ observations corresponding to different cells [40] (dataset also used in [39]). More precisely, we are given a set $\{\mathbf{x}_r\}_{r=1}^R$ of observations where $x_r \in \mathbb{R}^{11}$ represents the levels of 11 different proteins in an individual cell r. Denoting by $\hat{\mathbf{C}}_x$ the sample covariance of the observations, graphical lasso estimates the associated graph shift $\mathbf{S}_\eta$ as

$$\mathbf{S}_\eta = \underset{\mathbf{S} \in \mathcal{S}_+^N}{\operatorname{argmin}} \; -\log|\mathbf{S}| + \operatorname{tr}(\hat{\mathbf{C}}_x \mathbf{S}) + \eta\|\mathbf{S}\|_1, \quad (49)$$

where $\mathcal{S}_+^N$ denotes the cone of $N \times N$ positive semi-definite matrices and $\eta \geq 0$ parametrizes the level of regularization in estimating $\mathbf{S}_\eta$. In Fig. 4a we portray the shift operator recovered when $\eta = 10^3$. Denoting the eigendecomposition of the shift as $\mathbf{S}_\eta = \mathbf{V}_\eta \mathbf{\Lambda}_\eta \mathbf{V}_\eta^H$, the form of $\mathbf{V}_\eta^H \hat{\mathbf{C}}_x \mathbf{V}_\eta$ determines if the process is stationary on $\mathbf{S}_\eta$. One can first check that $\mathbf{V}_0^H \hat{\mathbf{C}}_x \mathbf{V}_0$ is a diagonal matrix, implying that the observed network process is stationary in $\mathbf{S}_0$. This is not surprising since $\mathbf{S}_0 = \hat{\mathbf{C}}_x^{-1}$, i.e., graphical lasso with no regularization returns the inverse sample covariance. Interestingly, for $\eta = 10^3$ we have that $\theta(\mathbf{V}_{10^3}^H, \hat{\mathbf{C}}_x) = 99\%$, demonstrating that the observed graph process is approximately stationary in $\mathbf{S}_{10^3}$, a graph that resembles the known dependencies between proteins [39], [40]; see Fig. 4b. Given that graphical lasso and related methods are widely used to unveil graph structures, the framework of stationary graph processes here developed is relevant to a broad range of applications.

VII. CONCLUSION

Different equivalent ways of generalizing the notion of stationarity to graph processes associated with a normal shift operator were studied. Given that graph stationary processes were shown to be diagonalized by the graph Fourier basis, the corresponding power spectral density was introduced and several estimation methods were studied in detail. We first generalized nonparametric methods including periodograms, window-based average periodograms, and filter banks. Their performance was analyzed, comparisons with the traditional time domain schemes were established and open issues, such as how to group nodes and frequencies in a graph, were identified. We then focused on parametric estimation, where we examined MA, AR, and

ARMA processes. We not only showed that those processes are useful to model linear diffusion dynamics, but also identified particular scenarios where the optimal parameter estimation problem is tractable.

APPENDICES

A. Proof of Prop. 2

To prove that the estimate is unbiased, we use the fact that $\hat{\mathbf{p}}_{\mathrm{pg}} = \hat{\mathbf{p}}_{\mathrm{cg}}$ combined with (13) to conclude that $\mathbb{E}[\hat{\mathbf{p}}_{\mathrm{pg}}] = \operatorname{diag}[\mathbf{V}^H R^{-1}(\sum_{r=1}^R \mathbb{E}[\mathbf{x}_r \mathbf{x}_r^H])\mathbf{V}] = \operatorname{diag}[\mathbf{V}^H R^{-1} R \mathbf{C}_x \mathbf{V}] = \mathbf{p}$. Hence, $\mathbf{b}_{\mathrm{pg}} = \mathbb{E}[\hat{\mathbf{p}}_{\mathrm{pg}}] - \mathbf{p} = \mathbf{0}$.

To compute the covariance, start by writing $\mathbf{\Sigma}_{\mathrm{pg}} = \mathbb{E}\left[\hat{\mathbf{p}}_{\mathrm{pg}} \hat{\mathbf{p}}_{\mathrm{pg}}^H\right] - \mathbf{p}\mathbf{p}^H$. We may expand the leftmost term as

$$\mathbb{E}\left[\hat{\mathbf{p}}_{\mathrm{pg}} \hat{\mathbf{p}}_{\mathrm{pg}}^H\right] = \mathbb{E}\left[\operatorname{diag}(\mathbf{V}^H \hat{\mathbf{C}}_x \mathbf{V})\operatorname{diag}(\mathbf{V}^H \hat{\mathbf{C}}_x \mathbf{V})^H\right] \quad (50)$$

$$= \frac{1}{R^2}\sum_{r=1}^R \sum_{r'=1}^R \mathbb{E}\left[\operatorname{diag}(\mathbf{V}^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{V})\operatorname{diag}(\mathbf{V}^H \mathbf{x}_{r'} \mathbf{x}_{r'}^H \mathbf{V})^H\right].$$

We may split the above summation into the terms where $r \neq r'$ and those where $r = r'$. For the former case, since $\mathbf{x}_r$ is assumed to be independent from $\mathbf{x}_{r'}$, we have that

$$\mathbb{E}\left[\operatorname{diag}(\mathbf{V}^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{V})\right]\mathbb{E}\left[\operatorname{diag}(\mathbf{V}^H \mathbf{x}_{r'} \mathbf{x}_{r'}^H \mathbf{V})^H\right] = \mathbf{p}\mathbf{p}^H. \quad (51)$$

By contrast, for the case where $r = r'$ we undertake an elementwise analysis of the elements in the expected value in (50). Notice that if we denote by $\mathbf{v}_i$ the ith column of $\mathbf{V}$ we have that $[\operatorname{diag}(\mathbf{V}^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{V})]_i = \mathbf{v}_i^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{v}_i$. Replacing this expression in (50), setting $r = r'$, and using the formula for the quartic form of a Gaussian, see e.g. [41, pp. 43], we obtain that

$$\left[\mathbb{E}\left[\operatorname{diag}(\mathbf{V}^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{V})\operatorname{diag}(\mathbf{V}^H \mathbf{x}_{r'} \mathbf{x}_{r'}^H \mathbf{V})^H\right]\right]_{ij}$$

$$= \mathbb{E}\left[\mathbf{v}_i^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{v}_i \mathbf{v}_j^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{v}_j\right] = \mathbf{v}_i^H \mathbb{E}\left[\mathbf{x}_r \mathbf{x}_r^H \mathbf{v}_i \mathbf{v}_j^H \mathbf{x}_r \mathbf{x}_r^H\right]\mathbf{v}_j$$

$$= p_i p_j + \mathbf{v}_i^H \mathbf{C}_x \mathbf{v}_j^* \mathbf{v}_i^T \mathbf{C}_x \mathbf{v}_j + \mathbf{v}_j^H \mathbf{C}_x \mathbf{v}_i \mathbf{v}_i^H \mathbf{C}_x \mathbf{v}_j. \quad (52)$$

When $\mathbf{S}$ is symmetric, hence $\mathbf{V}$ is real, the above expression can be further simplified to obtain

$$\mathbb{E}\left[\operatorname{diag}(\mathbf{V}^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{V})\operatorname{diag}(\mathbf{V}^H \mathbf{x}_r \mathbf{x}_r^H \mathbf{V})^H\right]$$

$$= 2\operatorname{diag}^2(\mathbf{p}) + \mathbf{p}\mathbf{p}^H. \quad (53)$$

Upon substituting (51) and (53) into (50), the result follows.

B. Proof of Prop. 4

From the definition of $\hat{\mathbf{p}}_{\mathcal{W}}$ in (20), it follows that

$$\hat{\mathbf{p}}_{\mathcal{W}} = \frac{1}{M}\sum_{m=1}^M \operatorname{diag}(\mathbf{V}^H \operatorname{diag}(\mathbf{w}_m)\mathbf{V}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H \mathbf{V}^H \operatorname{diag}(\mathbf{w}_m^*)\mathbf{V}). \quad (54)$$

Thus, using the definition of $\tilde{\mathbf{W}}_m$, we may write

$$\mathbb{E}[\hat{\mathbf{p}}_{\mathcal{W}}] = \frac{1}{M}\sum_{m=1}^M \operatorname{diag}(\tilde{\mathbf{W}}_m \mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right]\tilde{\mathbf{W}}_m^H). \quad (55)$$

Leveraging the fact that $\mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right] = \operatorname{diag}(\mathbf{p})$ is diagonal and recalling that $\tilde{\mathbf{W}}_{mm} = \tilde{\mathbf{W}}_m \circ \tilde{\mathbf{W}}_m^*$, the result in (22) follows.

In order to show (23), notice that only the diagonal elements of $\boldsymbol{\Sigma}_{\mathcal{W}}$ are needed. Each of them can be found as

$$[\boldsymbol{\Sigma}_{\mathcal{W}}]_{k,k} = \mathbb{E}\left[[\hat{\mathbf{p}}_{\mathcal{W}}]_k[\hat{\mathbf{p}}_{\mathcal{W}}]_k\right] - \mathbb{E}\left[[\hat{\mathbf{p}}_{\mathcal{W}}]_k\right]^2. \qquad (56)$$

Rewriting $\hat{\mathbf{p}}_{\mathcal{W}}$ as a sum across the M windows, it holds that

$$\mathbb{E}\left[[\hat{\mathbf{p}}_{\mathcal{W}}]_k[\hat{\mathbf{p}}_{\mathcal{W}}]_k\right] = \mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M}[|\tilde{\mathbf{W}}_m\tilde{\mathbf{x}}|^2]_k \frac{1}{M}\sum_{m'=1}^{M}[|\tilde{\mathbf{W}}_{m'}\tilde{\mathbf{x}}|^2]_k\right]$$

$$= \frac{1}{M^2}\sum_{m=1,m'=1}^{M}\mathbb{E}\left[[|\tilde{\mathbf{W}}_m\tilde{\mathbf{x}}|^2]_k[|\tilde{\mathbf{W}}_{m'}\tilde{\mathbf{x}}|^2]_k\right]. \qquad (57)$$

Denoting the kth row of $\tilde{\mathbf{W}}_m$ by $\tilde{\mathbf{w}}_{k|m}^T$, the term $[|\hat{\mathbf{W}}_m\tilde{\mathbf{x}}|^2]_k$ can be written as $\tilde{\mathbf{w}}_{k|m}^T\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\tilde{\mathbf{w}}_{k|m}^*$. Consequently, we have

$$\mathbb{E}\left[[|\hat{\mathbf{W}}_m\tilde{\mathbf{x}}|^2]_k[|\hat{\mathbf{W}}_{m'}\tilde{\mathbf{x}}|^2]_k\right] = \mathbb{E}\left[\tilde{\mathbf{w}}_{k|m}^T\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\tilde{\mathbf{w}}_{k|m}^*\tilde{\mathbf{w}}_{k|m'}^T\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\tilde{\mathbf{w}}_{k|m'}^*\right]$$

$$= \tilde{\mathbf{w}}_{k|m}^T\mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\tilde{\mathbf{w}}_{k|m}^*\tilde{\mathbf{w}}_{k|m'}^T\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right]\tilde{\mathbf{w}}_{k|m'}^*, \qquad (58)$$

where the middle factor is a quartic form of a Gaussian with covariance $\mathrm{diag}(\mathbf{p})$ (cf. Property 2). Solving this fourth moment, see e.g. [41, pp. 43], it follows that

$$\mathbb{E}\left[[|\hat{\mathbf{W}}_m\tilde{\mathbf{x}}|^2]_k[|\hat{\mathbf{W}}_{m'}\tilde{\mathbf{x}}|^2]_k\right] \qquad (59)$$

$$= |\tilde{\mathbf{w}}_{k|m}^T|^2\mathbf{p}|\tilde{\mathbf{w}}_{k|m'}^T|^2\mathbf{p} + |\tilde{\mathbf{w}}_{k|m}^T\mathrm{diag}(\mathbf{p})\tilde{\mathbf{w}}_{k|m'}^*|^2$$

$$+ \tilde{\mathbf{w}}_{k|m}^T\mathrm{diag}(\mathbf{p})\mathbf{V}^H\mathbf{V}^*\tilde{\mathbf{w}}_{k|m'}^*\tilde{\mathbf{w}}_{k|m}^H\mathbf{V}^T\mathbf{V}\mathrm{diag}(\mathbf{p})\tilde{\mathbf{w}}_{k|m'}^*.$$

When $\mathbf{S}$ is symmetric, hence $\mathbf{V}$ is real, the third summand in (59) is equal to the second one. Thus, substituting first (59) into (57) and then, (57) into (56) yields

$$[\boldsymbol{\Sigma}_{\mathcal{W}}]_{k,k} = \frac{2}{M^2}\sum_{m=1,m'=1}^{M}|\tilde{\mathbf{w}}_{k|m}^T\mathrm{diag}(\mathbf{p})\tilde{\mathbf{w}}_{k|m'}^*|^2. \qquad (60)$$

Since $\tilde{\mathbf{W}}_{mm'} = \tilde{\mathbf{W}}_m \circ \tilde{\mathbf{W}}_{m'}^*$, it follows that $|\tilde{\mathbf{w}}_{k|m}^T\mathrm{diag}(\mathbf{p})\tilde{\mathbf{w}}_{k|m'}^*| = [\tilde{\mathbf{W}}_{mm'}\mathbf{p}]_k$, thus

$$[\boldsymbol{\Sigma}_{\mathcal{W}}]_{k,k} = \frac{2}{M^2}\sum_{m=1,m'=1}^{M}|[\tilde{\mathbf{W}}_{mm'}\mathbf{p}]_k|^2. \qquad (61)$$

Finally, using (61) to write $\mathrm{tr}[\boldsymbol{\Sigma}_{\mathcal{W}}] = \sum_{k=1}^{N}[\boldsymbol{\Sigma}_{\mathcal{W}}]_{k,k}$, we obtain (23) and the proof concludes.

C. Proof of Prop. 6

Rewriting the norm in (27) as the trace of the corresponding outer product we get that

$$\mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right] = \mathrm{tr}[\mathrm{diag}(\tilde{\mathbf{q}}_k)\mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right]\mathrm{diag}(\tilde{\mathbf{q}}_k^*)]. \qquad (62)$$

Expression (28) follows from replacing $\mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right]$ by $\mathrm{diag}(\mathbf{p})$ and noting that the trace of the product of diagonal matrices equals the sum of the entrywise products of the diagonals.

To prove (29), we first find $\mathbb{E}[\hat{p}_{\tilde{\mathbf{q}}_k}\hat{p}_{\tilde{\mathbf{q}}_k}]$. Since $\hat{p}_{\tilde{\mathbf{q}}_k} = \mathrm{tr}[\mathrm{diag}(\tilde{\mathbf{q}}_k)\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\mathrm{diag}(\tilde{\mathbf{q}}_k)^H] = \mathrm{tr}[\tilde{\mathbf{x}}^H\mathrm{diag}(|\tilde{\mathbf{q}}_k|^2)\tilde{\mathbf{x}}]$ then, since the argument of the trace is a scalar, we can write

$$\mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\hat{p}_{\tilde{\mathbf{q}}_k}\right] = \mathbb{E}\left[\mathrm{tr}[\tilde{\mathbf{x}}^H\mathrm{diag}(|\tilde{\mathbf{q}}_k|^2)\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\mathrm{diag}(|\tilde{\mathbf{q}}_k|^2)\tilde{\mathbf{x}}]\right]$$

$$= \mathrm{tr}[\mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\mathrm{diag}(|\tilde{\mathbf{q}}_k|^2)\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right]\mathrm{diag}(|\tilde{\mathbf{q}}_k|^2)], \qquad (63)$$

where we have a quartic form of a Gaussian. From the expression of this quartic form, see e.g. [41, pp. 43], we obtain

$$\mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\hat{p}_{\tilde{\mathbf{q}}_k}\right] = \left\|\mathrm{diag}\left(|\tilde{\mathbf{q}}_k|^2\right)\mathbf{p}\right\|_2^2 + ((|\tilde{\mathbf{q}}_k|^2)^T\mathbf{p})^2 \qquad (64)$$

$$+ \mathrm{tr}[\mathrm{diag}(\mathbf{p})\mathbf{V}^H\mathbf{V}^*\mathrm{diag}(|\tilde{\mathbf{q}}_k|^2)\mathbf{V}^T\mathbf{V}\mathrm{diag}(\mathbf{p})\mathrm{diag}(|\tilde{\mathbf{q}}_k|^2)].$$

When $\mathbf{S}$ is symmetric, hence $\mathbf{V}$ is real, the trace in (64) is equal to $\left\|\mathrm{diag}\left(|\tilde{\mathbf{q}}_k|^2\right)\mathbf{p}\right\|_2^2$ and, since $\mathrm{var}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right] = \mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\hat{p}_{\tilde{\mathbf{q}}_k}\right] - \mathbb{E}\left[\hat{p}_{\tilde{\mathbf{q}}_k}\right]^2$, using (64) and (28), the expression (29) follows.

REFERENCES

[1] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes: Nonparametric power spectral density estimation," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, Jul. 2016, pp. 1–5.

[2] D. I. 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 Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[3] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[4] W. Huang *et al.*, "Graph frequency analysis of brain signals," *J. Sel. Topics Signal Process.*, vol. 10, no. 7, pp. 1189–1203, Oct. 2016.

[5] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, "Blind identification of graph filters," *IEEE Trans. Signal Process.*, vol. 65, no. 5, pp. 1146–1159, Mar. 2017.

[6] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning with Laplacian constraints: Modeling attractive Gaussian Markov random fields," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 1470–1474.

[7] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[8] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York, NY, USA: Wiley, 2009.

[9] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall 2005.

[10] B. Girault, "Stationary graph signals using an isometric graph translation," in *Proc. Eur. Signal Process. Conf.*, 2015, pp. 1516–1520.

[11] A. Gavili and X.-P. Zhang, "On the shift operator, graph frequency and optimal filtering in graph signal processing," arXiv: 1511.03512v4, 2015.

[12] B. Girault, P. Gonçalves, and E. Fleury, "Translation on graphs: An isometric shift operator," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2416–2420, Dec. 2015.

[13] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," arXiv: 1601.02522v2, 2016.

[14] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmonic Anal.*, vol. 40, no. 2, pp. 260–291, 2016.

[15] S. Segarra, A. G. Marques, and A. Ribeiro, "Distributed implementation of linear network operators using graph filters," in *Proc. 53rd Allerton Conf. Commun., Control, Comput.*, Univ. Illinois at Urbana-Champaign, Monticello, IL, USA, Sep. 30–Oct. 2, 2015, pp. 1406–1413.

[16] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from spectral templates," in *Proc. IEEE Int. Workshop Stat. Signal Process.*, Jun. 2016, pp. 1–5.

[17] A. Sandryhaila, S. Kar, and J. M. F. Moura, "Finite-time distributed consensus through graph filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2014, pp. 1080–1084.

[18] 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. 16, pp. 4363–4378, Aug. 2016.

[19] C. Godsil and G. Royle, *Algebraic Graph Theory* (Graduate Texts in Mathematics). New York, NY, USA: Springer-Verlag, 2001, vol. 207.

[20] 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.

[21] K. Conrad, *The Minimal Polynomial and Some Applications*. Dept. of Mathematics. Univ. of Connecticut, Storrs, CT, USA.

[22] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2077–2092, Apr. 2017.

[23] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[24] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 40–52, Mar. 2014.

[25] K. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, 2003.

[26] T. Nakano, M. Moore, F. Wei, A. Vasilakos, and J. Shuai, "Molecular communication and networking: Opportunities and challenges," *IEEE Trans. Nanobiosci.*, vol. 11, no. 2, pp. 135–148, Jun. 2012.

[27] S. P. Chepuri and G. Leus, "Subsampling for graph power spectrum estimation," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, Jul. 2016, pp. 1–5.

[28] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, Aug. 2017.

[29] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.

[30] 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, Apr. 2015.

[31] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 467–483, Sept. 2017.

[32] B. Bollobás, *Random Graphs*. New York, NY, USA: Springer, 1998.

[33] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. New York, NY, USA: Springer, 2009.

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

[35] A. Jain and R. C. Dubes, *Algorithms for Clustering Data* (ser. Prentice Hall Advanced Reference Series). Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.

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

[37] J. D. Medaglia, W. Huang, E. A. Karuza, S. L. Thompson-Schill, A. Ribeiro, and D. S. Bassett, "Functional alignment with anatomical networks is associated with cognitive flexibility," *Nature*, Nov. 2016. arXiv: 1611.08751v1.

[38] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. pp. 452–473, 1977.

[39] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[40] K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, and G. P. Nolan, "Causal protein-signaling networks derived from multiparameter single-cell data," *Science*, vol. 308, no. 5721, pp. 523–529, 2005.

[41] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Kongens Lyngby, Denmark: Technical Univ. Denmark, Nov. 2012.

Santiago Segarra (M'16) received the B.Sc. (highest Hons.) degree in industrial engineering (Valedictorian) from the Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina, in 2011, the M.Sc. degree in electrical engineering and the Ph.D. degree in electrical and systems engineering both from the University of Pennsylvania (Penn), Philadelphia, PA, USA, in 2014 and 2016, respectively. Since 2016, he has been working as a Postdoctoral Research Associate with the Institute for Data, Systems, and Society at the Massachusetts Institute of Technology. His research interests include network theory, data analysis, machine learning, and graph signal processing. He received the ITBA's 2011 Best Undergraduate Thesis Award in industrial engineering, the 2011 Outstanding Graduate Award granted by the National Academy of Engineering of Argentina, the Best Student Paper Awards at the 2015 Asilomar Conference, and the 2016 IEEE Statistical Signal Processing Workshop, and the Best Paper Award at the 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop.

Geert Leus (M'01–SM'05–F'13) received the M.Sc. and Ph.D. degrees in electrical engineering from the KU Leuven, Leuven, Belgium, in June 1996 and May 2000, respectively. He is currently an "Antoni van Leeuwenhoek" Full Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands. His research interests include the broad area of signal processing. He received a 2002 IEEE Signal Processing Society Young Author Best Paper Award and a 2005 IEEE Signal Processing Society Best Paper Award. He is a Fellow of IEEE and EURASIP. He was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, a Member-at-Large to the Board of Governors of the IEEE Signal Processing Society, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, THE IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, THE IEEE SIGNAL PROCESSING LETTERS, and the *EURASIP Journal on Advances in Signal Processing*. He is currently a member of the IEEE Sensor Array and Multichannel Technical Committee and serves as the Editor-in-Chief of the *EURASIP Journal on Advances in Signal Processing*.

Antonio G. Marques (SM'13) received the telecommunications engineering degree and the Doctorate degree, both with highest honors, from the Carlos III University of Madrid, Madrid, Spain, in 2002 and 2007, respectively. In 2007, he became a faculty in the Department of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain, where he currently develops his research and teaching activities as an Associate Professor. From 2005 to 2015, he held different visiting positions at the University of Minnesota, Minneapolis, MN, USA. In 2015 and 2016, he was a Visitor Scholar in the University of Pennsylvania, Philadelphia, PA, USA. His research interests lie in the areas of signal processing, networking and communications. His current research focuses on stochastic optimization of wireless and power networks, signal processing for graphs, and nonlinear network optimization. He has served the IEEE in a number of posts (currently, he is an Associate Editor of the SIGNAL PROCESSING LETTERS and a member of the IEEE Signal Processing Theory and Methods Technical Committee) and his work has been awarded in several conferences and workshops.

Alejandro Ribeiro (M'07) received the B.Sc. degree in electrical engineering from the Universidad de la República, Oriental del Uruguay, Montevideo, Uruguay, in 1998 and the M.Sc. and Ph.D. degrees in electrical engineering from the Department of Electrical and Computer Engineering, the University of Minnesota, Minneapolis, Minneapolis, MN, USA, in 2005 and 2007. From 1998 to 2003, he was a member of the technical staff at Bellsouth Montevideo. After M.Sc. and Ph.D studies, in 2008 he joined the University of Pennsylvania (Penn), Philadelphia, PA, USA, where he is currently the Rosenbluth Associate Professor in the Department of Electrical and Systems Engineering. His research interests include the applications of statistical signal processing to the study of networks and networked phenomena. His focus is on structured representations of networked data structures, graph signal processing, network optimization, robot teams, and networked control. He received the 2014 O. Hugo Schuck Best Paper Award, the 2012 S. Reid Warren, Jr. Award presented by Penn's undergraduate student body for outstanding teaching, the NSF CAREER Award in 2010, and paper awards at the 2016 SSP Workshop, 2016 SAM Workshop, 2015 Asilomar SSC Conference, ACC 2013, ICASSP 2006, and ICASSP 2005. He is a Fulbright scholar and a Penn Fellow.