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, {j.hara, higashi, ytanaka}@comm.eng.osaka-u.ac.jp

Abstract-This paper presents a filter transfer method for estimating time-varying signals, i.e., Kalman filtering between two different networks. In many sensor networks, observed signals 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 statespace 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 intracluster analysis based on the statistics in the cluster [10]. Since



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

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.

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 alternately¹. 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 graph. 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 the previous signal to the current signal within a graph. Finally, we transfer the state-space model of the source into the target graph by reflecting their statistics. In this transfer, we assume

This work is supported in part by JSPS KAKENHI under Grants 23K26110, 23K17461, and 24K21314, and JST AdCORP under Grant JPMJKB2307.

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

that the statistics of the two graphs 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 W for representing the connection between nodes, where its (m, n)element $[\mathbf{W}]_{mn} \geq 0$ is the edge weight between the *m*th and *n*th nodes; $[\mathbf{W}]_{mn} = 0$ for unconnected nodes. The degree matrix 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 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 *n*th node. Since 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 *i*th column is the eigenvector \mathbf{u}_i and $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$ is their diagonal eigenvalue matrix. Without loss of generality, we can assume $0 = \lambda_0 \leq \lambda_1, \dots, \lambda_{N-1} = \lambda_{\max}$ since L is a positive semidefinite matrix [11]. 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 [11].

B. Graph Wide Sense Stationarity

A graph signal 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:

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

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

where Σ 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 x are ergodic: The ensemble mean in Definition 1 is identical to the temporal mean. Later, we will 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. Then, 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 the 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]:

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

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

where $\mathbf{x}_t \in \mathbb{R}^N$ and $\mathbf{y}_t \in \mathbb{R}^M$ represent the current signal and its observation, respectively, and $\mathbf{u}_t \in \mathbb{R}^N$ indicates the control input. 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 observation matrices, respectively. The vector $\mathbf{v}_t \in \mathbb{R}^N$ represents system noise, and $\mathbf{w}_t \in \mathbb{R}^M$ denotes observation noise, conforming to white Gaussian noise with the standard deviations σ_v and σ_w , respectively. 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 $\tilde{x}_{t|t-1}$ and denote the posterior one by \tilde{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.

$$\mathbf{P}_{t|t-1} = \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^{\top} + \sigma_{v}\mathbf{I}, \qquad (4b)$$

where \mathbf{P}_{t-1} is the posterior error covariance at t-1. iii Deriving the optimal filter (Kalman gain) in (3) from

observations up to t - 1.

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

Update step

i Estimating the current signal using the Kalman gain.

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

ii Updating the posterior error covariance matrix.

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

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

At the initial time instance, t = 1, \mathbf{P}_{t-1} in (4b) is usually set to the scaled identity matrix, $\mathbf{P}_0 = \delta \mathbf{I}$, where the $\delta > 0$ is a scaling factor.

In our setting, A may not be known a priori and needs to be estimated based on statistics of x_t . To this end, we introduce the optimal transport in the next subsection.

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 $\mathbf{x}_1 \sim \alpha$ to a subsequent signal $\mathbf{x}_2 \sim \beta$, where α and β are two different probabilistic measures on \mathbb{R}^N . Then, the optimal transport seeks the assignment γ 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:

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

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 $T_{\#\alpha}$ denotes the push-forward measure of α 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 α and β are Gaussian distributions, i.e., $\alpha = \mathcal{N}(\mu_1, \Sigma_1)$ and $\beta = \mathcal{N}(\mu_2, \Sigma_2)$, the unique solution is obtained by the 2-Wasserstein distance $W_2^2(\alpha, \beta)$ [16]:

$$W_{2}^{2}(\alpha,\beta) = \|\boldsymbol{\mu}_{1} - \boldsymbol{\mu}_{2}\|_{2}^{2} + \operatorname{tr}\left(\boldsymbol{\Sigma}_{1} + \boldsymbol{\Sigma}_{2} - 2\left(\boldsymbol{\Sigma}_{1}^{1/2}\boldsymbol{\Sigma}_{2}\boldsymbol{\Sigma}_{1}^{1/2}\right)^{1/2}\right).$$
 (6)

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

$$T(\mathbf{x}_1) = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_1^{-1/2} (\boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{1/2})^{1/2} \boldsymbol{\Sigma}_1^{-1/2} (\mathbf{x}_1 - \boldsymbol{\mu}_1), \quad (7)$$

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

In this section, we review an existing study on graph filter transfer [17], which is mainly related to the proposed method. The graph filter transfer aims to transfer parameters of a graph filter for the source graph to the target one having different (but similar) statistics. Specifically, we introduce the method employed for probabilistic filters, such as the graph Wiener filter, Kalman filter, and Bayesian filter [17].

Hereafter, we denote the source and target graphs by src and trg, respectively. Let the subscript dom \in {src, trg} represent one of the two graphs, and let $\mathbf{x}_{\text{dom},t}$ and $\mathbf{y}_{\text{dom},t}$ denote an *unknown* graph signal at time *t* and its corresponding *known* observation, defined in (2b). Now, we consider estimating $\mathbf{x}_{\text{src},t}$ using a set of *K* historical observations, denoted by $\mathbf{Y}_{\text{src},t} = [\mathbf{y}_{\text{src},t}, \dots, \mathbf{y}_{\text{src},t-K+1}]$. We assume that the original graph signal $\mathbf{x}_{\text{dom},t}$ satisfies the GWSS conditions in Definition 1. Under the assumption, the probabilistic filters are generally formulated through maximum a posteriori (MAP) estimation [14], [17], which is expressed as:

$$\tilde{\mathbf{x}}_{\text{src},t} = \arg\max_{\mathbf{x}_{\text{src},t}} \mathcal{P}(\mathbf{x}_{\text{src},t} | \mathbf{Y}_{\text{src},t}, \mathbf{p}_{\text{src}}), \quad (8)$$

where \mathcal{P} denotes a probability distribution function, and \mathbf{p}_{src} represents the PSD of $\mathbf{x}_{src,t}$. Since \mathbf{p}_{src} governs the behavior of the estimator in (8), transferring (8) to the target graph results in adjusting \mathbf{p}_{src} to adapt the target graph.

In general, the eigenvalue distributions of different graph variation operators are different. This discrepancy presents a challenge of estimating \mathbf{p}_{trg} directly from \mathbf{p}_{src} . To address the issue, a graph filter transfer method in [17] is performed in the following three steps, as visualized in Fig. 2:

- 1) Approximating \mathbf{p}_{src} as a *continuous* ARMA graph filter [21] based on the least-squares (LS). The approximated PSD is denoted by $p_{\text{ARMA}}(\lambda; \boldsymbol{\alpha}_{\text{src}})$, where $\lambda \in [0, +\infty)$ and $\boldsymbol{\alpha}_{\text{dom}}$ represents a sequence of parameters of the ARMA graph filter.
- 2) Adapting $\alpha_{\rm src}$ in $p_{\rm ARMA}(\lambda; \alpha_{\rm src})$ to the target graph using Bayesian inference. The adapted parameters in the target graph is represented as $\alpha_{\rm trg}$.
- Discretizing p_{ARMA}(λ; α_{trg}) according to the eigenvalues of L_{trg} and estimating p_{trg}.

In the three steps, \mathbf{p}_{trg} is indirectly estimated from \mathbf{p}_{src} through the parameters of the ARMA filter. This approach can transfer a probabilistic filter from the source graph to the target graph having different eigenvalue distributions. Please see [17] for more details.

In the next section, we propose a cooperative Kalman filter for time-varying graph signals between two 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.



Fig. 2. Overview of a graph filter transfer method for the PSD estimation in the target graph in [17].

A. Signal Model

Here, we assume that \mathbf{x}_{src} and \mathbf{x}_{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:

$$\mathbb{E}\left[\mathbf{x}_{t \,(\mathrm{mod}\,P)}\right] = \mu_{t \,(\mathrm{mod}\,P)} = \mathrm{const},\tag{9a}$$

$$\mathbb{E}\left[\mathbf{x}_{t,(\text{mod }P)}\mathbf{x}_{t\,(\text{mod }P)}^{\top}\right] = \mathbf{\Sigma}_{t\,(\text{mod }P)} = \mathbf{U}\text{diag}(\mathbf{p}_{t\,(\text{mod }P)})\mathbf{U}^{\top}$$
(9b)

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 static².

In this paper, we assume two graphs are disconnected but the signals on them, $\mathbf{x}_{\mathrm{src},t-1}$ and $\mathbf{x}_{\mathrm{trg},t}$ have similar PSDs (see Fig. 1). Specifically, we define their similarity in terms of the graph filter kernels of the two PSDs, i.e., $p_{\mathrm{dom},t}(\lambda_i)$, where $p_{\mathrm{dom},t}(\lambda_i)$ can be obtained by smoothly interpolating $\mathbf{p}_{\mathrm{dom},t}$ such that $p_{\mathrm{dom},t}(\lambda_i) := [\mathbf{p}_{\mathrm{dom},t}]_i$ (see Section III). Formally, we assume $\sup_{\lambda} (p_{\mathrm{src},t-1}(\lambda) - p_{\mathrm{trg},t}(\lambda))^2 < C$ where C is a small constant.

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

B. Control System

We consider the following state-space model similar to (2a) and (2b).

$$\mathbf{x}_{\text{dom},t} = T(\mathbf{x}_{\text{dom},t-2}) + \mathbf{B}\mathbf{u}_{\text{dom},t-2} + \mathbf{v}_{\text{dom}}, \quad \mathbf{v}_{\text{dom}} \sim \mathcal{N}(0, \sigma_v^2 \mathbf{I})$$
(10a)

$$\mathbf{y}_{\text{dom},t} = \mathbf{C}\mathbf{x}_{\text{dom},t} + \mathbf{w}_{\text{dom}}, \ \mathbf{w}_{\text{dom}} \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I}).$$
 (10b)

In (10a), we utilize the transition function T, which is not necessarily to be time-invariant instead of the time-invariant matrix **A** in (2a). In (10b), we employ the same model as (2b).

LIST OF SIGNALS. ✓ CORRESPONDS TO ACCESSIBLE (KNOWN) DATA. THE BLANK IS UNKNOWN DATA. THE COLORED CELL INDICATES THE SIGNAL WE ESTIMATE.

	\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		1			1			1	
trg	1			1		1	1		

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 (10a) and Σ_2 in (7), is unknown. Therefore, we first estimate $\mathbf{p}_{\text{trg},t}$ from $\mathbf{p}_{\text{src},t-1}$.

In the standard Kalman filter introduced in Section III-A, its estimator is derived in one cluster. In contrast, we perform the estimation alternately in time between two clusters (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 (10a) and (10b). For simplicity, we replace the subscripts $\cdot_{\text{trg},t-2}$ and $\cdot_{\text{trg},t}$ with \cdot_1 and \cdot_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:

$$T(\mathbf{x}_1) = \boldsymbol{\mu}_2 + \mathbf{Q}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \qquad (11)$$

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

To calculate the RHS in (11), the current PSD \mathbf{p}_2 (= $\mathbf{p}_{trg,t}$) is required but unknown (see Table I). Therefore, we estimate \mathbf{p}_2 from $\mathbf{p}_{src,t-1}$ by using the graph filter transfer method introduced in Section III-C. Consequently, the control laws of the proposed Kalman filter is described as follows:

Preprocessing step

- i. Estimating p_2 from $p_{src,t-1}$ by using the graph filter transfer method in Section III-C.
- ii. Calculating the optimal transport map in (11) by using the estimated \mathbf{p}_2 . We denote the transport map as $T_{\text{trg}|\text{src}}(\mathbf{x}_1)$.

Prediction step

i. Calculating the prior estimation of signals.

$$\tilde{\mathbf{x}}_{2|1} = T_{\text{trg}|\text{src}}(\tilde{\mathbf{x}}_1) + \mathbf{B}\mathbf{u}_1, \quad (12a)$$

ii. Determining the prior error covariance matrix.

$$\mathbf{P}_{2|1} = \mathbf{Q}\mathbf{P}_1\mathbf{Q}^{\top} + \sigma_v^2\mathbf{I},\tag{12b}$$

iii. Deriving the Kalman gain.

$$\mathbf{K}_{2|1} = \mathbf{P}_{2|1} \mathbf{C}^{\top} (\mathbf{C} \mathbf{P}_{2|1} \mathbf{C}^{\top} + \sigma_w^2 \mathbf{I})^{-1}, \qquad (12c)$$

Update step

i. Estimating the current signals using the Kalman gain.

$$\tilde{\mathbf{x}}_2 = \tilde{\mathbf{x}}_{2|1} + \mathbf{K}_{2|1}(\mathbf{y}_2 - \mathbf{C}\tilde{\mathbf{x}}_{2|1}), \qquad (12d)$$

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

ii. Updating the posterior error covariance matrix.

$$\mathbf{P}_2 = (\mathbf{I} - \mathbf{K}_{2|1}\mathbf{C})\mathbf{P}_{2|1}.$$
 (12e)

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

In the initial estimation for each graph, \mathbf{P}_1 in (12b) is set to $\mathbf{P}_1 = \delta \mathbf{I}$, similar to the setting in Section III-A.

V. EXPERIMENT

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

1) Synthetic Graph Signals: We construct two different random sensor (RS) graphs³ \mathcal{G}_A and \mathcal{G}_B with $N_A = 90$ and $N_B = 45$. The period of CGWSS signals is set to P = 8and their PSDs are given by four different low-pass filters $\{\mathbf{p}_{dom,p}\}_{p=0,\dots,P-1}$ as functions in \mathbf{L}_{dom} :

$$[\mathbf{p}_{\text{dom},p}]_{i} = \begin{cases} 1 - \lambda_{i}/\lambda_{\text{max}} & (p = 0, 4), \\ \exp(-\lambda_{i}/\lambda_{\text{max}}) & (p = 1, 5), \\ 1/(1 + \lambda_{i}) & (p = 2, 6), \\ \cos(\pi \lambda_{i}/2\lambda_{\text{max}}) & (p = 3, 7). \end{cases}$$
(13)

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

Then, we consider a data update process for the sequential signal estimation experiment. Let $\mathbf{X}_{\text{dom},p}^{(l)} \in \mathbb{R}^{N_{\text{dom}} \times K}$ be a data slot at p, where $K = T_{\text{train}}/P$ 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:

$$\mathbf{X}_{\mathrm{src},p}^{(l+1)} = \begin{cases} \mathbf{X}_{\mathrm{src},p}^{\mathrm{train}} & \text{if } l = 0, \\ \begin{bmatrix} \tilde{\mathbf{x}}_{\mathrm{trg},t}, \begin{bmatrix} \mathbf{X}_{\mathrm{src},p}^{(l)} \end{bmatrix}_{:,1:K-1} \end{bmatrix} & \text{otherwise}, \end{cases}$$
(14)

where $\begin{bmatrix} \mathbf{X}_{\text{src},p}^{(l)} \end{bmatrix}_{:,1:K-1}$ denotes the submatrix of $\mathbf{X}_{\text{src},p}^{(l)}$ whose columns from 1 to K - 1. In (14), We divide $\mathbf{X}_{\text{dom}}^{\text{train}}$ into P periods and set it as the initial value, $\mathbf{X}_{\text{dom},p}^{\text{train}}$.

2) Experimental Setup: The observation matrix C in (10b) is set to a random sampling matrix with $M_A = 85$ and $M_B = 43$, respectively. For the initial estimation, we estimate $\mathbf{p}_{\text{src},0}$ and $\mathbf{p}_{\text{trg},1}$ (= \mathbf{p}_1) in (11), from the training data by using a PSD estimation method [14], [22]. We set the scaling factor of the initial error covariance $\delta = 1$, and the initial estimation of signal $\tilde{\mathbf{x}}_1 = [\mathbf{X}_{\text{trg},p}^{\text{train}}]_{:1}$ in (12a). We use a proportional control strategy for the control input [23]: $\mathbf{B} = \mathbf{I}$ and $\mathbf{u}_1 = \eta(\tilde{\mathbf{x}}_1 - \bar{\mathbf{x}}_{\text{trg},p})$, where η is a factor of proportionality and $\bar{\mathbf{x}}_{\text{trg},p}$ is a mean vector in the latest data slot. We empirically set $\eta = 5.0 \times 10^{-2}$.

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.

 TABLE II

 EXPERIMENTAL RESULTS ON SYNTHETIC DATASET

	Average MSE (10^{-2})						
0w	Proposed	RRTK	TGWF				
0.05	0.33	0.92	2.38				
0.10	1.00	1.31	2.94				
0.15	2.26	2.34	4.29				

Ridge regression with Tikhonov regularization: The first baseline method is signal estimation based on Tikhonov regularization (RRTK) [12]. By using the space model in (10b), the estimated signal can be written as:

$$\widetilde{\mathbf{x}}_{\text{trg},t}^{\text{RRTK}} = \arg\min_{\mathbf{x}_{\text{trg},t}} \|\mathbf{y}_{\text{trg},t} - \mathbf{C}\mathbf{x}_{\text{trg},t}\|_{2}^{2} + \zeta \mathbf{x}_{\text{trg},t}^{\top} \mathbf{L}_{\text{trg}} \mathbf{x}_{\text{trg},t}$$
$$= (\mathbf{C}^{\top}\mathbf{C} + \zeta \mathbf{L}_{\text{trg}})^{-1} \mathbf{C}^{\top} \mathbf{y}_{\text{trg},t},$$
(15)

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

Transferred graph Wiener filter: The Second one is signal estimation based on a transferred graph Wiener filter (TGWF) proposed in [17]. The estimated signal can be written as:

$$\tilde{\mathbf{x}}_{\text{trg},t}^{\text{TGWF}} = \mathbf{H}\mathbf{y} + \mathbf{b},\tag{16}$$

where $\mathbf{H} = \boldsymbol{\Sigma} \mathbf{C}^{\top} (\mathbf{C} \boldsymbol{\Sigma} \mathbf{C}^{\top})^{-1}$, $\mathbf{b} = (\mathbf{I} - \mathbf{H} \mathbf{C}) \boldsymbol{\mu}$. The covariance matrix $\boldsymbol{\Sigma}$ is calculated from the estimated PSD $\mathbf{p}_{\text{trg},t}$ (see Definition 2 and Section III-C).

Both of the baseline methods only use the space equation in (10b). In contrast, the proposed method utilizes Kalman filter derived from the state-space model.

The regularization term in (15) imposes the smoothness of signals on the target graph and therefore RRTK does not rely on GWSS assumptions. In contrast, TGWF assumes GWSS, which requires the PSD for every time instance, as stated in Sec. III-C. Therefore, in this experiment, TGWF estimates the PSD using the method in [14] at each time instance and then performs signal estimation for the target graph.

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 (10a) to conduct the experiments for a fair comparison. We consider additive white Gaussian noise on observations 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. 3. Fig. 4 plots MSEs over time.

In Table II, the proposed method outperforms alternative methods for all σ_w . In Fig. 4, 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

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 [22]).



Fig. 3. Signal estimation experiments for noisy graph signals on two different RS graphs with $N_A = 90$, $N_B = 45$ nodes at time t = 8, t = 7 within T_{test} , respectively. The color of a node represents the absolute error between original and estimated signals.



Fig. 4. Comparison of signal estimation MSEs for the RS graph ($\sigma_w = 0.10$). The MSEs for each of the methods is plotted as a horizontal line.

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 but having similar features. Initially, we assume that the timevarying graph signals conform to CGWSS for both graphs in our Kalman filter. We then formulate the control system based on the state-space mode 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

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