Graph Filter Transfer for Time-Varying Signal Estimation Between Two Networks

Tsutahiro Fukuhara, Junya Hara, Hiroshi Higashi, Yuichi Tanaka
Graduate School of Engineering, Osaka University, Japan
E-mail: t.fukuhara@msp-lab.org, jhara@msp-lab.org, higashi@comm.eng.osaka-u.ac.jp, ytanaka@comm.eng.osaka-u.ac.jp

***Abstract*—This paper presents a filter transfer method for estimating time-varying graph signals, i.e., Kalman filtering between two different networks. In many sensor networks, signals observed are associated with nodes (i.e., sensors), and edges of the network represent the inter-node connectivity. For a large sensor network, measuring the signal values at all nodes requires huge resources, particularly in terms of energy consumption. To alleviate the issue, one may extract one cluster from the network and perform intra-cluster analysis based on the statistics in the cluster. The statistics are then utilized to estimate the signals from another cluster. This leads to the requirement for transferring a set of parameters in the Kalman filter from one cluster to another. In this paper, we propose a cooperative Kalman filter between two networks. The proposed Kalman filter alternately estimates signals in time between the two networks. We formulate a state-space model in the source cluster and transfer it to the target cluster on the basis of optimal transport. In the signal estimation experiments, we validate the effectiveness of the proposed method.**

I. Introduction

Sensor networks are used in various disciplines to analyze sensor data using the interrelationships among sensors [1], [2], where their nodes observe the signals of the corresponding sensors and their edges represent the inter-node connectivity. Their applications include bottleneck detection in traffic networks and leakage detection in infrastructure networks [3], [4].

A network often has a sub-network having similar statistics called a *community* or *cluster*. In many clustered sensor networks, sensor data in one cluster could impact on those in different clusters as well as those obtained at the previous time instances. Therefore, predictive control (PC) for networks among clusters is crucial and has been extensively studied in many application fields [5], [6].

Kalman filter is the most popular PC method for networks [7], [8]. It linearly tracks and estimates *time-varying* signals on a *static* network by minimizing the mean squared error (MSE) between estimated and original signals. The system of Kalman filter is modeled by a state-space model and its estimator is performed in prediction and update steps.

We often encounter large networks, however, observing all of their signals at every time instance may be costly in terms of storage burdens and energy consumption, which may also shorten the lifetime of sensors [9]. To alleviate this, one may extract one cluster from the network and perform an intra-cluster analysis based on the statistics in the cluster [10]. Since one cluster could affect another as previously mentioned, a transfer method for a set of parameters on the PC is required, especially for clusters having different sizes.

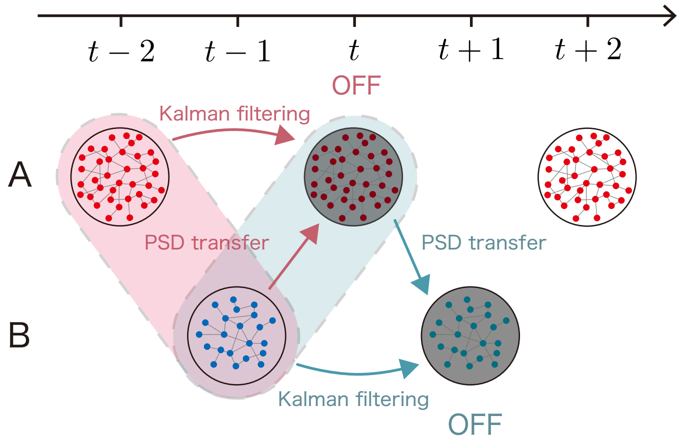

Fig. 1. Overview of a cooperative Kalman filter for two different time-varying graphs. Colored areas denote the set of the source and target at each time instance.

In this paper, we propose a cooperative Kalman filter for estimating time-varying signals between two networks, i.e., graphs. In the estimation, *graph signals*, defined as signals whose domain is the nodes in a graph [11], [12], are used to model the signals observed in a network. We illustrate the overview of the proposed method in Fig. 1. The proposed Kalman filter performs its estimation alternately in time between two graphs. Therefore, the two graphs work as the source and target alternately[1]. The set of parameters in the source is transferred to the target and then it is used in the Kalman estimator, which estimates the current target signal.

In the proposed method, first, we assume a *cyclic* graph wide sense stationarity (CGWSS) of time-varying graph signals for both of the source and target graphs. CGWSS is an extension of *static* GWSS [13], [14], [15] to the case where its power spectral density (PSD) changes periodically over time. Second, we formulate the state-space model in the source domain. The state equation is derived based on the optimal transport, which is a mathematical tool to determine an efficient mapping between two sets of random signals [16]. In our model, we utilize the optimal transport as the signal transition filter from

This work is supported in part by JSPS KAKENHI under Grant 23K26110 and 23K17461, and JST AdCORP under Grant JPMJKB2307.

[1] While this paper focuses on estimation between two graphs for simplicity, it could be generalized to three or more graphs.

the previous signal to the current signal in a graph. Finally, we transfer the state-space model of the source into the target domain by reflecting their statistics. In this transfer, we assume that the statistics of the two domains essentially differ in their PSDs. To compensate for the gap of the PSDs, we utilize a transfer method based on Bayesian inference [17]. As a result, we obtain the state-space model of the target and derive the corresponding control laws.

Our experiments on synthetic data demonstrate that the proposed method effectively estimates time-varying graph signals on two graphs.

II. Preliminaries

In this section, we introduce preliminaries of random signals defined on a graph, i.e., graph signals. Firstly, we review the basics of graph signal processing (GSP). Second, we define the notion of a graph wide sense stationarity (GWSS).

A. Basics of Graph Signal Processing

A weighted undirected graph is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, in which $\mathcal{V}$ and $\mathcal{E}$ are sets of nodes and edges, respectively. The number of nodes and edges are denoted by $N = |\mathcal{V}|$ and $E = |\mathcal{E}|$, respectively. We use a weighted adjacency matrix $\mathbf{W}$ for representing the connection between nodes, where its (m, n)-element $[\mathbf{W}]_{mn} \geq 0$ is the edge weight between the mth and nth nodes; $[\mathbf{W}]_{mn} = 0$ for unconnected nodes. The degree matrix $\mathbf{D}$ is a diagonal matrix whose element is defined as $[\mathbf{D}]_{mm} = \sum_n [\mathbf{W}]_{mn}$. Using $\mathbf{D}$ and $\mathbf{W}$, the graph Laplacian matrix is given by $\mathbf{L} = \mathbf{D} - \mathbf{W}$. A graph signal $\mathbf{x} \in \mathbb{R}^N$ is defined as $\mathbf{x} : \mathcal{V} \to \mathbb{R}^N$ where $[\mathbf{x}]_n$ corresponds to the signal value at the nth node. Since $\mathbf{L}$ is a real symmetric matrix, it has orthogonal eigenvectors and can be diagonalized as $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U} = [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{N-1}]$ is a matrix whose ith column is the eigenvector $\mathbf{u}_i$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$ is their diagonal eigenvalue matrix. Without loss of generality, we can assume $0 = \lambda_0 \leq \lambda_1, \dots, \lambda_{N-1} = \lambda_{\max}$. In GSP, λ_i is referred to as a *graph frequency*. Then, spectra of $\mathbf{x}$ in the graph frequency domain are defined as $\hat{\mathbf{x}} = \mathbf{U}^\top \mathbf{x}$: It is called *graph Fourier transform*.

B. Graph Wide Sense Stationarity

A graph signal $\mathbf{x}$ is a graph wide sense stationary (GWSS) process if the following two conditions are satisfied:

Definition 1 (GWSS [13])**.** *Let $\mathbf{x}$ be a random signal on graph $\mathcal{G}$. The signal $\mathbf{x}$ follows a GWSS process if and only if the following conditions are satisfied:*

$$\mathbb{E}[\mathbf{x}] = \mu = \text{const}, \tag{1a}$$

$$\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] = \mathbf{\Sigma} = \mathbf{U}\mathrm{diag}(\mathbf{p})\mathbf{U}^\top, \tag{1b}$$

where $\mathbf{\Sigma}$ and $\mathbf{p}$ are referred to as the covariance matrix and power spectral density (PSD), respectively, and μ indicates the mean.

In this paper, we assume GWSS signals $\mathbf{x}$ are ergodic: The ensemble mean in Definition 1 is identical to the temporal mean. Later, we consider a time-varying version of the GWSS process, whose PSD varies periodically over time.

III. Related Work

In this section, we review the preliminary studies on a transfer of Kalman filter. First, we introduce the standard (linear) Kalman filter, which is related to the main part of our framework. Second, we review optimal transport and a graph filter transfer method.

A. Kalman Filter

Kalman filter is an online algorithm for sequentially tracking and estimating dynamic signals from given observations [7], [8], [18]. In the following, we revisit the system of the standard Kalman filter and its control laws.

1) The System of Kalman Filter: Let $\mathbf{x}_t$ be the current signal, whose subscript t indicates a time instance. The system of Kalman filter is modeled by a state-space model. Generally, it is formulated as follows [18], [19]:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_{t-1} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(0, \sigma_v^2\mathbf{I}), \tag{2a}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, \sigma_w^2\mathbf{I}), \tag{2b}$$

where $\mathbf{v}_t \in \mathbb{R}^N$ represents system noise and $\mathbf{w}_t \in \mathbb{R}^M$ denotes observation noise, conforming to Gaussian white noise with the standard deviations σ_v and σ_w, respectively. In (2a) and (2b), the matrices $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{B} \in \mathbb{R}^{N \times N}$, and $\mathbf{C} \in \mathbb{R}^{M \times N}$ represent the signal transition, input, and measuring matrices, respectively; The vector $\mathbf{y}_t \in \mathbb{R}^M$ represents the observation and $\mathbf{u}_t \in \mathbb{R}^N$ indicates the control input. Note that (2a) describes the transition of the signals over time and (2b) represents the observation model at t.

In the Kalman filtering, the objective is to minimize the mean squared error (MSE) of the estimated signal for all t. The optimization problem is formalized as:

$$\min_{\tilde{\mathbf{x}}_t} \mathbb{E}[\|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|_2^2], \tag{3}$$

where $\tilde{\mathbf{x}}_t$ represents the estimated signal at t.

In the following, we show the control laws of Kalman filter to estimate the current signal $\tilde{\mathbf{x}}_t$ in (3).

2) Control Laws of Kalman Filter: Hereafter, we denote the estimated signal conditioned on observations up to t, i.e., the prior estimation, by $\tilde{\mathbf{x}}_{t|t-1}$ and denote the posterior one by $\tilde{\mathbf{x}}_t$. The control laws are given as follows [18]:

Prediction step

i Calculating the prior signal estimation.

$$\tilde{\mathbf{x}}_{t|t-1} = \mathbf{A}\tilde{\mathbf{x}}_{t-1} + \mathbf{B}\mathbf{u}_{t-1}, \tag{4a}$$

ii Determining the error covariance matrix of the prior signal estimation.

$$\mathbf{P}_{t|t-1} = \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^\top + \sigma_v\mathbf{I}, \tag{4b}$$

iii Deriving the optimal filter (Kalman gain) in (3) from observations up to $t-1$.

$$\mathbf{K}_{t|t-1} = \mathbf{P}_{t|t-1}\mathbf{C}^\top(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}^\top + \sigma_w\mathbf{I})^{-1}, \tag{4c}$$

Update step

i Estimating the current signal using the Kalman gain.

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_{t|t-1}(\mathbf{y}_t - \mathbf{C}\hat{\mathbf{x}}_{t|t-1}), \tag{4d}$$

ii Updating the posterior error covariance matrix.

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_{t|t-1}\mathbf{C})\mathbf{P}_{t|t-1}. \tag{4e}$$

iii Returning to the prediction step with $t \leftarrow t+1$.

In our setting, $\mathbf{A}$ may not be known a priori and needs to be estimated based on statistics of $\mathbf{x}_t$. To this end, we introduce the optimal transport in the next section.

B. Optimal Transport

Optimal transport theory is a mathematical tool to determine the most efficient mapping between two sets of random signals having different probability distributions [20]. Let us consider the transport from an input signal $\mathbf{x}_1 \sim \mu$ to a subsequent signal $\mathbf{x}_2 \sim \nu$, where μ and ν are two different probabilistic measures on $\mathbb{R}^N$. Then, the optimal transport seeks the assignment π such that it minimizes some transport cost from $\mathbf{x}_1$ to $\mathbf{x}_2$. Given a transport cost function $c: \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}_+$, the optimal transport problem can be formulated as:

$$\inf_{\pi \in \Pi(\mu,\nu)} \mathop{\mathbb{E}}_{(\mathbf{x}_1, T(\mathbf{x}_1)) \sim \pi} c(\mathbf{x}_1, T(\mathbf{x}_1)) \quad \text{s.t. } T_{\#}\mu = \nu, \tag{5}$$

where $\mathbf{x}_2 = T(\mathbf{x}_1)$ and Π represents a subset of joint distributions on $\mathbb{R}^N \times \mathbb{R}^N$. The mapping $T: \mathbb{R}^N \to \mathbb{R}^N$ is referred to as the optimal transport map, and $T_{\#}\mu$ denotes the push-forward measure of μ under T [16].

The cost function often employs the ℓ_p-norm. If the ℓ_p-norm is selected as the cost, the minimum value in (5) is referred to as the p-Wasserstein distance [16]. Obtaining the optimal transport map T analytically from (5) is generally challenging due to its non-uniqueness [16]. Nevertheless, if both of μ and ν are Gaussian distributions, i.e., $\mu = \mathcal{N}(\mu_1, \Sigma_1)$ and $\nu = \mathcal{N}(\mu_2, \Sigma_2)$, the unique solution is obtained by the 2-Wasserstein distance $W_2^2(\mu,\nu)$[16]:

$$W_2^2(\mu,\nu) = \|\mu_1 - \mu_2\|_2^2 + \mathrm{tr}\left(\Sigma_1 + \Sigma_2 - 2\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\right)^{1/2}\right). \tag{6}$$

Simultaneously, the optimal transport T is obtained [16] as

$$T(\mathbf{x}_1) = \mu_2 + \Sigma_1^{-1/2}(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}\Sigma_1^{-1/2}(\mathbf{x}_1 - \mu_1), \tag{7}$$

where we suppose that Σ_1 is non-singular.

In our proposed method, we use T in (7) as the signal transition filter, which represents the optimal mapping from $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$ in a graph, i.e., $\mathbf{A}$ in (2a). However, Σ_2 in (7) is not known in our setting. Therefore, we need to find it from the known statistics in the other domain.

In the next subsection, we introduce a transfer learning method to find Σ_2 from the statistics in the other graph.

C. Graph Filter Transfer

We review an existing graph filter transfer learning method. In particular, we introduce the method utilized for the minimum MSE estimator, i.e., graph Wiener filter [13], [14], [15].

First of all, we assume the original signals satisfy GWSS conditions in Definition 1 and there is no prior knowledge about the covariance and PSD of the graph signals, i.e., Σ_2 and $\mathbf{p}_2$ in the target graph. The graph filter transfer is formulated by the following three recursive optimizations [17].

1. Solving an optimization problem to reconstruct random original signals from the observed signals. The solution is given by a linear minimum mean square error (LMMSE) estimator.
2. Computing the PSD by using an autoregressive moving average (ARMA) graph filter based on convex optimization. Substituting the approximated ARMA filter into the covariance matrix obtained in Step 1.
3. Estimating the PSD in the target graph from the known samples of the other graph (source graph). Substituting the estimated PSD into the ARMA filter in Step 2.

When signals are GWSS, the solution in Step 1 is obtained from the estimated covariance of the target, i.e., Σ_2 in (7). In Step 2, there exist three sub-steps: Step 2-1) The PSD in the source graph is parametrically approximated by graph ARMA filter [21]. Step 2-2) The learned parameters of the graph ARMA filter is tranferred to the target graph [22]. Step 2-3) The PSD in the target graph is calculated based on convex optimization. In Step 3, we estimate the PSD in the target domain from the known samples of the source graph based on Bayes' theorem. Due to the limitation of space, we focus on the transfer method for the PSD estimation in Step 3, which is the main part of the filter transfer used in this paper.

We denote the source graph for transfer by $\mathcal{G}_{\text{src}}$ and the target one by $\mathcal{G}_{\text{trg}}$ along with their graph Laplacian matrices $\mathbf{L}_{\text{src}}$ and $\mathbf{L}_{\text{trg}}$, and their signals $\mathbf{x}_{\text{src}}$ and $\mathbf{x}_{\text{trg}}$. We assume that no statistical samples of the random signal $\mathbf{x}_{\text{trg}}$ are available. Therefore, we need to estimate the PSD of $\mathcal{G}_{\text{trg}}$ from that of $\mathcal{G}_{\text{src}}$. It is formulated as

$$[\mathbf{p}_{\text{trg};x}]_i = \frac{1}{K}\sum_k \beta(\mathbf{x}_{\text{src};k})[\hat{\mathbf{x}}_{\text{src};k}]_i^2, \tag{8}$$

where β denotes the density ratio [17]. By Bayes' theorem, the density ratio can be transformed into the following form:

$$\beta(\mathbf{x}) := \frac{q(\mathbf{x}|\mathbf{L}_{\text{src}})}{q(\mathbf{x}|\mathbf{L}_{\text{trg}})} = \frac{q(\mathbf{y}|\mathbf{L}_{\text{src}})q(\mathbf{x}|\mathbf{y})/q(\mathbf{y}|\mathbf{x})}{q(\mathbf{y}|\mathbf{L}_{\text{trg}})q(\mathbf{x}|\mathbf{y})/q(\mathbf{y}|\mathbf{x})} = \frac{q(\mathbf{y}|\mathbf{L}_{\text{src}})}{q(\mathbf{y}|\mathbf{L}_{\text{trg}})}. \tag{9}$$

Let $\tilde{\beta}$ denote the estimated density ratio. By using the relationship in (9), the estimation of the density ratio based on the least squares can be formulated as the following optimization problem [17], [23].

$$\arg\min_{\tilde{\beta}} \frac{1}{2K}\sum_{k=1}^{K}\tilde{\beta}(\mathbf{y}_{\text{src};k})^2 - \frac{1}{K}\sum_{k=1}^{K}\tilde{\beta}(\mathbf{y}_{\text{trg};k}) + \frac{1}{2K}\sum_{k=1}^{K}(\mathbf{y}_{\text{trg};k}). \tag{10}$$

The solution in this problem can be found by using the parametric approximation, which is referred to as a least-squares embedding [24].

In the next section, we propose a cooperative Kalman filter for time-varying graph signals between two sub-networks utilizing techniques introduced in this section.

IV. COOPERATIVE KALMAN FILTER

In this section, we derive a graph filter transfer method using Kalman filter. Initially, we design a stochastic signal model for time-varying graph signals and then formulate a state-space model. We derive Kalman filter from the state-space model and perform the transfer learning across two different graphs.

A. Signal Model

Here, we assume that $\mathbf{x}_{\text{src}}$ and $\mathbf{x}_{\text{trg}}$ satisfy the conditions defined below:

Definition 2 (Cyclic graph wide sense stationary (CGWSS))**.** *Let $\mathbf{x}$ be a random signal on the graph $\mathcal{G}$. The signal $\mathbf{x}$ is a cyclic graph wide sense stationary process if and only if the two following conditions are satisfied:*

$$\mathbb{E}\left[\mathbf{x}_{t \pmod P}\right] = \mu_{t \pmod P} = \text{const}, \tag{11a}$$

$$\mathbb{E}\left[\mathbf{x}_{t,\pmod P}\mathbf{x}_{t \pmod P}^{\top}\right] = \Sigma_{t \pmod P} = \mathbf{U}\operatorname{diag}(\mathbf{p}_{t \pmod P})\mathbf{U}^{\top}, \tag{11b}$$

where P denotes the period of CGWSS, and Σ_t and $\mathbf{p}_t$ are the periodically varying covariance and PSD, respectively.

The PSD of CGWSS thus periodically changes over time, while that of the GWSS in Definition 1 is static[2].

We assume that $\mathbf{x}_{\text{src},t-1}$ and $\mathbf{x}_{\text{trg},t}$ have similar PSDs: We assume the similarity as that of the kernel function $\phi_{\text{dom},t}(\lambda_i) := [\mathbf{p}_{\text{dom},t}]_i$. Specifically, we assume $\phi_{\text{src},t-1}(\lambda_i) \approx \phi_{\text{trg},t}(\lambda_i)$, i.e., $\sup(\phi_{\text{src},t-1}(\lambda) - \phi_{\text{trg},t}(\lambda))^2 < C$ where C is a small constant.

In the following, we formulate a state-space model based on Definition 2.

B. Control System

Let the subscript dom $\in \{\text{src}, \text{trg}\}$ indicate one of the two graphs. We consider the following state-space model similar to (2a) and (2b).

$$\mathbf{x}_{\text{dom},t} = T(\mathbf{x}_{\text{dom},t-2}) + \mathbf{B}\mathbf{u}_{\text{dom},t-2} + \mathbf{v}_{\text{dom}}, \quad \mathbf{v}_{\text{dom}} \sim \mathcal{N}(0, \sigma_v^2\mathbf{I}), \tag{12a}$$

$$\mathbf{y}_{\text{dom},t} = \mathbf{C}\mathbf{x}_{\text{dom},t} + \mathbf{w}_{\text{dom}}, \quad \mathbf{w}_{\text{dom}} \sim \mathcal{N}(0, \sigma_w^2\mathbf{I}). \tag{12b}$$

In (12a), we utilize the transition function T, which is not necessarily to be time-invariant instead of the time-invariant matrix $\mathbf{A}$ in (2a). In (12b), we employ the same model as (2b). Since we estimate the signals between two graphs alternately and asynchronously, we can reduce the sampling frequency for each subgraph in half.

We summarize which variables are known or unknown in Table I. Our aim is to estimate the unknown signals $\mathbf{x}_{\text{trg},t}$ from known signals. Note that $\Sigma_{\text{trg},t}$, which corresponds to $T(\cdot)$ in (12a) and Σ_2 in (7), is unknown. Therefore, we first estimate $\mathbf{p}_{\text{trg},t}$ from $\mathbf{p}_{\text{src},t-1}$.

TABLE I

OBSERVED AND UNKNOWN SIGNALS. ✓ AND THE *blank* MEAN ACCESSIBLE (KNOWN) AND INACCESSIBLE (UNKNOWN) DATA, RESPECTIVELY.

	$\mathbf{x}_{t-2}$	$\mathbf{x}_{t-1}$	$\mathbf{x}_t$	$\mathbf{y}_{t-2}$	$\mathbf{y}_{t-1}$	$\mathbf{y}_t$	$\mathbf{p}_{t-2}$	$\mathbf{p}_{t-1}$	$\mathbf{p}_t$
src	–	✓	–	–	✓	–	–	✓	–
trg	✓	–		✓	–	✓	✓	–	

In the standard Kalman filter introduced in Sec. III-A, its estimator is derived in one cluster. In contrast, we perform the estimation alternately in time between two graphs. In the following, we derive the proposed Kalman filter.

C. Kalman Filter Transfer

We derive the cooperative Kalman filter between two graphs based on (12a) and (12b). For simplicity, we replace the subscripts $\text{trg},t-2$ and trg,t with 1 and 2, respectively. Following from Definition 2, Σ_1 and Σ_2 can be jointly diagonalized. Therefore, the RHS in (7) in the target graph can be rewritten as follows:

$$T(\mathbf{x}_1) = \mu_2 + \mathbf{Q}(\mathbf{x}_1 - \mu_1), \tag{13}$$

where $\mathbf{Q} = \mathbf{U}_{\text{trg}}\operatorname{diag}(\mathbf{p}_2)\operatorname{diag}(\mathbf{p}_1)^{-1}\mathbf{U}_{\text{trg}}^{\top}$.

To calculate the RHS in (13), the current PSD $\mathbf{p}_{\text{trg},t}$ is required but unknown (see Table I). Therefore, we estimate $\mathbf{p}_{\text{trg},t}$ from $\mathbf{p}_{\text{src},t-1}$ by using the graph filter transfer method introduced in Sec. III-C. Consequently, the control laws of the proposed Kalman filter is described as follows:

Preprocessing step

i. Estimating $\mathbf{p}_{\text{trg},t}$ from $\mathbf{p}_{\text{src},t-1}$ by using the graph filter transfer method in Sec. III-C.
ii. Calculating $T_{\text{trg}|\text{src}}$ in (13) by using estimated $\mathbf{p}_{\text{trg},t}$ where $T_{\text{trg}|\text{src}}$ is the transport map estimated from the source.

Prediction step

i. Calculating the prior estimation of signals.

$$\hat{\mathbf{x}}_{2|1} = T_{\text{trg}|\text{src}}(\hat{\mathbf{x}}_1) + \mathbf{B}\mathbf{u}_1, \tag{14a}$$

ii. Determining the prior error covariance matrix.

$$\mathbf{P}_{2|1} = \mathbf{Q}\mathbf{P}_1\mathbf{Q}^{\top} + \sigma_v^2\mathbf{I}, \tag{14b}$$

iii. Deriving the Kalman gain.

$$\mathbf{K}_{2|1} = \mathbf{P}_{2|1}\mathbf{C}^{\top}(\mathbf{C}\mathbf{P}_{2|1}\mathbf{C}^{\top} + \sigma_w^2\mathbf{I})^{-1}, \tag{14c}$$

Update step

i. Estimating the current signals using the Kalman gain.

$$\hat{\mathbf{x}}_2 = \hat{\mathbf{x}}_{2|1} + \mathbf{K}_{2|1}(\mathbf{y}_2 - \mathbf{C}\hat{\mathbf{x}}_{2|1}), \tag{14d}$$

ii. Updating the posterior error covariance matrix.

$$\mathbf{P}_2 = (\mathbf{I} - \mathbf{K}_{2|1}\mathbf{C})\mathbf{P}_{2|1}. \tag{14e}$$

iii. Swapping src with trg and returning to the preprocessing step with $t \leftarrow t+1$.

[2] When $P = 1$, CGWSS is identical to GWSS.

V. EXPERIMENT

In this section, we perform signal estimation experiments for synthetic data.

1) Synthetic Graph Signals: We construct two different random sensor (RS) graphs[3] G_A and G_B with $N_A = 90$ and $N_B = 45$. The period of CGWSS signals is set to $P = 8$ and their PSDs are given by four different low-pass filters $\{\mathbf{p}_{\mathrm{dom},p}\}_{p=0,\dots,P-1}$ as functions in $\mathbf{L}_{\mathrm{dom}}$:

$$[\mathbf{p}_{\mathrm{dom},p}]_i = \begin{cases} 1-\lambda_i/\lambda_{\max} & (p=0,4), \\ \exp(-\lambda_i/\lambda_{\max}) & (p=1,5), \\ 1/(1+\lambda_i) & (p=2,6), \\ \cos(\pi\lambda_i/2\lambda_{\max}) & (p=3,7). \end{cases} \tag{15}$$

Accordingly, we generate samples of the signals conforming to $\mathcal{N}(\mathbf{1}, \mathbf{U}_{\mathrm{dom}}\mathrm{diag}(\mathbf{p}_{\mathrm{dom},p})\mathbf{U}_{\mathrm{dom}}^{\top})$, where $p = t \pmod P$ for $t = 1, \dots, T$. We denote the training and test datasets by $\mathbf{X}_{\mathrm{dom}}^{\mathrm{train}} \in \mathbb{R}^{N_{\mathrm{dom}} \times T_{\mathrm{train}}}$ and $\mathbf{X}_{\mathrm{dom}}^{\mathrm{test}} \in \mathbb{R}^{N_{\mathrm{dom}} \times T_{\mathrm{test}}}$, consisting of $T_{\mathrm{train}} = 200$ and $T_{\mathrm{test}} = 40$ samples, respectively.

Then, we consider a data update process for the sequential signal estimation experiment. Let $\mathbf{X}_{\mathrm{dom},p}^{(l)} \in \mathbb{R}^{N_{\mathrm{dom}} \times K}$ be a data slot at p, where $K = T_{\mathrm{train}}/P$ and l indicates the lth cycle, i.e., it satisfies $t = lP + p$. We update the data slot at every time instance in a warm-start manner:

$$\mathbf{X}_{\mathrm{src},p}^{(l+1)} = \begin{cases} \mathbf{X}_{\mathrm{src},p}^{\mathrm{train}} & \text{if } l = 0, \\ \left[\mathbf{x}_{\mathrm{trg},t}, \left[\mathbf{X}_{\mathrm{src},p}^{(l)}\right]_{:,1:K-1} \right] & \text{otherwise}, \end{cases} \tag{16}$$

where $\left[\mathbf{X}_{\mathrm{src},p}^{(l)}\right]_{:,1:K-1}$ denotes the submatrix of $\mathbf{X}_{\mathrm{src},p}^{(l)}$ whose columns from 1 to $K-1$. In (16), We divide $\mathbf{X}_{\mathrm{dom}}^{\mathrm{train}}$ into P periods and set it as the initial value, $\mathbf{X}_{\mathrm{dom},p}^{\mathrm{train}}$.

2) Experimental Setup: In the proposed method, we set the initial values of the error covariance in (14b) to $\mathbf{P}_{\mathrm{dom},0} = \mathbf{I}$, and the input matrix in (14a) to $\mathbf{B} = \mathbf{I}$. We use a proportional (PI) control strategy for the control input [26]: $\mathbf{u}_{t-2} = \alpha(\hat{\mathbf{x}}_{\mathrm{trg},t-2} - \bar{\mathbf{x}}_{\mathrm{trg},p})$, where α is a factor of proportionality and $\bar{\mathbf{x}}_{\mathrm{trg},p}$ is a mean vector in the latest data slot at the corresponding period number. We empirically set $\alpha = 5.0 \times 10^{-2}$ in the control input.

The measuring matrix is set to a random sampling matrix with $M_A = 85$ and $M_B = 43$, respectively.

Since there is no prior work on the cooperative Kalman filter to the best of our knowledge, we use the following well-known methods as baseline methods.

Ridge regression with Tikhonov regularization: The signal estimation based on Tikhonov regularization [12]. By using the space model in (12b), the estimated signal can be written as:

$$\begin{aligned} \mathbf{x}_{\mathrm{trg},t}^{\mathrm{TK}} &= \arg\min_{\mathbf{x}_{\mathrm{trg},t}} \|\mathbf{y}_{\mathrm{trg},t} - \mathbf{C}\mathbf{x}_{\mathrm{trg},t}\|_2^2 + \gamma \mathbf{x}_{\mathrm{trg},t}^{\top}\mathbf{L}_{\mathrm{trg}}\mathbf{x}_{\mathrm{trg},t} \\ &= (\mathbf{C}^{\top}\mathbf{C} + \gamma\mathbf{L}_{\mathrm{trg}})^{-1}\mathbf{C}^{\top}\mathbf{y}_{\mathrm{trg},t}, \end{aligned} \tag{17}$$

where γ is a parameter, which controls the intensity of the regularization term and is set to $\gamma = 0.05$.

Transferred graph Wiener filter: The signal estimation based on a transferred graph Wiener filter mentioned in Sec. III-C. The estimated signal can be written as [17]:

$$\mathbf{x}_{\mathrm{trg},t}^{\mathrm{GW}} = \mathbf{H}\mathbf{y} + \mathbf{b}, \tag{18}$$

where

$$\mathbf{H} = \mathbf{\Sigma}\mathbf{C}^{\top}(\mathbf{C}\mathbf{\Sigma}\mathbf{C}^{\top})^{-1}, \quad \mathbf{b} = (\mathbf{I} - \mathbf{H}\mathbf{C})\boldsymbol{\mu}. \tag{19}$$

The covariance matrix $\mathbf{\Sigma}$ is calculated from the estimated PSD $\mathbf{p}_{\mathrm{trg},t}$ in Sec. III-C.

There are two major differences between the proposed method and baseline methods:

1) Both of the baseline methods only use the space equation in (12b). In contrast, the proposed method utilizes Kalman filter, which is derived from state-space model.
2) In the proposed method, signals are assumed to conform to CGWSS, where the PSD can change periodically over time. In contrast, in the baseline methods, the signals are GWSS in Definition 1, whose PSD is static or fundamentally does not conform to GWSS.

To perform the two baseline methods in our time-varying setting, we repeat the baseline methods with alternately switching the source with the target at every time instance. We also set $\sigma_v = 0$ in (12a) to conduct the experiments for a fair comparison. We consider additive white Gaussian noise on measured signals with three different standard deviations σ_w.

3) Results: We evaluate the estimation performance of the proposed method with the MSE and compare it with the above alternative methods. Table II summarizes the average MSE in test datasets. We also visualize an example of absolute errors between the original and estimated signals in Fig. 2. Fig. 3 plots MSEs over time.

TABLE II
EXPERIMENTAL RESULTS ON SYNTHETIC DATASET

σ_w	Average MSE(10^{-2})		
	Proposed	Tikhonov	Wiener
0.05	**0.33**	0.92	2.38
0.10	**1.00**	1.31	2.94
0.15	**2.26**	2.34	4.29

In Table II, the proposed method outperforms alternative methods for all σ_w. In Fig. 3, we observe that the proposed method shows consistent estimation performance for all t, while those of the other methods oscillate significantly over time. This is because the proposed method can compensate its own estimation using signals on the previous time instance and that on the different graph with similar statistics, while other methods perform the estimation independently at all time instances.

VI. CONCLUSION

In this paper, we propose a Kalman filter transfer method for estimating time-varying signals between two different graphs

[3] Random sensor graphs are implemented by k nearest neighbor graphs whose nodes are randomly distributed in 2-D space $[0,1] \times [0,1]$ (See [25]).

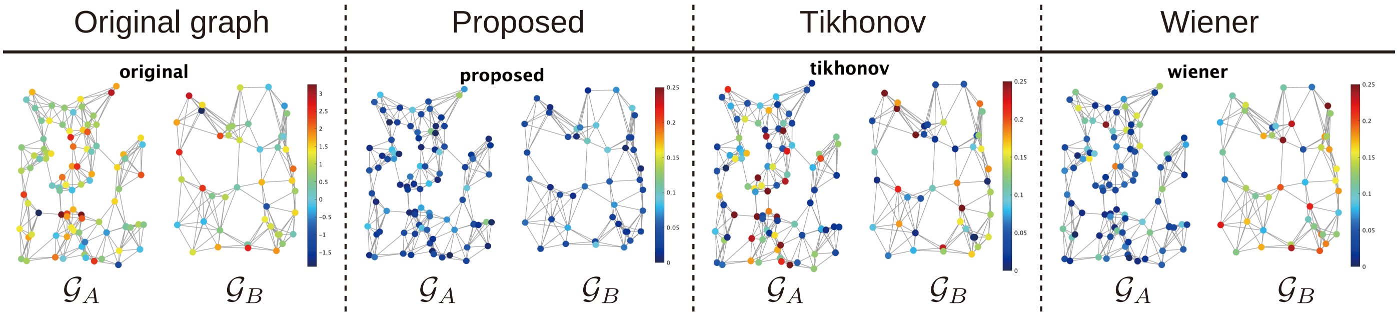

Fig. 2. Signal estimation experiments for noisy graph signals on two different RS graphs with $N_A = 90$, $N_B = 45$ nodes at time $t = 7$, $t = 8$, respectively. The color of a node represents the absolute error between original and estimated signals.

MSE

t

Fig. 3. Comparison of signal estimation MSEs for the RS graph ($_w$ = 0:10). The MSEs for each of the methods is plotted as a holizeontal line.

but having similar features. Initially, we assume that the time-varying graph signals conform to CGWSS for both graphs in our Kalman filter. We then formulate the control system based on the state-space model and optimal transport. To alternately perform signal estimation over time in the source and target, we transfer a set of parameters in the Kalman filter from the source to the target. Our experiments demonstrate that the proposed method effectively estimates the time-varying graph signals.

References

[1] C.-Y. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proc. IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Comput. netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[3] W.-H. Lee, S.-S. Tseng, J.-L. Shieh, and H.-H. Chen, "Discovering traffic bottlenecks in an urban network by spatiotemporal data mining on location-based services," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1047–1056, 2011.

[4] A. Martini, M. Troncossi, and A. Rivola, "Automatic leak detection in buried plastic pipes of water supply networks by means of vibration measurements," *Shock Vib.*, vol. 2015, no. 1, p. 165304, 2015.

[5] B. Kouvaritakis and M. Cannon, *Model Predictive Control: Classical, Robust and Stochastic*. Springer, 2016.

[6] A. Mesbah, "Stochastic model predictive control: An overview and perspectives for future research," *IEEE Control Syst. Mag.*, vol. 36, no. 6, pp. 30–44, 2016.

[7] G. Sagi, N. Shlezinger, and T. Routtenberg, "Extended kalman filter for graph signals in nonlinear dynamic systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 1–5, IEEE, 2023.

[8] M. Ramezani-Mayiami and B. Beferull-Lozano, "Joint graph learning and signal recovery via kalman filter for multivariate auto-regressive processes," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, pp. 907–911, IEEE, 2018.

[9] C. Zhu, V. C. Leung, L. Shu, and E. C.-H. Ngai, "Green internet of things for smart world," *IEEE Access*, vol. 3, pp. 2151–2162, 2015.

[10] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, 2018.

[11] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[12] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[13] J. Hara, Y. Tanaka, and Y. C. Eldar, "Graph signal sampling under stochastic priors," *IEEE Trans. Signal Process.*, vol. 71, pp. 1421–1434, 2023.

[14] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, 2017.

[15] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5911–5926, 2017.

[16] J. Delon and A. Desolneux, "A wasserstein-type distance in the space of gaussian mixture models," *SIAM J. Imaging Sci.*, vol. 13, no. 2, pp. 936–970, 2020.

[17] K. Yamada, "Graph filter transfer via probability density ratio weighting," *arXiv:2210.14633*, 2022.

[18] G. Welch, G. Bishop, *et al.*, "An introduction to the kalman filter," *Proc. ACM SIGGRAPH*, 1995.

[19] S. Li, K. Y. Lim, and D. G. Fisher, "A state space formulation for model predictive control," *AIChE J.*, vol. 35, no. 2, pp. 241–249, 1989.

[20] C. Villani *et al.*, *Optimal transport: old and new*, vol. 338. Springer, 2009.

[21] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 274–288, 2017.

[22] A. Kroizer, T. Routtenberg, and Y. C. Eldar, "Bayesian estimation of graph signals," *IEEE Trans. Signal Process.*, vol. 70, pp. 2207–2223.

[23] M. Sugiyama, T. Suzuki, and T. Kanamori, *Density Ratio Estimation in Machine Learning*. Cambridge University Press, 1 ed., 2012.

[24] T. Kanamori, S. Hido, and M. Sugiyama, "A least-squares approach to direct importance estimation," *J. Machine Learn. Res.*, 2010.

[25] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," *arXiv:1408.5781*, 2014.

[26] K. Ogata, *Modern Control Engineering*. Prentice Hall PTR, 4th ed., 2001.