

# Advances in Distributed Graph Filtering

Mario Coutino , *Student Member, IEEE*, Elvin Isufi , *Student Member, IEEE*, and Geert Leus , *Fellow, IEEE*

**Abstract**—Graph filters are one of the core tools in graph signal processing. A central aspect of them is their direct distributed implementation. However, the filtering performance is often traded with distributed communication and computational savings. To improve this tradeoff, this paper generalizes state-of-the-art distributed graph filters to filters where every node weights the signal of its neighbors with different values while keeping the aggregation operation linear. This new implementation, labeled as edge-variant graph filter, yields a significant reduction in terms of communication rounds while preserving the approximation accuracy. In addition, we characterize a subset of shift-invariant graph filters that can be described with edge-variant recursions. By using a low-dimensional parameterization, these shift-invariant filters provide new insights in approximating linear graph spectral operators through the succession and composition of local operators, i.e., fixed support matrices. A set of numerical results shows the benefits of the edge-variant graph filters over current methods and illustrates their potential to a wider range of applications than graph filtering.

**Index Terms**—Consensus, distributed beamforming, distributed signal processing, edge-variant graph filters, FIR, IIR, ARMA, graph filters, graph signal processing.

## I. INTRODUCTION

**F**ILTERING is one of the core operations in signal processing. The necessity to process large amounts of data defined over non-traditional domains characterized by a graph triggers advanced signal processing of the complex data relations embedded in that graph. Examples of the latter include biological, social, and transportation network data. The field of graph signal processing (GSP) [2]–[4] has been established to incorporate the underlying structure in the processing techniques.

Through a formal definition of the graph Fourier transform (GFT), harmonic analysis tools employed for filtering in traditional signal processing have been adapted to deal with signals defined over graphs [5]–[11]. Similarly to time-domain filtering, graph filters manipulate the signal by selectively amplifying/attenuating its graph Fourier coefficients. Graph filters have seen use in applications such as signal analysis [12], [13], classification [14], [15], reconstruction [7], [16], [17], denoising [8],

Manuscript received July 8, 2018; revised November 19, 2018 and January 23, 2019; accepted February 28, 2019. Date of publication March 13, 2019; date of current version April 1, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yuichi Tanaka. This work was supported by the ASPIRE project (project 14926 within the STW OTP programme), financed by the Netherlands Organization for Scientific Research (NWO). The work of M. Coutino was supported by CONACYT. This paper was presented in part at the IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Curacao, Netherlands Antilles, December 2017 [1]. (*Corresponding author: Mario Coutino.*)

The authors are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology 2826 CD Delft, The Netherlands (e-mail: m.a.coutinominguez@tudelft.nl; e.isufi-1@tudelft.nl; g.j.t.leus@tudelft.nl).

Digital Object Identifier 10.1109/TSP.2019.2904925

[18]–[20] and clustering [21]. Furthermore, they are the central block in graph filterbanks [22], [23], wavelets [24], and convolutional neural networks [25], [26].

Distributed implementations of graph filters emerged as a way to deal with the ubiquity of big data applications and to improve the scalability of computation. By allowing nodes to exchange only local information, finite impulse response (FIR) [4], [6], [9] and infinite impulse response (IIR) [10], [11], [27] architectures have been devised to implement a variety of responses.

However, being inspired by time-domain filters, the above implementations do not fully exploit the structure in the graph data. The successive signal aggregations are locally weighted with similar weights often leading to high orders in approximating the desired response. To overcome this challenge, this paper proposes a generalization of the distributed graph filtering concept by applying edge-based weights to the information coming from different neighbors. While the detailed contributions are provided in Section I-B, let us here highlight that the above twist yields graph filters that are flexible enough to capture complex responses with much lower complexity.

### A. Related Works

Driven by the practical need to implement a linear function with only few local operations, the works in [9], [28] propose to ease the communication and computational cost of graph filters (GF).

In [9], the polynomial graph filters (i.e., the FIR structure) are extended to graph filters with node-dependent weights. This architecture, referred to as a node-variant (NV) FIR graph filter, assigns different weights to different nodes and yields the same implementation as the classical FIR graph filter [4], [6]. The NV FIR filter addresses a broader family of linear operators that goes beyond the class of shift-invariant graph filters. However, the NV FIR filter still uses the same weight for all signals arriving at a particular node, ignoring the affinity between neighbors. As we show next, this limits the ability of the NV FIR filter to approximate the desired linear operator with very low orders.

The work in [28] introduces stochastic sparsification to reduce the cost of a distributed implementation. This method considers random edge sampling in each aggregation step to implement the filter output with a lower complexity by sacrificing the accuracy of the graph filter. In addition, although conceptually similar to this work, the filter following [28] is stochastic in nature and hence its performance guarantees only hold in expectation. Moreover, since this approach applies only to shift-invariant filters, such as the FIR [4], [6] and the IIR [10], [11], it cannot address linear operators that are not shift invariant.

Another related problem, which can be interpreted by means of graph filtering, is the multilayer sparse approximation of matrices [29]. Different from the previous two approaches, here a dense linear transform (matrix) is approximated through a sequence of sparse matrix multiplications to obtain a computational speedup. While this framework can be considered as a series of diffusion steps over a network, the support of such sparse matrices differs in each iteration. This in practice can be a limitation since it often requires information from non-adjacent nodes within an iteration. Finally, in [30] the problem of optimal subspace projection by local interactions is studied. This paper designs the weights of a network in order to achieve the fastest convergence to this kind of linear projection operators. However, this methods does not address the more general GSP setup of interest: the implementation of graph filters or more general linear operators.

### B. Paper Contributions

The main contribution of this work is the extension of the state-of-the-art graph filters to edge-variant (EV) graph filters. Due to the increased degrees of freedom (DoF), these filters allow for a complexity reduction of the distributed implementation while maintaining the approximation accuracy of current approaches. The salient points that broaden the existing literature are listed below.

- We present edge-variant architectures to implement FIR and IIR graph filtering. This framework extends the state-of-the-art graph filters by finding optimal edge weights to perform local aggregation of data. These new edge weights can differ per communication round and are generally distinct from the original weights of the graph matrix, thereby effectively reducing the number of exchanges. As for classical graph filter implementations, only local exchanges are required in each communication round, thus yielding an efficient distributed implementation. Three forms are analyzed: First, the general class of linear edge-variant FIR filters is presented and its distributed implementation is discussed. Then, the constrained edge-variant FIR graph filter is introduced, which maintains a similar distributed implementation as the general form, yet allowing a simple least-squares design. Finally, the family of edge-variant autoregressive moving average graph filters of order one (ARMA<sub>1</sub>) is treated. This new distributed IIR architecture allows a better trade-off between approximation accuracy and convergence rate compared to current IIR approaches.
- Through the definition of the filter modal response, we give a Fourier interpretation to a particular family of edge-variant graph filters. This subfamily shows a shift-invariant nature and links the filtering operation with a scaling of the graph modes (e.g., the graph shift eigenvectors). Further, we demonstrate the connection between the constrained edge-variant graph filters and per-tone filtering in traditional time-domain systems.
- Considering the general problem of approximating any linear operator through local operations, we demonstrate the applicability of our methods to problems beyond graph signal processing.

- Besides outperforming state-of-the-art graph filters in GSP tasks such as approximating a user-provided frequency response, distributed consensus, and Tikhonov denoising, we present two new applications that could be addressed distributively with the proposed edge-variant graph filters: a distributed solution to an inverse problem and distributed beamforming.

### C. Outline and Notation

This paper is organized as follows. Section II reviews the preliminaries of GSP and distributed graph filtering, and further defines the modal response of a graph filter. Section III generalizes the FIR graph filters to their edge-variant version. Here, we also introduce the shift-invariant edge-variant graph filter and characterize its graph modal response. Section IV analyzes a particular subfamily of edge-variant FIR graph filters, which enjoys a similar distributed implementation and a least-squares design strategy. In Section V, we generalize the idea of edge-variant filtering to the class of IIR graph filters. Section VI corroborates our findings with numerical results and Section VII discusses some concluding remarks.

Throughout this paper, we adopt the following notation. Scalars, vectors, matrices, and sets are denoted by lowercase letters ( $x$ ), lowercase boldface letters ( $\mathbf{x}$ ), uppercase boldface letters ( $\mathbf{X}$ ), and calligraphic letters ( $\mathcal{X}$ ), respectively.  $[\mathbf{X}]_{i,j}$  denotes the  $(i, j)$ th entry of the matrix  $\mathbf{X}$  whereas  $[\mathbf{x}]_i$  represents the  $i$ th entry of the vector  $\mathbf{x}$ .  $\mathbf{X}^\top$ ,  $\mathbf{X}^H$ , and  $\mathbf{X}^{-1}$  are respectively the transpose, the Hermitian, and inverse of  $\mathbf{X}$ . The Moore-Penrose pseudoinverse of  $\mathbf{X}$  is  $\mathbf{X}^\dagger$ . The Khatri-Rao product between  $\mathbf{X}$  and  $\mathbf{Y}$  is written as  $\mathbf{X} * \mathbf{Y}$ , while their Hadamard product as  $\mathbf{X} \odot \mathbf{Y}$ .  $\mathbf{1}$  and  $\mathbf{I}$  are the all-one vector and identity matrix of appropriate size, respectively.  $\text{vec}(\cdot)$  is the vectorization operation.  $\text{diag}(\cdot)$  refers to a diagonal matrix with its argument on the main diagonal.  $\text{null}\{\cdot\}$  and  $\text{span}\{\cdot\}$  denote the nullspace and span of their argument.  $\text{nnz}(\mathbf{X})$  and  $\text{supp}\{\mathbf{X}\}$  are the number of nonzero entries and the support of  $\mathbf{X}$ . Finally, we define the set  $[K] = \{1, 2, \dots, K\}$ .

## II. PRELIMINARIES

This section recalls the preliminary material that will be useful in the rest of the paper. It starts with the definition of the graph Fourier transform (GFT) and graph filtering. Then, two distributed recursions that implement FIR and IIR filtering operations on graphs are presented. Finally, the modal response of a graph filter is defined.

**Graph Fourier transform.** Consider an  $N$ -dimensional signal  $\mathbf{x}$  residing on the vertices of a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with  $\mathcal{V} = \{v_1, \dots, v_N\}$  the set of  $N$  vertices, i.e.,  $|\mathcal{V}| = N$ , and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  the set of  $M$  edges, i.e.,  $|\mathcal{E}| = M$ . Let  $\mathbf{W}$  be the weighted graph adjacency matrix with  $[\mathbf{W}]_{i,j} \neq 0$  if  $(v_j, v_i) \in \mathcal{E}$  and  $[\mathbf{W}]_{i,j} = 0$ , otherwise. For an undirected graph, the combinatorial graph Laplacian matrix is  $\mathbf{L} = \text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}$ . Both  $\mathbf{W}$  and  $\mathbf{L}$  are valid candidates for the so-called graph shift operator  $\mathcal{S}$ , an  $N \times N$  matrix that carries the notion of *shift* in the graph setting [2]–[5] but alternative shift operators are also applicable. Throughout this work, we assume the following

- (A.0)  $\mathbf{S}$  is diagonalizable.

This implies that the shift matrix,  $\mathbf{S}$ , is nondefective. This is a reasonable assumption as the set of diagonalizable  $N \times N$  matrices is *dense* on the set of  $N \times N$  matrices. Under (A.0) we can factor  $\mathbf{S}$  using the eigenvalue decomposition as  $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ . The GFT of  $\mathbf{x}$  is defined then as  $\hat{\mathbf{x}} = \mathbf{U}^{-1}\mathbf{x}$  and the inverse GFT is  $\mathbf{x} = \mathbf{U}\hat{\mathbf{x}}$ . Following the GSP convention, the eigenvectors  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N]$  represent the graph modes whereas the eigenvalues  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$  are referred to as the graph frequencies.

**Graph filtering.** A linear shift-invariant graph filter is an operation on the graph signal with graph frequency domain output

$$\hat{\mathbf{y}} = h(\mathbf{\Lambda})\hat{\mathbf{x}}. \quad (1)$$

Here,  $h(\mathbf{\Lambda})$  is a diagonal matrix with the filter frequency response on its diagonal. More formally, the frequency response of a graph filter is a function

$$h : \mathbb{C} \mapsto \mathbb{R}, \quad \lambda_i \rightarrow h(\lambda_i), \quad (2)$$

that assigns a particular value  $h(\lambda_i)$  to each graph frequency  $\lambda_i$ . This definition is akin to the one used in traditional signal processing. However, depending on the underlying graph topology, some shift operators might not have a simple spectrum, i.e., the multiplicity of some eigenvalues is greater than one. So, there is no one-to-one mapping between the graph frequencies  $\lambda_i$  and the graph modes  $\mathbf{u}_i$ . For this reason, at the end of this section, we will introduce the notion of graph modal response which treats the graph filters from the graph shift eigenvector perspective. Finally, by applying the inverse GFT on both sides of (1), we have the vertex domain filter output

$$\mathbf{y} = \mathbf{H}\mathbf{x}, \quad (3)$$

with  $\mathbf{H} = \mathbf{U}h(\mathbf{\Lambda})\mathbf{U}^{-1}$ .

**FIR graph filters.** A popular form of  $\mathbf{H}$  is its expression as a polynomial of the graph shift operator [4]–[6], i.e.,

$$\mathbf{H}_c \triangleq \sum_{k=0}^K \phi_k \mathbf{S}^k, \quad (4)$$

which we refer to as the *classical* FIR graph filter. It is possible to run the FIR filter (4) distributively due to the locality of  $\mathbf{S}$  [6], [9]. In particular, since  $\mathbf{S}^k \mathbf{x} = \mathbf{S}(\mathbf{S}^{k-1} \mathbf{x})$  the nodes can compute locally the  $k$ th shift of  $\mathbf{x}$  from the former  $(k-1)$ th shift. Overall, an FIR filter of order  $K$  requires  $K$  local exchanges between neighbors and amounts to a computational and communication complexity of  $\mathcal{O}(MK)$ .

To expand the possible set of operations that can be implemented distributively through FIR recursions, [9] proposed the NV FIR graph filter. These filters have the node domain form

$$\mathbf{H}_{\text{nv}} \triangleq \sum_{k=0}^K \text{diag}(\phi_k) \mathbf{S}^k, \quad (5)$$

where the vector  $\phi_k = [\phi_{k,1}, \dots, \phi_{k,N}]^T$  contains the node dependent coefficients applied at the  $k$ th shift. Note that for  $\phi_k = \phi_k \mathbf{1}$ , the NV FIR filter (5) reduces to the classical FIR filter (4). Furthermore, from (5), we can notice that the NV FIR

filter preserves also the efficient implementation of (4) since it relies on the same distributed implementation of the shift operator  $\mathbf{S}$ . Thus this filter also has a computational and communication complexity of  $\mathcal{O}(MK)$ .

In case that a linear operator  $\tilde{\mathbf{H}}$  is desired to be approximated by a matrix polynomial as in (4), the filter order  $K$  of a classical FIR filter might become large if a high accuracy is required. As the computational complexity scales with  $K$ , implementing large-order graph filters incurs high costs. The NV graph filter provides a first approach to tackle this issue. Starting from Section III, we generalize these ideas towards an *edge-variant* (EV) graph filter alternative, which due to its enhanced DoF can approximate  $\tilde{\mathbf{H}}$  with an even lower order  $K$ . Therefore, it leads to a more efficient implementation. One of the main benefits of both the NV and the EV graph filters is that they address a broader class of operators  $\tilde{\mathbf{H}}$  which do not necessarily share the eigenvectors with  $\mathbf{S}$ , such as analog network coding [9] as well as other applications that we will detail later on.

**IIR graph filters.** In [10], [11], the authors introduced an ARMA recursion on graphs to distributively implement IIR graph filtering, i.e., a filtering operation characterized by a rational frequency response. The building block of this filter is the so-called ARMA graph filter of order one (ARMA<sub>1</sub>). This filter is obtained as the steady-state of the first-order recursion

$$\mathbf{y}_t = \psi \mathbf{S} \mathbf{y}_{t-1} + \varphi \mathbf{x}, \quad (6)$$

with arbitrary  $\mathbf{y}_0$  and scalar coefficients  $\psi$  and  $\varphi$ . The operation (6) is a distributed recursion on graphs, where neighbors now exchange their former output  $\mathbf{y}_{t-1}$  rather than the input  $\mathbf{x}$ . The per-iteration complexity of such a recursion is  $\mathcal{O}(M)$ . Given that  $\psi$  satisfies the convergence conditions for (6) [11], the steady-state output of the ARMA<sub>1</sub> is

$$\begin{aligned} \mathbf{y} &\triangleq \lim_{t \rightarrow \infty} \mathbf{y}_t = \varphi \sum_{\tau=0}^{\infty} (\psi \mathbf{S})^\tau \mathbf{x} = \varphi (\mathbf{I} - \psi \mathbf{S})^{-1} \mathbf{x} \\ &\triangleq \mathbf{H}_{\text{arma}_1} \mathbf{x}. \end{aligned} \quad (7)$$

Such a filter addresses several GSP tasks including Tikhonov denoising, graph signal interpolation under a smoothness prior [11], and aggregate graph signal diffusion [31]. In Section V, we extend (6) to an edge-variant implementation with the aim to improve its convergence speed without heavily affecting the same approximation accuracy.

**Graph modal response.** Before moving to the main contributions of this work, we define next the *modal response* of a graph filter.

*Definition 1:* (Graph modal response) The modal response of a linear shift-invariant graph filter

$$\mathbf{H} = \mathbf{U} \text{diag}(h_1, \dots, h_N) \mathbf{U}^{-1}, \quad (8)$$

is defined as the function

$$h : [N] \rightarrow \mathbb{C}, \quad i \mapsto h_i,$$

where  $h_i$  is the scaling experienced by the  $i$ th graph mode.

This definition provides a notion for the scaling that the graph modes experience when a graph signal undergoes a linear shift-invariant graph filtering operation. Notice that this kind of response is equivalent to the graph frequency response (2) when the shift operator has a simple spectrum. Since this is not always the case, the modal response provides a description that resembles better the frequency response of time-domain filters. Hence, we will use this terminology in the rest of the paper.

### III. EDGE-VARIANT FIR GRAPH FILTERS

Let us assume a scenario in which each node *trusts* differently the information coming from different neighbors, e.g., a person is likely to value more the opinion of his/her partner than that of a colleague. So, it is reasonable to treat this case as a graph filter, where each node weights differently the information of its neighbors. Here, we formalize the above intuition in terms of EV FIR graph filters. First, we introduce the general form of these filters and then in Section III-B we focus on the class of shift-invariant edge-variant (SIEV) FIR graph filters. The filter design strategy is discussed in Section III-C.

#### A. General Form

Consider an extension of the above edge-dependent fusion to several diffusion steps (signal shifts) where in each shift a different set of weights is used. At the  $k$ th diffusion, node  $v_i$  weights its neighbouring node  $v_l$  with the weight  $\phi_{i,l}^{(k)}$ . Hence, in each shift  $k \in [K]$ , and for each node  $v_i \in \mathcal{V}$ , there is a set of coefficients  $\{\phi_{i,l}^{(k)}\}$  for  $l \in \mathcal{N}_{v_i}$ . Here,  $\mathcal{N}_{v_i}$  denotes the set of nodes adjacent to  $v_i$  as well as the node  $v_i$  itself, and  $K$  is the number of shifts. Mathematically, the above behavior can be written through an order- $K$  general EV FIR graph filter defined as

$$\mathbf{H}_{\text{ev}} \triangleq \Phi_1 + \Phi_2 \Phi_1 + \cdots + \Phi_K \Phi_{K-1} \cdots \Phi_1$$

$$= \sum_{k=1}^K \Phi_{k:1}, \quad (9)$$

where  $\Phi_{k:1} = \Phi_k \Phi_{k-1} \cdots \Phi_1$  and  $\Phi_j \in \mathbb{C}^{N \times N}$  is an edge-weighting matrix constructed from the coefficient set  $\{\{\phi_{1,l}^{(j)}\}, \dots, \{\phi_{N,l}^{(j)}\}\}$ , more specifically  $[\Phi_j]_{i,l} = \phi_{i,l}^{(j)}$ . The fact that  $l \in \mathcal{N}_{v_i}$  can be formalized by the following assumption.

- (A.1) Each  $\Phi_j$ ,  $j \in [K]$  shares the support with  $\mathbf{S} + \mathbf{I}$ .

Notice that this assumption possibly allows each node to use also its own information when  $\mathbf{S}$  has zero entries on its diagonal, e.g., when  $\mathbf{S} = \mathbf{W}$ . Note that definition (9) does not impose any symmetry on the coefficient matrices  $\Phi_j$ . In fact, depending on how adjacent nodes trust each other, the applied weights can be different. From this point on, assumption (A.1) will extend throughout the paper.

The above filter description can also be interpreted differently through time-varying shift operators [32], [33], where  $\Phi_j$  is the weighted, possibly directed shift operator for the  $j$ th diffusion step with the same support as  $\mathbf{S} + \mathbf{I}$ . Therefore, the general EV FIR filter accounts for signals that are generated through time-varying systems in directed subgraphs of the original graph. In this interpretation, the filter coefficient matrix only allows for *edge deletion* or a re-weighting of graph flows.

Note that recursion (9) is a distributed graph filter. For computing the output  $\mathbf{y} = \mathbf{H}_{\text{ev}} \mathbf{x}$ , each node is only required to track the following quantities:

- the shifted signal output  $\mathbf{x}^{(k)} = \Phi_k \mathbf{x}^{(k-1)}$ ,  $\mathbf{x}^{(0)} = \mathbf{x}$ ,
- the accumulator output  $\mathbf{y}^{(k)} = \mathbf{y}^{(k-1)} + \mathbf{x}^{(k)}$ ,  $\mathbf{y}^{(0)} = \mathbf{0}$ .

From the locality of  $\Phi_k$ , the output  $\mathbf{x}^{(k)}$  can be computed locally in each node by combining only neighboring data. Hence, (9) preserves the efficient distributed implementation of the classical FIR graph filter (4). The final filter output is  $\mathbf{y} = \mathbf{y}^{(K)}$  which yields a complexity of  $\mathcal{O}(MK)$ . This stems from the fact that at every step, similarly to the classical FIR, only a single sparse matrix-vector multiplication is performed with complexity  $\mathcal{O}(M)$ . Hence, the complexity scales linearly in  $K$  for a fixed number of edges  $M$ .

Before addressing the design strategy of the filter (9), in the sequel, we introduce a particular structure of EV FIR graph filters that enjoys a graph Fourier domain interpretation.

#### B. Shift-Invariant Edge-Variant Graph Filters

An important family of graph filters is that of *shift-invariant graph filters*, i.e., filters that commute with the graph shift operator  $\mathbf{S}$ . That is, given the shift  $\mathbf{S}$  and the filter matrix  $\mathbf{H}$ , the following holds

$$\mathbf{S}\mathbf{H} = \mathbf{H}\mathbf{S}. \quad (10)$$

For a non-defective shift operator  $\mathbf{S}$  and filter  $\mathbf{H}$ , i.e., the matrices accept an eigenvalue decomposition, condition (10) is equivalent to saying that the matrices  $\mathbf{S}$  and  $\mathbf{H}$  are jointly diagonalizable, or that their eigenbases coincide.

There is no reason to believe that graph filters of the form (9) are shift invariant. However, it is possible to characterize a subset of edge-variant graph filters that satisfy this property. To do so, we consider (A.1) and the following assumption on the coefficient matrices of the EV FIR graph filter:

- (A.2) Each  $\Phi_j$ ,  $j \in [K]$  is diagonalizable with the eigenbasis of  $\mathbf{S}$ .

Given the above assumption holds, we can rewrite (9) as

$$\mathbf{H}_{\text{ev}} = \sum_{k=1}^K \Phi_{k:1} = \mathbf{U} \left[ \sum_{k=1}^K \prod_{j=1}^k \Lambda_j \right] \mathbf{U}^{-1}, \quad (11)$$

where we substituted  $\Phi_j = \mathbf{U} \Lambda_j \mathbf{U}^{-1}$ . To provide a closed-form expression for the effect of such graph filters on the graph modes, we first describe the set of *fixed-support matrices that are diagonalizable with a particular eigenbasis* (i.e., matrices that meet (A.1) and (A.2)). Mathematically, this set is defined as

$$\mathcal{J}_{\mathcal{U}}^{\mathcal{A}} = \{\mathbf{A} : \mathbf{A} = \mathbf{U} \Omega \mathbf{U}^{-1}, [\text{vec}(\mathbf{A})]_i = 0, \forall i \in \mathcal{A}\}, \quad (12)$$

where  $\mathcal{A}$  is the index set defining the zero entries of  $\mathbf{S} + \mathbf{I}$  and  $\Omega$  is diagonal. The fixed-support condition in  $\mathcal{J}_{\mathcal{U}}^{\mathcal{A}}$  can be expressed in the linear system form

$$\Phi_{\mathcal{A}} \text{vec}(\mathbf{A}) = \mathbf{0}, \quad (13)$$

with  $\Phi_{\mathcal{A}} \in \{0, 1\}^{|\mathcal{A}| \times N^2}$  denoting the selection matrix whose rows are the rows of an  $N^2 \times N^2$  identity matrix indexed by the set  $\mathcal{A}$ . By leveraging the vectorization operation properties and

the knowledge of the eigenbasis of  $\mathbf{A}$ , we can rewrite (13) as

$$\Phi_{\mathcal{A}} \text{vec}(\mathbf{A}) = \Phi_{\mathcal{A}}(\mathbf{U}^{-\top} * \mathbf{U})\boldsymbol{\omega} = \mathbf{0}, \quad (14)$$

where “\*” represents the Kathri-Rao product and  $\boldsymbol{\omega} = [[\Omega]_{1,1}, [\Omega]_{2,2}, \dots, [\Omega]_{N,N}]^{\top}$  is the vector containing the eigenvalues of  $\mathbf{A}$ . From (14), we see that  $\boldsymbol{\omega}$  characterizes the intersection of the nullspace of  $\Phi_{\mathcal{A}}$  and the range of  $\mathbf{U}^{-\top} * \mathbf{U}$ . More formally,

$$\boldsymbol{\omega} \in \text{null}\{\mathbf{T}_{\mathcal{U}}^{\mathcal{A}}\}, \quad (15)$$

with  $\mathbf{T}_{\mathcal{U}}^{\mathcal{A}} = \Phi_{\mathcal{A}}(\mathbf{U}^{-\top} * \mathbf{U})$  and assuming  $d = \dim(\text{null}\{\mathbf{T}_{\mathcal{U}}^{\mathcal{A}}\})$ .

With this in place, the following proposition characterizes the matrices that belong to the set  $\mathcal{J}_{\mathcal{U}}^{\mathcal{A}}$ .

*Proposition 1:* (Graph shift nullspace property) Given an orthonormal basis  $\mathbf{U}$  and a sparsity pattern defined by the set  $\mathcal{A}$ , the matrices within the set  $\mathcal{J}_{\mathcal{U}}^{\mathcal{A}}$  are of the form  $\mathbf{A} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{-1}$  and have eigenvalues given by

$$\boldsymbol{\Omega} = \text{diag}(\mathbf{B}_{\mathcal{U}}^{\mathcal{A}}\boldsymbol{\alpha}), \quad (16)$$

where the  $N \times d$  matrix  $\mathbf{B}_{\mathcal{U}}^{\mathcal{A}}$  is a basis for the nullspace of  $\mathbf{T}_{\mathcal{U}}^{\mathcal{A}}$ , i.e.,

$$\text{span}\{\mathbf{B}_{\mathcal{U}}^{\mathcal{A}}\} = \text{null}\{\mathbf{T}_{\mathcal{U}}^{\mathcal{A}}\},$$

and  $\boldsymbol{\alpha}$  is the basis expansion coefficient vector.

*Proof:* The proof follows from (14)–(15). ■

The above result has been used for assessing the uniqueness of the graph shift operator in topology identification [34]. Here, we leverage Proposition 1 for interpreting the response of a particular class of SIEV FIR filters. Specifically, under (A.1) and (A.2) we can express each  $\Phi_j$  of (9) as

$$\Phi_j = \mathbf{U} \text{diag}(\mathbf{B}_{\mathcal{U}}^{\mathcal{A}}\boldsymbol{\alpha}_j)\mathbf{U}^{-1}, \quad (17)$$

and write our SIEV FIR filter as

$$\mathbf{H}_{\text{siev}} = \mathbf{U} \left[ \sum_{k=1}^K \prod_{j=1}^k \text{diag}(\mathbf{B}_{\mathcal{U}}^{\mathcal{A}}\boldsymbol{\alpha}_j) \right] \mathbf{U}^{-1}. \quad (18)$$

The following proposition formally characterizes the frequency interpretation of such filters in terms of the modal response.

*Proposition 2:* (Modal Response of SIEV FIR) An FIR graph filter of the form (9) satisfying (A.1) and (A.2) has  $i$ th modal response

$$h_i = \sum_{k=1}^K \prod_{j=1}^k (\mathbf{b}_{\mathcal{U},i}^{\mathcal{A}})^{\top} \boldsymbol{\alpha}_j \quad (19)$$

where  $(\mathbf{b}_{\mathcal{U},i}^{\mathcal{A}})^{\top}$  is the  $i$ th row of  $\mathbf{B}_{\mathcal{U}}^{\mathcal{A}}$ .

*Proof:* The proof follows directly from (18). ■

An interesting outcome of Proposition 2 is that the filter response is independent of the graph frequencies. This is clear from (19), where we see that the eigenvalue  $\lambda_i$  does not appear in the expression of  $h_i$ . Therefore, we can interpret SIEV FIR graph filters as *eigenvector filters*, since they act on the eigenmodes of the graph. That is, for each graph eigenmode (eigenvector)  $\mathbf{u}_i$ ,  $\mathbf{H}_{\text{siev}}$  might apply a different gain given by (19) (independent of  $\lambda_i$ ) to the component of the input signal  $\mathbf{x}$  in the direction of  $\mathbf{u}_i$ . This is in contrast to classical FIR graph filters which apply the same polynomial expression in  $\lambda_i$  to all modes  $\{\mathbf{u}_i\}_{i \in [N]}$ . As a

result, the classical FIR graph filter will always filter different graph modes with the same graph frequency in the same way.

The following section introduces methods for designing EV FIR graph filters in the node domain and SIEV FIR graph filters using the parametrization in (19).

### C. Filter Design

**General form.** Given a desired operator  $\tilde{\mathbf{H}}$ , we design an EV FIR filter  $\mathbf{H}_{\text{ev}}$  [cf. (9)] that approximates  $\tilde{\mathbf{H}}$  as the solution of the optimization problem

$$\begin{aligned} & \underset{\{\Phi_k\}}{\text{minimize}} \quad \left\| \tilde{\mathbf{H}} - \sum_{k=1}^K \Phi_{k:1} \right\| \\ & \text{subject to} \quad \text{supp}\{\Phi_k\} = \text{supp}\{\mathbf{S} + \mathbf{I}\}, \quad \forall k \in [K], \end{aligned} \quad (20)$$

where  $\|\cdot\|$  is an appropriate distance measure, e.g., the Frobenius norm ( $\|\cdot\|_F$ ), or the spectral norm ( $\|\cdot\|_2$ ).

Unfortunately, (20) is a high-dimensional nonconvex problem and hard to optimize. An approach to finding a local solution for it is through block coordinate descent methods, which provide local convergence guarantees when applied to such problems [35]. In fact, the cost in (20) is a *block multi-convex* function, i.e., the cost function is a convex function of  $\Phi_i$  with all the other variables fixed.

Starting then with an initial set of matrices  $\{\Phi_j^{(0)}\}_{j \in [K]}$  (potentially initialized with an order- $K$  classical FIR filter), we solve a sequence of optimization problems where at the  $i$ th step, the matrix  $\Phi_i$  is found. That is, at the  $i$ th iteration, we fix the matrices  $\{\Phi_j^{(0)}\}_{j \in [K] \setminus \{i\}}$  and solve the convex problem

$$\begin{aligned} & \underset{\Phi_i}{\text{minimize}} \quad \left\| \tilde{\mathbf{H}} - \sum_{k=1}^K \Phi_{k:(i+1)}^{(0)} \Phi_i \Phi_{(i-1):1}^{(0)} \right\| \\ & \text{subject to} \quad \text{supp}\{\Phi_i\} = \text{supp}\{\mathbf{S} + \mathbf{I}\}, \end{aligned} \quad (21)$$

where  $\Phi_{a:b}^{(0)} = \Phi_a^{(0)} \Phi_{a-1}^{(0)} \dots \Phi_{b+1}^{(0)} \Phi_b^{(0)}$  for  $a \geq b$  and  $\Phi_{a:b}^{(0)} = \mathbf{I}$ , otherwise. Then, the matrix  $\Phi_i^{(0)}$  is updated with its solution and the procedure is repeated for all  $\{\Phi_j\}_{j \in [K]}$ . If the final fitting error is large, the whole process can be repeated until the desired performance is reached, or until a local minimum is found.

Although filter (9) is the most general EV FIR filter form, the non-convexity encountered in the above design strategy may often lead to a local solution with an unacceptable performance. To tackle such issue, in Section IV, we introduce a constrained EV FIR filter which provides a higher flexibility than the state-of-the-art graph filters while accepting a simple least-squares design.

**SIEV form.** Besides enjoying the modal response interpretation, the SIEV FIR filter also has a simpler design than the general form (9). For  $\{\tilde{h}_i\}_{i=1}^N$  being the desired graph modal response,<sup>1</sup> the SIEV FIR filter design consists of solving the

<sup>1</sup>This can be for instance a low-pass form if we want to keep only the eigenvector contributions associated with the slowly varying graph modes.

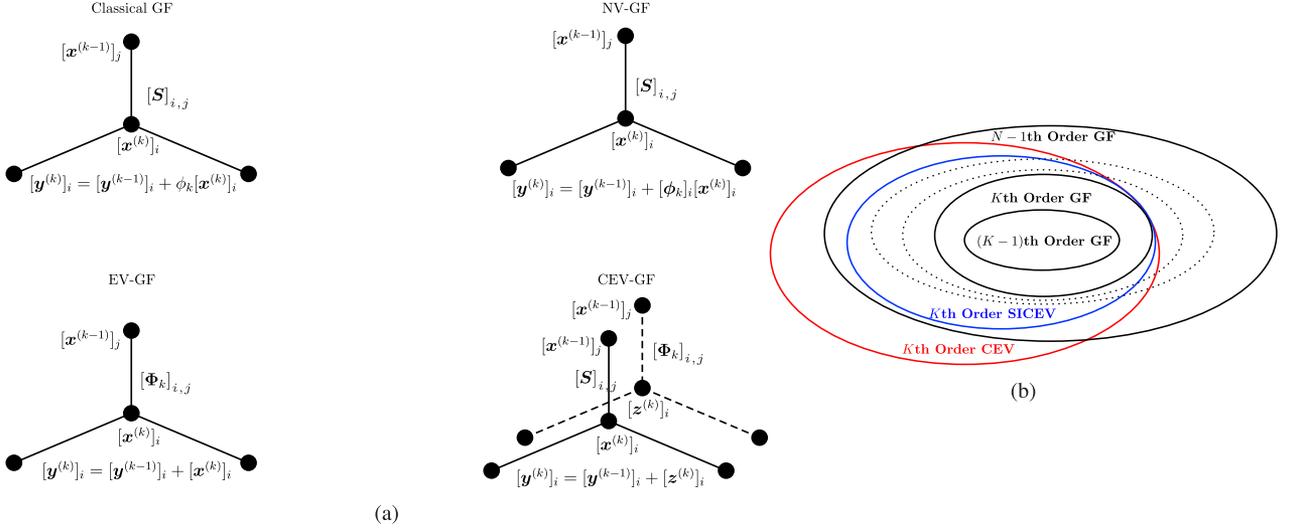


Fig. 1. (a) Illustration of the required communication, scaling, and recursion performed by the different graph filters. (b) Relation between the classical and CEV FIR graph filters. This figure depicts the possibility of obtaining higher-order polynomial graph filters with reduced order CEV graph filters.

optimization problem

$$\underset{\{\alpha_j\}}{\text{minimize}} \sum_{i=1}^N \left\| \tilde{h}_i - \sum_{k=1}^K \prod_{j=1}^k (\mathbf{b}_{\mathcal{U},i}^A)^T \alpha_j \right\|_2^2. \quad (22)$$

Similarly to (20), problem (22) is nonconvex and cannot in general be solved to global optimality with standard convex optimization methods. However, (22) is also a block multi-convex function in each  $\alpha_i$ ,  $i \in [K]$  and, therefore, the block coordinate descent methods [35] can again be employed to find a local minimum, yet the number of unknowns is smaller than for the general EV form. Alternatively, the straightforward analytical expression of the gradient of the cost function allows the use of off-the-shelf solvers for global optimization, such as MATLAB's built-in `fmincon` function [36].

#### IV. CONSTRAINED EDGE-VARIANT FIR GRAPH FILTERS

To overcome the design issues of the general EV FIR filter, here we present a constrained version of it that retains both the distributed implementation and the edge-dependent weighting. This reduction of the DoF will, in fact, allow us to design the filter coefficients in a least-squares fashion. The structure of these filters along with their distributed implementation is presented in the next section. In Section IV-B we provide a modal response interpretation of these filters, and finally in Section IV-C we present the design strategy.

##### A. General Form

The *constrained* EV (CEV) FIR graph filter is defined as

$$\mathbf{H}_{\text{cev}} = \Phi_1 + \Phi_2 \mathbf{S} + \dots + \Phi_K \mathbf{S}^{K-1} \triangleq \sum_{k=1}^K \Phi_k \mathbf{S}^{k-1}, \quad (23)$$

where the edge-weighting matrices  $\{\Phi_k\}_{k \in [K]}$  share again the support with  $\mathbf{S} + \mathbf{I}$ . These filters enjoy the same distributed implementation of the general form (9). In fact, each node can

compute locally the filter output by tracking the following quantities:

- the regular shift output  $\mathbf{x}^{(k)} = \mathbf{S}\mathbf{x}^{(k-1)}$ ,  $\mathbf{x}^{(0)} = \mathbf{x}$ ,
- the weighted shift output  $\mathbf{z}^{(k)} = \Phi_k \mathbf{x}^{(k-1)}$ ,
- the accumulator output  $\mathbf{y}^{(k)} = \mathbf{y}^{(k-1)} + \mathbf{z}^{(k)}$ ,  $\mathbf{y}^{(0)} = \mathbf{0}$ .

From the locality of  $\mathbf{S}$  and  $\Phi_k$ , both  $\mathbf{x}^{(k)}$  and  $\mathbf{z}^{(k)}$  require only neighboring information. The final filter output is  $\mathbf{y} = \mathbf{y}^{(K)}$  and is obtained with the same computational complexity of  $\mathcal{O}(MK)$ .

Note that the construction (23) still applies different weights to the signal coming from different edges. However, instead of adopting a different *diffusion* matrix at every step, the signal diffusion occurs through the graph shift  $\mathbf{S}$ . The additional extra step mixes locally  $\mathbf{x}^{(k-1)}$  using edge-dependent weights, which are allowed to vary for each term  $k$ . We adopt the term constrained for this implementation since the first  $k-1$  diffusion steps in every term  $k$  are constrained to be performed by the graph shift  $\mathbf{S}$ . Note though that the CEV FIR graph filter is *not* a special case of the EV FIR graph filter. Fig. 1(a) visually illustrates the differences between the different graph filters, analyzed so far. We conclude this section with the following remark.

*Remark 1:* Note that the NV graph filter of order  $K$  from [9] [cf. (5)] is a particular case of the CEV graph filter of order  $K$ . The local matrices  $\{\Phi_k\}_{k=1}^K$  are then in fact substituted as  $\Phi_1 = \text{diag}(\phi_0) + \text{diag}(\phi_1)\mathbf{S}$  and  $\Phi_k = \text{diag}(\phi_k)\mathbf{S}^k \forall k > 1$ . Also, the classical graph filter of order  $K$  from [4]–[6] [cf. (4)] is a particular case of the CEV graph filter of order  $K$ . In that case, the local matrices  $\{\Phi_k\}_{k=1}^K$  are just substituted as  $\Phi_1 = \phi_0 \mathbf{I} + \phi_1 \mathbf{S}$  and  $\Phi_k = \phi_k \mathbf{S}^k \forall k > 1$ .

##### B. Shift-Invariant Constrained Edge-Variant Graph Filters

Following the same lines of Section III-B, we can use the set  $\mathcal{J}_{\mathcal{U}}^A$  (12) to characterize the graph modal response of the CEV FIR graph filter when the matrices  $\{\Phi_k\}_{k=1}^K$  satisfy (A.1) and (A.2). This subset of CEV FIR graph filters, which we refer to as shift-invariant CEV (SICEV) FIR graph filters, can again be

expressed in terms of  $B_U^A$  and  $\{\alpha_k\}_{k=0}^K$  as

$$\mathbf{H}_{\text{sicev}} = \mathbf{U} \left[ \sum_{k=1}^K \text{diag}(B_U^A \alpha_k \odot \lambda^{\odot(k-1)}) \right] \mathbf{U}^{-1}, \quad (24)$$

where  $\lambda^{\odot k}$  denotes the  $k$ th element-wise power of the eigenvalue vector of the shift operator  $\mathbf{S}$ . The subsequent proposition formalizes the modal response of these filters.

**Proposition 3:** (Modal Response of SICEV FIR) An FIR graph filter of the form (23) satisfying (A.1) and (A.2) has  $i$ th modal response

$$h_i = \sum_{k=1}^K \gamma_{ik} \lambda_i^{k-1}, \quad (25)$$

where  $\gamma_{ik} = (\mathbf{b}_{U,i}^A)^T \alpha_k$  is the  $k$ th polynomial coefficient for the  $i$ th graph frequency and  $(\mathbf{b}_{U,i}^A)^T$  is the  $i$ th row of  $B_U^A$ .

*Proof:* The proof follows directly from (24). ■

From (25), we see that there is a substantial difference between the SICEV FIR graph filters and the more general SIEV FIR graph filters. Here, the modal response is a polynomial in the graph frequencies. This is similar to the classical FIR filter (4), but now each frequency has a different set of coefficients. In other words, the modal response of the SICEV FIR filter is a *mode-dependent polynomial*. For readers more familiar with traditional discrete-time processing, this behavior can be interpreted as applying different polynomial filters to each frequency bin (see e.g., [37]).

**Remark 2:** The particular form of the SICEV FIR filter allows it to match all shift-invariant polynomial responses of order  $K$  as well a subset of higher-order polynomials of order up to  $N - 1$  which includes all higher order responses. The latter property follows from the observation that any shift-invariant graph filter is a polynomial of the graph shift operator [4] and from the filter response in (25). In fact, the SICEV FIR filter is still a polynomial of the shift  $\mathbf{S}$ , though with a different polynomial response per graph frequency. This additional freedom extends the set of functions that can be approximated by a SICEV FIR filter of order  $K$ . Fig. 1(b) further illustrates the relation among different graph filters.

### C. Filter Design

**General form.** Following a similar approach as in Section III-C, we can approximate a desired operator  $\tilde{\mathbf{H}}$  with a CEV FIR filter by solving the problem

$$\begin{aligned} & \underset{\{\Phi_k\}}{\text{minimize}} \quad \left\| \tilde{\mathbf{H}} - \sum_{k=1}^K \Phi_k \mathbf{S}^{k-1} \right\|_F^2 \\ & \text{subject to} \quad \text{supp}\{\Phi_k\} = \text{supp}\{\mathbf{S} + \mathbf{I}\}, \forall k \in [K]. \end{aligned} \quad (26)$$

Exploiting then the properties of the vectorization operator and the Frobenius norm, we can transform (26) into

$$\begin{aligned} & \underset{\{\phi_k\}}{\text{minimize}} \quad \left\| \tilde{\mathbf{h}} - \sum_{k=1}^K (\mathbf{S}^{k-1} \otimes \mathbf{I}) \phi_k \right\|_2 \\ & \text{subject to} \quad \text{supp}\{\Phi_k\} = \text{supp}\{\mathbf{S} + \mathbf{I}\}, \forall k \in [K], \end{aligned} \quad (27)$$

where  $\tilde{\mathbf{h}} \triangleq \text{vec}(\tilde{\mathbf{H}})$  and  $\phi_k \triangleq \text{vec}(\Phi_k)$ .

Since the support of the weighting matrices is known, problem (27) can be written in the reduced-size form

$$\underset{\{\phi_k\}}{\text{minimize}} \quad \|\tilde{\mathbf{h}} - \Psi \boldsymbol{\theta}\|_2^2 \quad (28)$$

where  $\Psi = [\mathbf{I} \check{\mathbf{S}} \cdots \check{\mathbf{S}}_K]$ ,  $\boldsymbol{\theta} = [\check{\phi}_0^T \check{\phi}_1^T \cdots \check{\phi}_K^T]^T$ ,  $\check{\phi}_k$  is the vector  $\phi_k$  with the zero entries removed and  $\check{\mathbf{S}}_k$  is the matrix  $(\mathbf{S}^k \otimes \mathbf{I})$  with the appropriate columns removed. In addition, if a regularized solution is desired, a natural penalization term might be the convex  $\ell_1$ -norm which induces sparsity in the solution yielding only few active coefficients.

Problem (27) has a unique solution as long as  $\Psi$  is full column rank, i.e.,  $\text{rank}(\Psi) = \text{nnz}(\mathbf{S}) \cdot K + N$ . Otherwise, regularization must be used to obtain a unique solution.

**Remark 3:** Besides leading to a simple least-squares problem, the design of the CEV FIR filter can also be computed distributively. Given that each node knows the desired filter response and the graph shift operator (i.e., the network structure), it can be shown that by reordering the columns of  $\Psi$  and the entries of  $\boldsymbol{\theta}$  the framework of splitting-over-features [38] can be employed for a decentralized estimation of  $\boldsymbol{\theta}$ .

**SICEV form.** Similar to the more general CEV FIR filter, the design of  $\{\alpha_k\}_{k=1}^K$  for the SICEV form can be performed in a least-squares fashion.

First, for a set of vectors  $\{\alpha_k\}_{k=1}^K$  the modal response of the SICEV FIR filter reads as

$$\mathbf{h}_\lambda = \sum_{k=1}^K [\mathbf{B}_U^A \alpha_k \odot \lambda^{\odot(k-1)}], \quad (29)$$

where  $\mathbf{h}_\lambda$  is obtained by stacking the modal responses, i.e.,  $\{h_i\}_{i=1}^N$ , in a column vector. By using the properties of the Hadamard product, (29) can be written as

$$\mathbf{h}_\lambda = \sum_{k=1}^K \text{diag}(\lambda^{\odot(k-1)}) \mathbf{B}_U^A \alpha_k = \sum_{k=1}^K \mathbf{M}_k \alpha_k, \quad (30)$$

with  $\mathbf{M}_k = \text{diag}(\lambda^{\odot(k-1)}) \mathbf{B}_U^A$ . Defining then  $\mathbf{M} = [\mathbf{M}_1, \dots, \mathbf{M}_K]$  and  $\boldsymbol{\alpha} = [\alpha_1^T, \dots, \alpha_K^T]^T$ , we obtain the linear relation

$$\mathbf{h}_\lambda = \mathbf{M} \boldsymbol{\alpha}. \quad (31)$$

Therefore, the approximation of a desired response  $\tilde{\mathbf{h}}_\lambda = [\tilde{h}_1, \dots, \tilde{h}_N]^T$  consists of solving the least-squares problem

$$\underset{\boldsymbol{\alpha} \in \mathbb{R}^{d(K+1)}}{\text{minimize}} \quad \|\tilde{\mathbf{h}}_\lambda - \mathbf{M} \boldsymbol{\alpha}\|_2, \quad (32)$$

which has a unique solution when  $\mathbf{M}$  is full column rank, i.e.,  $\text{rank}(\mathbf{M}) = d(K+1) \leq N$ .

## V. EDGE-VARIANT IIR GRAPH FILTERS

We now extend the edge-variant filtering concept to the class of IIR graph filters. As stated in Section II, we focus on the basic building block of these filters, i.e., the ARMA<sub>1</sub> recursion (6). We follow the same organization of the former two sections, by introducing the edge-variant ARMA<sub>1</sub> structure in Section V-A, the shift-invariant version in Section V-B, and the design strategies in Section V-C.

### A. Edge-Variant ARMA<sub>1</sub>

We build an edge-variant ARMA<sub>1</sub> (EVA<sub>1</sub>) recursion on graphs by modifying (6) as

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_0 \mathbf{x}, \quad (33)$$

where  $\Phi_0$  and  $\Phi_1$  are the edge-weighting matrices having the support of  $\mathbf{S} + \mathbf{I}$  that respectively weight locally the entries of  $\mathbf{y}_{t-1}$  and  $\mathbf{x}$ . Proceeding similarly as in [11], for  $\|\Phi_1\|_2 < 1$ , the steady-state output of (33) is

$$\mathbf{y} = \lim_{t \rightarrow \infty} \mathbf{y}_t = (\mathbf{I} - \Phi_1)^{-1} \Phi_0 \mathbf{x} \triangleq \mathbf{H}_{\text{eva}_1} \mathbf{x}, \quad (34)$$

where we notice the inverse relation w.r.t. the edge-weighting matrix  $\Phi_1$ . Recursion (33) converges to (34) linearly with a rate governed by  $\|\Phi_1\|_2$ . The classical form (6) can be obtained by substituting  $\Phi_1 = \psi \mathbf{S}$  and  $\Phi_0 = \varphi \mathbf{I}$ .

The EVA<sub>1</sub> filter presents the same frequency interpretation challenges as the FIR filter counterpart. Therefore, we next analyze the shift-invariant version of it and we will see a rational modal response.

### B. Shift-Invariant EV ARMA<sub>1</sub>

By limiting the choices of  $\{\Phi_0, \Phi_1\}$  to the ones that satisfy (A.1) and (A.2), we obtain the shift-invariant edge-variant ARMA<sub>1</sub> (SIEVA<sub>1</sub>) graph filter

$$\mathbf{H}_{\text{sieva}_1} = \mathbf{U}[(\mathbf{I} - \text{diag}(\mathbf{B}_{\mathcal{U}}^A \boldsymbol{\alpha}_1))^{-1} \text{diag}(\mathbf{B}_{\mathcal{U}}^A \boldsymbol{\alpha}_0)] \mathbf{U}^{-1}, \quad (35)$$

where  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\alpha}_1$  are the respective basis expansion vectors of  $\Phi_0$  and  $\Phi_1$  onto the nullspace of  $\mathbf{T}_{\mathcal{U}}^A$  (see Proposition 1). From (35), we see that the inverse relation that appears in (34) indeed appears as a function affecting the graph eigenmodes. The following proposition concludes this section by stating this finding in a formal way.

**Proposition 4:** (Modal Response of SIEVA<sub>1</sub>) An ARMA<sub>1</sub> graph filter of the form (34) satisfying (A.1) and (A.2) for  $K = 1$  has  $i$ th modal response

$$h_i = \frac{(\mathbf{b}_{\mathcal{U},i}^A)^T \boldsymbol{\alpha}_0}{1 - (\mathbf{b}_{\mathcal{U},i}^A)^T \boldsymbol{\alpha}_1} \quad (36)$$

where  $(\mathbf{b}_{\mathcal{U},i}^A)^T$  is the  $i$ th row of the matrix  $\mathbf{B}_{\mathcal{U}}^A$ .

*Proof:* The proof follows directly from (35). ■

### C. Filter Design

**EVA<sub>1</sub> form.** Here, we extend the design approach of [39] and design  $\{\Phi_0, \Phi_1\}$  by using Prony's method. For  $\tilde{\mathbf{H}}$  being the desired operator, we can define the fitting error matrix

$$\mathbf{E} = \tilde{\mathbf{H}} - (\mathbf{I} - \Phi_1)^{-1} \Phi_0, \quad (37)$$

which presents nonlinearities in the denominator coefficients, i.e., in  $\Phi_1$ . To tackle this issue, we consider the modified fitting error matrix

$$\mathbf{E}' = \tilde{\mathbf{H}} - \Phi_1 \tilde{\mathbf{H}} - \Phi_0, \quad (38)$$

which is obtained by multiplying both sides of (37) by  $\mathbf{I} - \Phi_1$ .

This way, the filter design problem is transformed to the convex optimization problem

$$\begin{aligned} & \underset{\Phi_0, \Phi_1}{\text{minimize}} && \|\tilde{\mathbf{H}} - \Phi_1 \tilde{\mathbf{H}} - \Phi_0\| \\ & \text{subject to} && \|\Phi_1\|_2 < \delta, \delta < 1, \\ & && \text{supp}\{\Phi_0\} = \text{supp}\{\Phi_1\} = \text{supp}\{\mathbf{S} + \mathbf{I}\}. \end{aligned} \quad (39)$$

The objective function in (39) aims at reducing the modified error  $\mathbf{E}'$ , while the first constraint trades the convergence rate of (33) with approximation accuracy.

Note that it is possible to improve the performance of the filter design (39) if a second step is performed after the matrices  $\{\Phi_0, \Phi_1\}$  have been obtained. Using the matrix  $\Phi_1$  obtained by solving (39), we can fit again  $\Phi_0$  by using the error expression in (37). That is, we can obtain a new matrix  $\Phi_0$  by solving

$$\begin{aligned} & \underset{\Phi_0}{\text{minimize}} && \|\tilde{\mathbf{H}} - (\mathbf{I} - \Phi_1)^{-1} \Phi_0\| \\ & \text{subject to} && \text{supp}\{\Phi_0\} = \text{supp}\{\mathbf{S} + \mathbf{I}\}, \end{aligned} \quad (40)$$

where  $\Phi_1$  is the coefficient matrix obtained from (39). This procedure is known as Shanks' method [40] and has also been adopted in [11], [39].

**SIEVA<sub>1</sub> form.** Following the same idea as in (37)–(39), the modified fitting error of a SIEVA<sub>1</sub> graph filter is

$$e'_i = \tilde{h}_i - \tilde{h}_i (\mathbf{b}_{\mathcal{U},i}^A)^T \boldsymbol{\alpha}_1 - (\mathbf{b}_{\mathcal{U},i}^A)^T \boldsymbol{\alpha}_0, \quad (41)$$

with  $\tilde{h}_i$ ,  $(\mathbf{b}_{\mathcal{U},i}^A)^T \boldsymbol{\alpha}_0$ , and  $(\mathbf{b}_{\mathcal{U},i}^A)^T \boldsymbol{\alpha}_1$  denoting respectively the desired modal response and the eigenvalues of  $\Phi_0$  and  $\Phi_1$  w.r.t. the  $i$ th mode. In vector form, (41) can be written as

$$\mathbf{e}' = \tilde{\mathbf{h}}_\lambda - \bar{\mathbf{M}} \bar{\boldsymbol{\alpha}}, \quad (42)$$

with  $\mathbf{e}' = [e'_1, \dots, e'_N]^T$ ,  $\tilde{\mathbf{h}}_\lambda = [\tilde{h}_1, \dots, \tilde{h}_N]^T$ ,  $\bar{\mathbf{M}} = [\mathbf{B}_{\mathcal{U}}^A, \text{diag}(\tilde{\mathbf{h}}_\lambda) \mathbf{B}_{\mathcal{U}}^A]$ , and  $\bar{\boldsymbol{\alpha}} = [\boldsymbol{\alpha}_0^T, \boldsymbol{\alpha}_1^T]^T$ . Then,  $\{\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1\}$  can be estimated as the solution of the constrained least-squares problem

$$\begin{aligned} & \underset{\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1 \in \mathbb{R}^d}{\text{minimize}} && \|\tilde{\mathbf{h}}_\lambda - \bar{\mathbf{M}} \bar{\boldsymbol{\alpha}}\|_2^2 \\ & \text{subject to} && \|\mathbf{B}_{\mathcal{U}}^A \boldsymbol{\alpha}_1\|_\infty < \delta, \delta < 1, \boldsymbol{\alpha} = [\boldsymbol{\alpha}_0^T, \boldsymbol{\alpha}_1^T]^T. \end{aligned} \quad (43)$$

As (39), problem (43) again aims at minimizing the modified fitting error, while tuning the convergence rate through  $\delta$ . Here again, Shanks' method can be considered.

Differently from the general EVA<sub>1</sub>, here the number of unknowns is reduced to  $2d$ , as now only the vectors  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\alpha}_1$  need to be designed. Due to this low dimensionality, one can also opt for global optimization solvers to find an acceptable local minimum of the true error (i.e., the equivalent of (37)).

**Remark 4:** The approximation accuracy of the EVA<sub>1</sub> and SIEVA<sub>1</sub> filters can be further improved by following the iterative least-squares approach proposed in [41]. This method has shown to improve the approximation accuracy of Prony's design by not only taking the modified fitting error into account but also the true one. However, as this idea does not add much to this work, interested readers are redirected to [41] for more details.

TABLE I  
SUMMARY OF THE DIFFERENT GRAPH FILTERS. (\*) INDICATES A CONTRIBUTION OF THIS WORK. HERE,  $I$  STANDS FOR THE MAXIMUM NUMBER OF ITERATIONS

Filter Type	Expression	Shift-Invariant	Design Strategy	Implementation Cost	Coefficients
Classical FIR [4]	$\mathbf{H}_c \triangleq \sum_{k=0}^K \phi_k \mathbf{S}^k$	always	LS [4], Chebyshev [5, 6]	$\mathcal{O}(MK)$	scalars: $\{\phi_k\}$
NV FIR [9]	$\mathbf{H}_{nv} \triangleq \sum_{k=0}^K \text{diag}(\phi_k) \mathbf{S}^k$	not in general	LS, convex program [9]	$\mathcal{O}(MK)$	vectors: $\{\phi_k\}$
EV FIR (*)	$\mathbf{H}_{ev} \triangleq \sum_{k=1}^K \Phi_k \dots \Phi_1$	not in general	iterative design [Sec. III-C]	$\mathcal{O}(MK)$	matrices: $\{\Phi_k\}$
SIEV FIR (*)	(18)	always	iterative design [Sec. III-C]	$\mathcal{O}(MK)$	vectors: $\{\alpha_k\}$
CEV FIR (*)	$\mathbf{H}_{cev} \triangleq \sum_{k=1}^K \Phi_k \mathbf{S}^{k-1}$	not in general	LS [Sec. IV-C]	$\mathcal{O}(MK)$	matrices: $\{\Phi_k\}$
SICEV FIR (*)	(23)	always	LS [Sec. IV-C]	$\mathcal{O}(MK)$	vectors: $\{\alpha_k\}$
Classical ARMA <sub>1</sub> [11]	$\mathbf{H}_{arma_1} \triangleq \varphi(\mathbf{I} - \psi \mathbf{S})^{-1}$	always	closed-form, iterative design [11]	$\mathcal{O}(I \cdot M)$	scalars: $\{\varphi, \psi\}$
EVA <sub>1</sub> (*)	$\mathbf{H}_{eva_1} \triangleq (\mathbf{I} - \Phi_1)^{-1} \Phi_0$	not in general	Shanks' Method [Sec. V-C]	$\mathcal{O}(I \cdot M)$	matrices: $\{\Phi_0, \Phi_1\}$
SIEVA <sub>1</sub> (*)	(35)	always	Shanks' Method [Sec. V-C]	$\mathcal{O}(I \cdot M)$	vectors: $\{\alpha_0, \alpha_1\}$

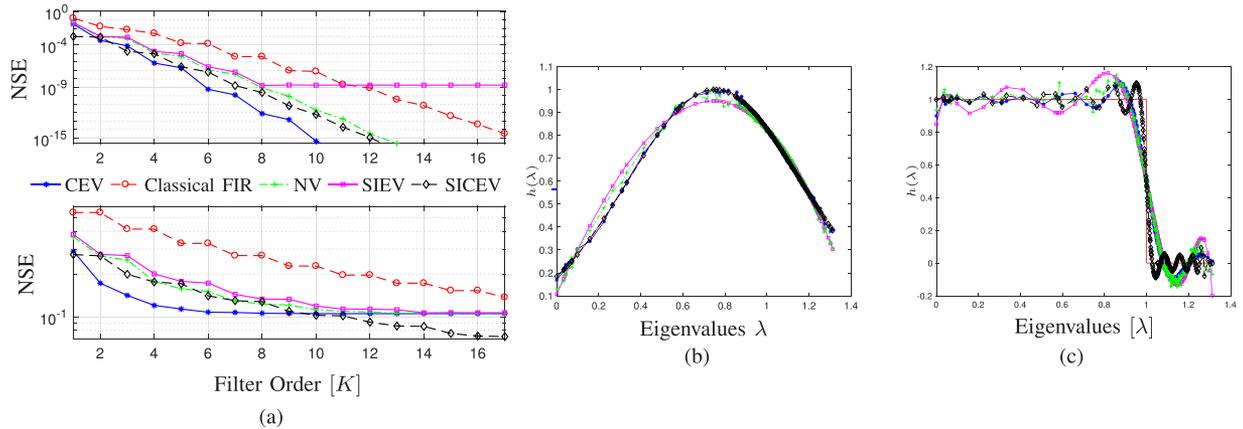


Fig. 2. (a) NSE vs. filter order for different FIR graph filters. (Top) Results in approximating the exponential kernel response. (Bottom) Results in approximating a low-pass response. (b) Frequency response of the graph filters when approximating an exponential kernel with parameters  $\mu = 0.75$  and  $\gamma = 3$ . (c) Frequency response of the graph filters when approximating a perfect low pass filter with  $\lambda_c = 1$ .

## VI. NUMERICAL RESULTS

We now present a set of numerical examples to corroborate the applicability of the proposed filters for several distributed tasks. Table I presents a summary of the different graph filters mentioned in this work along with their specifications. In our simulations,<sup>2</sup> we made use of the GSP toolbox [42].

### A. Graph Filter Approximation

We here test the proposed FIR graph filters in approximating a user-provided frequency response. We consider a random community graph of  $N = 256$  nodes and shift operator  $\mathbf{S} = \mathbf{L}$ . The frequency responses of interest are two commonly used responses in the GSP community, i.e.,

- i) the exponential kernel

$$\tilde{h}(\lambda) := e^{-\gamma(\lambda-\mu)^2},$$

with  $\gamma$  and  $\mu$  being the spectrum decaying factor and the central parameter, respectively;

- ii) the ideal low-pass filter

$$\tilde{h}(\lambda) = \begin{cases} 1 & 0 \leq \lambda \leq \lambda_c \\ 0 & \text{otherwise,} \end{cases}$$

with  $\lambda_c$  being the cut-off frequency.

The approximation accuracy of the different filters is evaluated in terms of the normalized squared error  $\text{NSE} = \|\tilde{\mathbf{H}} - \mathbf{H}_{\text{fit}}\|_F^2 / \|\tilde{\mathbf{H}}\|_F^2$ .  $\mathbf{H}_{\text{fit}}$  stands for the filter matrix of the fitted filters. Fig. 2 illustrates the performances of the different filters. In the exponential kernel scenario, we observe that the CEV FIR filter outperforms the other alternatives by showing a performance improvement of several orders of magnitude. A similar result is also seen in the low-pass example, where the CEV FIR filter achieves the error floor for  $K = 8$ , while the NV graph filter for  $K = 13$  and the classical FIR filter for  $K = 17$ . Additionally, we observe that the SIEV FIR filter achieves a similar performance as the NV FIR filter. This result suggests that despite the additional DoF of the SIEV FIR filter, the nonconvex design strategy (22) yields a local minimum that does not exploit the full capabilities of the filter. This local minimality effect can be seen in the stagnation of the error of the SIEV FIR filter for the exponential kernel case after  $K \geq 8$ . Finally, we notice that the SICEV filter achieves a performance similar to the NV Filter of the same order, while having less DoF. This characteristic of the SICEV shows its benefits as the order increases. Having to estimate less parameters, the error stagnation for the step response

<sup>2</sup>The code to reproduce the figures in this paper can be found at <https://gitlab.com/fruzti/graphFilterAdvances>

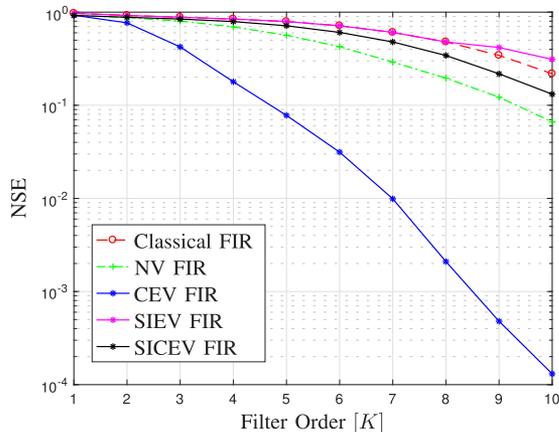


Fig. 3. NSE versus filter order for different distributed FIR filter implementations when approximating the consensus operator  $\tilde{\mathbf{H}} = 1/N\mathbf{1}\mathbf{1}^T$ .

is achieved at a higher filter order, hence a better approximation can be obtained.

The above observations further motivate the use of the CEV FIR filter, which trades better the simplicity of the design and the available DoF. In fact, even though the CEV FIR filter is conceptually simpler than the SIEV graph filter, it performs better than the latter. In addition, the larger DoF of the CEV FIR filter compared to the NV FIR filter (i.e.,  $\text{nnz}(\mathbf{S}) \cdot K + N$  vs  $N \cdot (K + 1)$ ) allows the CEV FIR filter to better approximate the desired response. In a distributed setting, these benefits translate into communication and computational savings.

### B. Distributed Linear Operator Approximation

Several distributed tasks of interest consist of performing a linear operation  $\tilde{\mathbf{H}} \in \mathbb{R}^{N \times N}$  over a network. This can be for instance a beamforming matrix over a distributed array or a consensus matrix. In most of these cases, such linear operators cannot be straightforwardly distributed. In this section, we illustrate the capabilities of the developed graph filters in addressing this task.

Given a desired linear operator  $\tilde{\mathbf{H}}$ , we aim at implementing this linear operator distributively through the solution of the optimization problem

$$\begin{aligned} & \underset{\theta}{\text{minimize}} && \|\tilde{\mathbf{H}} - \mathbf{H}(\mathbf{S}, \theta)\| \\ & \text{subject to} && \theta \in \Theta, \end{aligned} \quad (44)$$

where  $\mathbf{H}(\mathbf{S}, \theta)$  stands for the considered graph filter parametrized by the shift  $\mathbf{S}$  and a set of parameters  $\theta$  living in the domain  $\Theta$ .

**Distributed consensus.** For distributed consensus, the operator  $\tilde{\mathbf{H}}$  has the form  $\tilde{\mathbf{H}} = \frac{1}{N}\mathbf{1}\mathbf{1}^T$ , which for  $\mathbf{S} = \mathbf{L}$  translates into a low-pass graph filter passing only the constant signal component.

Fig. 3 compares the fitting  $\text{NSE} = \|\tilde{\mathbf{H}} - \mathbf{H}_{\text{fit}}\|_F^2 / \|\tilde{\mathbf{H}}\|_F^2$  for the different FIR graph filters. We notice once again that the CEV implementation offers the best approximation accuracy among the contenders achieving an NSE of order  $10^{-4}$  in only

10 exchanges. These results yield also different insights about the SIEV and SICEV graph filters. In this example, we consider a community network consisting of  $N = 512$  nodes generated using the GSP Toolbox [42].

First, both the SIEV and the SICEV implementations fail to compare well with the CEV, though the linear operator  $\tilde{\mathbf{A}}$  is shift invariant. We attribute this degradation of performance to assumption (A.1), which is a sufficient (not necessary) condition for these filters to have a modal response interpretation. In fact, forcing each filter coefficient matrix to be shift invariant seems limiting the filter ability to match well the consensus operator.

Second, the different design strategies used in SIEV and SICEV further discriminate the two filters. We can see that the least-squares design of the SICEV implementation is more beneficial, though the SIEV filter has more DoF. Unfortunately, this is the main drawback of the latter graph filter, which due to the nonconvexity of the design problem leads to suboptimal solutions. However, we remark that both these filters outperform (or compare equally with) the classical FIR filter. Further investigation in this direction is needed to understand if the SIEV and/or SICEV structures can be used to achieve finite-time consensus as carried out in [43], [44].

**Wiener-based denoising.** When the statistics of the graph signal and noise signal are available, a typical approach for performing denoising is the Wiener filter. This filter is obtained by minimizing the mean-squared error, i.e.,

$$\tilde{\mathbf{H}} = \underset{\mathbf{H} \in \mathbb{R}^{N \times N}}{\text{argmin}} \mathbb{E} [\|\mathbf{H}\mathbf{z} - \mathbf{x}\|_2^2], \quad (45)$$

where  $\mathbf{z} = \mathbf{x} + \mathbf{n}$  is the graph signal corrupted with additive noise. For the case of zero-mean signals  $\mathbf{x}$  and  $\mathbf{n}$  with covariance matrices  $\Sigma_{\mathbf{x}}$  and  $\Sigma_{\mathbf{n}}$ , respectively, the solution for (45) is

$$\tilde{\mathbf{H}} = \Sigma_{\mathbf{x}}(\Sigma_{\mathbf{x}} + \Sigma_{\mathbf{n}})^{-1}, \quad (46)$$

given  $\Sigma_{\mathbf{x}} + \Sigma_{\mathbf{n}}$  is not singular. When the covariance matrices  $\Sigma_{\mathbf{x}}$  and  $\Sigma_{\mathbf{n}}$  share the eigenvectors with the graph shift operator, the optimal filter  $\tilde{\mathbf{H}}$  can be approximated by classical graph filters. However, in many instances, this is not the case.

We illustrate an example where instead of approximating the Wiener filter through a classical FIR graph filter, we employ a CEV FIR filter. For this example we consider the Molene dataset,<sup>3</sup> where the temperature data of several cities in France has been recorded. The graph employed is taken from [45] and the graph signal has been corrupted with white Gaussian noise. The results in terms of NSE for the different fitted graph filters are shown in Fig. 4. From this plot we observe that the CEV FIR filter outperforms all the other alternatives. This is due to the fact that the optimal Wiener filter is not jointly diagonalizable with the eigenbasis of the shift operator, i.e., the covariance matrix of data is not shift invariant, hence classical graph filters are not appropriate to approximate the filter.

<sup>3</sup>Access to the raw data through the link [donneespubliques.meteofrance.fr/donnees\\_libres/Hackathon/RADOMEH.tar.gz](https://donneespubliques.meteofrance.fr/donnees_libres/Hackathon/RADOMEH.tar.gz)

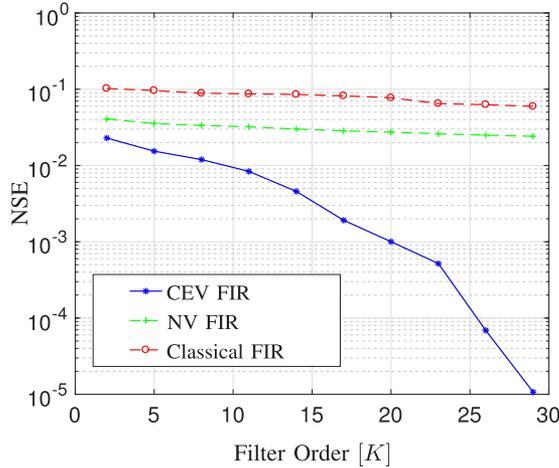


Fig. 4. NSE versus filter order for different distributed FIR filter implementations when approximating the Wiener filter for the Molene temperature dataset.

**Distributed Beamforming.** We here consider the task of applying a beamforming matrix  $\mathbf{W}^H$  to signals acquired via a distributed array. More specifically, we aim at obtaining the output

$$\mathbf{y} = \mathbf{W}^H \mathbf{x}, \quad (47)$$

where  $\mathbf{x}$  is the data acquired in a distributed way. Since  $\mathbf{W}^H$  might often be a dense matrix, e.g., in zero-forcing beamforming, the operation (47) cannot be readily distributed. To obtain the output at each node, we approximate the beamforming matrix with different graph filters.

We illustrate this scenario in a distributed 2D sensor array. The network is generated using  $N = 1000$  random locations on a 2D plane where the communication network is an  $\lceil N/5 \rceil$ -nearest neighbor graph. The beamforming matrix is the matched filter [46] matrix for a uniform angular grid of 1000 points in the range  $(-180, 180)$ . In other words, every node will see the information from a sector of approximately 0.36 degrees. Since in general  $\mathbf{W}^H$  does not share the eigenbasis with  $\mathbf{S}$ , classical graph filters fail to address this task. Therefore, here we compare only the CEV FIR filter and the NV FIR filter. Fig. 5 shows two output beampatterns obtained by solving (44) with  $\tilde{\mathbf{H}} = \mathbf{W}^H$  for the two considered filters with order  $K = 5$ . We notice that the CEV outperforms the NV FIR filter as it follows more closely the desired beampattern.

To demonstrate that our proposed designs can be directly extended to *asymmetric* shift matrices, we also plot, in Fig. 5, the beampatterns obtained when we consider an asymmetric version of the previous communication network. This asymmetric network is generated by converting 323 random edges of the original nearest neighbor graph into directed edges. We refer to these beampatterns in the figure as CEV<sub>Asym</sub> or the CEV filter and NV<sub>Asym</sub> for the NV filter. From Fig. 5, we observe that despite the performance changes in the sidelobe region, in the mainlobe region the CEV and CEV<sub>Asym</sub> perform similarly both outperforming their NV graph filter counterparts.

Note that the above framework treats the distributed beamforming differently from approaches based on distributed optimization tools [47]. The latter methods usually aim at computing

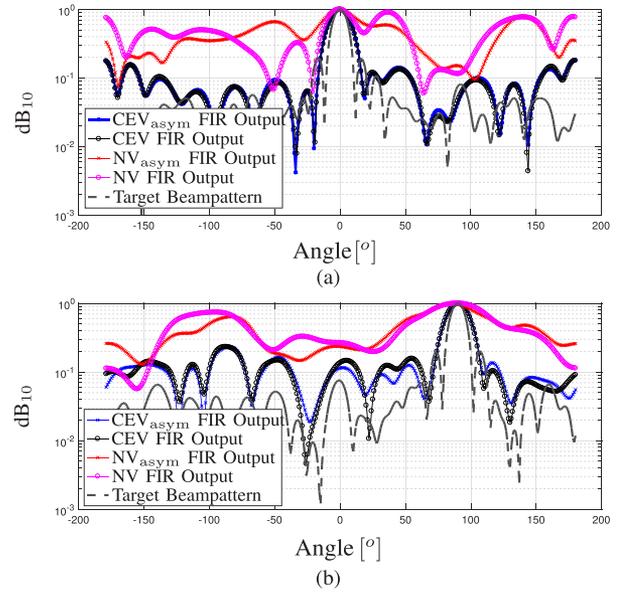


Fig. 5. Comparison of beampatterns for the node with desired steering angle,  $\theta_0$ . (a)  $\theta_0 = 0^\circ$  and (b)  $\theta_0 = 90^\circ$ .

the beamforming matrix (i.e., the weighting matrix is data dependent) and then perform consensus. On the other hand, we assume that  $\mathbf{W}^H$  is fixed and that it must be applied to the array data. Still, this problem can be solved through distributed convex optimization tools by solving the least-squares problem

$$\underset{\mathbf{y}}{\text{minimize}} \|\mathbf{x} - (\mathbf{W}^H)^\dagger \mathbf{y}\|_2^2. \quad (48)$$

Our formulation avoids the computation of the pseudo-inverse and the graph-filtering based approach requires only five iterations to compute the final output.

We next compare the CEV and the NV graph filters with distributed optimization tools for solving a general inverse problem.

### C. Comparison With Distributed Optimization

We now compare the proposed graph filters with the primal-dual method of multipliers (PDMM)<sup>4</sup> [48] to solve the following least-squares problem in a distributed fashion:

$$\underset{\mathbf{x}}{\text{minimize}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2. \quad (49)$$

Without loss of generality, we consider  $\mathbf{A}$  to be an  $N \times N$  matrix. The baseline assumption for all distributed optimization methods is that  $v_i$  knows its own data, i.e.,  $i$ th element of  $\mathbf{y}$ ,  $y_i$ , and its own regressor, i.e., the  $i$ th row of  $\mathbf{A}$ ,  $\mathbf{a}_i^\top$ . The task is then for each node to retrieve the full vector  $\mathbf{x}_{\text{ls}} = \mathbf{A}^\dagger \mathbf{y}$  by means of local communications. Here, the communication graph is a community graph with  $N = 512$  with  $\lceil \sqrt{N}/2 \rceil$  communities generated using the GSP toolbox [42].

For the graph filter-based approaches, we approximate  $\mathbf{A}^\dagger$  through a set of rank-one matrices  $\{\tilde{\mathbf{H}}_i \triangleq \mathbf{1}\tilde{\mathbf{a}}_i^\top\}_{i=1}^N$  with  $\tilde{\mathbf{a}}_i^\top$  being the  $i$ th row of  $\mathbf{A}^\dagger$ . This means that in contrast to distributed

<sup>4</sup>PDMM is an alternative to the classical alternating direction method of multipliers (ADMM), and is often characterized by a faster convergence [48].

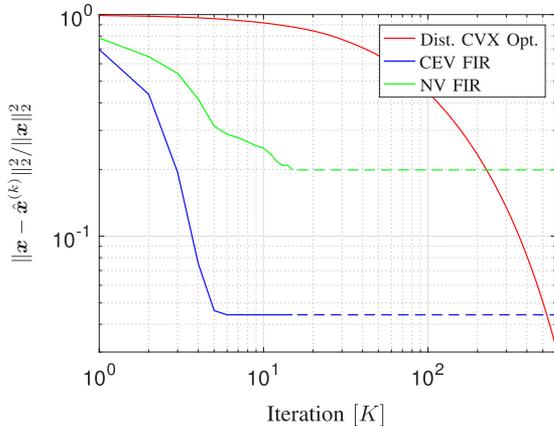


Fig. 6. Convergence error versus the number of iterations for the NV and the CEV graph filters and for the PDMM solver [48]. Dashed lines indicates the saturation floor of the NV and CEV FIRs.

optimization methods, here every node  $v_i$  needs to know the full  $\mathbf{A}$ . Each  $\tilde{\mathbf{H}}_i$  is then fitted with the NV and CEV recursions to approximate  $\mathbf{x}_{ls}$  as the output after filtering the graph signal  $\mathbf{y}$ . It must be noticed that the number of rounds between adjacent nodes does not scale with  $N$ . In fact, both the NV and the CEV will shift the signal only  $K$  times and the nodes can locally apply the respective coefficients to obtain the outputs.

To quantify the performance, we perform 100 Monte Carlo simulations with a randomly generated system matrix and solution vector. Fig. 6 compares the graph filter approaches with the distributed optimization methods in terms of the NSE =  $\|\mathbf{x} - \hat{\mathbf{x}}^{(k)}\|_2^2 / \|\mathbf{x}\|_2^2$ . The graph filter methods achieve a faster decay compared to the distributed optimization method in the first hundred iterations. However, due the ill-conditioning of the system matrices perfect approximation of the desired response is not achieved and both graph filters exhibit an error floor. PDMM, on the other hand, does not run into this issue and guarantees convergence to the true solution. Despite this difference in performance, the graph filter approaches can be employed for cases where the accuracy requirements are not strict, or as *warm starts* for distributed optimization methods. The above comparison, besides proposing graph filters as an alternative for solving distributed least-squares problems, raises the question *how graph filters relate to distributed convex optimization*. Further research is needed to relate the design and implementation of distributed EV graph filters with the well-established theory of distributed optimization.

#### D. Tikhonov-Based Denoising

One of the central problems in GSP is that of recovering an unknown signal  $\mathbf{x}$  from a noisy realization  $\mathbf{z} = \mathbf{x} + \mathbf{n}$  given that  $\mathbf{x}$  is smooth w.r.t. the underlying graph [3]. Differently known as the Tikhonov denoiser, the estimation of  $\mathbf{x}$  can be obtained by solving the regularized least-squares problem:

$$\mathbf{x}_{\text{tik}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{z} - \mathbf{x}\|_2^2 + \mu \mathbf{x}^T \mathbf{S} \mathbf{x}, \quad (50)$$

for  $\mathbf{S} = \mathbf{L}$  and where  $\mu$  trades off the noise removal with the smoothness prior. Problem (50) has the well-known solution  $\mathbf{x}_{\text{tik}} = (\mathbf{I} + \mu \mathbf{S})^{-1} \mathbf{z}$ , which in terms of the terminology used

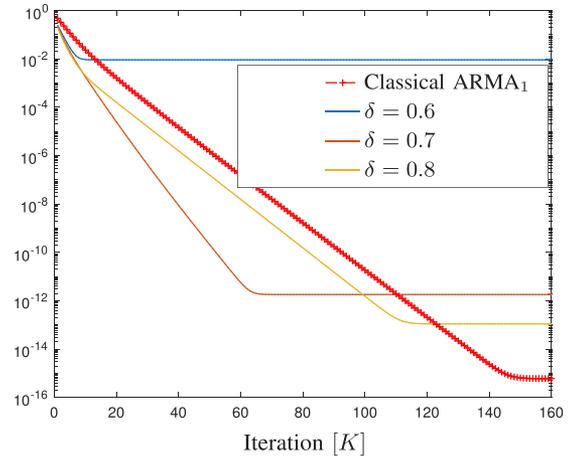


Fig. 7. Convergence error versus the number of iterations for the Tikhonov denoising problem. The  $\text{EVA}_1$  results are plotted for different values of  $\delta$  in (39) to highlight the tradeoff between convergence speed and approximation accuracy.

in Section II is an  $\text{ARMA}_1$  graph filter with  $\varphi = 1$  and  $\psi = -\mu$  (see also [11] for further analysis). While recursion (6) can implement this problem distributively, the convergence of the Neumann series in (7) cannot be controlled as the rate is fixed by  $|\mu| \lambda_{\max}\{\mathbf{S}\}$ .

Here, we show that through the  $\text{EVA}_1$  in (33) it is possible to improve the convergence speed of the  $\text{ARMA}_1$  graph filter by exploiting the additional DoF given by the edge-weighting matrices  $\{\Phi_0, \Phi_1\}$ . However, since now the design is not exact and involves the modified error [cf. (39)], this speed benefit will come at the expense of accuracy. To illustrate this, we consider an example of problem (50) with  $\mu = 0.8$  and  $\mathbf{S} = \lambda_{\max}^{-1}(\mathbf{L})\mathbf{L}$ , such that  $\mathbf{S}$  has unitary spectral norm. Here, the network is generated using  $N = 300$  random locations on a 2D plane where the communication network is an  $\lceil N/5 \rceil$ -nearest neighbor graph. Fig. 7 shows the convergence error of the  $\text{EVA}_1$  for different values of  $\delta$  in (39) and compares it with the classical  $\text{ARMA}_1$ .

We make the following observations. First, low values of  $\delta$  are preferred to improve the convergence speed. However, values below 0.7 should in general be avoided since this restricts too much the feasible set of (39), hence leading to a worse approximation error. Second, values of  $\delta \approx 0.7$  seem to give the best tradeoff, since the convergence speed is doubled w.r.t the  $\text{ARMA}_1$  and the approximation error is close to machine precision. Finally, we did not plot the classical FIR filter for solving this problem, since its performance is identical to the  $\text{ARMA}_1$  for the same distributed cost [11].

## VII. CONCLUSION

In this work, a generalization of distributed graph filters was proposed. These filters, that we referred to as edge-variant graph filters, have the ability to assign different weights to the information coming from different neighbors. Through the design of edge-weighting matrices, we have shown that it is possible to weight, possibly in an asymmetric fashion, the information propagated in the network and improve the performance of state-of-the-art graph filters.

By introducing the notion of filter modal response, we showed that a subclass of the edge-variant graph filters have a graph Fourier interpretation that illustrates the filter action on the graph modes. Despite that the most general edge-variant graph filter encounters numerical challenges in the design phase, a constrained version of it was introduced to tackle this issue. This so-called constrained edge-variant graph filter enjoys a similar distributed implementation, generalizes the state-of-the-art approaches, and is characterized by a simple least-squares design. For the constrained version, we also showed that there exists a subclass which has a modal response interpretation.

Finally, we extended the edge-variant idea to the family of IIR graph filters, particularly to the ARMA<sub>1</sub> graph filter. We showed that by adopting the same local structure a distributed rational filter can be achieved, yet with a much faster convergence speed. Several numerical tests corroborate our findings and show the potential of the proposed filters to improve state-of-the-art techniques.

Future research in this direction should concern the following points: *i*) improve the design strategy of the more general edge-variant version; *ii*) improve the saturation accuracy of the proposed methods when dealing with a distributed implementation of linear operators; *iii*) conciliate the world of GSP with that of distributed optimization and exploit the latter to design distributed graph filters; and *iv*) extend the edge-variant concept beyond the ARMA<sub>1</sub> implementation to the global family of IIR graph filters.

## REFERENCES

- [1] M. Coutino, E. Isufi, and G. Leus, "Distributed edge-variant graph filters," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Proc.*, 2017, pp. 1–5.
- [2] G. Taubin, "Geometric signal processing on polygonal meshes," in *Eurographics 2000 State of the Art Rep. (STAR)*, Sep. 2000.
- [3] 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.
- [4] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [5] G. Taubin, T. Zhang, and G. Golub, "Optimal surface smoothing as filter design," in *Proc. Eur. Conf. Comput. Vision*, 1996, pp. 283–292.
- [6] D. I. Shuman, P. Vandergheynst, D. Kressner, and P. Frossard, "Distributed signal processing via Chebyshev polynomial approximation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 4, pp. 736–751, Dec. 2018.
- [7] S. K. Narang, A. Gadda, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 5445–5449.
- [8] M. Onuki, S. Ono, M. Yamagishi, and Y. Tanaka, "Graph signal denoising via trilateral filter on graph spectral domain," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 2, pp. 137–148, Jun. 2016.
- [9] S. Segarra, A. 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, 1 Aug. 2017.
- [10] 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.
- [11] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 274–288, Jan. 2017.
- [12] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, 2016.
- [13] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.
- [14] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, no. Nov., pp. 2399–2434, 2006.
- [15] J. Ma, W. Huang, S. Segarra, and A. Ribeiro, "Diffusion filtering of graph signals and its use in recommendation systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 4563–4567.
- [16] B. Girault, P. Gonçalves, E. Fleury, and A. S. Mor, "Semi-supervised learning for graph to signal mapping: A graph signal wiener filter interpretation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 1115–1119.
- [17] E. Isufi, P. Di Lorenzo, P. Banelli, and G. Leus, "Distributed wiener-based reconstruction of graph signals," in *Proc. IEEE Statist. Signal Workshop*, 2018, pp. 21–25.
- [18] F. Zhang and E. R. Hancock, "Graph spectral image smoothing using the heat kernel," *Pattern Recognit.*, vol. 41, no. 11, pp. 3328–3342, 2008.
- [19] A. C. Yağın and M. T. Özgen, "A spectral graph wiener filter in graph fourier domain for improved image denoising," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2016, pp. 450–454.
- [20] E. Isufi and G. Leus, "Distributed sparsified graph filters for denoising and diffusion tasks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 5865–5869.
- [21] N. Tremblay, G. Puy, R. Gribonval, and P. Vandergheynst, "Compressive spectral clustering," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1002–1011.
- [22] D. B. Tay and Z. Lin, "Design of near orthogonal graph filter banks," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 701–704, Jun. 2015.
- [23] O. Teke and P. P. Vaidyanathan, "Extending classical multirate signal processing theory to graphs—Part II: M-channel filter banks," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 423–437, Jan. 2017.
- [24] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, 2011.
- [25] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Adv. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
- [26] F. Gama, A. G. Marques, G. Leus, and A. Ribeiro, "Convolutional neural networks architectures for signals supported on graphs," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1034–1049, 2019.
- [27] X. Shi, H. Feng, M. Zhai, T. Yang, and B. Hu, "Infinite impulse response graph filters in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1113–1117, Aug. 2015.
- [28] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Filtering random graph processes over random time-varying graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4406–4421, Aug. 2017.
- [29] L. L. Magoarou and R. Gribonval, "Flexible multilayer sparse approximations of matrices and applications," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 688–700, Jun. 2016.
- [30] S. Barbarossa, G. Scutari, and T. Battisti, "Distributed signal subspace projection algorithms with maximum convergence rate for sensor networks with topological constraints," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 2893–2896.
- [31] E. Isufi and G. Leus, "Distributed sparsified graph filters for denoising and diffusion tasks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 5865–5869.
- [32] V. Kalofolias, A. Loukas, D. Thanou, and P. Frossard, "Learning time varying graphs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 2826–2830.
- [33] M. Kolar, L. Song, A. Ahmed, and E. P. Xing, "Estimating time-varying networks," *Ann. App. Statist.*, vol. 4, pp. 94–123, 2010.
- [34] M. Coutino, S. Chepuri, and G. Leus, "Sparsest network support estimation: A submodular approach," in *Proc. IEEE Data Sci. Workshop*, 2018, pp. 200–204.
- [35] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.
- [36] "MATLAB Optimization Toolbox," The MathWorks, Inc., Natick, MA, USA, 2018.
- [37] K. Van Acker, G. Leus, M. Moonen, O. Van de Wiel, and T. Pollet, "Per tone equalization for DMT-based systems," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 109–119, Jan. 2001.

- [38] C. Manss, D. Shutin, and G. Leus, "Distributed splitting-over-features sparse Bayesian learning with alternating direction method of multipliers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 3654–3658.
- [39] E. Isufi, A. Loukas, and G. Leus, "Autoregressive moving average graph filters a stable distributed implementation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 4119–4123.
- [40] J. L. Shanks, "Recursion filters for digital processing," *Geophysics*, vol. 32, no. 1, pp. 33–51, 1967.
- [41] J. Liu, E. Isufi, and G. Leus, "Filter design for autoregressive moving average graph filters," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 1, pp. 47–60, Mar. 2019.
- [42] N. Perraudin *et al.*, "GSPBOX: A toolbox for signal processing on graphs," Aug. 2014, arXiv preprint arXiv:1408.5781.
- [43] L. Wang and F. Xiao, "Finite-time consensus problems for networks of dynamic agents," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 950–955, Apr. 2010.
- [44] A. Sandryhaila, S. Kar, and J. M. Moura, "Finite-time distributed consensus through graph filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 1080–1084.
- [45] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, Jul. 2017.
- [46] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2002.
- [47] T. Sherson, W. B. Kleijn, and R. Heusdens, "A distributed algorithm for robust LCMV beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 101–105.
- [48] G. Zhang and R. Heusdens, "Distributed optimization using the primal-dual method of multipliers," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 173–187, Mar. 2018.



**Mario Coutino** (S'17) received the M.Sc. (*cum laude*) degree in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 2016. Since October 2016, he has been working toward the Ph.D. degree in signal processing with Delft University of Technology, Delft, The Netherlands. He was with Thales Netherlands, during summer 2015, and Bang & Olufsen, during 2015–2016. His research interests include signal processing on networks, submodular and convex optimization, and numerical linear algebra. He was the recipient of the Best Student

Paper Award for his publication at the CAMSAP 2017 conference in Curacao.



**Elvin Isufi** (S'15) was born in Vlore, Albania, in 1989. He received the Ph.D. degree from Delft University of Technology, Delft, The Netherlands, in 2019, and the M.Sc. degree from the University of Perugia, Perugia, Italy, in 2014. He is currently a Postdoctoral Researcher with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. His current research interest focuses on the intersection of signal processing, mathematical modeling, machine learning, and network theory.



**Geert Leus** (F'12) received the M.Sc. and Ph.D. degrees in electrical engineering from the KU Leuven, Leuven, Belgium, in 1996 and 2000, respectively. He is currently an "Antoni van Leeuwenhoek" Full Professor with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands. His research interests include the broad area of signal processing, with a specific focus on wireless communications, array processing, sensor networks, and graph signal processing. He is a Fellow of *EURASIP*. He was a

Member-at-Large of the Board of Governors of the IEEE Signal Processing Society, the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, a Member of the IEEE Sensor Array and Multichannel Technical Committee, and the Editor-in-Chief for the *EURASIP Journal on Advances in Signal Processing*. He was also on the Editorial Boards of the *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, *IEEE SIGNAL PROCESSING LETTERS*, and *EURASIP Journal on Advances in Signal Processing*. He is currently the Vice-Chair of the *EURASIP* Special Area Team on Signal Processing for Multisensor Systems, an Associate Editor for *Foundations and Trends in Signal Processing*, and the Editor-in-Chief for *EURASIP Signal Processing*. He was the recipient of the 2002 IEEE Signal Processing Society Young Author Best Paper Award and the 2005 IEEE Signal Processing Society Best Paper Award.