Delft University of Technology

Observing and tracking bandlimited graph processes from sampled measurements

Isufi, Elvin; Banelli, Paolo; Di Lorenzo, Paolo; Leus, Geert

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Observing and tracking bandlimited graph processes from sampled measurements☆

Elvin Isufi[a],*, Paolo Banelli[b], Paolo Di Lorenzo[c], Geert Leus[a]

[a] *Faculty of EEMCS, Delft University of Technology, Delft 2628 CD, the Netherlands*
[b] *Department of Engineering, University of Perugia, 06125 Perugia, Italy*
[c] *Department of Information Engineering, Electronics, and Telecommunications, Sapienza University of Rome, 00184 Rome, Italy*

ABSTRACT

A critical challenge in graph signal processing is the sampling of bandlimited graph signals; signals that are sparse in a well-defined graph Fourier domain. Current works focused on sampling time-invariant graph signals and ignored their temporal evolution. However, time can bring new insights on sampling since sensor, biological, and financial network signals are correlated in both domains. Hence, in this work, we develop a sampling theory for time varying graph signals, named *graph processes*, to observe and track a process described by a linear state-space model. We provide a mathematical analysis to highlight the role of the graph, process bandwidth, and sample locations. We also propose sampling strategies that exploit the coupling between the topology and the corresponding process. Numerical experiments corroborate our theory and show the proposed methods trade well the number of samples with accuracy.

© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

Graph signal processing (GSP) provides new tools to analyze signals on networks such as measurements in sensor networks, fMRI recordings in brain networks, and preferences in opinion networks [1]. By introducing the notion of graph frequency, GSP allows for a harmonic decomposition of these signals [1,2]. Recent contributions have therefore developed concepts like graph signal filtering [2–6] and sampling [7–11] to support this analysis. Sampling theory for graph signals assumes the latter are often sparse (or bandlimited) in the graph frequency domain. This assumption is satisfied in temperature measurements in sensor networks [8], fMRI data in brain networks [12], and user ratings in recommendation networks [13], to name a few. The works in [8,10,11,14,15] exploit signal bandlimitedness to propose sampling and reconstruction strategies for graph signals.

While seminal, the above works ignore the signal temporal evolution and focus on a single snapshot. In our opinion, this is limited because time varying graph signals, referred to as *graph processes*, are often encountered in practice, e.g., in consecutive sensor measurements, biological signal evolution prone to stimuli, and information diffusion over networks. We wish, therefore, to propose a graph and time sampling framework to observe and track a bandlimited graph process–a time varying graph signal that has contribution only on a fixed subset of graph frequencies. Before detailing our contributions, we highlight the differences with earlier works that dealt with graph processes.

1.1. Related works

The importance to approach graph processes with GSP was considered in [16], which formalized a graph-time harmonic analysis. The works in [17,18] discuss the two dimensional graph-time filters, while [19,20] used these filters to predict the process evolution. The work in [21] considered sampling strategies for adaptive reconstruction of a bandlimited graph signal from a stream of noisy samples. We build on the latter and propose a more involved sampling theory to observe and track a graph process that follows a linear model.

Observability of network processes was considered in [22,23] for sensor placement and in [24] for designing observable

topologies. However, these works do not discuss sampling. We here do so, where by exploiting the bandlimited prior, we derive theoretical guarantees for when a graph process is observable from a few samples. We also propose sampling schemes to collect the measurements over the graph and the temporal dimension.

Other works that exploited the bandlimited prior for graph processes are [21,25–27]. These works propose adaptive algorithms to track slowly time varying graph signals from a few nodes. But since their main goal is the adaptive signal reconstruction, tracking yields as a byproduct without theoretical guarantees. We instead leverage Kalman filtering and propose sampling strategies to optimally track the process. Our motivation to use the Kalman filter comes from its proven success in applications like network dynamics data fusion, and target tracking –see e.g., [28] and references therein. Kalman filtering for tracking graph processes has also been used in [29]. This work formulated tracking from a few measurements as a regularized interpolation problem on a larger graph; the Kalman filer is used here to alleviate the computational burden. We identify three main differences with [29]: (*i*) we derive conditions on the minimum number of samples required to track the process; (*ii*) we propose sampling schemes optimized to improve tracking and not uniformly at random sampling; (*iii*) we use the Kalman filter to match the process model and avoid cross-validating the regularizer.

Kalman filtering has been used to track network dynamics in [30,31]. The graph topology embeds the communication links between sensors and is disconnected from the process over it. To the best of our knowledge, this work represents the first attempt to conciliate tracking with graph sampling theory.

1.2. Contribution

We group the paper contributions into two parts.

(*i*) *Observability of graph processes* (Section 4): We develop a sampling theory to observe graph processes. We derive *necessary* and *sufficient* conditions to observe the process from a few samples. We propose two approaches for observability: (*i-a*) *observability with deterministic sampling, i.e., when the nodes are sampled deterministically* (Section 4.1); (*i-b*) *observability with random sampling, i.e., when the nodes are sampled with a given probability* (Section 4.2). We also perform a mean-square error (MSE) analysis of the estimated state to connect the graph topology with the process bandwidth and sampling set. Finally, we develop sparse sensing techniques to pick the minimum number of samples that guarantee a target MSE performance.

(*ii*) *Tracking of graph processes (Section V):* We propose Kalman filtering to track bandlimited graph processes that follow a linear dynamic model. First, we consider time varying models (Section 5.1) and then we specialize our derivations to time-invariant ones (Section 5.2). We derive *necessary* conditions on the minimum number of nodes required to track the process and conciliate these conditions with those for observability. The MSE tracking performance, given by the posterior (or steady-state) error covariance matrix, highlights the role played by the graph topology, process bandwidth, and sampling set. A sparse sensing approach is developed to collect samples with guaranteed tracking MSE.

These theoretical findings put focus for the first time on the sampling of time varying graph signals and are corroborated with synthetic experiments as well as employed in the ETEX data set [32].

Notation. We denote scalars, column vectors, and matrices with plain letters a (A), bold lowercase letters $\boldsymbol{a}$ and bold uppercase letters $\boldsymbol{A}$, respectively. A_{ij} is the (i, j)th element of $\boldsymbol{A}$, $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and $\boldsymbol{1}_N$ ($\boldsymbol{0}_N$) is the $N \times 1$ vector of all ones (zeros). diag($\cdot$) denotes the diagonal operator, i.e., $\boldsymbol{A} = \text{diag}(\boldsymbol{a})$ is a diagonal matrix with $\boldsymbol{a}$ on the main diagonal and $\boldsymbol{a} = \text{diag}(\boldsymbol{A})$ stores

the diagonal of $\boldsymbol{A}$ in $\boldsymbol{a}$. Similarly, $\boldsymbol{A} = \text{blkdiag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N)$ is a block diagonal matrix containing $\boldsymbol{A}_i$ as the ith diagonal block. The pseudoinverse of $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^\dagger$, its trace as $\text{Tr}(\boldsymbol{A})$, and the matrix spectral norm as $\|\boldsymbol{A}\|$. For two sets $\mathcal{R}$ and $\mathcal{S}$, $|\mathcal{R}|$ denotes the cardinality of $\mathcal{R}$, $\mathcal{R} \subset \mathcal{S}$ the subset operation, $\mathcal{R} \cup \mathcal{S}$ the union of the sets, and $\mathcal{S}^c = \mathcal{R} \backslash \mathcal{S}$ the complementary set of $\mathcal{S}$ w.r.t. $\mathcal{R}$. Vector $\boldsymbol{1}_\mathcal{R}$ is the set indicator vector, whose rth entry is equal to one if $r \in \mathcal{R}$ and zero otherwise. $\lceil \cdot \rceil$ indicates the ceiling operator.

The remaining parts of this paper are organized as follows. Section 2 covers the background information, while Section 3 the problem formulation. Section 6 contains the numerical experiments and Section 7 concludes the paper. The proofs are collected in the appendix.

2. Background

Basics of GSP. Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node set $\mathcal{V} = \{1, \ldots, N\}$ and edge set $\mathcal{E}$. The nodes' connectivity is represented by the weighted adjacency matrix $\boldsymbol{W}$, where $W_{nm} > 0$ is the edge weight connecting nodes n and m and $W_{nm} = 0$ means the nodes are disconnected. The graph Laplacian matrix is $\boldsymbol{L} = \text{diag}(\boldsymbol{W}\boldsymbol{1}_N) - \boldsymbol{W}$. Each node $n \in \mathcal{V}$ is associated with a graph signal $x_n \in \mathbb{R}$, which are collected in vector $\boldsymbol{x} = [x_1, \ldots, x_N]^\top$.

Both $\boldsymbol{W}$ and $\boldsymbol{L}$ are candidates for the graph shift operator matrix $\boldsymbol{S}$, an $N \times N$ matrix that allows for a frequency analysis of graph signals [1,2,33]. Due to its symmetry, $\boldsymbol{S}$ can be eigendecomposed as $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ with eigenvector matrix $\boldsymbol{U}$ and diagonal eigenvalue matrix $\boldsymbol{\Lambda}$. This eigendecomposition carries the notion of frequency in the graph setting. The projection of $\boldsymbol{x}$ onto the eigenspace of $\boldsymbol{S}$, i.e., $\hat{\boldsymbol{x}} = \boldsymbol{U}^\top \boldsymbol{x}$, is named the graph Fourier transform (GFT) and its inverse is $\boldsymbol{x} = \boldsymbol{U}\hat{\boldsymbol{x}}$. The main diagonal of $\boldsymbol{\Lambda}$ contains the spectral support of $\boldsymbol{x}$, named the graph frequencies.

Signal reconstruction. Reconstructing a graph signal $\boldsymbol{x}$ from its sampled version relates to the joint localization properties of $\boldsymbol{x}$ in the vertex and graph frequency domain [11]. Consider a subset of vertices $\mathcal{S} \subseteq \mathcal{V}$ and the respective set projection matrix $\boldsymbol{C}_\mathcal{S} = \text{diag}(\boldsymbol{1}_\mathcal{S})$. The graph signal $\boldsymbol{x}$ is localized over the set $\mathcal{S}$ if $\boldsymbol{C}_\mathcal{S}\boldsymbol{x} = \boldsymbol{x}$. Similarly, consider a subset of graph frequency indices $\mathcal{F} \subseteq \{1, \ldots, N\}$ and define $\boldsymbol{U}_\mathcal{F} \in \mathbb{R}^{N \times |\mathcal{F}|}$ as the matrix containing the columns of $\boldsymbol{U}$ relative to set $\mathcal{F}$. The matrix $\boldsymbol{C}_\mathcal{F} = \boldsymbol{U}_\mathcal{F}\boldsymbol{U}_\mathcal{F}^\top$ is a bandlimiting operator and the graph signal $\boldsymbol{x}$ is localized over $\mathcal{F}$, or $\mathcal{F}$–bandlimited, if $\boldsymbol{C}_\mathcal{F}\boldsymbol{x} = \boldsymbol{x}$. The latter coupled with the localization over $\mathcal{S}$ implies $\boldsymbol{C}_\mathcal{S}\boldsymbol{C}_\mathcal{F}\boldsymbol{x} = \boldsymbol{x}$, which, in turn, means $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{C}_\mathcal{S}\boldsymbol{C}_\mathcal{F}$ (or equivalently of $\boldsymbol{C}_\mathcal{S}\boldsymbol{U}_\mathcal{F}$) associated with the unit eigenvalue. The unit eigenvalue is by construction the largest admissible eigenvalue of $\boldsymbol{C}_\mathcal{S}\boldsymbol{C}_\mathcal{F}$ and $\boldsymbol{x}$ is the respective eigenvector if $\|\boldsymbol{C}_\mathcal{S}\boldsymbol{U}_\mathcal{F}\| = 1$. Then, an $\mathcal{F}$–bandlimited graph signal can be recovered from the samples in $\mathcal{S}$ iff

$$\|\boldsymbol{C}_{\mathcal{S}^c}\boldsymbol{U}_\mathcal{F}\| < 1 \tag{1}$$

where $\boldsymbol{C}_{\mathcal{S}^c} = \boldsymbol{I}_N - \boldsymbol{C}_\mathcal{S}$ is the projection matrix onto the complementary vertex set $\mathcal{S}^c = \mathcal{V} \backslash \mathcal{S}$ [11]. I.e., an $\mathcal{F}$–bandlimited graph signal can be recovered from samples in $\mathcal{S}$ if it is not localized over the complementary vertex set. Indeed, if $\|\boldsymbol{C}_{\mathcal{S}^c}\boldsymbol{U}_\mathcal{F}\| = 1$ were true, then there exists a nontrivial $\mathcal{F}$–bandlimited graph signal localized onto the complementary set $\mathcal{S}^c$. This signal cannot be recovered from the samples in $\mathcal{S}$ which are null due to the perfect localization over the complementary sampling set; therefore, making condition (1) necessary and sufficient for perfect recovery– refer to [11] for further detail.

3. Problem formulation

A graph process $\boldsymbol{x}_t$ is a time varying graph signal in which entry x_{tn} is a time series evolving on node n. In the sequel, we first formulate the graph process as a linear system on a graph. Then, we

define a bandlimited graph process and analyze systems on graphs in the graph Fourier domain.

3.1. Systems on graphs

Consider the $N-$state discrete linear time varying system

$$\boldsymbol{x}_t = \boldsymbol{A}_{t-1}\boldsymbol{x}_{t-1} + \boldsymbol{B}_{t-1}\boldsymbol{u}_{t-1} \tag{2a}$$

$$\boldsymbol{y}_t = \boldsymbol{C}_{\mathcal{S}_t}(\boldsymbol{x}_t + \boldsymbol{v}_t) \tag{2b}$$

where $\boldsymbol{x}_t$ is the $N \times 1$ state representing the graph signal at time t, $\boldsymbol{u}_t$ is the $N \times 1$ input signal, and $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ are the $N \times N$ time varying state-transition and input matrices, respectively. The measurement is $\boldsymbol{y}_t \in \mathbb{R}^N$ and $\boldsymbol{C}_{\mathcal{S}_t} = \mathrm{diag}(c_{t1}, \ldots, c_{tN})$ denotes the sampling matrix with entry $c_{tn} = 1$ if node n is sampled at time t. The set of nodes sampled at time t is $\mathcal{S}_t = \{(n, t) \mid n \in \{1, \ldots, N\}, c_{tn} = 1\}$. Vector $\boldsymbol{v}_t$ denotes a zero-mean Gaussian noise with covariance matrix $\boldsymbol{\Sigma}_v = \sigma_v^2 \boldsymbol{I}_N$. Throughout this work, we shall consider given the graph and the system matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$. Readers interested in how to estimate the graph and the system matrices in (2) from data are directed to [34–36] and references therein. Model (2) comprises the following processes.[1]

Signal diffusion. For $\boldsymbol{x}_0$ being the initial graph signal state, its instantaneous diffused state at time t is given by the exponential matrix product

$$\boldsymbol{x}_t = e^{-w\boldsymbol{L}t}\boldsymbol{x}_0 = e^{-w\boldsymbol{L}}e^{-w\boldsymbol{L}(t-1)}\boldsymbol{x}_0 \triangleq \boldsymbol{A}\boldsymbol{x}_{t-1} \tag{3}$$

where $w > 0$ is the diffusion rate and $\boldsymbol{A} = e^{-w\boldsymbol{L}}$ is the time-invariant state-transition matrix [37]. This recursion is also known as the heat diffusion model. If the graph is the mesh of a surface, the initial state $\boldsymbol{x}_0$ may be a heat source starting at some node and $\boldsymbol{x}_t$ the heat state at time t. We can also incorporate an input signal $\boldsymbol{u}_{t-1}$ in (3) to represent additional heat sources available only at time $t - 1 > 0$. Besides modeling heat diffusion on meshes, model (3) is also used to analyze chemical substance dispersion, opinion and influence propagation [38,39], brain signal connections [40], and protein interactions [41].

Wave propagation. For an initial state $\boldsymbol{z}_0$ and speed c, the discretised wave equation on graphs follows the two-step recursion

$$\boldsymbol{z}_t = (2\boldsymbol{I}_N - c^2\boldsymbol{L})\boldsymbol{z}_{t-1} - \boldsymbol{z}_{t-2}. \tag{4}$$

This recursion can be formulated in form (2a) by defining the state vector $\boldsymbol{x}_t = [\boldsymbol{z}_{t-1}, \boldsymbol{z}_t]^\top$, the state transition matrix $\boldsymbol{A} = [[\boldsymbol{I}_N, (2\boldsymbol{I}_N - c^2\boldsymbol{L})]^\top, [\boldsymbol{0}_N\boldsymbol{0}_N^\top, \boldsymbol{I}_N]^\top]$, and $\boldsymbol{z}_{-1} = \boldsymbol{0}_N$ [42]. The latter is of interest in seismic data [16].

ARMA graph processes. We denote an ARMA graph process as

$$\boldsymbol{x}_t = f(\boldsymbol{S})\boldsymbol{x}_{t-1} + g(\boldsymbol{S})\boldsymbol{u}_{t-1} \tag{5}$$

where $f(\boldsymbol{S})$ and $g(\boldsymbol{S})$ are matrix functions that share the eigenvectors with the shift operator such as polynomials in $\boldsymbol{S}$ [19,20,43]. A particular form of (5) is the first-order recursion

$$\boldsymbol{x}_t = -w\boldsymbol{S}\boldsymbol{x}_{t-1} + \boldsymbol{u}_0 \text{ with } \quad \boldsymbol{x}_0 = \boldsymbol{0}_N \tag{6}$$

which has the steady-state

$$\boldsymbol{x} = \lim_{t \to \infty} \boldsymbol{x}_t = (\boldsymbol{I}_N + w\boldsymbol{S})^{-1}\boldsymbol{u}_0 \tag{7}$$

if the parameter w satisfies $0 < w < 1/\lambda_{\max}(\boldsymbol{S})$. Expression (7) solves the aggregate diffusion model used in image smoothing [44], Tikhonov denoising [1], and rating prediction [45].

---

[1] We can also consider an additive Gaussian noise $\boldsymbol{w}_{t-1}$ to the state evolution in (2a) to capture uncertainties in the graph process model. However, the latter leads to upper bounds in the observability MSE analysis rather than closed-form expressions, see Proposition 3. Since this is not the case for tracking using the Kalman filter, we shall consider model noise in Section 5. Remark that also in the current form the system in (2) matches models (3)-(7).

3.2. Bandlimited systems on graphs

To proceed with the graph Fourier analysis of (2), we define the following.

Definition 1. A graph process $\boldsymbol{x}_t$ with GFT $\hat{\boldsymbol{x}}_t = \boldsymbol{u}^\top \boldsymbol{x}_t$ is $\mathcal{F}-$bandlimited if $\hat{\boldsymbol{x}}_t$ has non-zero frequency content only on the subset of graph frequency indices $\mathcal{F}$.

The set $\mathcal{F} = \{n \in \{1, \ldots, N\} | \exists t \text{ for which } \hat{x}_{tn} \neq 0\}$ is common for all $\boldsymbol{x}_t$ and $t \geq 0$. In other words, $\mathcal{F}$ is the union of all instantaneous sets $\mathcal{F}_t = \{n \in \{1, \ldots, N\} | \hat{x}_{tn} \neq 0\}$. We can then write the graph signal at time t as $\boldsymbol{x}_t = \boldsymbol{U}_{\mathcal{F}}\tilde{\boldsymbol{x}}_t$ where the vector $\tilde{\boldsymbol{x}}_t \in \mathbb{R}^{|\mathcal{F}|}$ contains the entries of $\hat{\boldsymbol{x}}_t$ indexed by $\mathcal{F}$.

We pose two assumptions to obtain an $\mathcal{F}-$bandlimited graph process.

Assumption 1. The system evolution matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ are diagonalizable by the eigenvectors of the graph shift operator $\boldsymbol{S}$.

Assumption 2. The input $\boldsymbol{u}_t$ is an $\mathcal{F}-$bandlimited graph process.

Assumption 1 considers linear time varying systems on time-invariant graphs This assumption focuses to graph processes whose system evolution matrices share the eigenvectors with the shift operator $\boldsymbol{S}$, which is the case for the processes in (3)-(7). Assumption 2 requires the input signal to have a sparse GFT over time. That is, $\boldsymbol{u}_t$ should be a smooth graph signal, or have properties similar to the signals studied in [8,10–12,14,15,19–21,25–27,46–49]. While in practice it is difficult to find a perfect bandlimited graph process, this assumption is made in GSP to derive theoretical results. But it is possible to find in practice approximately bandlimited graph processes; thus, if we treat them as bandlimited, the theoretical findings are close to the empirical ones. In our numerical results, we will never impose exact bandlimitedness.

With the above assumptions, we want to link the graph process, its GFT, and the sampling strategy. If one of the two assumptions is violated, similar sampling strategies can be derived by ignoring the limited frequency support and consider the process as $N-$bandlimited. Since the latter is a particular case of our framework and since it has a looser connection with the graph spectrum, we shall not discuss it further. We will nevertheless evaluate the full bandwidth case in Section 6.

We can now claim the following.

Proposition 1. Let $\boldsymbol{x}_t$ be a graph process following model (2a) and let Assumptions 1 and 2 hold. The graph process $\boldsymbol{x}_t$ is an $\mathcal{F}-$bandlimited graph process if and only if $\boldsymbol{x}_0$ is an $\mathcal{F}-$bandlimited graph signal.

Proposition 1 implies that we can consider the in-band process evolution $\tilde{\boldsymbol{x}}_t$ to have an equivalent representation for $\boldsymbol{x}_t$. By denoting the GFT coefficients of $\boldsymbol{u}_{t-1}$ indexed by $\mathcal{F}$ with $\tilde{\boldsymbol{u}}_{t-1} \in \mathbb{R}^{|\mathcal{F}|}$, we can write the evolution of $\tilde{\boldsymbol{x}}_t$ as

$$\tilde{\boldsymbol{x}}_t = \tilde{\boldsymbol{A}}_{t-1}\tilde{\boldsymbol{x}}_{t-1} + \tilde{\boldsymbol{B}}_{t-1}\tilde{\boldsymbol{u}}_{t-1} \tag{8a}$$

$$\boldsymbol{y}_t = \boldsymbol{C}_{\mathcal{S}_t}(\boldsymbol{U}_{\mathcal{F}}\tilde{\boldsymbol{x}}_t + \boldsymbol{v}_t) \tag{8b}$$

where $\tilde{\boldsymbol{A}}_t = \boldsymbol{U}_{\mathcal{F}}^\top \boldsymbol{A}_t \boldsymbol{U}_{\mathcal{F}}$ and $\tilde{\boldsymbol{B}}_t = \boldsymbol{U}_{\mathcal{F}}^\top \boldsymbol{B}_t \boldsymbol{U}_{\mathcal{F}}$ are diagonal matrices that contain the in-band spectrum of $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$, respectively. The $\mathcal{F}-$bandlimited graph processes considered in this work follow evolutions (2a) and (8a) in the vertex and spectral domain, respectively. We shall refer to systems in the form (2) that can be written in the form (8) as $\mathcal{F}-$bandlimited systems on graphs.

Problem statement. Given an $\mathcal{F}-$bandlimited graph process $\boldsymbol{x}_t$ following model (2) in the vertex domain and model (8) in the spectral domain; the goal is to find conditions on and design the sampling matrices $\boldsymbol{C}_{\mathcal{S}_t}$ for $t \geq 0$ to observe and track the graph process.

4. Observing graph processes

In this section, we devise a sampling theory for observing the graph process. We generalize condition (1) from reconstructing an $\mathcal{F}-$bandlimited graph signal to observing an $\mathcal{F}-$bandlimited graph process. We start by adapting the definition of observability [50] to our setting.

Definition 2. An $\mathcal{F}-$bandlimited system on a graph is observable over the set $\mathcal{S}_{0:T} = cup_{t=0}^{T} \mathcal{S}_t = \{(n,t) \mid n \in \{1,\ldots,N\}; t \in \{0,\ldots,T\}, c_{tn} = 1\}$ if for any $\mathcal{F}-$bandlimited initial state $\boldsymbol{x}_0$ and some final time T, the initial state $\boldsymbol{x}_0$ can be uniquely determined in the absence of noise by the knowledge of the input $\boldsymbol{u}_t$ and measurement $\boldsymbol{y}_t$ collected over the sets $\mathcal{S}_t$ for all $t \in \{0,\ldots,T\}$.

Definition 2 follows the same baseline idea of conventional observability, i.e., describe when the initial state of a system $\mathbf{x}_0$ is uniquely determined in the absence of noise. Since we now deal with $\mathcal{F}$-bandlimited systems on graphs and samples collected only on the set $\mathcal{S}_{0:T}$, it specialises the conventional definition in [46] to also account for these quantities. The set $\mathcal{S}_{0:T}$ specifies all graph-time locations where and when to collect samples in the interval $\{0,\ldots,T\}$. With the above formulation in place, we want to answer the questions: (Q1) *Under which conditions is an $\mathcal{F}-$bandlimited graph process observable from a few samples?* (Q2) *How should we collect samples to estimate $\mathbf{x}_0$ with a controlled accuracy?*

We start by writing the measurement $\boldsymbol{y}_t$ as a function of the initial $\mathcal{F}-$bandlimited signal $\tilde{\boldsymbol{x}}_0$ as

$$\boldsymbol{y}_t = \boldsymbol{C}_{\mathcal{S}_t}\boldsymbol{U}_{\mathcal{F}}\tilde{\boldsymbol{A}}_{t:0}\tilde{\boldsymbol{x}}_0 + \boldsymbol{C}_{\mathcal{S}_t}\boldsymbol{U}_{\mathcal{F}}\sum_{\tau=0}^{t-1}\tilde{\boldsymbol{A}}_{t:\tau+1}\tilde{\boldsymbol{B}}_{\tau}\tilde{\boldsymbol{u}}_{\tau} + \boldsymbol{C}_{\mathcal{S}_t}\boldsymbol{v}_t \quad (9)$$

with

$$\tilde{\boldsymbol{A}}_{t:\tau} = \begin{cases} \tilde{\boldsymbol{A}}_{t-1}\tilde{\boldsymbol{A}}_{t-2}\ldots\tilde{\boldsymbol{A}}_{\tau} & t > \tau \\ \boldsymbol{I}_{|\mathcal{F}|} & t = \tau \\ \boldsymbol{0}_{|\mathcal{F}|}\boldsymbol{0}_{|\mathcal{F}|}^{\top} & t < \tau \end{cases}. \quad (10)$$

By collecting the measurements taken in the interval $\{0,\ldots,T\}$ into vector $\boldsymbol{y}_{0:T} = [\boldsymbol{y}_0^{\top}, \boldsymbol{y}_1^{\top}, \ldots, \boldsymbol{y}_t^{\top}]^{\top}$, we can write (9) as

$$\boldsymbol{y}_{0:T} = \boldsymbol{O}_{0:T}\tilde{\boldsymbol{x}}_0 + \boldsymbol{J}_{0:T}\tilde{\boldsymbol{u}}_{0:T-1} + \boldsymbol{C}_{\mathcal{S}_{0:T}}\boldsymbol{v}_{0:t} \quad (11)$$

where we defined the observability matrix

$$\boldsymbol{O}_{0:T} = [(\boldsymbol{C}_{\mathcal{S}_0}\boldsymbol{U}_{\mathcal{F}}\tilde{\boldsymbol{A}}_{0:0})^{\top}, (\boldsymbol{C}_{\mathcal{S}_1}\boldsymbol{U}_{\mathcal{F}}\tilde{\boldsymbol{A}}_{1:0})^{\top}, \ldots, (\boldsymbol{C}_{\mathcal{S}_T}\boldsymbol{U}_{\mathcal{F}}\tilde{\boldsymbol{A}}_{T:0})^{\top}]^{\top}$$
$$= \boldsymbol{C}_{\mathcal{S}_{0:T}}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\tilde{\boldsymbol{A}}_{0:T} \quad (12)$$

the block matrices $\boldsymbol{C}_{\mathcal{S}_{0:T}} = \text{blkdiag}(\boldsymbol{C}_{\mathcal{S}_0},\ldots,\boldsymbol{C}_{\mathcal{S}_T})$, $\tilde{\boldsymbol{A}}_{0:T} = \left[\boldsymbol{I}_{|\mathcal{F}|}, \tilde{\boldsymbol{A}}_{1:0}^{\top}, \ldots, \tilde{\boldsymbol{A}}_{T:0}^{\top}\right]^{\top}$ and vectors $\tilde{\boldsymbol{u}}_{0:T-1} = [\tilde{\boldsymbol{u}}_0^{\top}, \tilde{\boldsymbol{u}}_1^{\top}, \ldots, \tilde{\boldsymbol{u}}_{T-1}^{\top}]^{\top}$, $\boldsymbol{v}_{0:T} = [\boldsymbol{v}_0^{\top}, \boldsymbol{v}_1^{\top}, \ldots, \boldsymbol{v}_T^{\top}]^{\top}$. The matrix $\boldsymbol{J}_{0:T}$ captures the input evolution in the interval $\{0,\ldots,T\}$ whose expression is not required for our derivations, but can be obtained from (9). Next, we answer questions (Q1) and (Q2) for $\boldsymbol{C}_{\mathcal{S}_t}$ being deterministic (called deterministic sampling), while in Section 4.2 we consider the case where the entries of $\boldsymbol{C}_{\mathcal{S}_t}$ follow a Bernoulli distribution (called random sampling).

4.1. Observability with deterministic sampling

Recall that $\boldsymbol{C}_{\mathcal{S}_{0:T}}$ is the set projection matrix for the set $\mathcal{S}_{0:T}$. System (8) is observable over the set $\mathcal{S}_{0:T}$ if and only if the observability matrix $\boldsymbol{O}_{0:T}$ in (12) is full rank, i.e., $\text{rank}(\boldsymbol{O}_{0:T}) = |\mathcal{F}|$. From the structure of $\boldsymbol{O}_{0:T}$, we deduce that a *sufficient* condition for observability is that at least one block matrix $\boldsymbol{C}_{\mathcal{S}_t}\boldsymbol{U}_{\mathcal{F}}\tilde{\boldsymbol{A}}_{t,0}$ has rank $|\mathcal{F}|$, which in turn requires to sample at least $|\mathcal{F}|$ nodes for a specific t, i.e., $|\mathcal{S}_t| \geq |\mathcal{F}|$. This condition is similar to that obtained in adaptive graph signal reconstruction [21,27]. Differently, we link the cardinality of the sampling set with the process evolution by bringing

$\tilde{\boldsymbol{A}}_{t:0}$ into the analysis. This link allows us to collect samples in a graph-time fashion, resulting in so-called *graph-time* samples. We claim the following.

Proposition 2. *An $\mathcal{F}-$bandlimited system on a graph [cf. (8)] is observable over the set $\mathcal{S}_{0:T}$ only if the cardinality of the sampling set is greater than or equal to the process bandwidth, i.e.,*

$$|\mathcal{S}_{0:T}| \geq |\mathcal{F}|. \quad (13)$$

That is, only if at least $|\mathcal{F}|$ graph-time samples are taken in the time interval $\{0,\ldots,T\}$. These samples can be taken by $|\mathcal{F}|$ nodes at a single time instant, one node in $|\mathcal{F}|$ time instants, or a combination of the two.

Note that Proposition 2 provides only a necessary condition: the observability matrix $\boldsymbol{O}_{0:T}$ [cf. (12)] may be ill-conditioned depending on the location of the samples and process spectral support. It is then paramount to pick the graph-time samples such that $\boldsymbol{O}_{0:T}$ has full rank $|\mathcal{F}|$, i.e., $\text{rank}(\boldsymbol{O}_{0:T}) = |\mathcal{F}|$. By exploiting then rank identity $\text{rank}(\boldsymbol{O}_{0:T}^{\top}\boldsymbol{O}_{0:T}) = \text{rank}(\boldsymbol{O}_{0:T}) = |\mathcal{F}|$ and expression (12) of the observability matrix, the sampling set $\mathcal{S}_{0:T}$ should satisfy

$$\text{rank}\left(\sum_{t=0}^{T}\tilde{\boldsymbol{A}}_{t,0}^{\top}\boldsymbol{U}_{\mathcal{F}}^{\top}\boldsymbol{C}_{\mathcal{S}_t}\boldsymbol{U}_{\mathcal{F}}\tilde{\boldsymbol{A}}_{t,0}\right) = |\mathcal{F}|. \quad (14)$$

For $T = 0$, Eq. (14) particularizes to the one-shot graph signal reconstruction condition derived in [11]. The following theorem generalizes result (1) to a necessary and sufficient condition for the observability of an $\mathcal{F}$-bandlimited graph process over a sampling set.

Theorem 1. *An $\mathcal{F}-$bandlimited system on a graph [cf. (8)] is observable over the set $\mathcal{S}_{0:T}$ if and only if*

$$\|\boldsymbol{C}_{\mathcal{S}_{0:T}^c}(\boldsymbol{I}_{T+1}\otimes\boldsymbol{U}_{\mathcal{F}})\| < \frac{s_{\min}^2(\tilde{\boldsymbol{A}}_{0:T})}{s_{\max}^2(\tilde{\boldsymbol{A}}_{0:T})} \quad (15)$$

where $\boldsymbol{C}_{\mathcal{S}_{0:T}^c} = \boldsymbol{I}_{N(T+1)} - \boldsymbol{C}_{\mathcal{S}_{0:T}}$ is the set projection matrix onto the complementary set $\mathcal{S}_{0:T}^c = \{(n,t) \mid n \in \{1,\ldots,N\}, t \in \{0,\ldots,T\}, c_{tn} = 0\}$ and $s_{\min}(\tilde{\boldsymbol{A}}_{0:T})$ and $s_{\max}(\tilde{\boldsymbol{A}}_{0:T})$ indicate the minimum and maximum singular values of $\tilde{\boldsymbol{A}}_{0:T}$, respectively.

Theorem 1 and Proposition 2 are our answer to question (Q1) for samples collected deterministically. Theorem 1 involves the localization properties of graph signals by accounting also for the process evolution through the matrix $\tilde{\boldsymbol{A}}_{0:T}$. In other words, Theorem 1 implies there are no $\mathcal{F}$-bandlimited graph processes localized on the complementary set $\mathcal{S}_{0:T}^c$ throughout their evolution. For $T = 0$, the result in (15) for observing a graph process specializes to the one-shot graph signal reconstruction condition in (1).

MSE analysis. When the measurements $\boldsymbol{y}_t$ are corrupted by noise $\boldsymbol{v}_t \neq 0$, we estimate the initial state $\tilde{\boldsymbol{x}}_0$ through least squares as

$$\tilde{\boldsymbol{x}}_0^o = \boldsymbol{O}_{0:T}^{\dagger}(\boldsymbol{y}_{0:T} - \boldsymbol{J}_{0:T}\boldsymbol{u}_{0:T-1}). \quad (16)$$

This implies that, besides ensuring full rank, the sampling set $\mathcal{S}_{0:T}$ should also contain those nodes that lead to the minimum reconstruction error from noisy measurements. The mean squared error is a standard metric to characterize the impact on noisy data and is formalized by the following proposition.

Proposition 3. *Consider an $\mathcal{F}-$bandlimited graph process following model (8) and let the sampling set $\mathcal{S}_{0:T}$ satisfy condition (15). Then, the MSE of the least squares estimated signal $\tilde{\boldsymbol{x}}_0^o$ is*

$$\text{MSE} = \mathbb{E}\{\|\tilde{\boldsymbol{x}}_0^o - \tilde{\boldsymbol{x}}_0\|^2\} = \mathbb{E}\left\{\text{Tr}\left[(\tilde{\boldsymbol{x}}_0^o - \tilde{\boldsymbol{x}}_0)(\tilde{\boldsymbol{x}}_0^o - \tilde{\boldsymbol{x}}_0)^\top\right]\right\}$$

$$= \sigma_v^2 \text{Tr}\left\{\left[\tilde{\boldsymbol{A}}_{0:T}^\top (\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_\mathcal{F})^\top \boldsymbol{C}_{\mathcal{S}_{0:T}} (\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_\mathcal{F}) \tilde{\boldsymbol{A}}_{0:T}\right]^{-1}\right\} \quad (17)$$

The proof follows from the covariance matrix of the least squares estimator [51, Chapter 8].

Besides characterizing the impact of the graph-time samples on the MSE, the result in (17) shows that their location is as important as the number of samples. This is due to the presence of the sampling matrix $\boldsymbol{C}_{\mathcal{S}_{0:T}}$ into the inverse covariance matrix in (17). Thus, we focus next on how to select these samples to guarantee a target MSE performance.

Sampling strategy. We follow sparse sensing to design the sampling set $\mathcal{S}_{0:T}$ [52]. This consists of finding the minimum number of samples that ensure the MSE in (17) is smaller than a target value $\gamma > 0$. Defining the sampling vector in the interval $\{0, \ldots, T\}$ as $\boldsymbol{c}_{0:T} = [\boldsymbol{c}_0^\top, \ldots, \boldsymbol{c}_T^\top]^\top$ with $\boldsymbol{c}_t = [c_{t1}, \ldots, c_{tN}]^\top$, we find the graph-time samples by solving the relaxed convex problem

$$\underset{\boldsymbol{c}_{0:T}}{\text{minimize}} \quad \mathbf{1}_{N \times (T+1)}^\top \boldsymbol{c}_{0:T}$$

$$\text{subject to} \quad \text{Tr}\left[\left(\boldsymbol{\Psi}_{0:T}^\mathsf{H} \boldsymbol{c}_{\mathcal{S}_{0:T}} \boldsymbol{\Psi}_{0:T}\right)^{-1}\right] \leq \frac{\gamma}{\sigma_v^2} \quad (18)$$
$$\boldsymbol{\Psi}_{0:T} = (\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_\mathcal{F}) \tilde{\boldsymbol{A}}_{0:T}$$
$$\boldsymbol{c}_{\mathcal{S}_{0:T}} = \text{diag}(\boldsymbol{c}_{0:T}), \quad \boldsymbol{c}_{0:T} \in [0, 1]^{N(T+1)}$$

where the objective function is the l_1-norm surrogate of the l_0-norm and imposes sparsity in $\mathcal{S}_{0:T}$, the first three constraints impose an upper bound γ to the MSE, and the last constraint relaxes the Boolean constraint $c_{0:T,i} \in \{0, 1\}$ to the box one.[2] As an alternative, we can adopt a greedy approach to build the set $\mathcal{S}_{0:T}$ [53]. Both approaches can be used to answer our question (Q2). We can also consider the opposite problem of (18), where we minimize the MSE but impose a fixed number of samples; this translates as well into a convex problem.

4.2. Observability with random sampling

One major limitation of deterministic observability is the design of the sampling set $\mathcal{S}_{0:T}$ by solving the suboptimal relaxed problem in (18). We tackle this limitation by considering the entries of $\boldsymbol{C}_{\mathcal{S}_t} = \text{diag}(\boldsymbol{c}_t)$ in (2b) to be i.i.d. in time Bernoulli random variables with expected value $\bar{\boldsymbol{C}} = \text{diag}(\bar{\boldsymbol{c}})$. Hence, the set $\mathcal{S}_{0:T}$ over which we want to observe the system becomes random. Our goal is to relate the sampling probabilities $\boldsymbol{c}$ with the observability of an $\mathcal{F}$−bandlimited system on a graph over a realization of $\mathcal{S}_{0:T}$. Depending on these probabilities, the set $\mathcal{S}_{0:T}$ may contain all samples ($\bar{c}_n = 1$ for all $n \in \{1, \ldots, N\}$) or be empty ($\bar{c}_n = 0$ for all $n \in \{1, \ldots, N\}$). Denote then by $\bar{\mathcal{S}} = \{n \in \{1, \ldots, N\}|\bar{c}_n > 0\}$ the set of nodes sampled with a probability greater than zero or the expected sampling set. Our task translates now into answering questions (Q1) and (Q2) w.r.t. the expected set $\bar{\mathcal{S}}$.

Given the measurements $\boldsymbol{y}_{0:T}$ and a realization of $\boldsymbol{C}_{\mathcal{S}_{0:T}}$, we define the measurement vector to which we subtracted the input signal as[3]

$$\boldsymbol{z}_{0:T} = \boldsymbol{y}_{0:T} - \boldsymbol{J}_{0:T} \tilde{\boldsymbol{u}}_{0:T} = \boldsymbol{O}_{0:T} \tilde{\boldsymbol{x}}_0 + \boldsymbol{C}_{\mathcal{S}_{0:T}} \boldsymbol{v}_{0:T}. \quad (19)$$

From Proposition 2, it follows that a necessary condition for the observability matrix $\boldsymbol{O}_{0:T}$ [cf. (12)] to be full rank is that the realization of the sampling set $\mathcal{S}_{0:T}$ has a cardinality greater than, or equal to, the signal bandwidth $|\mathcal{F}|$. From the structure of $\boldsymbol{O}_{0:T}$

($\boldsymbol{C}_{\mathcal{S}_{0:T}}$), it follows that the rank of $\boldsymbol{O}_{0:T}$ ($\boldsymbol{C}_{\mathcal{S}_{0:T}}$) depends on the rank of $\bar{\boldsymbol{C}}$. That is, the rank of the observability matrix depends on the cardinality of the expected sampling set $\bar{\mathcal{S}}$. We formalize this observation in the following proposition.

Proposition 4. *Given an $\mathcal{F}$−bandlimited graph process following model (8) and given the diagonal sampling matrix $\boldsymbol{C}_{\mathcal{S}_t}$ in (2) has i.i.d. in time Bernoulli entries with expected value $\bar{\boldsymbol{C}}$. A necessary condition for the observability of the process from a realization of $\mathcal{S}_{0:T}$ [cf. Def. 2] is that at least $\lceil |\mathcal{F}|/(T+1) \rceil$ nodes are sampled with a probability greater than zero.*

Proposition 4 provides a necessary condition on the expected sampling set $\bar{\mathcal{S}}$, which is equivalent to

$$|\bar{\mathcal{S}}| \geq \lceil |\mathcal{F}|/(T+1) \rceil. \quad (20)$$

It implies that for $T \geq |\mathcal{F}|$, we could observe an $\mathcal{F}$−bandlimited graph process by taking random measurements only at one node. This result is not entirely surprising: the authors in [10] have shown that a graph signal can be reconstructed also by sampling successive aggregations of a single node. Hence, we can observe the process form a node that has linearly independent measurements in time; but since the sampling is random in the finite interval $\{0, \ldots, T\}$, there is a possibility that a realization of the set $\mathcal{S}_{0:T}$ has a cardinality smaller than $|\mathcal{F}|$. We quantify the latter in the following proposition.

Proposition 5. *Given an $\mathcal{F}$−bandlimited graph process following model (8) and given the diagonal sampling matrix $\boldsymbol{C}_{\mathcal{S}_t}$ in (2) has i.i.d. in time Bernoulli entries with expected value $\bar{\boldsymbol{C}}$. The cardinality of a realization of $\mathcal{S}_{0:T}$ is an integer $\varepsilon \geq 0$ smaller than the process bandwidth $|\mathcal{F}|$ with probability*

$$\Pr(|\mathcal{S}_{0:T}| < |\mathcal{F}| - \varepsilon) = \sum_{k=0}^{|\mathcal{F}|-1-\varepsilon} \frac{\alpha^k e^{-\alpha}}{k!} \quad (21)$$

where $\alpha = (T+1)\mathbf{1}_N^\top \bar{\boldsymbol{c}}$ is the mean of a Poisson binomial distribution.

Propositions 4 and 5 are our answer to question (Q1) for samples collected randomly. The interesting part of (21) is for $\varepsilon = 0$, which quantifies the probability of the sampling set being smaller than the process bandwidth. The latter drops in fact to zero even for moderate sampling probabilities $\bar{\boldsymbol{c}}$ if $N \sim 100$ and $T \gg 0$. In Section 6.1, we test the observability with random sampling for the ETEX data set and show the probability in (21) drops below machine precision. Next, we perform a mean squared error analysis on the least squares estimated state to quantify the impact of the sampling probabilities $\bar{\boldsymbol{C}} = \text{diag}(\bar{\boldsymbol{c}})$ on the process observability.

MSE analysis. The least squares estimate of the initial state $\tilde{\boldsymbol{x}}_0$ is $\tilde{\boldsymbol{x}}_0^o = \boldsymbol{O}_{0:T}^\dagger \boldsymbol{z}_{0:T}$. To render the MSE analysis tractable, we follow the procedure used for the Cramér-Rao lower bound[4] (CRLB) [51] and propose a lower bound on the achievable MSE. The following proposition quantifies this finding.

Proposition 6. *Given an $\mathcal{F}$−bandlimited graph process following model (8) and given the diagonal sampling matrix $\boldsymbol{C}_{\mathcal{S}_t}$ in (2) has i.i.d. in time Bernoulli entries with expected value $\bar{\boldsymbol{C}}$. The MSE of the least squares estimate observed signal $\tilde{\boldsymbol{x}}_0^o = \boldsymbol{O}_{0:T}^\dagger \boldsymbol{z}_{0:T}$ is lower-bounded by*

$$\text{MSE} \geq \sigma_v^2 \text{Tr}\left\{\left[\tilde{\boldsymbol{A}}_{0:T}^\top \left(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_\mathcal{F}^\top \bar{\boldsymbol{C}} \boldsymbol{U}_\mathcal{F}\right) \tilde{\boldsymbol{A}}_{0:T}\right]^{-1}\right\}. \quad (22)$$

Besides expressing the lowest achievable MSE as a function of $\bar{\boldsymbol{C}}$, the bound in (22) is useful to design these sampling probabilities.

---

[2] Given the solution of (18), we can use randomized rounding or thresholding to project it onto the set $\{0, 1\}^{N(T+1)}$ [52].

[3] This is analogous to observability in linear system theory where the input signal should be known. This is not a problem since the realization of $\boldsymbol{C}_{\mathcal{S}_{0:T}}$ is also known.

[4] Since the measurements are the product of a Bernoulli and a Gaussian random variable, the joint pdf does not satisfy the CRLB regularity condition.

Sampling strategy. We design the expected sampling set $\bar{\mathcal{S}}$ by following the sparse sensing principle; but instead of using the CRLB as a design criterion, we use the lower bound in (22). This consists of finding the minimum sampling rate such that the MSE lower bound is smaller than a target value $\gamma > 0$. In other words, we find $\bar{c}$, therefore $\bar{\mathcal{S}}$, by solving the convex problem

$$
\begin{aligned}
\underset{\bar{c}}{\text{minimize}} \quad & \mathbf{1}_N^{\mathsf{T}} \bar{c} \\
\text{subject to} \quad & \text{Tr}\left\{ \left[\tilde{\boldsymbol{A}}_{0:T}^{\mathsf{T}} \left(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}}^{\mathsf{H}} \bar{\boldsymbol{C}} \boldsymbol{U}_{\mathcal{F}} \right) \tilde{\boldsymbol{A}}_{0:T} \right]^{-1} \right\} \leq \frac{\gamma}{\sigma_v^2} \\
& \bar{\boldsymbol{C}} = \text{diag}(\bar{c}), \ \bar{c} \in [c_{\min}, c_{\max}]^N, \ 0 \leq c_{\min} \leq c_{\max} \leq 1
\end{aligned}
\tag{23}
$$

Although conceptually similar to the deterministic sampling design in (18), problem (23) differs in two main aspects that preserve the optimality of the solution. First, the objective is the true function (i.e., the overall sampling rate) we want to minimize. Second, the convex box constraint $\bar{c} \in [c_{\min}, c_{\max}]^N$ is not a relaxation, since we optimize over the sampling probabilities for some rate bounds $c_{\min}$ and $c_{\max}$. Problem (23) preserves optimality because we changed the problem setting: the samples are collected randomly. The price to pay is that condition (20) is necessary but not sufficient for observability and holds in probability as quantified by (21). While by solving (23) we can answer question (Q2) for random sampling, we can also add a constraint on the probability criterion in (21) to control outliers. But the latter should be relaxed by an upper bound since it leads to a non-convex problem. As we show in Section 6, this constraint is unnecessary since a small enough γ on the MSE trades well the sampling probabilities with performance.

5. Tracking graph processes

We now consider tracking an $\mathcal{F}$−bandlimited graph process from deterministic samples. Given the process (8), we use the Kalman filter to optimally track it [50]. First, we discuss the time varying scenario and then we move to time-invariant models [cf. (3)–(7)]. For both cases, we provide conditions on the sampling set and design strategies to track the process.

5.1. Kalman filtering for time varying models

The Kalman filter on graphs for (8) evolves as described in Algorithm 1. Here, we also consider uncertainty into the process

---

Algorithm 1: Kalman filtering on graphs.

Initialize the a posteriori estimate $\tilde{\boldsymbol{x}}_0^+$ and the a posteriori error covariance matrix $\boldsymbol{P}_0^+ = \epsilon \boldsymbol{I}_N$ for a small $\epsilon > 0$. For $t > 0$ repeat:
(i) Update the a priori state estimate $\tilde{\boldsymbol{x}}_t^- = \tilde{\boldsymbol{A}}_{t-1} \tilde{\boldsymbol{x}}_{t-1}^+ + \tilde{\boldsymbol{B}}_{t-1} \tilde{\boldsymbol{u}}_{t-1}$;
(ii) Update the a priori error covariance matrix $\boldsymbol{P}_t^- = \tilde{\boldsymbol{A}}_{t-1} \boldsymbol{P}_{t-1}^+ \tilde{\boldsymbol{A}}_{t-1}^{\mathsf{T}} + \tilde{\boldsymbol{\Sigma}}_w$;
(iii) Compute the Kalman gain matrix

$$
\boldsymbol{K}_t = \boldsymbol{P}_t^- \boldsymbol{U}_{\mathcal{F}}^{\mathsf{T}} \boldsymbol{C}_{\mathcal{S}_t} \left(\boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{U}_{\mathcal{F}} \boldsymbol{P}_t^- \boldsymbol{U}_{\mathcal{F}}^{\mathsf{T}} \boldsymbol{C}_{\mathcal{S}_t} + \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{\Sigma}_v \boldsymbol{C}_{\mathcal{S}_t} \right)^{\dagger};
$$

(iv) Update the a posteriori state estimate $\tilde{\boldsymbol{x}}_t^+ = \tilde{\boldsymbol{x}}_t^- + \boldsymbol{K}_t \left(\boldsymbol{y}_t - \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{U}_{\mathcal{F}} \tilde{\boldsymbol{x}}_t^- \right)$; (v) Update the a posteriori error covariance matrix

$$
\begin{aligned}
\boldsymbol{P}_t^+ = {} & \boldsymbol{P}_t^- - \boldsymbol{P}_t^- \boldsymbol{U}_{\mathcal{F}}^{\mathsf{T}} \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{K}_t^{\mathsf{T}} - \boldsymbol{K}_t \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{U}_{\mathcal{F}} \boldsymbol{P}_t^- \\
& + \boldsymbol{K}_t \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{\Sigma}_v \boldsymbol{K}_t^{\mathsf{T}} + \boldsymbol{K}_t \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{U}_{\mathcal{F}} \boldsymbol{P}_t^- \boldsymbol{U}_{\mathcal{F}}^{\mathsf{T}} \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{K}_t^{\mathsf{T}}.
\end{aligned}
$$

---

evolution [cf. (8a)] through an additive zero-mean Gaussian noise $\tilde{\boldsymbol{w}}_{t-1}$ with covariance matrix $\tilde{\boldsymbol{\Sigma}}_w$ independent from $\boldsymbol{v}_t$. The Kalman filtering initializes the a posteriori state estimate $\tilde{\boldsymbol{x}}_0^+$ and the a posteriori error covariance matrix $\boldsymbol{P}_0^+$. The a priori error covariance matrix $\boldsymbol{P}_t^-$ is updated in step (ii), while the Kalman gain matrix $\boldsymbol{K}_t$ in step (iii). The a posteriori estimate $\tilde{\boldsymbol{x}}_t^+$ is updated in step (iv) and the corresponding error covariance matrix $\boldsymbol{P}_t^+$ in step (v). The Kalman gain matrix $\boldsymbol{K}_t$ depends on the sampling matrix $\boldsymbol{C}_{\mathcal{S}_t}$ and leads to the minimum a posteriori MSE, $\text{Tr}(\boldsymbol{P}_t^+)$; hence, $\boldsymbol{C}_{\mathcal{S}_t}$ should be carefully designed to improve tracking.

The sampling set $\boldsymbol{C}_{\mathcal{S}_t}$ in the Kalman gain requires a minimum number of samples to be fully exploited. Consider $R < |\mathcal{F}|$ samples at time t and $\text{rank}(\boldsymbol{P}_t^-) = |\mathcal{F}|$. Then $\text{rank}(\boldsymbol{C}_{\mathcal{S}_t}) < |\mathcal{F}|$ and the rank of the pseudo-inverse part in the Kalman gain matrix $\boldsymbol{K}_t$ is

$$
\text{rank}(\boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{U}_{\mathcal{F}} \boldsymbol{P}_t^- \boldsymbol{U}_{\mathcal{F}}^{\mathsf{H}} \boldsymbol{C}_{\mathcal{S}_t} + \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{\Sigma}_v \boldsymbol{C}_{\mathcal{S}_t}) = R.
$$

This implies the $|\mathcal{F}| \times N$ matrix $\boldsymbol{K}_t$ is rank deficient ($\text{rank}(\boldsymbol{K}_t) \leq R < |\mathcal{F}|$). Thus, a necessary condition for $\boldsymbol{K}_t$ to be full rank is that

$$
|\mathcal{S}_t| \geq |\mathcal{F}|
\tag{24}
$$

samples are collected at time t. In other words, the Kalman filter on graphs fully exploits the Kalman gain only if the number of sampled nodes is greater than, or equal to, the signal bandwidth *for each t*. Condition (24) generalizes the observability condition in (13) to tracking. Moreover, the sampled nodes impact also the a posteriori error covariance matrix $\boldsymbol{P}_t^+$. We will exploit this connection to design set $\mathcal{S}_t$ at time t that guarantees a target a posteriori MSE.

Sampling strategy. Condition (24) and the a posteriori error covariance matrix $\boldsymbol{P}_t^+$ highlight the role played by the location of the samples for tracking. To optimally select these samples, we consider a sparse sensing approach with the posterior Cramér-Rao bound (PCRB) as design criterion [54,55].

Given the log-likelihood of the measurements satisfies the regularity condition $\boldsymbol{E}[\ln p(\boldsymbol{y}_t; \tilde{\boldsymbol{x}}_t)/\tilde{\boldsymbol{x}}_t] = \mathbf{0}_N \ \forall t$, the PCRB satisfies

$$
\boldsymbol{P}_t^+ = \boldsymbol{E}\left[\left(\tilde{\boldsymbol{x}}_t^+ - \tilde{\boldsymbol{x}}_t \right) \left(\tilde{\boldsymbol{x}}_t^+ - \tilde{\boldsymbol{x}}_t \right)^{\mathsf{T}} \right] \succeq \boldsymbol{F}_t^{-1}(\tilde{\boldsymbol{x}}_t)
\tag{25}
$$

where $\boldsymbol{F}_t(\tilde{\boldsymbol{x}}_t)$ is the posterior Fisher information matrix (FIM) and $\succeq$ stands in for the matrix inequality in a positive semidefinite sense. For linear systems in additive Gaussian noise, $\boldsymbol{P}_t^+$ is independent of $\tilde{\boldsymbol{x}}_t$ and expression (25) holds with equality, i.e., the Kalman filter is optimal. Thus, by designing the sampling set $\mathcal{S}_t$ w.r.t. the FIM $\boldsymbol{F}_t(\tilde{\boldsymbol{x}}_t)$, we achieve the same result as working with the a posteriori error covariance matrix $\boldsymbol{P}_t^+$. The posterior FIM for the Kalman filter in Algorithm 1 is

$$
\boldsymbol{F}_t(\tilde{\boldsymbol{x}}_t) = \left(\tilde{\boldsymbol{A}}_t \boldsymbol{F}_{t-1}^{-1}(\tilde{\boldsymbol{x}}_t) \tilde{\boldsymbol{A}}_t^{\mathsf{T}} + \boldsymbol{\Sigma}_{\tilde{w}} \right)^{-1} + \sum_{n=1}^N c_{tn} \boldsymbol{F}_{tn}^o(\tilde{\boldsymbol{x}}_t)
\tag{26}
$$

where c_{tn} is the nth diagonal entry of sampling matrix $\boldsymbol{C}_{\mathcal{S}_t}$ and $\boldsymbol{F}_{tn}^o(\tilde{\boldsymbol{x}}_t) = \sigma_v^{-2} c_{tn} \boldsymbol{u}_{\mathcal{F}n} \boldsymbol{u}_{\mathcal{F}n}^{\mathsf{T}}$ is the FIM related to the nth node measurement at time t. The first term on the right-hand side of (26) is the FIM of tracking up to $t-1$, while the second term accounts for the innovation at time t. By substituting $\boldsymbol{F}_{tn}^o(\tilde{\boldsymbol{x}}_t)$ and rearranging the sum, we obtain the a posteriori FIM

$$
\boldsymbol{F}_t(\tilde{\boldsymbol{x}}_t) = \left(\tilde{\boldsymbol{A}}_t \boldsymbol{F}_{t-1}^{-1}(\tilde{\boldsymbol{x}}_t) \tilde{\boldsymbol{A}}_t^{\mathsf{T}} + \boldsymbol{\Sigma}_{\tilde{w}} \right)^{-1} + \boldsymbol{U}_{\mathcal{F}}^{\mathsf{H}} \boldsymbol{C}_{\mathcal{S}_t} \boldsymbol{\Sigma}_v^{-1} \boldsymbol{U}_{\mathcal{F}}.
\tag{27}
$$

Following then the sparse sensing idea as in (18) and (23), we design the instantaneous sampling set $\mathcal{S}_t$ by solving the relaxed convex problem

$$
\begin{aligned}
\underset{c_t}{\text{minimize}} \quad & \mathbf{1}^{\mathsf{T}} c_t \\
\text{subject to} \quad & \boldsymbol{F}_t(\tilde{\boldsymbol{x}}_t) \succeq \gamma \boldsymbol{I}_{|\mathcal{F}|} \\
& \boldsymbol{C}_{\mathcal{S}_t} = \text{diag}(c_t), \ c_t \in [0, 1]^N
\end{aligned}
\tag{28}
$$

Problem (28) is the counterpart of the relaxed problem (18) to tracking and carries on all the remarks about solution optimality.

5.2. Steady-state Kalman filtering on graphs

We specialize the above derivations to time-invariant $\mathcal{F}$−bandlimited systems on graphs as those for models (3)–(7). These systems converge often to a steady-state that can be exploited to design a fixed sampling set $\mathcal{S} = \mathcal{S}_t \; \forall t$.

Differently from the time varying models, the sampling set $\mathcal{S}$ affects now the entire system evolution. The *a priori* error covariance matrix $\boldsymbol{P}_t^-$ converges to the unique limit point $\boldsymbol{P}_\infty$ if: (i) the pair $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}})$ is stabilizable; and (ii) the pair $(\tilde{\boldsymbol{A}}, \boldsymbol{C}_{\mathcal{S}} \boldsymbol{U}_{\mathcal{F}})$ is detectable. The first condition is a characteristic of the graph process and is application specific. For stable time-invariant processes as the ones we discuss, this condition is satisfied. The second condition imposes the sampling set to guarantee steady-state convergence. As it follows from system theory, an observable system is also detectable; thus, the Kalman filter on graphs converges if the limiting observability matrix

$$\boldsymbol{O}_\infty = \lim_{T \to \infty} \boldsymbol{O}_{0:T} = \lim_{T \to \infty} (\boldsymbol{I}_{T+1} \otimes \boldsymbol{C}_{\mathcal{S}} \boldsymbol{U}_{\mathcal{F}}) \tilde{\boldsymbol{A}}_{0,T} \tag{29}$$

is full rank. Defining the complementary sampling set $\mathcal{S}^c = \mathcal{V}/\mathcal{S}$ and following similar arguments as in Theorem 1, the Kalman filtering on graphs converges if

$$\lim_{T \to \infty} \|\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}}^\mathsf{T} \boldsymbol{C}_{\mathcal{S}^c} \boldsymbol{U}_{\mathcal{F}}\| \le \lim_{T \to \infty} \frac{s_{\min}^2(\tilde{\boldsymbol{A}}_{0:T})}{s_{\max}^2(\tilde{\boldsymbol{A}}_{0:T})}. \tag{30}$$

Condition (30) accounts for the localization properties of the sampling set throughout the entire temporal evolution (in practice for $T \gg 0$). For a set $\mathcal{S}$ satisfying (30), the steady-state error covariance matrix $\boldsymbol{P}_\infty$ leads to the discrete algebraic Riccatti equation (DARE)

$$\boldsymbol{P}_\infty = \tilde{\boldsymbol{A}} \boldsymbol{P}_\infty \tilde{\boldsymbol{A}}^\mathsf{T} + \tilde{\boldsymbol{\Sigma}}_w - \tilde{\boldsymbol{A}} \boldsymbol{P}_\infty \boldsymbol{U}_{\mathcal{F}}^\mathsf{T} \boldsymbol{C}_{\mathcal{S}}$$
$$\times \left(\boldsymbol{C}_{\mathcal{S}} \boldsymbol{U}_{\mathcal{F}} \boldsymbol{P}_\infty \boldsymbol{U}_{\mathcal{F}}^\mathsf{T} \boldsymbol{C}_{\mathcal{S}} + \boldsymbol{C}_{\mathcal{S}} \boldsymbol{\Sigma}_v \boldsymbol{C}_{\mathcal{S}} \right)^\dagger \boldsymbol{C}_{\mathcal{S}} \boldsymbol{U}_{\mathcal{F}} \boldsymbol{P}_\infty \tilde{\boldsymbol{A}}^\mathsf{T}. \tag{31}$$

which can be solved numerically [50]. The steady-state Kalman gain matrix is

$$\boldsymbol{K}_\infty = \boldsymbol{P}_\infty \boldsymbol{U}_{\mathcal{F}}^\mathsf{T} \boldsymbol{C}_{\mathcal{S}} \left(\boldsymbol{C}_{\mathcal{S}} \boldsymbol{U}_{\mathcal{F}} \boldsymbol{P}_\infty \boldsymbol{U}_{\mathcal{F}}^\mathsf{T} \boldsymbol{C}_{\mathcal{S}} + \boldsymbol{C}_{\mathcal{S}} \boldsymbol{\Sigma}_v \boldsymbol{C}_{\mathcal{S}} \right)^\dagger \tag{32}$$

with posterior state estimate

$$\tilde{\boldsymbol{x}}_t^+ = \left(\boldsymbol{I}_{|\mathcal{F}|} - \boldsymbol{K}_\infty \boldsymbol{C}_{\mathcal{S}} \boldsymbol{U}_{\mathcal{F}} \right) \tilde{\boldsymbol{A}} \tilde{\boldsymbol{x}}_{t-1}^+ + \boldsymbol{K}_\infty \boldsymbol{y}_t. \tag{33}$$

The steady-state Kalman filter is only asymptotically optimal. As it follows from (32), a necessary condition to exploit the Kalman gain is that the cardinality of the sampling set satisfies $|\mathcal{S}| > |\mathcal{F}|$. That is, we need to fix at least $|\mathcal{F}|$ nodes. We discuss next how these nodes are sampled.

Sampling strategy. By following the sparse sensing idea, we design $\mathcal{S}$ as the set that minimizes the steady-state performance. This translates into solving the non-convex problem

$$\begin{array}{ll} \underset{\boldsymbol{c}}{\text{minimize}} & \text{Tr}(\boldsymbol{P}_\infty) \\ \text{subject to} & \boldsymbol{C}_{\mathcal{S}} = \text{diag}(\boldsymbol{c}) \\ & \|\boldsymbol{c}\|_0 = |\mathcal{S}|, \; \boldsymbol{c} \in \{0, 1\}^N \end{array} \tag{34}$$

Problem (34) finds the sampling set $\mathcal{S}$ that leads to the best steady-state tracking accuracy. This problem is intractable since a closed-form solution for the DARE (31) misses. We tackle this issue by adapting the greedy selection strategy from [56] as illustrated in Algorithm 2. The sampling strategy starts with an empty set; solves the DARE in *ii)* numerically for all nodes; and adds greedily the node with the smallest steady-state error increment in step *iii)*. The algorithm stops when $|\mathcal{S}|$ nodes are selected.

The greedy Algorithm 2 is optimal with respect to problem (34) if: *i)* the measurement noise $\boldsymbol{v}_t$ is uncorrelated; *ii)* the set of

---

Algorithm 2: Greedy node sampling algorithm for problem (34).

Start with an empty sampling set $\mathcal{S} = \emptyset$, a cardinality $|\mathcal{S}|$, and a counter $c = 0$:
i) while $c \le |\mathcal{S}|$
ii) Compute $\text{Tr}(\boldsymbol{P}_\infty(\mathcal{S} \cup \{n\}))$ as in (31) for all $n \in \mathcal{S}^c$
iii) Select n as $\text{argmin}_n \text{Tr}(\boldsymbol{P}_\infty(\mathcal{S} \cup \{n\}))$;
iv) Update the sampling set $\mathcal{S} = \mathcal{S} \cup \{n\}$;
v) Update the counter $c = c + 1$.

---

node information matrices $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_N$ with $\boldsymbol{F}_n = \sigma_v^{-2} \boldsymbol{u}_{\mathcal{F}n} \boldsymbol{u}_{\mathcal{F}n}^\mathsf{T}$ is ordered w.r.t. the order relation of positive semidefiniteness [56]. The first condition is easily met in practice. The second connects the graph topology and process bandwidth to the optimal sampling set $\mathcal{S}$. Consider then the information matrix for all nodes in $\mathcal{S}$

$$\boldsymbol{F}(\mathcal{S}) = \sum_{n=1}^{|\mathcal{S}|} \sigma_v^{-2} \boldsymbol{u}_{\mathcal{F}n} \boldsymbol{u}_{\mathcal{F}n}^\mathsf{T} = \boldsymbol{U}_{\mathcal{F}}^\mathsf{T} \boldsymbol{C}_{\mathcal{S}} \boldsymbol{\Sigma}_v^{-1} \boldsymbol{U}_{\mathcal{F}}.$$

Condition *ii)* implies that for two sets $\mathcal{S}'$ and $\mathcal{S}''$, the cost function in (34) satisfies $\text{Tr}(\boldsymbol{P}_\infty(\mathcal{S}')) \le \text{Tr}(\boldsymbol{P}_\infty(\mathcal{S}''))$ if $\boldsymbol{F}(\mathcal{S}') \succeq \boldsymbol{F}(\mathcal{S}'')$. That is, the node sampled in $\mathcal{S}'$ is a better choice than the node sampled in $\mathcal{S}''$ [56]. This guarantees Algorithm 2 builds the sampling set with the best steady-state tracking accuracy.

6. Numerical evaluation

We corroborate the above theory with numerical results on both synthetic and real scenarios. First, we test observability and then tracking. We used the GSP box [57] and CVX [58].

6.1. Observing graph processes

We test the observability with deterministic sampling on the Molene weather data set[5] and the observability with random sampling on the European tracer experiment (ETEX) data set[6] [32].

Deterministic sampling. The Molene weather data set contains $R = 744$ hourly temperature recordings collected in January 2014 over $N = 32$ cities in Brest (France). We used these recordings to build a synthetic diffusion scenario for model (3). Our rationale is to corroborate the sampling approaches when the bandlimited assumption on the state signal is violated. The graph is a k-nearest neighbor (kNN) with $k = 3$. Each temperature recording is treated as a separate instance; it represents the initial state $\boldsymbol{x}_0$ for model (3) and it is diffused with a weight $w = 1.5$ for $T = 10$ time instances.

First, we analyzed the effect of the sampling set $\mathcal{S}_{0:T}$ when the process has contribution on the entire bandwidth $|\mathcal{F}| = N = 32$. We corrupted the measurements $\boldsymbol{y}_t$ in (2b) with a zero-mean Gaussian noise with variance $\sigma_v^2 = 10^{-1}$. For $\boldsymbol{r}_\tau$ being the τth recording, this corresponds to an average signal-to-noise ratio (SNR) of 19.3dB computed as

$$\overline{\text{SNR}} = 10 \log_{10} \left[\frac{\sum_{\tau=1}^R \|\boldsymbol{r}_\tau\|_2^2}{N R \sigma_v^2} \right]. \tag{35}$$

The $|\mathcal{S}_{0:T}|$ samples are the ones that minimize the MSE in (17) in a sparse sense manner. The performance is evaluated through the normalized MSE (NMSE) between the estimated recording $\boldsymbol{r}_\tau^o$ and the true one $\boldsymbol{r}_\tau$ defined as

$$\text{NMSE} = \frac{\sum_{\tau=1}^R \|\boldsymbol{r}_\tau^o - \boldsymbol{r}_\tau\|^2}{\sum_{\tau=1}^R \|\boldsymbol{r}_\tau\|^2}. \tag{36}$$

---

[5] https://donneespubliques.meteofrance.fr/donnees_libres/Hackathon/RADOMEH.tar.gz.
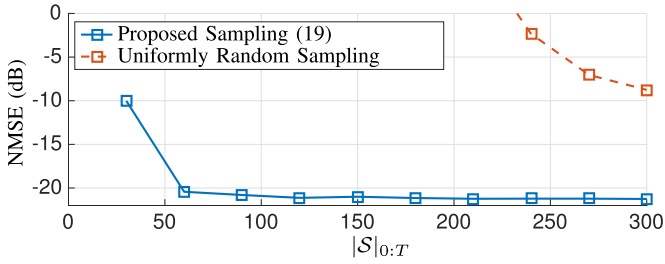
[6] https://rem.jrc.ec.europa.eu/RemWeb/etex/.

Fig. 1. NMSE versus the number of samples for the sampling strategy in (18) and uniformly random sampling whose NMSE is averaged over additional 100 iterations. The graph process has a localized spectrum on the entire bandwidth $|\mathcal{F}| = N = 32$.
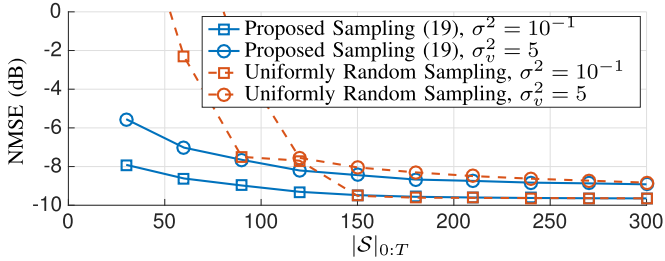

Fig. 2. NMSE versus the number of samples for the sampling strategy in (18) and uniformly random sampling. The state spectral evolution is localized on the first $|\mathcal{F}| = 8$ frequencies.

Table 1
Theoretical and empirical NMSE and cardinality of the sampling set $\mathcal{S}_{0:T}$ for different MSE requirements γ in (18).

	$\gamma = 2.05$	$\gamma = 2.5$	$\gamma = 3$	$\gamma = 3.5$
Theo. (18)	−21.26dB	−20.42dB	−19.64dB	−19.32dB
Emp.	−21.22dB	−20.37dB	−19.57dB	−19.28dB
$|\mathcal{S}_{0:T}|$	277	61	37	32

Fig. 1 shows the NMSE as a function of the number of samples $|\mathcal{S}_{0:T}|$. We observe that with 60 out of 320 samples the proposed approach achieves an NMSE of −20dB, while the uniformly random sampling requires far more measurements. *This finding suggests that sparse observability is possible also for graph processes that have a contribution on the entire bandwidth.* This is because of the sampling extrapolation to the temporal dimension.

To provide further insight, we show in Table 1 the theoretical and empirical NMSE as a function of the target value γ in (18). We also show the cardinality of the sampling set $\mathcal{S}_{0:T}$. These results suggest that a stricter NMSE requirement (smaller γ) leads to a larger number of nodes in $\mathcal{S}_{0:T}$ and, vice-versa, a looser NMSE requirement (larger γ) leads to a smaller number of nodes in $\mathcal{S}_{0:T}$.

Second, we analyzed the effect of the sampling set $\mathcal{S}_{0:T}$ when the process bandwidth is restricted to the first $|\mathcal{F}| = 8$ graph frequencies. This will inevitably cause a loss of information about $\boldsymbol{x}_0$. We tested two different noise variances $\sigma_v^2 = \{10^{-1}, 5\}$ ($\overline{\text{SNR}} = \{19.3\text{dB}, 2.3\text{dB}\}$). The $|\mathcal{S}_{0:T}|$ samples are collected as in the previous scenario.

Fig. 2 shows the proposed sampling approach yields a lower NMSE than uniformly random sampling. The NMSE has a higher floor when compared with the full bandwidth (Fig. 1) and its value does not reduce even by increasing the sampling set cardinality $|\mathcal{S}_{0:T}|$. We attribute this limitation to the restricted bandwidth: the out-of-band process contribution plays a role to improve the performance. We corroborate this observation in Fig. 3 where we show the average SNR as a function of the graph frequency index. We observe that in a low noise regime it is beneficial to consider a larger bandwidth since the average SNR per frequency is high. We
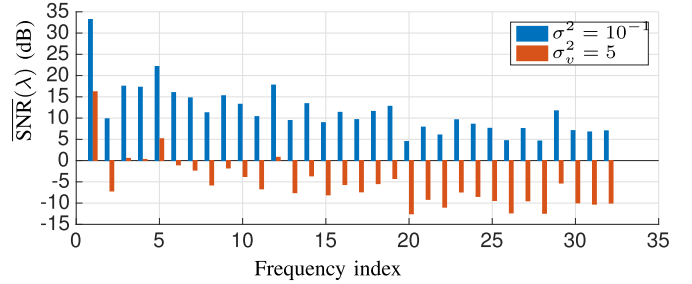

Fig. 3. Average SNR as a function of the graph frequency index n for two different noise powers. The graph process has a negative $\overline{\text{SNR}}$ in most of the graph frequencies.
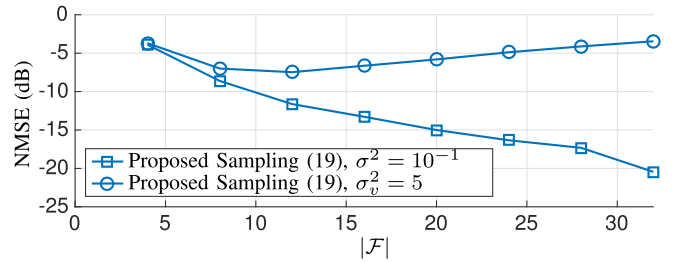

Fig. 4. NMSE versus the signal bandwidth $|\mathcal{F}|$ for different noise powers. The sampling set has cardinality $|\mathcal{S}_{0:T}| = 100$ and it is chosen by minimizing the MSE in (18). A larger bandwidth is unfavourable in high noise regimes.

do not observe this behavior in a high noise regime since the average SNR becomes negative at the higher graph frequencies. Hence, the SNR–not the signal energy alone–defines the signal bandwidth.

Third, we analyzed the effect of the signal bandwidth. We fixed the number of samples to $|\mathcal{S}_{0:T}| = 100$ and computed the NMSE for different bandwidths $|\mathcal{F}|$ and noise powers σ_v^2.

Fig. 4 shows an increasing trend of the NMSE in the high noise regime ($\sigma_v^2 = 5$), implying the meaningful information is concentrated in the first few frequencies and, therefore, the Molene recordings are naturally bandlimited. By increasing the bandwidth $|\mathcal{F}|$, the average SNR degrades and we end up bringing in more noise; hence, the performance degrades. This observation is reinforced in the low noise regime, where a larger bandwidth is preferred to exploit the SNR at the higher frequencies and improve observability. *This result further corroborates that the signal-to-noise ratio determines the process bandwidth.*

Random sampling. The ETEX experiment contains measurements of an identifiable perfluorocarbon concentration released near Rennes (France) and diffused over Europe [32]. Thirty measurements were collected over a period of 72 h at $N = 168$ ground-level stations. These stations are the nodes of a kNN graph and the $T = 30$ temporal measurements are the graph process. For different reasons, the measurements are often unavailable and, in these cases, the tracer concentration is set to zero. We considered model (3) to capture the tracer evolution over time. Set $\mathcal{F}$ contains the frequency indices where 30% of the process energy concentrates with cardinality $|\mathcal{F}| = 6$. Since diffused graph signals are often bandlimited, our intuition is that also in this experiment most of the frequencies will not have useful information. With this setup, we found heuristically that $k = 3$ and $w = 3.5$ lead to the smallest estimation error by using all the 168×72 recordings. This result will serve as a baseline to test observability with random sampling. To account for the measurement noise, we corrupted the signal with a zero-mean Gaussian noise of variance $\sigma_v^2 = 10^{-4}$ ($\overline{\text{SNR}} = 29.2\text{dB}$). The obtained results are averaged over 2000 iterations.
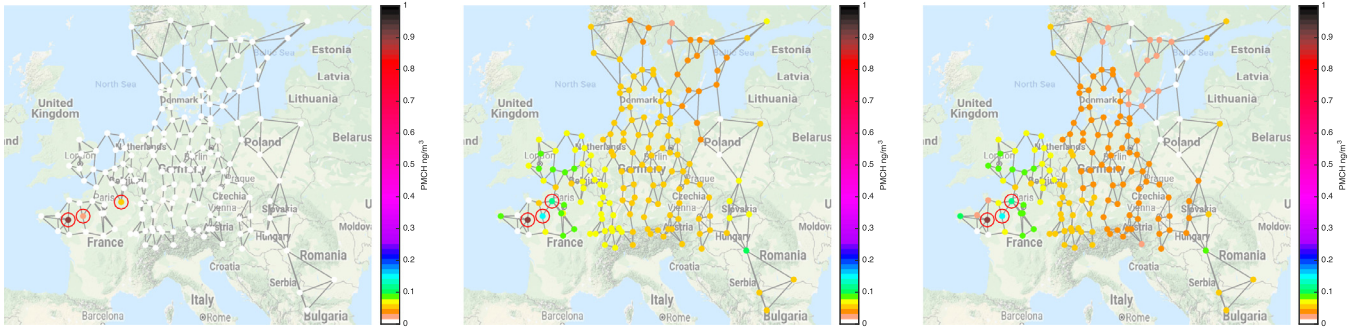
Fig. 5. PMCH tracer concentration at 168 stations. The three nodes with the highest concentration are circled in red. The graph is the 3NN. (Left) Ground truth concentration at $t = 0$. (Center) Observed tracer concentration with all samples by the diffusion model in (3) with diffusion weight $w = 3.5$. (Right) Mean observed tracer concentration following observability with random sampling with an overall sampling rate of $\mathbf{1}_N^\top \bar{c} = 60$ (out of 168). The NMSE between the observed state with random sampling (right) and the observed ground truth (center) is -16.2 dB and has a variance of -48.1dB. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
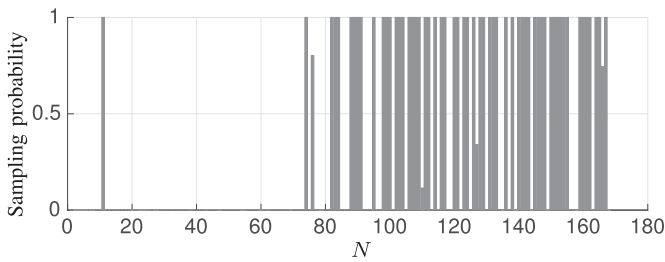
Fig. 6. Optimal sampling probabilities over the nodes obtained from solving the opposite of problem (23). We observe that several nodes are sampled with probability one and that the overall solution is highly sparse.

Table 2
Theoretical and empirical NMSE, overall sampling rate, and parameter α in (21) for different values of MSE requirements γ ($\times 10^{-4}$) in (18).

	$\gamma = 3.12$	$\gamma = 3.15$	$\gamma = 3.18$	$\gamma = 3.21$
Theo. Lower Bound (22)	-36.47dB	-36.42dB	-36.38dB	-36.34dB
Emp.	-24.52dB	-18.77dB	-16.48dB	-16.13dB
$\mathbf{1}_N^\top \bar{c}$	130.4	83.9	62.7	50.2
α	3910	2515	1881	1505

Fig. 5 (left) shows the tracer concentration at $t = 0$, wherein the red circles are the three nodes with the highest concentration. The middle plot shows the estimate of the observed state by using model (3) with all nodes collecting measurements. The graph model is able to identify the region where the tracer was released, but, at the same time, they yield a tracer concentration of around 0.04ng/m³ on all nodes. We attribute this leakage to the missing values that are set to zero and the absence of wind information on those days. For the scope of this work, we will use this observed state (Fig. 5 (center)) as a baseline since it is the best result with the fitted model. The right plot shows the average observed state with a sampling rate of 60; the sampling probabilities are obtained by solving the opposite of problem (23) and are illustrated in Fig. 6. The proposed sampling approach achieves an average NMSE of -16.2dB with a variance of -48.1dB.

We observe the following: (i) the deviation of a particular realization from the averaged observed state is in general negligible, yielding a good practical result; and (ii) similarly to the approaches that use the CRLB to perform sparse sampling, the lower bound (22) is a suitable cost function to design a sparse sampler. These observations should nevertheless be considered insightful and not conclusive since random sampling is validated only in a specific scenario.

Table 2 shows the impact of the MSE requirement γ in (23) on the lower bound (22), empirical NMSE, overall sampling rate $\mathbf{1}_N^\top \bar{c}$,
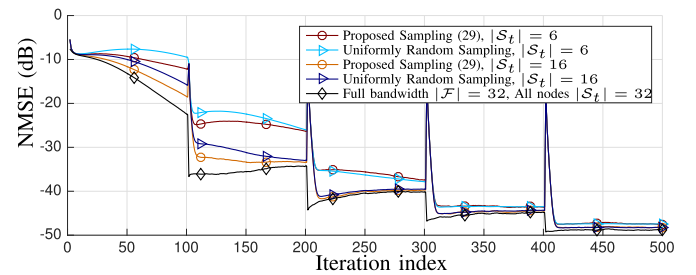

Fig. 7. NMSE versus the iteration index for different numbers of sampled nodes. The results are shown for the sampling approach in (28), uniformly random sampling, and all nodes sampled.

and α in (21). Despite the gap between the lower bound and the empirical NMSE, a looser requirement on (22) induces a lower sampling rate; and all the reported values of α lead to a probability in (21) below machine precision. This corroborates criterion (22) to design the node sampling probabilities for observability with random sampling.

6.2. Tracking graph processes

We now corroborate the performance of the time varying and steady-state Kalman filtering discussed in Section 5. All results are averaged over 500 different realizations.

Time varying. We considered a synthetic tracking scenario where the graph process follows the diffusion model in (3) with a zero initial state and inputs $\boldsymbol{u}_t$ for $t \in \{1, 101, \ldots, 401\}$ taken from the recordings of the Molene data set. We considered real (non synthetic) measurements as input to represent approximately bandlimited processes. The graph is a 3NN, the process bandwidth $\mathcal{F}$ consists of the first 16 graph frequencies, the diffusion weight is $w = 1$, and the input matrix is $\boldsymbol{B}_t = \boldsymbol{I}_N$. In a nutshell, the state evolution considers temperature diffusion for 100 iterations and then a new input is introduced. The model and measurement noise are zero-mean with covariance matrixes $\boldsymbol{\Sigma}_w = 10^{-4}\boldsymbol{I}_N$ and $\boldsymbol{\Sigma}_v = 10^{-1}\boldsymbol{I}_N$, respectively. We initialized the Kalman filter with *a posteriori* estimate $\tilde{\boldsymbol{x}}_0^+ = \mathbf{1}_{|\mathcal{F}|}$ and error covariance matrix $\boldsymbol{P}_0^+ = \boldsymbol{\Sigma}_{\tilde{w}}$. We compared the proposed sparse sensing approach (the opposite problem of (28) that selects $|\mathcal{S}|$ nodes with minimum MSE) with the uniformly random sampling whose performance is averaged over 500 additional realizations.

Fig. 7 shows the tracking NMSE as a function of the iteration index for different values of sampling set cardinalities $|\mathcal{S}_t|$. When more nodes are sampled a smaller NMSE is achieved. This is evidenced in the first iterations. These results show also that 50% of
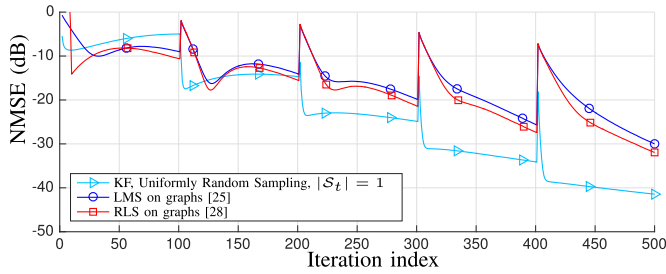
Fig. 8. NMSE versus iteration index for Kalman filter (KF), LMS ($\mu = 0.0875$) [21] and RLS ($\beta = 0.95$) [27]. The sampling set $\mathcal{S}_t$ of the Kalman filter is composed of one node, while the sampling rate for the LMS and RLS on graphs is 16.08 (greater than $|\mathcal{F}| = 16$) and with five nodes sampled with probability one.
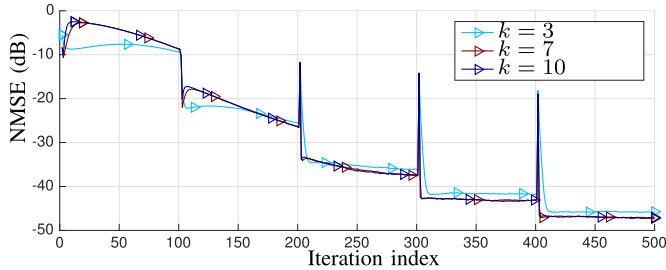

Fig. 9. NMSE versus the iteration index for graph topologies when $|\mathcal{S}_t| = 6$ nodes are sampled uniformly at random.
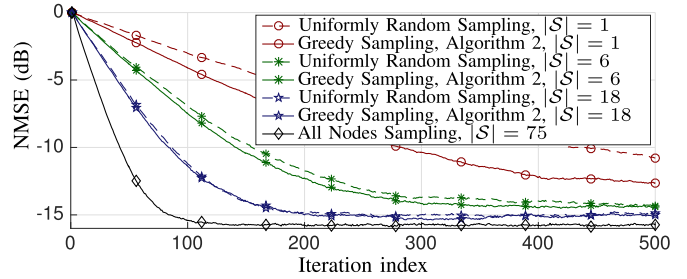

Fig. 10. NMSE versus iteration index for steady-state Kalman filter for different cardinalities of the sampling set. The results are shown for the sampling approach in Algorithm 2, uniformly random sampling, and all nodes sampled.
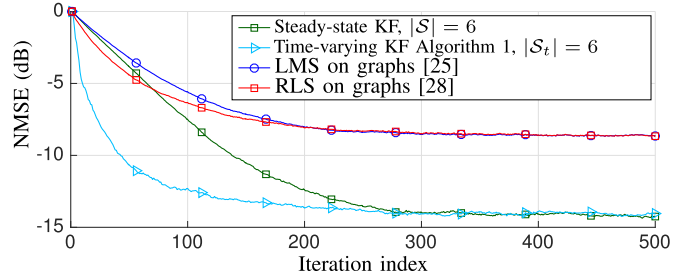

Fig. 11. NMSE versus iteration index for the steady-state Kalman filter (KF) ($|\mathcal{S}| = 6$ chosen with Algorithm 2), time varying Kalman filter ($|\mathcal{S}_t| = |\mathcal{S}| = 6$ chosen randomly for each iteration), RLS ($\beta_{RLS} = 0.99$), and LMS ($\mu_{LMS} = 0.041$) with sampling rate $|\mathcal{S}_t| = |\mathcal{F}| = 18$.

the samples can be saved compared with the full bandwidth case, yet without hindering the NMSE. Uniformly random sampling is an option to track the process only for large t. We also remark that problem (28) may often give a dense solution for higher t and, since it is an SDP relaxation, it might also lead to sampling rates that are far from the minimum MSE. Finally, remark the spikes in the estimated NMSE represent the input signal and are common for both sampling strategies.

Next, we compare the Kalman filter with the LMS [21] and RLS [27] on graphs. The sampling set $\mathcal{S}_t$ consists of one node sampled uniformly at random. The sampling probabilities for the RLS on graphs are found with $\beta_{RLS} = 0.95$ and $\gamma_{RLS} = 7 \times 10^{-2}$ [27]. These parameters result in an average sampling rate of 16.08 (greater than $|\mathcal{F}| = 16$) for each t and with five nodes sampled with probability one. These sampling probabilities are also used for the LMS on graphs and its step size is set to $\mu_{LMS} = 0.0875$ to meet the RLS steady-state MSE. Both algorithms are initialized as the Kalman filter. Fig. 8 shows the Kalman filter suffers in the first iterations, but it outperforms both LMS and RLS (hence, other state-of-the-art algorithms with which LMS and RLS compare well [21,26,27]) when the system evolution is learned better. This result highlights the potential of the proposed approach to track the process by collecting only one sample at a time.

In Fig. 9, we further analyze how the graph topology affect the Kalman filter. We changed the number of neighbors, k, in the kNN graph, while kept fixed all other parameters. The sampling set is composed of $|\mathcal{S}_t| = 6$ nodes selected uniformly at random. The Kalman filter tracks faster in less connected graphs but when the initialization effect vanished (after more iterations) it tracks better in denser graphs. For $k \geq 5$, we have not seen any significant different in tracking accuracy.

Steady-state. We test the steady-state Kalman filter to track a heat diffusion process on a rectangular grid of $N = 75$ nodes (5×15) with unitary edge weights. We aim to provide insights into how GSP tools can be useful in temperature monitoring systems. The initial state $\boldsymbol{x}_0$ is one at the five nodes of the leftmost column of the grid and zero elsewhere. This signal is diffused fol-

lowing the heat propagating model (3) with weight $w = 10$ for $T = 500$ instances. Set $\mathcal{F}$ comprises the frequency indices where 99% of the energy of $\boldsymbol{x}_0$ is concentrated and resulted into $|\mathcal{F}| = 18$ active frequencies. We considered a model and measurement noise with covariance matrices $\boldsymbol{\Sigma}_w = 10^{-4}\boldsymbol{I}_N$ and $\boldsymbol{\Sigma}_\nu = 10^{-1}\boldsymbol{I}_N$, respectively.

Fig. 10 shows the tracking NMSE as a function of the diffusion time for different cardinalities of the sampling set. A larger number of samples improves the steady-state NMSE and convergence rate. The sampling strategy in Algorithm 2 is also beneficial when a few samples are collected, while uniformly random sampling can be adopted for larger $|\mathcal{S}|$.

We then compared the steady-state Kalman filter with the LMS and RLS on graphs algorithms. We chose the parameters of the adaptive algorithms as in Fig. 8 and found that $\mu_{LMS} = 0.041$ (LMS), $\beta_{RLS} = 0.99$ (RLS), and a sampling rate of 18 samples per iteration (equalling the bandwidth). The Kalman filtering with time varying sampling in Algorithm 1 is considered as baseline with sampling set $|\mathcal{S}_t| = |\mathcal{S}| = 6$ chosen uniformly at random.

Fig. 11 shows that the Kalman filter outperforms the adaptive algorithms in both the steady-state performance and convergence speed. Both Kalman filter approaches reach an identical steady-state NMSE, but with a sampling rate that is three times smaller compared with the adaptive algorithms. The convergence improvement of the time varying Kalman filter comes at the expense of complexity, i.e., the updates for each t of the Kalman gain matrix, the *a priori*, and *a posteriori* error covariance matrices.

In Fig. 12, we finally show how different grid connectivities affect the steady-state Kalman filter. In addition to the original rectangular grid, we considered also a star grid (i.e., nodes have also diagonal connections) and a multihop grid (i.e., nodes are connected also with the two hops away neighbors of the rectangular grid). The tracker is quite robust to different grid connectivities.[7]

---

[7] For each graph, the bandwidth is formed by those frequencies where 99% of the energy is concentrated.
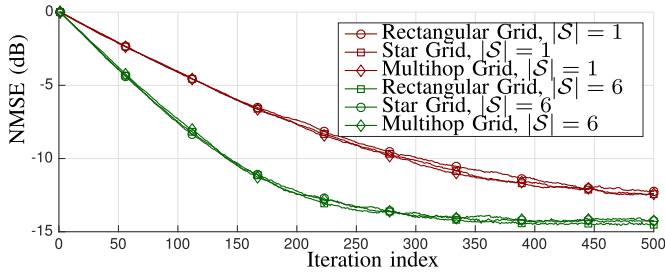
Fig. 12. NMSE versus the iteration index for the steady-state Kalman filter over different graph topologies. The sampling set is build following Algorithm 2.

Differently from the time varying scenario, we considered the greedy sampling strategy which has a bigger impact in the performance. The proposed method adapts the sampled nodes to achieve the optimal steady-state performance which is seen to have also a similar convergence rate.

7. Conclusion

This work generalized graph signal sampling to the temporal dimension for observing and tracking graph processes with contribution only on a fixed set of graph frequencies. We derived necessary and sufficient conditions to observe a bandlimited graph process from a subset of nodes. These conditions relate the graph structure with the process bandwidth and the location of the samples on the graph; and they treat sampling theory for graph signals and adaptive algorithms as particular cases. We also developed a mathematical framework for observability with random sampling: we proposed conditions to observe the graph process and sampling strategies. The proposed findings extend naturally to tracking the graph process with a Kalman filter. We derived conditions on the minimum number of nodes such that the Kalman gain is fully leveraged and provided sampling strategies for tracking. The theory is corroborated with numerical experiments, which show the usefulness of the bandlimited prior to reduce the number samples collected over graph and time.

We identify three directions that require additional research. First, we considered the process bandwidth time-invariant. This is necessary for our derivations to find recovery conditions; however, it limits the applicability to a few models. It should be investigated if similar or different conditions can be obtained when this assumption is violated. Second, our derivations concern graph topologies that are time-invariant. It is an interesting direction to generalize some of these findings to time varying graphs. Third, from a more practical perspective, future research is needed to adapt the proposed sampling methods to nonlinear dynamics over graphs which alter the state bandwidth over time

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

A1. Proof of Proposition 1

From Definition 1, we have that an $\mathcal{F}$−bandlimited graph process has contribution only on the set $\mathcal{F}$ of graph frequencies. Then, under Assumption 1, we can write $\boldsymbol{x}_t$ as

$$\boldsymbol{x}_t = \boldsymbol{U}\hat{\boldsymbol{A}}_{t-1}\boldsymbol{U}^{\top}\boldsymbol{x}_{t-1} + \boldsymbol{U}\hat{\boldsymbol{B}}_{t-1}\boldsymbol{U}^{\top}\boldsymbol{u}_{t-1} \tag{37}$$

where $\hat{\boldsymbol{A}}_{t-1}$ and $\hat{\boldsymbol{B}}_{t-1}$ are $N \times N$ diagonal matrices containing the eigenvalues of $\boldsymbol{A}_{t-1}$ and $\boldsymbol{B}_{t-1}$, respectively. Then, since $\hat{\boldsymbol{x}}_{t-1} = \boldsymbol{U}^{\top}\boldsymbol{x}_{t-1}$ and $\hat{\boldsymbol{u}}_{t-1} = \boldsymbol{U}^{\top}\boldsymbol{u}_{t-1}$ are the GFTs of $\boldsymbol{x}_{t-1}$ and $\boldsymbol{u}_{t-1}$, respectively, we have that

$$\hat{\boldsymbol{x}}_t = \hat{\boldsymbol{A}}_{t-1}\hat{\boldsymbol{x}}_{t-1} + \hat{\boldsymbol{B}}_{t-1}\hat{\boldsymbol{u}}_{t-1}. \tag{38}$$

Since matrices $\hat{\boldsymbol{A}}_{t-1}$ and $\hat{\boldsymbol{B}}_{t-1}$ are diagonal, we can split (38) into N independent models with nth model representing the process evolution in the nth graph frequency. Then, $\hat{\boldsymbol{x}}_t$ evolves in the frequency set $\mathcal{F}$ only if the previous state $\hat{\boldsymbol{x}}_{t-1}$ or the input $\hat{\boldsymbol{u}}_{t-1}$ have a contribution on this set. Since $\hat{\boldsymbol{u}}_{t-1}$ is $\mathcal{F}$−bandlimited [cf. Assumption 2], it is zero in the complementary frequency set $\mathcal{F}^c$. Hence, state $\hat{\boldsymbol{x}}_t$ evolves in $\mathcal{F}^c$ only if the former state $\hat{\boldsymbol{x}}_{t-1}$ has a contribution on this set. The latter forms a recursion and implies that state $\hat{\boldsymbol{x}}_t$ evolves on $\mathcal{F}^c$ only if the initial state $\hat{\boldsymbol{x}}_0$ in non-zero on this set.

For the sufficiency part, observe that if $\hat{\boldsymbol{x}}_0$ is $\mathcal{F}$−bandlimited, it is zero on the complementary set $\mathcal{F}^c$. This implies that $\hat{\boldsymbol{x}}_t$ has a contribution only in $\mathcal{F}$; hence, $\mathcal{F}$−bandlimited. For the necessary part, observe that if $\hat{\boldsymbol{x}}_0$ is not $\mathcal{F}$−bandlimited, it is non-zero on the complementary set $\mathcal{F}^c$. This implies that $\hat{\boldsymbol{x}}_t$ has a contribution also on $\mathcal{F}^c$; hence, not $\mathcal{F}$−bandlimited. □

A2. Proof of Proposition 2

By applying the rank inequality $\text{rank}(\boldsymbol{A}\boldsymbol{B}) \leq \min\{\text{rank}(\boldsymbol{A}), \text{rank}(\boldsymbol{B})\}$ to $\boldsymbol{O}_{0:T} = \boldsymbol{C}_{\mathcal{S}_{0:T}}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\tilde{\boldsymbol{A}}_{0:T}$ in (12), $\boldsymbol{O}_{0:T}$ is full column rank $|\mathcal{F}|$ only if

$$\text{rank}(\boldsymbol{C}_{\mathcal{S}_{0:T}}) \geq |\mathcal{F}| \tag{39}$$

which from the structure of $\boldsymbol{C}_{\mathcal{S}_{0:T}}$ is true when the claim is satisfied. □

A3. Proof of Theorem 1

By substituting $\boldsymbol{C}_{\mathcal{S}_{0:T}} = \boldsymbol{I}_{N(T+1)} - \boldsymbol{C}_{\mathcal{S}_{0:T}}^c$ into the rank argument of (14) we can write the matrix form expression

$$\tilde{\boldsymbol{A}}_{0:T}^{\top}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}}^{\top})\boldsymbol{C}_{\mathcal{S}_{0:T}}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\tilde{\boldsymbol{A}}_{0:T} = \tilde{\boldsymbol{A}}_{0:T}^{\top}\tilde{\boldsymbol{A}}_{0:T} \\ -\tilde{\boldsymbol{A}}_{0:T}^{\top}(\boldsymbol{I}_{(T+1)} \otimes \boldsymbol{U}_{\mathcal{F}}^{\top})\boldsymbol{C}_{\mathcal{S}_{0:T}^c}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\tilde{\boldsymbol{A}}_{0:T}, \tag{40}$$

which is invertible if

$$\|\tilde{\boldsymbol{A}}_{0:T}^{\top}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}}^{\top})\boldsymbol{C}_{\mathcal{S}_{0:T}^c}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\tilde{\boldsymbol{A}}_{0:T}\| < \lambda_{\min}(\tilde{\boldsymbol{A}}_{0:T}^{\top}\tilde{\boldsymbol{A}}_{0:T}), \tag{41}$$

where $\lambda_{\min}(\boldsymbol{A})$ is the minimum eigenvalue of $\boldsymbol{A}$. Here, we exploited the positive semi-definiteness of both matrices on the right-hand side of (40). Then, from the norm sub-multiplicativity we have

$$\|\tilde{\boldsymbol{A}}_{0:T}^{\top}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}}^{\top})\boldsymbol{C}_{\mathcal{S}_{0:T}^c}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\tilde{\boldsymbol{A}}_{0:T}\| \leq$$

$$\|(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}}^{\top})\|\|\boldsymbol{C}_{\mathcal{S}_{0:T}^c}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\|\|\tilde{\boldsymbol{A}}_{0:T}\|^2 < \lambda_{\min}(\tilde{\boldsymbol{A}}_{0:T}^{\top}\tilde{\boldsymbol{A}}_{0:T}),$$

which then leads to ($\|(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}}^{\top})\| = 1$)

$$\|\boldsymbol{C}_{\mathcal{S}_{0:T}^c}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\| < \frac{\lambda_{\min}(\tilde{\boldsymbol{A}}_{0:T}^{\top}\tilde{\boldsymbol{A}}_{0:T})}{\|\tilde{\boldsymbol{A}}_{0:T}\|^2} = \frac{s_{\min}^2(\tilde{\boldsymbol{A}}_{0:T})}{s_{\max}^2(\tilde{\boldsymbol{A}}_{0:T})}. \tag{42}$$

The equality in (42) derives from the definition of the spectral norm and the relation between the singular and the eigenvalues of a matrix.

To prove that (15) (i.e., (42)) is a necessary and sufficient condition we follow similar arguments as in [11,21]. For the sufficiency part, consider that an $\mathcal{F}$−bandlimited system on graph is observable if and only if $\boldsymbol{O}_{0:T}$ is full rank and, therefore, if and only if (14) holds. Then, from (40)–(42), we have (14) always holds true if (15) (i.e., (42)) is true. To prove (15) (i.e., (42)) is necessary, we proceed by contradiction. When (15) (i.e., (42)) does not hold we can write

$$\|\boldsymbol{C}_{\mathcal{S}_{0:T}^c}(\boldsymbol{I}_{T+1} \otimes \boldsymbol{U}_{\mathcal{F}})\| = \lambda_{\min}(\tilde{\boldsymbol{A}}_{0:T}^{\top}\tilde{\boldsymbol{A}}_{0:T})/\|\tilde{\boldsymbol{A}}_{0:T}\|^2 + \epsilon$$

for some scalar $\epsilon \geq 0$. Then, for $T = 0$ and thus $\tilde{A}_{0:0} = I_N$ we have $\|C_{\mathcal{S}_0^c}(I_1 \otimes U_{\mathcal{F}})\| = 1 + \epsilon$, which contradicts the conventional observability (recovery) of bandlimited graph signals in (1). This proves that (15) is also necessary. □

A4. Proof of Proposition 4

From the structure of $C_{\mathcal{S}_{0:T}}$, we have that

$$\mathrm{rank}(C_{\mathcal{S}_{0:T}}) \leq \mathrm{rank}(E[C_{\mathcal{S}_{0:T}}]) = \mathrm{rank}(I_{T+1} \otimes \bar{C}). \tag{43}$$

A necessary condition then for $\mathrm{rank}(C_{\mathcal{S}_{0:T}})$ to be $|\mathcal{F}|$ is

$$\mathrm{rank}(I_{T+1} \otimes \bar{C}) \geq |\mathcal{F}|. \tag{44}$$

From $\mathrm{rank}(A \otimes B) = \mathrm{rank}(A)\mathrm{rank}(B)$, (44) writes as

$$\mathrm{rank}(\bar{C}) \geq |\mathcal{F}|/(T+1). \tag{45}$$

Then, since $\bar{C}$ is diagonal, it means that at least $\lceil |\mathcal{F}|/(t+1) \rceil$ nodes must be sampled with a probability different from zero.

A5. Proof of Proposition 5

Denote by $c_t = \mathrm{diag}(C_{\mathcal{S}_t})$ the random sampling vector with expectation $\bar{c}$ for $t \in \{0, \dots, T\}$. Let also $d = |\mathcal{S}_{0:T}| = \sum_{t=0}^{T} \sum_{n=1}^{N} c_{tn}$ be an auxiliary variable that characterises the cardinality of realization of the set $\mathcal{S}_{0:T}$. Then, d is a Poisson binomial random variable being it the sum of $N(T+1)$ independent Bernoulli random variables [59]. The claim (21) follows by simple statistical properties.

A6. Proof of Proposition 6

By rewriting the MSE as

$$\mathrm{MSE} = \mathbb{E}_C\left\{ \mathbb{E}_\nu\left[\mathrm{Tr}\left[(\tilde{x}_0^o - \tilde{x}_0)(\tilde{x}_0^o - \tilde{x}_0)^\top \right] \right] \right\} \tag{46}$$

and from (17) we have

$$\mathrm{MSE} = \sigma_\nu^2 \mathbb{E}_C\left\{ \mathrm{Tr}\left[\left(\tilde{A}_{0:T}^\top (I_{T+1} \otimes U_{\mathcal{F}})^\top C_{\mathcal{S}_{0:T}}(I_{T+1} \otimes U_{\mathcal{F}})\tilde{A}_{0:T} \right)^{-1} \right] \right\}. \tag{47}$$

Then, since function $\varphi : X \to \mathrm{Tr}[X^{-1}]$ is convex for a positive semidefinite matrix X, we apply Jensen's inequality $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$ to lower bound (47) as in (22).

CRediT authorship contribution statement

Elvin Isufi: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Paolo Banelli:** Conceptualization, Writing - review & editing, Funding acquisition. **Paolo Di Lorenzo:** Writing - review & editing, Funding acquisition. **Geert Leus:** Writing - review & editing, Supervision.

References

[1] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, 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. 30 (3) (2013) 83–98.
[2] A. Sandryhaila, J.M.F. Moura, Discrete signal processing on graphs, Trans. Signal Process. 61 (7) (2013) 1644–1656.
[3] S. Segarra, A. Marques, A. Ribeiro, Optimal graph-filter design and applications to distributed linear network operators, IEEE Trans. Signal Process. (2017).
[4] E. Isufi, A. Loukas, A. Simonetto, G. Leus, Autoregressive moving average graph filtering, IEEE Trans. Signal Process. 65 (2017) 274–288.
[5] E. Isufi, A. Loukas, A. Simonetto, G. Leus, Filtering random graph processes over random time-varying graphs, IEEE Trans. Signal Process. 65 (16) (2017) 4406–4421.
[6] M. Coutino, E. Isufi, G. Leus, Advances in distributed graph filtering, IEEE Trans. Signal Process. 67 (9) (2019) 2320–2333.
[7] S.K. Narang, A. Gadde, A. Ortega, Signal processing techniques for interpolation in graph structured data, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2013, pp. 5445–5449.
[8] S. Chen, R. Varma, A. Sandryhaila, J. Kovacevic, Discrete signal processing on graphs: sampling theory, IEEE Trans. Signal Process. 63 (24) (2015) 6510–6523.
[9] I. Pesenson, Sampling in Paley-Wiener spaces on combinatorial graphs, Trans. Am. Math. Soc. 360 (10) (2008) 5603–5627.
[10] A.G. Marques, S. Segarra, G. Leus, A. Ribeiro, Sampling of graph signals with successive local aggregations, IEEE Trans. Signal Process. 64 (7) (2016) 1832–1843.
[11] M. Tsitsvero, S. Barbarossa, P. Di Lorenzo, Signals on graphs: uncertainty principle and sampling, IEEE Trans. Signal Process. 64 (18) (2016) 4845–4860.
[12] C. Hu, J. Sepulcre, K.A. Johnson, G.E. Fakhri, Y.M. Lu, Q. Li, Matched signal detection on graphs: theory and application to brain imaging data classification, Neuroimage 125 (2016) 587–600.
[13] W. Huang, A.G. Marques, A.R. Ribeiro, Rating prediction via graph signal processing, IEEE Trans. Signal Process. 66 (19) (2018) 5066–5081.
[14] X. Wang, P. Liu, Y. Gu, Local-set-based graph signal reconstruction, IEEE Trans. Signal Process. 63 (9) (2015) 2432–2444.
[15] S.K. Narang, A. Gadde, E. Sanou, A. Ortega, Localized iterative methods for interpolation in graph structured data, in: Proceedings of the Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2013.
[16] F. Grassi, A. Loukas, N. Perraudin, B. Ricaud, A time-vertex signal processing framework, IEEE Trans. Signal Process. 66 (3) (2018) 817–829.
[17] E. Isufi, A. Loukas, A. Simonetto, G. Leus, Separable autoregressive moving average graph-temporal filters, in: Proceedings of the Signal Processing Conference (EUSIPCO), 2016 24th European, IEEE, 2016.
[18] E. Isufi, G. Leus, P. Banelli, 2-Dimensional Finite Impulse Response Graph-Temporal Filters, in: Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP), Washington D.C, USA, 2016.
[19] J. Mei, J.M. Moura, Signal processing on graphs: causal modeling of unstructured data, IEEE Trans. Signal Process. 65 (8) (2017) 2077–2092.
[20] A. Loukas, E. Isufi, N. Perraudin, Predicting the evolution of stationary graph signals, in: Proceedings of the Asilomar Conference on Signals, Systems and Computations, 2017.
[21] P. Di Lorenzo, S. Barbarossa, P. Banelli, S. Sardellitti, Adaptive least mean squares estimation of graph signals, IEEE Trans. Signal Inf. Process. Netw. 2 (4) (2016) 555–568.
[22] Y. Xue, S. Pequito, J.R. Coelho, P. Bogdan, G.J. Pappas, Minimum number of sensors to ensure observability of physiological systems: a case study, in: Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, 2016, pp. 1181–1188.
[23] S. Pequito, P. Bogdan, G.J. Pappas, Minimum number of probes for brain dynamics observability, in: Proceedings of the IEEE 54th Annual Conference on Decision and Control (CDC), IEEE, 2015, pp. 306–311.
[24] S. Pequito, F. Rego, S. Kar, A.P. Aguiar, A. Pascoal, C. Jones, Optimal design of observable multi-agent networks: a structural system approach, in: Control Conference (ECC), IEEE, 2014.
[25] P. Di Lorenzo, P. Banelli, S. Barbarossa, S. Sardellitti, Distributed adaptive learning of graph signals, IEEE Trans. Signal Process. 65 (16) (2016) 4193–4208.
[26] P. Di Lorenzo, E. Isufi, P. Banelli, S. Barbarossa, G. Leus, Distributed recursive least squares strategies for adaptive reconstruction of graph signals, in: Proceedings of the EURASIP European Signal Processing Conference (EUSIPCO)., 2017.
[27] P. Di Lorenzo, P. Banelli, E. Isufi, S. Barbarossa, G. Leus, Adaptive graph signal processing: algorithms and optimal sampling strategies, IEEE Trans. Signal Process. 66 (13) (2017) 3584–3598.
[28] H. Geng, Z. Wang, Y. Cheng, Distributed federated Tobit Kalman filter fusion over a packet-delaying network: a probabilistic perspective, IEEE Trans. Signal Process. 66 (17) (2018) 4477–4489.
[29] D. Romero, V.N. Ioannidis, G.B. Giannakis, Kernel-based reconstruction of space-time functions on dynamic graphs, IEEE Trans. Signal Process. 11 (6) (2017) 856–869.
[30] A. Soule, K. Salamatian, A. Nucci, N. Taft, Traffic matrix tracking using Kalman filters, ACM SIGMETRICS Perform. Eval. Rev. 33 (3) (2005) 24–31.
[31] F.S. Cattivelli, A.H. Sayed, Diffusion strategies for distributed Kalman filtering and smoothing, IEEE Trans. Autom. Control 55 (9) (2010) 2069–2084.
[32] K. Nodop, R. Connolly, F. Girardi, The field campaigns of the european tracer experiment (etex): overview and results, Atmos. Environ. 32 (24) (1998) 4095–4108.
[33] G. Taubin, Geometric Signal Processing on Polygonal Meshes, EUROGRAPHICS, 2000.
[34] M. Coutino, E. Isufi, T. Maehara, G. Leus, State-space network topology identification from partial observations, arXiv:1906.10471 (2019).
[35] S. Segarra, A.G. Marques, G. Mateos, A. Ribeiro, Network topology inference from spectral templates, IEEE Trans. Signal Inf. Process. Netw. 3 (3) (2017) 467–483.
[36] D. Thanou, X. Dong, D. Kressner, P. Frossard, Learning heat diffusion graphs, IEEE Trans. Signal Inf. Process. Netw. 3 (3) (2017) 484–499.
[37] R.I. Kondor, J. Lafferty, Diffusion kernels on graphs and other discrete input spaces, in: Proceedings of the ICML, 2, 2002, pp. 315–322.
[38] J.C. Dittmer, Consensus formation under bounded confidence, Nonlinear Anal.: Theory. Methods Appl. 47 (7) (2001) 4615–4621.

[39] H. Ma, H. Yang, M.R. Lyu, I. King, Mining social networks using heat diffusion processes for marketing candidates selection, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, ACM, 2008, pp. 233–242.

[40] A. Tarun, D. Van De Ville, Extrapolating functional mri data into white matter via structurally-informed graph diffusion, in: Proceedings of the Organization for Human Brain Mapping (OHBM) Meeting, 2018.

[41] F. Escolano, E.R. Hancock, M.A. Lozano, Heat diffusion: thermodynamic depth complexity of networks, Phys. Rev. E 85 (3) (2012) 036206.

[42] J. Friedman, J.-P. Tillich, Wave equations for graphs and the edge-based Laplacian, Pacific J. Math. 216 (2) (2004) 229–266.

[43] E. Isufi, A. Loukas, G. Leus, Autoregressive moving average graph filtering: a stable distributed implementation, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.

[44] F. Zhang, E.R. Hancock, Graph spectral image smoothing using the heat kernel, Pattern Recognit. 41 (11) (2008) 3328–3342.

[45] J. Ma, W. Huang, S. Segarra, A. Ribeiro, Diffusion filtering of graph signals and its use in recommendation systems, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016.

[46] C.A. Gomez-Uribe, N. Hunt, The Netflix recommender system: algorithms, business value, and innovation, ACM Trans. Manag. Inf. Syst. (TMIS) 6 (4) (2016) 13.

[47] Y. Yamanishi, J.-P. Vert, M. Kanehisa, Protein network inference from multiple genomic data: a supervised approach, Bioinformatics 20 (suppl_1) (2004) i363–i370.

[48] S. Barbarossa, S. Sardellitti, A. Farina, On sparse controllability of graph signals, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016, pp. 4104–4108.

[49] F. Gamma, E. Isufi, G. Leus, A. Ribeiro, Control of graph signals over random time-varying graphs, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018.

[50] D. Simon, Optimal State Estimation: Kalman, h Infinity, and Nonlinear Approaches, John Wiley & Sons, 2006.

[51] S.M. Kay, Fundamentals of Statistical Signal Processing: Practical Algorithm Development, 3, Pearson Education, 2013.

[52] S. Joshi, S. Boyd, Sensor selection via convex optimization, IEEE Trans. Signal Process. 57 (2) (2009) 451–462.

[53] L.F. Chamon, A. Ribeiro, Greedy sampling of graph signals, IEEE Trans. Signal Process. 66 (1) (2018) 34–47.

[54] L. Zuo, R. Niu, P.K. Varshney, Posterior CRLB based sensor selection for target tracking in sensor networks, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2, IEEE, 2007, pp. II–1041.

[55] P. Tichavsky, C.H. Muravchik, A. Nehorai, Posterior Cramér-Rao bounds for discrete-time nonlinear filtering, IEEE Trans. Signal Process. 46 (5) (1998) 1386–1396.

[56] H. Zhang, R. Ayoub, S. Sundaram, Sensor selection for Kalman filtering of linear dynamical systems: complexity, limitations and greedy algorithms, Automatica 78 (2017) 202–210.

[57] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, D.K. Hammond, Gspbox: a toolbox for signal processing on graphs, arXiv:1408.5781 (2014).

[58] M. Grant, S. Boyd, Y. Ye, CVX: Matlab software for disciplined convex programming, 2008,

[59] S.X. Chen, J.S. Liu, Statistical applications of the poisson-binomial and conditional bernoulli distributions, Stat. Sin (1997) 875–892.