Stochastic Graph Filtering on Time-Varying Graphs -EXTENDED VERSION-

Elvin Isufi*, Andrea Simonetto*, Andreas Loukas§, and Geert Leus*
* Faculty of EEMCS, Delft University of Technology, 2826 CD Delft, The Netherlands
§ TU Berlin, Germany
e-mails: {e.isufi-1, a.simonetto, g.j.t.leus}@tudelft.nl, a.loukas@tu-berlin.de

***Abstract*—We have recently seen a surge of work on distributed graph filters, extending classical results to the graph setting. State of the art filters have however only been examined from a deterministic standpoint, ignoring the impact of stochasticity in the computation (e.g., temporal fluctuation of links) and input (e.g., the value of each node is a random process). Initiating the study of stochastic graph signal processing, this paper shows that a prominent class of graph filters, namely autoregressive moving average (ARMA) filters, are suitable for the stochastic setting. In particular, we prove that an ARMA filter that operates on a stochastic signal over a stochastic graph is equivalent, in the mean, to the same filter operating on the expected signal over the expected graph. We also characterize the variance of the output and we provide an upper bound for its average value among different nodes. Our results are validated by numerical simulations.**

I. INTRODUCTION

Signal processing on graphs [2]–[4] has been developed recently as a tool that extends the classical concept of signal processing on time and space signals to signals indexed by nodes of an irregular graph. The definition of a graph Fourier transform offers the possibility to analyze the graph signal not only in the node domain, but also in the graph frequency domain. Making use of graph filters, the graph signal can be filtered keeping a desired part of the spectrum, while attenuating the other frequencies. Notable applications are signal denoising [5], [6] and event boundary detection [7].

Since the graph signal is indexed by the nodes of the graph, i.e., it is distributed across the graph, distributed implementations of graph filters are preferred. Distributed implementations of finite impulse response (FIR) graph filters are considered in [8]–[10]. FIR graph filters have the benefit that they can be easily implemented in the node domain, due to their polynomial frequency response. However, the polynomial form limits their performance, and they have been shown to be more sensitive to graph changes. To improve robustness, distributed infinite impulse response (IIR) graph filters have been proposed in [11], [12]. These filters have a frequency response that is a rational function and they offer better performance than FIR graph filters. Furthermore, these IIR filters are designed for a continuous range of frequencies, thus the knowledge of the graph spectrum is not necessary.

In this paper, our starting point is the autoregressive moving average (ARMA) graph filter, a type of IIR filter proposed in [12]. We aim at analyzing the effects of *stochasticity in the graph and signal* on the filter performance. Stochasticity is unavoidable in real applications: it occurs for instance when the signal defined on the node of a graph is randomly distributed (e.g., a noisy graph signal), or when the graph topology changes with a certain probability, i.e., link failures occur or new nodes appear. In this paper, we derive the expected value and an upper bound on the variance of the steady state signal in the graph, based only on statistical knowledge of the graph and input signal. We show that in the mean our filters reflect the same graph frequency response as deterministic ARMA filters. For 1st order filters in particular, the expected value of the output of the filter is a filtered version of the expected value of the graph signal. This is an important result since it shows that the output signal is robust to graph changes and it can handle random processes as graph signals.

II. GRAPH FILTERS AND STOCHASTIC MODELING

Let us consider an undirected and connected graph G composed of N nodes, and denote by V and E the vertex and edge set, respectively. Let the vector $\boldsymbol{x} \in \mathbb{R}^N$ represent the signal on the graph G, where the i-th entry x_i is the signal component relative to node i. The *graph Fourier transform* (GFT) expands the signal $\boldsymbol{x}$ into the graph frequency domain: the forward and inverse GFTs of $\boldsymbol{x}$ are $\hat{x}_n = \langle \boldsymbol{x}, \boldsymbol{\phi}_n \rangle$ and $x_n = \langle \hat{\boldsymbol{x}}, \boldsymbol{\phi}_n \rangle$, where $\langle , \rangle$ denotes the inner product. The vectors $\{\boldsymbol{\phi}_n\}_{n=1}^N$ form an orthonormal basis and are commonly chosen as the eigenvectors of a graph Laplacian. To avoid any restrictions on the generality of our approach, in the following we present our results for a *general basis matrix* $\boldsymbol{L}$. We only require that $\boldsymbol{L}$ is *symmetric* and *1-local*: for all $i \neq j$, $\boldsymbol{L}_{ij} = 0$ whenever the couple (i, j) is not in the edge set E, i.e., there is no link between node i and node j.

A *graph filter* is a linear operator that acts on the graph frequency component of the input signal attenuating some of the frequency components and amplifying others. The output signal $\boldsymbol{y}$ can be expressed as

$$\boldsymbol{y} = \sum_{n=1}^{N} h(\lambda_n)\, \hat{x}_n \boldsymbol{\phi}_n, \tag{1}$$

where $h(\lambda_n)$ is the graph frequency response of the graph filter for a given frequency λ_n. In this paper, we are interested in implementing a desired *response* $h^*(\lambda)$ in a distributed fashion. In particular, we start our analysis from the distributed

This paper present the extended version of [1]. This research was supported in part by STW under the D2S2 project from the ASSYS program (project 10561).

ARMA filter design in [12], and we take a step further, by considering stochastic time-varying realizations of both the graph and input signal.

Stochastic model. We assume the following stochastic model:

Assumption 1: (Signal) The input signal $x_{i,k}$ at each node i and time instant k is a realization of an i.i.d. random process in time $\mathcal{P}_i$ with first order moment $\bar{x}_i$ and second order moment σ_i^2. The random process at different nodes does not have to be i.i.d for a fixed time instant.

Assumption 2: (Graph) The probability that a link (i,j) in the edge set E is activated at time k is p, with $0 < p \leq 1$. The edges are activated independently across time, with $\boldsymbol{L}_k$ denoting the graph Laplacian at time instant k. Graph realizations are considered mutually independent with the graph signal process.

Assumption 3: (Basis) There exist lower/upper bounds, uniform in time, on the eigenvalues of $\{\boldsymbol{L}_k\}$, i.e., $\lambda_{\text{min}} \leq \lambda(\boldsymbol{L}_k) \leq \lambda_{\text{max}}$ for all k.

Assumption 1 is a weak assumption on the nature of the input signal, which generalizes the deterministic signals analyzed in current literature. Assumption 2 is quite standard in the literature on network algorithms [13], [14]; it means that, at each time step k, we draw a realization of the edge set $E_k \subseteq E$ generated via an i.i.d. Bernoulli process. Let us refer from now on by $\boldsymbol{L}$ to the Laplacian relative to the graph E, and by $\boldsymbol{L}_k$ to the Laplacian of E_k. Given realization E_k, each node locally derives the instantaneous basis $\boldsymbol{L}_k$ by communicating with its neighbors. For convenience, denote the expected basis $\mathbb{E}[\boldsymbol{L}_k]$ as $\bar{\boldsymbol{L}}$. Assumption 3, which concerns the basis realization $\boldsymbol{L}_k$, is not restrictive; lower and upper bounds are usually easy to find: For normalized Laplacians, $\lambda_{\text{min}} = 0$ and $\lambda_{\text{max}} = 2$, whereas for standard Laplacians[1], $\lambda_{\text{min}} = 0$ and $\lambda_{\text{max}} = \lambda_{\text{max}}(\boldsymbol{L})$. Several (finite) upper bounds for $\lambda_{\text{max}}(\boldsymbol{L})$ are known: for instance, $\lambda_{\text{max}}(\boldsymbol{L}) \leq \max\{d(u) + d(v) | (u,v) \in E\}$, where $d(u)$ is the degree at node u.

ARMA filters. With this in place, we study the recursion

$$\boldsymbol{y}_k = \psi \boldsymbol{M}_k \boldsymbol{y}_{k-1} + \varphi \boldsymbol{x}_k \text{ and } \boldsymbol{y}_0 \text{ arbitrary}, \tag{2}$$

which is a stochastic extension of the potential kernel [16] and is indexed as an ARMA_1 graph filter in [12]. Graph signals $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$ are the input and output of the filter at time k. The coefficients ψ and φ are arbitrary real numbers which influence the filter graph frequency response, i.e., how well we approximate $h^*(\lambda)$. Matrix $\boldsymbol{M}_k$ is a shifted version of the instantaneous basis $\boldsymbol{L}_k$: $\boldsymbol{M}_k = \frac{\lambda_{\text{max}}+\lambda_{\text{min}}}{2}\mathbf{I} - \boldsymbol{L}_k$, such that $\boldsymbol{M}_k$ has a reduced maximum eigenvalue (which helps for the filter design). From Assumption 3, all the realizations $\boldsymbol{M}_k$ enjoy lower and upper bounds on their eigenvalues as $\mu_{\text{min}}\mathbf{I} \leq \boldsymbol{M}_k \leq \mu_{\text{max}}\mathbf{I}$. And in particular, their spectral norm is bounded as

$$\|\boldsymbol{M}_k\| \leq M = \max\{|\mu_{\text{min}}|, |\mu_{\text{max}}|\}. \tag{3}$$

[1]From the interlacing properties of the standard Laplacian [15], it follows that $\lambda_{\text{max}}(\boldsymbol{L}_k) \leq \lambda_{\text{max}}(\boldsymbol{L})$.

For example, if the $\boldsymbol{L}_k$'s are normalized Laplacians, we can set $M = 1$. We further indicate with $\bar{\boldsymbol{M}}$ the expected value of $\boldsymbol{M}_k$. A word is in order for our choice of the iteration matrix $\boldsymbol{M}_k$. The design of $\boldsymbol{M}_k$ in a shifted version does not influence our results. By Sylvester's matrix theorem, matrices $\boldsymbol{M}_k$ and $\boldsymbol{L}_k$ have the same eigenvectors and the eigenvalues $\mu_{n,k}$ of $\boldsymbol{M}_k$ are related to the eigenvalues $\lambda_{n,k}$ of $\boldsymbol{L}_k$ by $\mu_{n,k} = (\lambda_{\text{max}} + \lambda_{\text{min}})/2 - \lambda_{n,k}$.

III. ARMA_1 FILTERS IN THE MEAN

For simplicity of presentation, we consider ARMA_1 filters, like the one in (2). Our approach however is straightforwardly extended to higher-order graph filters: as shown in [12], one implements a Kth order filter by running (and linearly combining) K ARMA_1 filters in parallel.

Central to our filter design is the stability condition

$$|\psi| M < 1, \tag{4}$$

imposed in the filter design phase, i.e., $|\psi| < 1/M$.

We are now ready to show that the ARMA_1 recursion in (2) under Assumptions 1, 2, and 3 and the stability condition (4) behaves as an ARMA_1 filter in the mean. This is encoded in the following theorem.

Theorem 1: Let Assumptions 1–3 hold as well as the stability condition (4). The steady state value of the expected value of the ARMA_1 recursion (2) is given by

$$\bar{\boldsymbol{y}} := \lim_{k\to\infty} \mathbb{E}[\boldsymbol{y}_k] = \varphi(\mathbf{I} - \psi\mathbb{E}[\boldsymbol{M}_k])^{-1}\mathbb{E}[\boldsymbol{x}_k] = \varphi(\mathbf{I} - \psi\bar{\boldsymbol{M}})^{-1}\bar{\boldsymbol{x}}, \tag{5}$$

where $\bar{\boldsymbol{x}}$ is the vector containing the first order moments $\bar{x}_i$.

Proof: Define the transition matrix $\boldsymbol{\Phi}(t', t) := \boldsymbol{M}_{t'} \cdots \boldsymbol{M}_t$, for $t' \geq t$, and $\boldsymbol{\Phi}(t', t) := \mathbf{I}$ if $t' < t$. We can then write recursion (2) at time instant k as

$$\boldsymbol{y}_k = \psi^k \boldsymbol{\Phi}(k, 1)\boldsymbol{y}_0 + \varphi \sum_{t=0}^{k-1} \psi^t \boldsymbol{\Phi}(k, k-t+1)\boldsymbol{x}_{k-t}, \tag{6}$$

while the steady state value of its expectation reads

$$\lim_{k\to\infty} \mathbb{E}[\boldsymbol{y}_k] = \lim_{k\to\infty} \mathbb{E}\Big[\psi^k \boldsymbol{\Phi}(k,1)\boldsymbol{y}_0 + \varphi \sum_{t=0}^{k-1} \psi^t \boldsymbol{\Phi}(k, k-t+1)\boldsymbol{x}_{k-t}\Big]. \tag{7}$$

By using linearity of the expectation operator and the limit, we expand (7) as

$$\lim_{k\to\infty} \mathbb{E}[\boldsymbol{y}_k] = \lim_{k\to\infty} \mathbb{E}\big[\psi^k \boldsymbol{\Phi}(k,1)\boldsymbol{y}_0\big] + \varphi \sum_{t=0}^{k-1} \lim_{k\to\infty} \mathbb{E}\big[\psi^t \boldsymbol{\Phi}(k, k-t+1)\boldsymbol{x}_{k-t}\big]. \tag{8}$$

We now use the independence of the graph realizations and graph signals, to simplify (8) into

$$\lim_{k\to\infty} \mathbb{E}[\boldsymbol{y}_k] = \lim_{k\to\infty} \psi^k \mathbb{E}[\boldsymbol{M}_k]^k \boldsymbol{y}_0 + \lim_{k\to\infty} \varphi \sum_{t=0}^{k-1} \psi^t \mathbb{E}[\boldsymbol{M}_k]^t \mathbb{E}[\boldsymbol{x}_{k-t}]. \tag{9}$$

We then make use of two properties of the spectral norm $\|\cdot\|$ of a square matrix, namely sub-multiplicativity, i.e., $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$ (See Chapter 3 in [17]), and convexity, i.e., $\|\mathbb{E}[\boldsymbol{A}]\| \leq \mathbb{E}[\|\boldsymbol{A}\|]$, to bound the maximum eigenvalue of the matrix $\psi^s \mathbb{E}[\boldsymbol{M}_k]^s$, for any exponent $s > 0$. This will allow us to derive the claim of the theorem. By the sub-multiplicativity property of the spectral norm, we can write

$$\|\psi^s \mathbb{E}[\boldsymbol{M}_k]^s\| \leq \|\psi \mathbb{E}[\boldsymbol{M}_k]\|^s \leq (|\psi| \|\mathbb{E}[\boldsymbol{M}_k]\|)^s; \tag{10}$$

by using now the convexity of the spectral norm, we can upper bound the last expression as

$$(|\psi| \|\mathbb{E}[\boldsymbol{M}_k]\|)^s \leq (|\psi| \mathbb{E}[\|\boldsymbol{M}_k\|])^s \leq (|\psi| M)^k < 1, \tag{11}$$

where in the last two inequalities, we have used the spectral condition (3), and the stability criterium (4), respectively. By putting together the bounds (11) and (10), we obtain

$$\|\psi^s \mathbb{E}[\boldsymbol{M}_k]^s\| < 1, \quad \text{for all } s > 0. \tag{12}$$

Due to (12) for $s = k$, the first term of the right-hand side of (9) vanishes in the limit, i.e.,

$$\lim_{k\to\infty} \psi^k \mathbb{E}[\boldsymbol{M}_k]^k \boldsymbol{y}_0 = \mathbf{0}. \tag{13}$$

As for the second term of the right-hand side of (9), due to the norm condition (12) for $s = 1$, we can compute the infinite sum as

$$\lim_{k\to\infty} \varphi \sum_{t=0}^{k-1} \psi^t \mathbb{E}[\boldsymbol{M}_k]^t \mathbb{E}[\boldsymbol{x}_{k-t}] = \varphi(\mathbf{I} - \psi \mathbb{E}[\boldsymbol{M}_k])^{-1} \mathbb{E}[\boldsymbol{x}_k]. \tag{14}$$

By substituting (13) and (14) into (9) the claim is proven. ■

Theorem 1 says that the expected value of the steady state output is only influenced by the expected value of the signal on the graph and by the expected value of the graph distribution. An important feature is that the steady state output of the graph filter is not influenced in the mean by the graph topology changes. Furthermore, given its final expression is (5), we can readily use the conclusions of Theorem 3 in [16], in which P is replaced by $\bar{\boldsymbol{M}}$ and $1 - \varphi$ by ψ, to conclude that the stochastic recursion (2) acts as an ARMA$_1$ graph filter in the mean.

IV. Variance of the expected ARMA$_1$ filter

We proceed to characterize the variance of the filter output in (2), and as such to quantify how far from the mean a given realization can be. First of all, we derive in closed form the variance in the case of a static deterministic graph and a stochastic time varying signal. For the general and more involved case where both the graph and the signal are stochastic, we derive an upper bound on the average variance across all nodes.

Deterministic graph, stochastic signal. In this scenario, recursion (2) (and (6)) simplifies into

$$\boldsymbol{y}_k = \psi^k \boldsymbol{M}^k \boldsymbol{y}_0 + \varphi \sum_{t=0}^{k-1} \psi^t \boldsymbol{M}^t \boldsymbol{x}_{k-t}, \tag{15}$$

which leads to the following result in terms of expected value and covariance matrix.

Theorem 2: Let Assumptions 1 and 3 hold true, as well as the stability condition (4). Consider the graph filter (15) with a static deterministic graph and a stochastic time-varying signal and denote by $\boldsymbol{\Sigma_{xx}}$ the covariance matrix of the input $\boldsymbol{x}_k$, which is diagonal with diagonal entries σ_i^2. The steady state of the expected value $\bar{\boldsymbol{y}}$, and the limiting covariance matrix $\boldsymbol{\Sigma_{yy}} = \lim_{k\to\infty} \mathbb{E}[\boldsymbol{y}_k \boldsymbol{y}_k^\mathsf{T}]$ are respectively

$$\bar{\boldsymbol{y}} = \varphi(\mathbf{I} - \psi \boldsymbol{M})^{-1} \bar{\boldsymbol{x}} \tag{16}$$

$$\boldsymbol{\Sigma_{yy}} = \varphi^2 \sum_{t=0}^{\infty} \psi^{2t} \boldsymbol{M}^t \boldsymbol{\Sigma_{xx}} (\boldsymbol{M}^t)^\mathsf{T}. \tag{17}$$

Proof: (Sketch) The claim can be derived by computing the covariance of $\boldsymbol{y}_k$ in (15) and remembering the independence of the initial condition $\boldsymbol{y}_0$ and the input signal $\boldsymbol{x}_k$. Then, by taking the limit for $k \to \infty$, equation (17) follows. ■

The results of Theorem 2 are not surprising, considering that the graph filter (2) is a linear operator. In fact, directly from linear system theory, the covariance expression (17) is the unique solution of the discrete Lyapunov equation

$$\boldsymbol{\Sigma_{yy}} = \psi^2 \boldsymbol{M} \boldsymbol{\Sigma_{yy}} \boldsymbol{M}^\mathsf{T} + \varphi^2 \boldsymbol{\Sigma_{xx}}. \tag{18}$$

It tells us that the covariance of the steady state is directly related to the covariance of the input signal with the graph matrix $\boldsymbol{M}$, and it is independent of the particular signal realizations.

Stochastic graph, stochastic signal. For the general case, the following theorem gives a constructive proof on how to upper bound the average variance of the output $\boldsymbol{y}$ at steady state.

Theorem 3: Under the same assumptions and definitions of Theorem 1, define the limiting average variance experienced at each node as

$$\lim_{k\to\infty} \overline{\mathrm{Var}}[\boldsymbol{y}_k] = \lim_{k\to\infty} \mathrm{tr}(\mathbb{E}[\boldsymbol{y}_k \boldsymbol{y}_k^\mathsf{T}] - \mathbb{E}[\boldsymbol{y}_k]\mathbb{E}[\boldsymbol{y}_k]^\mathsf{T})/N, \tag{19}$$

where $\mathrm{tr}(\cdot)$ indicates the trace operator. Let $\boldsymbol{\Sigma_{xx}}$ be the covariance matrix of the input $\boldsymbol{x}_k$. Then, $\lim_{k\to\infty} \overline{\mathrm{Var}}[\boldsymbol{y}_k]$ is upper bounded as

$$\lim_{k\to\infty} \overline{\mathrm{Var}}[\boldsymbol{y}_k] \leq \frac{\varphi^2}{N} \mathrm{tr}\left(\frac{\boldsymbol{\Sigma_{xx}} + \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^\mathsf{T}}{(1 - |\psi| M)^2} - \bar{\boldsymbol{y}}\bar{\boldsymbol{y}}^\mathsf{T} \right). \tag{20}$$

Proof: To ease notation, define $\boldsymbol{\Phi}_q := \boldsymbol{\Phi}(k, k-q+1)$. Due to the expression of $\boldsymbol{y}_k$ given in (6), the limiting trace of the covariance matrix $\lim_{k\to\infty} \mathrm{tr}(\mathbb{E}[\boldsymbol{y}_k \boldsymbol{y}_k^\mathsf{T}])$ can then be written as

$$\lim_{k\to\infty} \mathrm{tr}(\mathbb{E}[\boldsymbol{y}_k \boldsymbol{y}_k^\mathsf{T}]) = \lim_{k\to\infty} \mathrm{tr}\Bigg(\mathbb{E}\Bigg[\Big(\psi^k \boldsymbol{\Phi}_k \boldsymbol{y}_0 + \varphi \sum_{t=0}^{k-1} \psi^t \boldsymbol{\Phi}_t \boldsymbol{x}_{k-t}\Big) \times \Big(\psi^k \boldsymbol{\Phi}_k \boldsymbol{y}_0 + \varphi \sum_{l=0}^{k-1} \psi^l \boldsymbol{\Phi}_l \boldsymbol{x}_{k-l}\Big)^\mathsf{T}\Bigg]\Bigg). \tag{21}$$

By expanding the right-hand side, and by using the linearity of the expectation, the trace, and the limit, we rewrite (21) as the sum of four terms, namely the limiting trace of the covariance of the initial state with itself

$$\lim_{k\to\infty} \mathrm{tr}(\mathbb{E}[\psi^{2k}\boldsymbol{\Phi}_k \boldsymbol{y}_0 \boldsymbol{y}_0^\mathsf{T} \boldsymbol{\Phi}_k^\mathsf{T}]), \tag{22a}$$

the limiting trace of the covariance of the initial state with the graph signal

$$\lim_{k\to\infty} \mathrm{tr}\Big(\mathbb{E}\Big[\varphi\psi^k \boldsymbol{\Phi}_k \boldsymbol{y}_0 \sum_{l=0}^{k} \psi^l \boldsymbol{x}_{k-l}^\mathsf{T} \boldsymbol{\Phi}_l^\mathsf{T}\Big]\Big), \tag{22b}$$

the transpose of the term above, and the limiting trace of the covariance of the graph signal with itself

$$\lim_{k\to\infty} \mathrm{tr}\Big(\mathbb{E}\Big[\varphi^2 \sum_{t=0}^{k} \psi^t \boldsymbol{\Phi}_t \boldsymbol{x}_{k-t} \sum_{l=0}^{k} \psi^l \boldsymbol{x}_{k-l}^\mathsf{T} \boldsymbol{\Phi}_l^\mathsf{T}\Big]\Big). \tag{22c}$$

By leveraging the cyclic property of the trace, i.e., $\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{CAB})$, the linearity of the expectation and the trace, and the independence of graph realizations, graph signals, and initial condition, we can rearrange the terms (22) as

$$\lim_{k\to\infty} \mathrm{tr}(\mathbb{E}[\psi^{2k}\boldsymbol{\Phi}_k^\mathsf{T} \boldsymbol{\Phi}_k]\mathbb{E}[\boldsymbol{y}_0 \boldsymbol{y}_0^\mathsf{T}]), \tag{23a}$$

$$\lim_{k\to\infty} \varphi \sum_{l=0}^{k} \mathrm{tr}\Big(\mathbb{E}[\psi^{l+k}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_k]\mathbb{E}[\boldsymbol{y}_0 \boldsymbol{x}_{k-l}^\mathsf{T}]\Big), \tag{23b}$$

$$\lim_{k\to\infty} \varphi^2 \sum_{t=0}^{k}\sum_{l=0}^{k} \mathrm{tr}\Big(\mathbb{E}[\psi^{t+l}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_t]\mathbb{E}[\boldsymbol{x}_{k-t} \boldsymbol{x}_{k-l}^\mathsf{T}]\Big). \tag{23c}$$

In order to simplify (23), we make use of the following property of the trace,

$$\mathrm{tr}(\boldsymbol{AB}) \le \frac{\|\boldsymbol{A}+\boldsymbol{A}^\mathsf{T}\|}{2}\mathrm{tr}(\boldsymbol{B}) \le \|\boldsymbol{A}\|\mathrm{tr}(\boldsymbol{B}) \tag{24}$$

valid for any square matrix $\boldsymbol{A}$ and positive semidefinite matrix $\boldsymbol{B}$ of appropriate dimension [18]. This inequality, along with the convexity and sub-multiplicative of the spectral norm $\|\cdot\|$ will help us to prove the claim.

We start with (23a). Since the matrix $\mathbb{E}[\boldsymbol{y}_0\boldsymbol{y}_0^\mathsf{T}]$ is positive semidefinite matrix, we can apply the inequality (24) yielding

$$\lim_{k\to\infty} \mathrm{tr}(\mathbb{E}[\psi^{2k}\boldsymbol{\Phi}_k^\mathsf{T} \boldsymbol{\Phi}_k]\mathbb{E}[\boldsymbol{y}_0 \boldsymbol{y}_0^\mathsf{T}]) \le \lim_{k\to\infty} \|\mathbb{E}[\psi^{2k}\boldsymbol{\Phi}_k^\mathsf{T} \boldsymbol{\Phi}_k]\|\mathrm{tr}(\mathbb{E}[\boldsymbol{y}_0 \boldsymbol{y}_0^\mathsf{T}]). \tag{25}$$

The spectral norm is convex, i.e., $\|\mathbb{E}[\boldsymbol{A}]\| \le \mathbb{E}[\|\boldsymbol{A}\|]$, and sub-multiplicative, i.e., $\|\boldsymbol{AB}\| \le \|\boldsymbol{A}\|\|\boldsymbol{B}\|$, thus we can bound the term above as

$$\lim_{k\to\infty} \|\mathbb{E}[\psi^{2k}\boldsymbol{\Phi}_k^\mathsf{T} \boldsymbol{\Phi}_k]\|\mathrm{tr}(\mathbb{E}[\boldsymbol{y}_0 \boldsymbol{y}_0^\mathsf{T}]) \le \lim_{k\to\infty} |\psi^{2k}|\mathbb{E}[\|\boldsymbol{M}_k\|]^{2k}\mathrm{tr}(\mathbb{E}[\boldsymbol{y}_0 \boldsymbol{y}_0^\mathsf{T}]) = 0, \tag{26}$$

where the last result comes from the spectral bounds (3) and stability condition (4).

We can now focus on (23b) (and its transpose). By using (24) and the fact that $\mathrm{tr}(\boldsymbol{A}) \le n\|\boldsymbol{A}\|$, if the matrix $\boldsymbol{A}$ is of dimension $n \times n$, then

$$\lim_{k\to\infty} \varphi \sum_{l=0}^{k} \mathrm{tr}\Big(\mathbb{E}[\psi^{l+k}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_k]\mathbb{E}[\boldsymbol{y}_0 \boldsymbol{x}_{k-l}^\mathsf{T}]\Big) \le \lim_{k\to\infty} \varphi \sum_{l=0}^{k} N\|\mathbb{E}[\psi^{l+k}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_k]\|\|\mathbb{E}[\boldsymbol{y}_0]\mathbb{E}[\boldsymbol{x}_{k-l}^\mathsf{T}]\|. \tag{27}$$

Once again by the sub-multiplicativity and convexity of the spectral norm, we can show that in the limit [due to (3) and (4)]

$$\lim_{k\to\infty} \|\mathbb{E}[\psi^{l+k}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_k]\| \le \lim_{k\to\infty} |\psi|^{l+k} M^{l+k} = 0, \tag{28}$$

which makes the term (27) and therefore (23b) and its transpose vanish.

We are now left with (23c). We study the matrix $\mathbb{E}[\boldsymbol{x}_{k-t}\boldsymbol{x}_{k-l}^\mathsf{T}]$ for different t and l. In particular,

$$\mathbb{E}[\boldsymbol{x}_{k-t}\boldsymbol{x}_{k-l}^\mathsf{T}] = \begin{cases} \boldsymbol{\Sigma}_{\boldsymbol{xx}}, & \text{if } l = t, \\ \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^\mathsf{T}, & \text{otherwise} \end{cases}, \tag{29}$$

which makes the matrix $\mathbb{E}[\boldsymbol{x}_{k-t}\boldsymbol{x}_{k-l}^\mathsf{T}]$ positive semidefinite. We can therefore make use of the trace inequality (24) bounding the term (23c) as

$$\lim_{k\to\infty} \varphi^2 \sum_{t=0}^{k}\sum_{l=0}^{k} \mathrm{tr}\Big(\mathbb{E}[\psi^{t+l}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_t]\mathbb{E}[\boldsymbol{x}_{k-t} \boldsymbol{x}_{k-l}^\mathsf{T}]\Big) \le \lim_{k\to\infty} \varphi^2 \sum_{t=0}^{k}\sum_{l=0}^{k} \|\mathbb{E}[\psi^{t+l}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_t]\|\mathrm{tr}\Big(\mathbb{E}[\boldsymbol{x}_{k-t} \boldsymbol{x}_{k-l}^\mathsf{T}]\Big). \tag{30}$$

By using once again the properties of the spectral norm and (3)-(4) as we did before, we bound,

$$\|\mathbb{E}[\psi^{t+l}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_t]\| \le \mathbb{E}[\|\psi^{t+l}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_t\|] \le (|\psi|M)^{t+l}, \tag{31}$$

while by leveraging (29), we upper bound

$$\mathrm{tr}\Big(\mathbb{E}[\boldsymbol{x}_{k-t} \boldsymbol{x}_{k-l}^\mathsf{T}]\Big) \le \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{xx}} + \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^\mathsf{T}), \tag{32}$$

and therefore

$$\lim_{k\to\infty} \varphi^2 \sum_{t=0}^{k}\sum_{l=0}^{k} \|\mathbb{E}[\psi^{t+l}\boldsymbol{\Phi}_l^\mathsf{T} \boldsymbol{\Phi}_t]\|\mathrm{tr}\Big(\mathbb{E}[\boldsymbol{x}_{k-t} \boldsymbol{x}_{k-l}^\mathsf{T}]\Big) \le \lim_{k\to\infty} \varphi^2 \sum_{t=0}^{k}\sum_{l=0}^{k} (|\psi|M)^{t+l}\mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{xx}} + \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^\mathsf{T}). \tag{33}$$

By computing the infinite sum (33), and putting together this result with the ones obtained for (23a), (23b), and its transpose, we conclude that

$$\lim_{k\to\infty} \mathrm{tr}(\mathbb{E}[\boldsymbol{y}_k \boldsymbol{y}_k^\mathsf{T}]) \le \varphi^2\, \mathrm{tr}\left(\frac{\boldsymbol{\Sigma}_{\boldsymbol{xx}} + \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^\mathsf{T}}{(1-|\psi|M)^2}\right). \tag{34}$$

By the definition of the limiting variance $\lim_{k\to\infty} \overline{\mathrm{Var}}[\boldsymbol{y}_k]$, the linearity of the trace, and the expression of $\lim_{k\to\infty} \mathbb{E}[\boldsymbol{y}_k]$ given in Theorem 1, the claim (20) follows. ■

Theorem 2 describes a bound on the average variance at node i. When the stochasticity is limited, meaning that the link

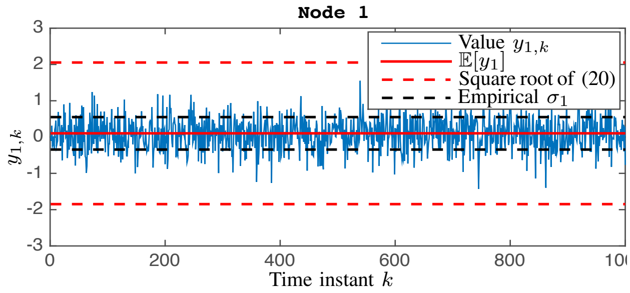

Fig. 1. Filter output of node 1 for different time instants k. The empirical standard deviation and the square root of the bound are centered with respect to the theoretical expected value, $\mathbb{E}[y_1] = 0.15$.

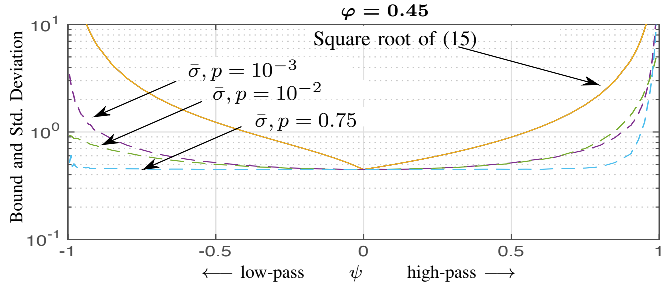

Fig. 2. Square root of (20), and the empirical average standard deviation, $\bar{\sigma}$, both considered for different values of p. The results are plotted as a function of $\psi \in [-1/M, 1/M]$. The ARMA$_1$ filter is low-pass for $\psi < 0$ and high-pass for $\psi > 0$.

activation probability p is close to 1 and the variance on the signal $\boldsymbol{x}_k$ is low, then the average variance bound is expected to be small (because M bounds in a tighter way matrices $\boldsymbol{M}_k$, and $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ is close to zero). On the other hand, when the stochasticity is high, the two terms on the right-hand side of (20) differ more and the average variance bound is expected to be higher. Note however that the result in (20) is still a bound and it does not go to zero for the deterministic case. Nonetheless, as shown in the numerical simulation section, this bound can be tight and can therefore be useful in selecting the value of ψ.

V. Numerical Evaluation

To illustrate our results, we consider an undirected graph G of $N = 100$ nodes, with edge set E. We analyze the graph for different link probabilities p constant for all edges in E, an initial state $\boldsymbol{y}_0 = \boldsymbol{0}$, and select the normalized Laplacian as our basis matrix (thus $M = 1$). The input signal is assumed normal distributed with $\mathbb{E}[x_{i,k}] = 1$ and a diagonal covariance matrix with $\sigma_i^2 = 1$.

We simulate the ARMA$_1$ filter (2), where the coefficients ψ and φ have been found according to the filter design proposed in [12] aiming at approximating an ideal low-pass filter with pass-band $[0, 1]$ and suppressing to zero all higher frequencies. Figure 1 displays the analytical expected value of the steady state and one realization of the output signal as a function of time, plotted for node 1. In this case p is considered 0.5. The empirical standard deviation of the output signal and the calculated bound are also shown. We can see that the output signal fluctuates around the theoretical mean, which is in line with the results of Theorem 1. Furthermore, in this case the bound is quite tight to the empirical standard deviation value.

To examine the influence of the graph stochasticity p, we consider two extreme values of graph connectivity $p = 10^{-3}, p = 10^{-2}$, and one case where the graph is mostly connected, $p = 0.75$. In Figure 2, we plot the square root of the bound (20) as a function of ψ, and the average standard deviation for the steady state calculated empirically. The figure illustrates that, though it is true that the variance can grow very large, in most cases of interest, the ARMA$_1$ filter is not significantly affected by variations. Note that in order to decrease the upper bound of the variance one could tune the parameters ψ and also φ in the filter design (e.g., by trading-off approximation accuracy, high $|\psi|$, with low variance, low $|\psi|$). This aspect, along with the generalization to higher-order and periodic ARMA filters, is left for future work.

References

[1] E. Isufi, A. Simonetto, A. Loukas, and G. Leus, "Stochastic Graph Filtering on Time-Varying Graphs," in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2015.

[2] A. Sandryhaila and J. M. Moura, "Discrete Signal Processing on Graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[3] ——, "Discrete Signal Processing on Graphs: Frequency Analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, 2014.

[4] 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 Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[5] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovacevic, "Signal Denoising on Graphs via Graph Filtering," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, USA, December 2014, pp. 872–876.

[6] F. Zhang and E. R. Hancock, "Graph Spectral Image Smoothing Using the Heat Kernel," *Pattern Recognition*, vol. 41, no. 11, pp. 3328–3342, 2008.

[7] A. Loukas, M. A. Zúñiga, I. Protonotarios, and J. Gao, "How to Identify Global Trends from Local Decisions? Event Region Detection on Mobile Networks," in *International Conference on Computer Communications (INFOCOM)*, 2014.

[8] A. Sandryhaila, S. Kar, and J. M. Moura, "Finite-Time Distributed Consensus Through Graph Filters," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 1080–1084.

[9] D. I. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev Polynomial Approximation for Distributed Signal Processing," in *IEEE International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, 2011, pp. 1–8.

[10] S. Safavi and U. A. Khan, "Revisiting Finite-Time Distributed Algorithms via Successive Nulling of Eigenvalues," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 54–57, 2015.

[11] X. Shi, H. Feng, M. Zhai, T. Yang, and B. Hu, "Infinite Impulse Response Graph Filters in Wireless Sensor Networks," *IEEE Signal Processing Letters*, Jan 2015.

[12] A. Loukas, A. Simonetto, and G. Leus, "Distributed Autoregressive Moving Average Graph Filters," *IEEE Signal Processing Letters*, 2015, arXiv:1508.05808.

[13] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, "Asynchronous Distributed Optimization using a Randomized Alternating Direction Method of Multipliers," in *52nd IEEE Conference on Decision and Control*, Firenze, Italy, December 2013.

[14] E. Wei and A. Ozdaglar, "On the $O(1/k)$ Convergence of Asynchronous Distributed Alternating Direction Method of Multipliers," *Submitted to Mathematical Programming*, 2013, available at arXiv:1307.8254.

[15] G. Chen, G. Davis, F. Hall, Z. Li, K. Patel, and M. Stewart, "An Interlacing Result on Normalized Laplacians," *SIAM Journal on Discrete Mathematics*, vol. 18, no. 2, pp. 353–361, 2004.

[16] A. Loukas, M. A. Zúñiga, M. Woehrle, M. Cattani, and K. Langendoen, "Think Globally, Act Locally: On the Reshaping of Information Landscapes," in *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2013.

[17] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, 1991.

[18] T. Mori, "Comments on "A Matrix Inequality Associated with Bounds on Solutions of Algebraic Riccati and Lyapunov Equation"," *IEEE Transactions on Automatic Control*, vol. 33, no. 11, p. 1088, 1988.