Digit. Signal Process. 162 (2025) 105156

Contents lists available at ScienceDirect

Digital Signal Processing

journal homepage: www.elsevier.com/locate/dsp

Wiener filters on graphs and distributed implementations

Cong Zheng [a], Cheng Cheng [b,*], Qiyu Sun [a]

[a] Department of Mathematics, University of Central Florida, Orlando, FL 32816, USA
[b] School of Mathematics, Sun Yat-sen University, Guangzhou, Guangdong 510275, China

ARTICLE INFO

Keywords:
Graph Wiener filter
Inverse filter
Stationary graph signal
Distributed algorithms
Graph signal processing

ABSTRACT

In this paper, we introduce Wiener filters to recover deterministic and (wide-band) stationary graph signals from their observations corrupted by random noises, and we propose distributed algorithms to implement the Wiener filters. Furthermore, the proposed algorithms can be implemented on distributed networks in which agents are equipped with a data processing subsystem for limited data storage and computation power, and with a one-hop communication subsystem for direct data exchange only with their adjacent agents. Our simulations indicate that the proposed Wiener filtering procedure works well on estimating synthetic (wide-band) stationary signals and real temperature datasets from their noisy observations.

1. Introduction

Massive data sets on networks are collected in numerous applications, such as (wireless) sensor networks, smart grids and social networks [7,8,24,25,44]. Graph signal processing provides an innovative framework to extract knowledge from (noisy) data sets residing on networks [4,10,13,19,28,33,34,38,40]. Graphs $\mathcal{G}=(V,E)$ are widely used to model the complicated topological structure of networks in engineering applications, where a vertex in V may represent an agent of the network and an edge in E between vertices could indicate that the corresponding agents have a peer-to-peer communication link between them and/or they are within a certain range in the spatial space. Data sets on a network can be represented by signals $\mathbf{x}=(x_i)_{i\in V}$ residing on the graph $\mathcal{G}$, where x_i represents the real/complex/vector-valued data at the vertex/agent $i\in V$. In this paper, the data x_i at each vertex $i\in V$ is assumed to be real-valued.

Graph signal processing has been extensively exploited, and many important concepts in classical signal processing, such as Fourier transform, shifts, graph filters, and wavelet filter banks have been extended to the graph setting [21,27,28,34,38]. Graph shift is a basic concept in graph signal processing. We say that a matrix $\mathbf{S}=(S(i,j))_{i,j\in V}$ on the graph $\mathcal{G}=(V,E)$ is a *graph shift* if $S(i,j)\neq 0$ only if either $j=i$ or $(i,j)\in E$. Illustrative examples are the adjacency matrix $\mathbf{A}$, Laplacian matrix $\mathbf{L}=\mathbf{D}-\mathbf{A}$, and symmetrically normalized Laplacian $\mathbf{L}^{\mathrm{sym}}:=\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is the degree matrix of the graph [13,15,21,34,36]. In [13], the notion of multiple commutative graph shifts $\mathbf{S}_1,\ldots,\mathbf{S}_d$ is introduced,

$$\mathbf{S}_k\mathbf{S}_{k'}=\mathbf{S}_{k'}\mathbf{S}_k,\ 1\le k,k'\le d. \tag{1.1}$$

One important property for the commutative graph shifts is that they can be upper-triangularized simultaneously,

$$\widehat{\mathbf{S}}_k=\mathbf{U}^{\mathrm{H}}\mathbf{S}_k\mathbf{U},\ 1\le k\le d, \tag{1.2}$$

where $\mathbf{U}$ is a unitary matrix and $\mathbf{U}^{\mathrm{H}}$ is its Hermitian, and $\widehat{\mathbf{S}}_k=(\widehat{S}_k(i,j))_{1\le i,j\le N}$, $1\le k\le d$, are upper triangular matrices [17, Theorem 2.3.3]. It is exemplified by the construction of multiple commutative graph shifts on circulant/Cayley graphs and Cartesian product graphs, accompanied by a meaningful physical interpretation [13]. Throughout this paper, we assume that $\mathbf{S}_1,\ldots,\mathbf{S}_d$ are **commutative symmetric** graph shifts on a simple graph $\mathcal{G}=(V,E)$, and call the set

$$\Lambda=\big\{\boldsymbol{\lambda}_i=\big(\widehat{S}_1(i,i),\ldots,\widehat{S}_d(i,i)\big),1\le i\le N\big\} \tag{1.3}$$

as the *joint spectrum* of $\mathbf{S}_1,\ldots,\mathbf{S}_d$, where $\widehat{S}_k(i,i),1\le i\le N$, are eigenvalues of $\mathbf{S}_k,1\le k\le d$. As the graph shifts $\mathbf{S}_1,\ldots,\mathbf{S}_d$ are symmetric, all $\widehat{S}_k(i,i),1\le i\le N,1\le k\le d$ are real-valued and the joint spectrum of graph shifts $\mathbf{S}_1,\ldots,\mathbf{S}_d$ is contained in some cube

$$\Lambda\subset[\boldsymbol{\mu},\boldsymbol{\nu}]:=[\mu_1,\nu_1]\times\cdots\times[\mu_d,\nu_d]\subset\mathbb{R}^d, \tag{1.4}$$

where for each $1\le k\le d$, $[\mu_k,\nu_k]$ is the (minimal) interval to contain the spectrum of the graph shift $\mathbf{S}_k$.

A *graph filter* on the graph $\mathcal{G}$ maps one graph signal linearly to another graph signal, and it is usually represented by a matrix indexed on the graph. Graph filters are fundamental in graph signal process-

* Corresponding author.
E-mail addresses: cong.zheng@ucf.edu (C. Zheng), chengch66@mail.sysu.edu.cn (C. Cheng), qiyu.sun@ucf.edu (Q. Sun).

https://doi.org/10.1016/j.dsp.2025.105156

Available online 14 March 2025
1051-2004/© 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

ing, and have been used in denoising, smoothing, consensus of multi-agent systems, the estimation of time series and many other applications [6,9,13,18,19,22,23,36,39,41]. Graph shift is an elementary graph filter. Graph filters are usually designed to be a polynomial of graph shifts, and we say a graph filter $\mathbf{H}$ is a *polynomial filter of graph shifts* if

$$\mathbf{H} = h(\mathbf{S}_1, \dots, \mathbf{S}_d) = \sum_{l_1=0}^{L_1} \cdots \sum_{l_d=0}^{L_d} h_{l_1,\dots,l_d} \mathbf{S}_1^{l_1} \cdots \mathbf{S}_d^{l_d}, \tag{1.5}$$

for a multivariate polynomial h in variables $t_1, \cdots, t_d$,

$$h(t_1, \dots, t_d) = \sum_{l_1=0}^{L_1} \cdots \sum_{l_d=0}^{L_d} h_{l_1,\dots,l_d} t_1^{l_1} \cdots t_d^{l_d}.$$

The commutativity of graph shifts $\mathbf{S}_1, \dots, \mathbf{S}_d$ guarantees that the polynomial graph filter $\mathbf{H}$ in (1.5) is independent on equivalent expressions of the multivariate polynomial h.

In this paper, we consider the filtering procedure

$$\mathbf{x} \longmapsto \mathbf{y} = \mathbf{H}\mathbf{x}, \tag{1.6}$$

where the inputs $\mathbf{x}$ are either (wide-band) stationary signals or deterministic signals with finite energy and the graph filter $\mathbf{H}$ is a polynomial filter of graph shifts satisfying (1.5). A significant advantage is that the corresponding filtering procedure (1.6) can be implemented at the vertex level in which each vertex is equipped with a **one-hop communication subsystem**, i.e., each agent has direct data exchange only with its adjacent agents, see [13, Algorithms 1 and 2]. Stationarity of random signals is a wide-used hypothesis in stochastic signal processing, which has been extended to the graph setting recently [16,19,30,35,43]. We say that a random graph signal $\mathbf{x}$ is *stationary* if its correlation matrix $\mathbb{E}(\mathbf{x}\mathbf{x}^T)$ is a polynomial of graph shifts $\mathbf{S}_1, \dots, \mathbf{S}_d$, and *wide-band stationary* if $\mathbb{E}\mathbf{x} = c\mathbf{1}$ for some $c \in \mathbb{R}$ and the covariance matrix $\mathbb{E}(\mathbf{x} - \mathbb{E}\mathbf{x})(\mathbf{x} - \mathbb{E}\mathbf{x})^T$ is a polynomial of graph shifts $\mathbf{S}_1, \dots, \mathbf{S}_d$. Wiener filter is extensively employed to filter out the random noise $\boldsymbol{\epsilon}$ and to estimate the underlying original signal $\mathbf{x}$ of interest from the noisy observation $\mathbf{y}$ by minimizing some mean square error between the estimated signal and the original signal [2,14,16,19,20,30,35,43].

In Sections 2 and 3, we introduce the stochastic/worst-case mean square errors in the stochastic/deterministic settings, and show the corresponding stochastic/worst-case Wiener filters are essentially the product of a polynomial filter and inverse of another polynomial filter, see Theorems 2.1, 2.5 and 3.1. The iterative Chebyshev polynomial algorithm in [5,6,13] is employed to benefit the implementation of the Wiener filtering on distributed networks, in which agents are equipped with a data processing subsystem for limited data storage and computing power, a one-hop communication subsystem for direct data exchange to its adjacent agents only, see Algorithms 1 and 2. Numerical experiments in Sections 4.1, 4.2, and 4.3 suggest that the Wiener filtering procedure may have better performance on denoising (wide-band) stationary signals, USPS dataset, a real temperature data set collected in the region of Brest (France) and ORL dataset than the conventional Tikhonov regularization approach does [13,37].

Notation: Let $\mathbb{Z}_+$ be the set of all nonnegative integers and set $\mathbb{Z}_+^d = \{(n_1, \dots, n_d),\ n_k \in \mathbb{Z}_+, 1 \le k \le d\}$. Define $\|\mathbf{x}\|_2 = (\sum_{i\in V} |x_i|^2)^{1/2}$ for a graph signal $\mathbf{x} = (x_i)_{i\in V}$ and $\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2$ for a graph filter $\mathbf{A}$. Denote the trace of a square matrix $\mathbf{A}$ by $\mathrm{tr}(\mathbf{A})$, and the transpose and Frobenius norm of a matrix $\mathbf{A}$ by $\mathbf{A}^T$ and $\|\mathbf{A}\|_{\mathrm{F}}$ respectively. As usual, we use $\mathbf{O}, \mathbf{I}, \mathbf{0}, \mathbf{1}$ to denote the zero matrix, identity matrix, zero vector, and vector of all 1s of appropriate sizes respectively.

2. Wiener filters for stationary graph signals

Wiener filter is a conventional approach to estimate the original signal $\mathbf{x}$ from the noisy observation $\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon}$ by minimizing some mean square error between the estimated signal and the original signal [2,14,16,19,20,30,35,43]. In this section, we find explicit expressions for Wiener filters to (wide-band) stationary graph signals. Additionally, we propose a distributed algorithm to execute the Wiener filtering procedure at the vertex level, incorporating one-hop communication.

2.1. Optimal recovery of stationary graph signals

In this subsection, we construct explicitly the optimal reconstruction filter $\mathbf{W}_{\mathrm{mse}}$ with respect to the stochastic mean squared error $F_{\mathrm{mse},P,\mathbf{K}}$ in (2.2), and we present the distributed implementation of the stochastic Wiener filtering procedure $\mathbf{y} \longmapsto \mathbf{W}_{\mathrm{mse}}\mathbf{y}$. We assume that the filtering procedure (1.6) has a polynomial filter

$$\mathbf{H} = h(\mathbf{S}_1, \dots, \mathbf{S}_d), \tag{2.1a}$$

the inputs $\mathbf{x}$ are stationary with the correlation matrix

$$\mathbf{R} = r(\mathbf{S}_1, \dots, \mathbf{S}_d) \tag{2.1b}$$

being a polynomial of graph shifts $\mathbf{S}_1, \dots, \mathbf{S}_d$, and the outputs

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon} \tag{2.1c}$$

are corrupted by some random noise $\boldsymbol{\epsilon}$ being independent with the input signal $\mathbf{x}$, and having zero mean and covariance matrix $\mathbf{G} = \mathrm{cov}(\boldsymbol{\epsilon})$ being a polynomial of graph shifts $\mathbf{S}_1, \dots, \mathbf{S}_d$, i.e.,

$$\mathbb{E}\boldsymbol{\epsilon} = \mathbf{0},\ \mathbb{E}\boldsymbol{\epsilon}\mathbf{x}^T = \mathbf{O} \ \text{ and } \ \mathbf{G} = g(\mathbf{S}_1, \dots, \mathbf{S}_d) \tag{2.1d}$$

for some multivariate polynomial g.

Write the weight at each vertex by $p(i), i \in V$, the regularization matrix on the graph by $\mathbf{K}$. Then we define the *stochastic mean squared error*, SMSE for abbreviation, of a reconstruction filter $\mathbf{W}$ by

$$F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W}) = \mathbb{E}\big[(\mathbf{W}\mathbf{y} - \mathbf{x})^T \mathbf{P}(\mathbf{W}\mathbf{y} - \mathbf{x}) + \mathbf{y}^T \mathbf{W}^T \mathbf{K}\mathbf{W}\mathbf{y}\big], \tag{2.2}$$

where $\mathbf{P}$ is the diagonal matrix with diagonal entries $p(i), i \in V$. The SMSE $F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W})$ in (2.2) contains the regularization term $\mathbb{E}\mathbf{y}^T \mathbf{W}^T \mathbf{K}\mathbf{W}\mathbf{y}$ and the fidelity term

$$\mathbb{E}(\mathbf{W}\mathbf{y} - \mathbf{x})^T \mathbf{P}(\mathbf{W}\mathbf{y} - \mathbf{x}) = \sum_{i\in V} p(i)\mathbb{E}|(\mathbf{W}\mathbf{y})(i) - x(i)|^2.$$

The SMSE $F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W})$ is discussed in [20,30] for the case that the filter $\mathbf{H}$, the covariance $\mathbf{G}$ of noises and the regularizer $\mathbf{K}$ are polynomials of the graph Laplacian $\mathbf{L}$, and that the probability measure P is the uniform probability measure P_U, i.e., $p_U(i) = 1/N, i \in V$. In the following theorem, we provide an explicit solution to the minimization $\min_{\mathbf{W}} F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W})$. See Section 5.1 for the proof.

Theorem 2.1. *Let the filter $\mathbf{H}$, the input $\mathbf{x}$, the noisy output $\mathbf{y}$ and additive noise $\boldsymbol{\epsilon}$ be as in (2.1), and let the SMSE $F_{\mathrm{mse},P,\mathbf{K}}$ be as in (2.2). Assume that $\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G}$ and $\mathbf{P} + \mathbf{K}$ are strictly positive definite (or equivalently, nonsingular as they are positive semi-definite), and define*

$$\mathbf{W}_{\mathrm{mse}} = (\mathbf{P} + \mathbf{K})^{-1}\mathbf{P}\mathbf{R}\mathbf{H}^T \big(\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G}\big)^{-1}. \tag{2.3}$$

Then $\mathbf{W}_{\mathrm{mse}}$ is the unique minimizer of the minimization problem

$$\mathbf{W}_{\mathrm{mse}} = \arg\min_{\mathbf{W}} F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W}). \tag{2.4}$$

Furthermore,

$$F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W}_{\mathrm{mse}}) = \mathrm{tr}\big(\mathbf{P}(\mathbf{I} - \mathbf{W}_{\mathrm{mse}}\mathbf{H})\mathbf{R}\big). \tag{2.5}$$

We call the optimal reconstruction filter $\mathbf{W}_{\mathrm{mse}}$ in (2.3) as the *stochastic Wiener filter*. For the case that the SMSE does not take the regularization term into account, i.e., $\mathbf{K} = \mathbf{O}$, we obtain from (2.3) that the corresponding stochastic Wiener filter $\mathbf{W}_{\mathrm{mse}}$ becomes

$$\mathbf{W}_{\mathrm{mse}}^0 = \mathbf{R}\mathbf{H}^T \big(\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G}\big)^{-1}, \tag{2.6}$$

which is independent of the weight of each vertex $p(i), i \in V$ on the graph $\mathcal{G}$. For the special case $\mathbf{H} = \mathbf{I}$, the Wiener filter $\mathbf{W}^0_{\text{mse}}$ in (2.6) is consistent with the one constructed in [20] in the frequency domain. If we assume that weight at each vertex is uniform, that is $p(i) = 1/N, i \in V$, denoted by P_U and the input signals $\mathbf{x}$ are i.i.d. with mean zero and variance $\delta_1 \mathbf{I}$, the stochastic Wiener filter turns into

$$\mathbf{W}^0_{\text{mse}} = \delta_1^2 \mathbf{H}^T (\delta_1^2 \mathbf{H}\mathbf{H}^T + \mathbf{G})^{-1}$$

and the corresponding SMSE is given by

$$F_{\text{mse},P_U}(\mathbf{W}^0_{\text{mse}}) = \frac{\delta_1^2}{N} \text{tr}\left((\delta_1^2 \mathbf{H}\mathbf{H}^T + \mathbf{G})^{-1}\mathbf{G}\right), \tag{2.7}$$

cf. (3.7) and (3.8), and [30, Eqn. (16)].

Denote the reconstructed signal via the stochastic Wiener filter $\mathbf{W}_{\text{mse}}$ by

$$\mathbf{x}_{\text{mse}} = \mathbf{W}_{\text{mse}}\mathbf{y}, \tag{2.8}$$

where $\mathbf{y}$ is given in (2.1c). The above estimator via stochastic Wiener filter $\mathbf{W}_{\text{mse}}$ is **biased** in general. For the case that $\mathbf{G}, \mathbf{H}, \mathbf{K}$ and $\mathbf{R}$ are polynomials of commutative symmetric graph shifts $\mathbf{S}_1, \ldots, \mathbf{S}_d$, one may verify that matrices $\mathbf{H}^T, \mathbf{H}, \mathbf{G}, \mathbf{R}, \mathbf{K}$ are commutative, and

$$\begin{aligned}\mathbb{E}(\mathbf{x} - \mathbf{x}_{\text{mse}}) = (\mathbf{P} + \mathbf{K})^{-1}\big(\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G}\big)^{-1}\mathbf{R}\mathbf{H}^T\mathbf{H}\mathbf{K}\mathbb{E}\mathbf{x} \\ + \big(\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G}\big)^{-1}\mathbf{G}\mathbb{E}\mathbf{x}.\end{aligned} \tag{2.9}$$

Therefore the estimator (2.8) is **unbiased** if

$$\mathbf{K}\mathbb{E}\mathbf{x} = \mathbf{G}\mathbb{E}\mathbf{x} = \mathbf{0}. \tag{2.10}$$

By (2.10), we see that the stochastic Wiener filter $\mathbf{W}_{\text{mse}}$ in (2.4) provides an unbiased estimate of stationary signals $\mathbf{x}$ with **zero** mean. For the case that the original signal $\mathbf{x}$ is wide-band stationary, we have that $\mathbb{E}(\mathbf{x}) = c\mathbf{1}$ for some constant c. Then the requirement (2.10) is met if $\mathbf{1}$ is the common eigenvector of the covariance matrix $\mathbf{G}$ and the regularization matrix $\mathbf{K}$ associated with eigenvalue zero. For polynomials $\mathbf{K} = k(\mathbf{S}_1, \ldots, \mathbf{S}_d)$ and $\mathbf{G} = g(\mathbf{S}_1, \ldots, \mathbf{S}_d)$ of graph shifts $\mathbf{S}_1, \ldots, \mathbf{S}_d$, one may verify that

$$\mathbf{K}\mathbf{1} = \mathbf{G}\mathbf{1} = \mathbf{0} \text{ if and only if } k(\mathbf{0}) = g(\mathbf{0}) = 0,$$

when the graph shifts $\mathbf{S}_1, \ldots, \mathbf{S}_d$ satisfy $\mathbf{S}_1\mathbf{1} = \cdots = \mathbf{S}_d\mathbf{1} = \mathbf{0}$. The specified condition on graph shifts remains applicable to our illustrative example of Laplacian matrices on the Cartesian product graph in the spatial/temporal domain, see Section 4.2. We note that in the case of the conventional i.i.d. Gaussian noise model, the covariance matrix $\mathbf{G}$ is a multiple of the identity matrix. Consequently, $\mathbf{G}\mathbf{1} \neq \mathbf{0}$, indicating that the corresponding Wiener filter $\mathbf{W}_{\text{mse}}$ is **biased** toward the recovery of wide-band stationary signals corrupted by the Gaussian noise.

2.1.1. Distributed implementation of the Wiener filtering procedure in (2.8)

By (2.3) and (2.6), the reconstructed signal $\mathbf{x}_{\text{mse}}$ in (2.8) can be obtained in two steps,

$$\mathbf{w} = \mathbf{W}^0_{\text{mse}}\mathbf{y} = \mathbf{R}\mathbf{H}^T\big(\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G}\big)^{-1}\mathbf{y} \tag{2.11a}$$

and

$$\mathbf{x}_{\text{mse}} = \mathbf{P}^{-1/2}\big(\mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}\big)^{-1}\mathbf{P}^{1/2}\mathbf{w}, \tag{2.11b}$$

where the first step (2.11a) is the Wiener filtering procedure without the regularization term taken into account, and the second step (2.11b) is the solution of the following Tikhonov regularization problem, i.e.,

$$\mathbf{x}_{\text{mse}} = \arg\min_{\mathbf{x}} (\mathbf{x} - \mathbf{w})^T\mathbf{P}(\mathbf{x} - \mathbf{w}) + \mathbf{x}^T\mathbf{K}\mathbf{x}. \tag{2.12}$$

By symmetry and commutativity assumptions on the graph shifts $\mathbf{S}_1, \ldots, \mathbf{S}_d$, and the polynomial assumptions (2.1a), (2.1b) and (2.1d), the Wiener filter $\mathbf{W}^0_{\text{mse}}$ in (2.6) is the product of a polynomial filter $\mathbf{R}\mathbf{H}^T = (hr)(\mathbf{S}_1, \ldots, \mathbf{S}_d)$ and the inverse of another polynomial filter

$$\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G} = (h^2 r + g)(\mathbf{S}_1, \ldots, \mathbf{S}_d). \tag{2.13}$$

Set $\mathbf{z}_1 = (\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G})^{-1}\mathbf{y}$. Then the iterative Chebyshev polynomial algorithm (ICPA for abbreviation) in [13, Algorithm 4] can be applied to the inverse filtering procedure $\mathbf{y} \longmapsto \mathbf{z}_1$, when

$$h^2(\mathbf{t})r(\mathbf{t}) + g(\mathbf{t}) > 0 \text{ for all } \mathbf{t} \in [\boldsymbol{\mu}, \boldsymbol{\nu}], \tag{2.14}$$

where $[\boldsymbol{\mu}, \boldsymbol{\nu}]$ is the cube in (1.4) containing the joint spectrum of graph shifts $\mathbf{S}_1, \cdots, \mathbf{S}_d$. And the filtering procedure $\mathbf{w} = \mathbf{R}\mathbf{H}^T\mathbf{z}_1$ can be implemented at the vertex level with one-hop communication using [13, Algorithms 1 and 2], since $\mathbf{R}\mathbf{H}^T$ is the product of two polynomial filters. See Part I of Algorithm 1 for the implementation of the Wiener filtering procedure (2.11a) without regularization at the vertex level. We remark that the requirement (2.14) can be considered as a strong version of the strictly positive definite assumption on $\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G}$ in Theorem 2.1, which is equivalent to

$$h^2(\mathbf{t})r(\mathbf{t}) + g(\mathbf{t}) > 0 \text{ for all } \mathbf{t} \in \Lambda \subset [\boldsymbol{\mu}, \boldsymbol{\nu}],$$

where Λ is the joint spectrum of the graph shifts $\mathbf{S}_1, \ldots, \mathbf{S}_d$ contained in the cube $[\boldsymbol{\mu}, \boldsymbol{\nu}]$.

Next, we consider the distributed implementation for the filtering process (2.11b). Set

$$\mathbf{z}_2 = \mathbf{P}^{1/2}\mathbf{w} \tag{2.15}$$

and

$$\mathbf{z}_3 = \big(\mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}\big)^{-1}\mathbf{z}_2. \tag{2.16}$$

As $\mathbf{P}$ is a diagonal matrix, the rescaling procedure $\mathbf{z}_2(i) = \sqrt{\mathbf{P}(i,i)}\mathbf{w}(i)$ and $\mathbf{x}_{\text{mse}}(i) = \frac{\mathbf{z}_3(i)}{\sqrt{\mathbf{P}(i,i)}}$ can be implemented at the vertex level $i \in V$. Then it remains to find a distributed algorithm to implement the inverse filtering procedure (2.16) at the vertex level.

We next propose a convergent algorithm to implement the inverse filtering procedure (2.16) at the vertex level with the positive semidefinite regularization matrix $\mathbf{K} = k(\mathbf{S}_1, \ldots, \mathbf{S}_d)$ being a polynomial form of graph shifts $\mathbf{S}_1, \ldots, \mathbf{S}_d$. Define the sequence $\mathbf{w}_m, m \geq 0$, by

$$\mathbf{w}_{m+1} = \mathbf{w}_0 + \frac{K}{K + p_{\min}}\mathbf{w}_m - \frac{p_{\min}}{K + p_{\min}}\mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}\mathbf{w}_m, \ m \geq 0 \tag{2.17}$$

with initial $\mathbf{w}_0 = \frac{p_{\min}}{K + p_{\min}}\mathbf{z}_2$. As $\mathbf{P}$ is a diagonal matrix and $\mathbf{K}$ is of polynomial form, then

$$\mathbf{w}_{m+1}(i) = \mathbf{w}_0(i) + \frac{K}{K + p_{\min}}\mathbf{w}_m(i) - \sum_{j, \rho(i,j) \leq K} \frac{\mathbf{K}(i,j)\tilde{\mathbf{w}}_m(j)}{\sqrt{\mathbf{P}(i,i)}},$$

where K is the polynomial order of regularization matrix $\mathbf{K}$, and $\tilde{\mathbf{w}}_m = \left(\frac{\mathbf{w}_m(i)}{\sqrt{\mathbf{P}(i,i)}}\right)_{i \in V}$.

Crucially, within each iteration in the algorithm (2.17) to implement the inverse filtering procedure (2.16), there are primarily two rescaling procedures and a filtering procedure associated with the polynomial filter $\mathbf{K}$. Hence the regularization procedure (2.11b) can be implemented at the vertex level with one-hop communication, see Part II of Algorithm 1. Then we show that the sequences $\mathbf{w}_m, m > 0$ in (2.17) converge to the $\mathbf{z}_3$ in (2.16) exponentially. Set

$$K = \sup_{\mathbf{t} \in [\boldsymbol{\mu}, \boldsymbol{\nu}]} k(\mathbf{t}) \text{ and } p_{\min} = \min_{i \in V} p(i). \tag{2.18}$$

Then by the positive semidefinite of $\mathbf{K}$, we have

$$\mathbf{I} \preceq \mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2} \preceq \frac{K + p_{\min}}{p_{\min}}\mathbf{I}, \tag{2.19}$$

where for symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, we use $\mathbf{A} \preceq \mathbf{B}$ to denote the positive semidefiniteness of $\mathbf{B} - \mathbf{A}$. Applying Neumann series expansion

$(1-t)^{-1} = \sum_{n=0}^{\infty} t^n$ with t replaced by $\mathbf{I} - \frac{p_{\min}}{K+p_{\min}}(\mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2})$, we obtain

$$\left(\mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}\right)^{-1} = \frac{p_{\min}}{K+p_{\min}} \sum_{n=0}^{\infty} \left(\frac{K\mathbf{I} - p_{\min}\mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}}{K+p_{\min}} \right)^n.$$

Then the sequence $\mathbf{w}_m, m \geq 0$ converges to $\mathbf{z}_3$ exponentially, since

$$\begin{aligned}
\|\mathbf{w}_m - \mathbf{z}_3\|_2 &= \frac{p_{\min}}{K+p_{\min}} \left\| \sum_{n=m+1}^{\infty} \left(\frac{K\mathbf{I} - p_{\min}\mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}}{K+p_{\min}} \right)^n \mathbf{z}_2 \right\|_2 \\
&\leq \frac{p_{\min}\|\mathbf{z}_2\|_2}{K+p_{\min}} \sum_{n=m+1}^{\infty} \left\| \frac{K\mathbf{I} - p_{\min}\mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}}{K+p_{\min}} \right\|^n \\
&\leq \frac{p_{\min}\|\mathbf{z}_2\|_2}{K+p_{\min}} \sum_{n=m+1}^{\infty} \left(\frac{K}{K+p_{\min}} \right)^n \\
&= \left(\frac{K}{K+p_{\min}} \right)^{m+1} \|\mathbf{z}_2\|_2, \ m \geq 1,
\end{aligned}$$

where the last inequality follows from (2.19).

Remark 2.2. For the case that the weight at each vertex $p(i), i \in V$ is uniform [30], $\mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2} = \mathbf{I} + N\mathbf{K}$ is a polynomial filter of $\mathbf{S}_1, \ldots, \mathbf{S}_d$ if $\mathbf{K} = k(\mathbf{S}_1, \ldots, \mathbf{S}_d)$ is, and hence ICPA in [13, Algorithm 4] can be directly applied to the inverse filtering procedure $\mathbf{z}_3 = \left(\mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}\right)^{-1}\mathbf{z}_2$ in (2.16) (and hence the Wiener filtering procedure in (2.8)) if $1 + Nk(\mathbf{t}) > 0$ for all $\mathbf{t} \in [\boldsymbol{\mu}, \boldsymbol{\nu}]$, which may have faster convergence than the iterative algorithm (2.17) (applicable for the weight P) does.

For the case that the weight at each vertex $p(i), i \in V$ is nonuniform, $\mathbf{P}^{-1/2}$ **may not commute** with the graph shifts $\mathbf{S}_1, \ldots, \mathbf{S}_d$, and the filter $\mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}$ is not necessarily a polynomial filter of some graph shifts even if $\mathbf{K} = k(\mathbf{S}_1, \ldots, \mathbf{S}_d)$ is, hence the ICPA in [13, Algorithm 4] **does not directly apply** to the above inverse filtering procedure. In that case, we may apply the iterative distributed algorithm (2.17) to implement (2.16) and (2.15), (and hence the Tikhonov regularization (2.11b) and the Wiener filtering procedure in (2.8)), provided that the positive semidefinite regularization matrix $\mathbf{K} = k(\mathbf{S}_1, \ldots, \mathbf{S}_d)$ is a polynomial of graph shifts $\mathbf{S}_1, \ldots, \mathbf{S}_d$.

Remark 2.3. Let K and $p_{\min}$ be as in (2.18). One may verify that the matrix $\mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}$ has its eigenvalues contained in $[0, K/p_{\min}]$. Then we can consider $\mathbf{S} = \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}$ as a whole (not necessarily a graph shift) and use one-dimensional polynomials, such as Chebyshev polynomial in [13,35]. Denote the polynomial approximation of $(1+t)^{-1}$ by g_M and set the corresponding polynomial filter by $\mathbf{G}_M = g_M(\mathbf{S})$. Then the ICPA in [13, Algorithm 4] can be used to implement the inverse filtering procedure (2.16), and each iteration in the proposed algorithm is still implementable at the vertex level in one-hop communication [9,13], as the filters $\mathbf{P}^{-1/2}, \mathbf{P}^{-1}$ and $\mathbf{K}$ can be implemented and

$$(\mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2})^l = \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1} \cdots \mathbf{K}\mathbf{P}^{-1}\mathbf{K}\mathbf{P}^{-1/2}, l \geq 1.$$

For a data set on a large network, it is often impractical to have a center (cloud) server to collect, store, manage and process. Even if feasible, this approach could have limitations for real-time and low-latency applications. This highlights the demand for designing distributed (fog) processing; see [12,32,42] and references therein. Distributed systems with one-hop communication have been developed to implement the filtering procedure (1.6) associated with a polynomial filter or its inverse [6,9,13,18,22,23,36,39,41]. For two positive quantities a and b, we denote $a = O(b)$ if $a \leq Cb$ for some absolute constant C depending only on d, the number of graph shifts in our setting. In the following remark, we discuss the distributed implementation in Algorithm 1 for the Wiener filtering procedure.

Algorithm 1 Polynomial approximation algorithm to implement the Wiener filtering procedure $\mathbf{x}_{\text{mse}} = \mathbf{W}_{\text{mse}}\mathbf{y}$ at a vertex $i \in V$.

Inputs: Polynomial coefficients of polynomial filters $\mathbf{H}, \mathbf{G}, \mathbf{K}, \mathbf{R}$ and $\mathbf{G}_M$, entries $S_k(i,j), j \in \mathcal{N}_i$ in the i-th row of the shifts $\mathbf{S}_k, 1 \leq k \leq d$, the value $y(i)$ of the input signal $\mathbf{y} = (y(i))_{i \in V}$ at the vertex i, the probability $p(i)$ at the vertex i, and numbers N_1 and N_2 of the first and second iteration, where $\mathbf{G}_M$ is a Chebyshev polynomial approximation in [13, Eqn. (4.15)] such that the spectral radius of $\mathbf{I} - \mathbf{G}_M(\mathbf{H}^2\mathbf{R} + \mathbf{G})$ is strictly less than 1.

Part I: Implementation of the Wiener filtering procedure (2.11a) at the vertex i

Pre-processing: Find the polynomial coefficients of polynomial filters $\mathbf{H}^2\mathbf{R} + \mathbf{G}$ and $\mathbf{R}\mathbf{H}$.

Initialization: $n = 0$ and zero initial $x^{(0)}(i) = 0$.

Iteration: Use [13, Algorithms 1 and 2] to implement the filtering procedures $\mathbf{e}^{(m)} = (\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G})\mathbf{x}^{(m-1)} - \mathbf{y}$ and $\mathbf{x}^{(m)} = \mathbf{x}^{(m-1)} - \mathbf{G}_M\mathbf{e}^{(m)}, 0 \leq m \leq L_1$ at the vertex i.

Output of the iteration: Denote the output of the N_1-th iteration by $z_1^{(N_1)}(i)$, which is the approximate value of the output data of the inverse filtering procedure $\mathbf{z}_1 = (\mathbf{H}^2\mathbf{R} + \mathbf{G})^{-1}\mathbf{y}$ at the vertex i.

Post-processing after the iteration: Use [13, Algorithms 1 and 2] to implement the filtering procedure $\mathbf{w} = \mathbf{R}\mathbf{H}\mathbf{z}_1 = \mathbf{W}_{\text{mse}}^0\mathbf{y}$ at the vertex i, where the input is $z_1^{(N_1)}(i)$ and the output denoted by $w^{(N_1)}(i)$, is the approximate value of the output data of the above filtering procedure.

Part II: Implementation of the regularization procedure (2.11b) at the vertex i

Pre-processing: Rescaling $z_2^{(N_1)}(i) = p(i)^{1/2}w^{(N_1)}(i)$, the approximate value of the output data of the rescaling procedure $\mathbf{z}_2 = \mathbf{P}^{1/2}\mathbf{w}$.

Iteration: Start from $\mathbf{w}_0(i) = z_2^{(N_1)}(i)$, and use [13, Algorithms 1 and 2] and rescaling $\mathbf{P}^{-1/2}$ to implement the procedure (2.17) for $0 \leq m \leq N_2$, with the output, denoted by $z_3^{(N_1,N_2)}(i)$, being the approximation value of the output data of the inverse filtering procedure $\mathbf{z}_3 = \left(\mathbf{I} + \mathbf{P}^{-1/2}\mathbf{K}\mathbf{P}^{-1/2}\right)^{-1}\mathbf{z}_2$ at the vertex i

Post-processing: $x_{\text{mse}}^{(N_1,N_2)}(i) = p(i)^{-1/2}z_3^{(N_1,N_2)}(i)$.

Output: $x_{\text{mse}}(i) \approx x_{\text{mse}}^{(N_1,N_2)}(i)$, the approximate value of the output data of the Wiener filtering procedure $\mathbf{x}_{\text{mse}} = (\mathbf{P} + \mathbf{K})^{-1}\mathbf{P}\mathbf{w} = \mathbf{W}_{\text{mse}}\mathbf{y}$ at the vertex i.

Remark 2.4. In Algorithm 1 to implement the Wiener filtering procedure $\mathbf{x}_{\text{mse}} = \mathbf{W}_{\text{mse}}\mathbf{y}$ at the vertex level, we equip each vertex i with a server to collect, store, manage and process data at the vertex i and vertices j in its neighborhood $\mathcal{N}_i$, and exchange data to its neighboring vertices $j \in \mathcal{N}_i$. The server at each vertex is only required to have limited data storage, computation power and data exchanging facility to its adjacent vertices. In particular, following the evaluation in [13] we see that the storage requirement for the server at each vertex i is about $O\left((L+1)^d(\#\mathcal{N}_i + 1)\right)$, the number of operations of addition and multiplication is about $O\left((N_1+N_2)(L+1)^d(\#\mathcal{N}_i+1)\right)$, and the total one-hop communication cost is about $O((N_1+N_2)(\#\mathcal{N}_i+1))$. Here L is the maximal degree of polynomial filters of $\mathbf{H}, \mathbf{G}, \mathbf{K}, \mathbf{R}$ and $\mathbf{G}_M$, $\#\mathcal{N}_i$ is the number of adjacent vertices of the vertex i chosen, the maximal iteration numbers N_1 and N_2 is so chosen that the outputs $z_1^{(N_1)}(i)$ and $z_3^{(N_1,N_2)}(i)$ of in the iterations of Part I and Part II are within certain given approximation error ϵ, i.e., $|z_1^{(N_1)}(i) - z_1^{(N_1-1)}(i)| \leq \epsilon$ and $|z_3^{(N_1,N_2)}(i) - z_3^{(N_1,N_2-1)}(i)| \leq \epsilon$. Consequently, the implementation in Algorithm 1 offers many advantages such as real-time interactions with low latency, low energy consumption, privacy and data security protection. In addition, the above distributed implementation is scalable, and in the scenario of adding/deleting vertices (and hence servers equipped at those vertices) in dynamically changing networks, we only need to adjust parameters of the servers of neighboring vertices.

In the implementation in Algorithm 1, there is no central facility to coordinate servers at each vertex, however the coefficients of polynomial filters $\mathbf{H}, \mathbf{G}, \mathbf{K}, \mathbf{R}$ and $\mathbf{G}_M$ and the numbers N_1 and N_2 of iterations in Part I and Part II respectively should be distributed to every server in some manner. The iterations in Algorithm 1 are required to be performed synchronously, even though there is no central coordinator or synchronized step. At the top of our agenda is the development

of asynchronous distributed algorithms to implement the Wiener filtering procedure. In future studies, we plan to integrate machine learning methods for enhanced noise estimation and tackle stochastic optimization problems with objective functions other than mean square error in (2.4), cf. [1,3,12,29,31,32,42].

2.2. Optimal recovery of wide-band stationary graph signals

In this subsection, we consider optimal **unbiased** reconstruction filters for the scenario that the filtering procedure satisfies (2.1a) and

$$\mathbf{H}\mathbf{1} = \tau\mathbf{1} \tag{2.20}$$

for some $\tau \neq 0$, the input signals $\mathbf{x}$ are wide-band stationary with

$$\mathbb{E}\mathbf{x} = c\mathbf{1} \quad \text{and} \quad \mathbb{E}(\mathbf{x} - \mathbb{E}\mathbf{x})(\mathbf{x} - \mathbb{E}(\mathbf{x}))^T = \widetilde{\mathbf{R}} = \tilde{r}(\mathbf{S}_1, \dots, \mathbf{S}_d) \tag{2.21}$$

for some (unknown) constant $c \in \mathbb{R}$ and multivariate polynomial $\tilde{r}$, the output $\mathbf{y}$ in (2.1c) are corrupted by some noise $\boldsymbol{\epsilon}$ satisfying (2.1d), and the covariance matrix $\mathbf{G}$ of the noise and the regularization matrix $\mathbf{K}$ satisfy

$$\mathbf{G}\mathbf{1} = \mathbf{K}\mathbf{1} = \mathbf{0}, \tag{2.22}$$

cf. (2.10) and see Remark 2.6 when the requirements (2.20) and (2.22) are not met.

Set $\tilde{\mathbf{x}} = \mathbf{x} - \mathbb{E}\mathbf{x} = \mathbf{x} - c\mathbf{1}$. By (2.1), (2.20), (2.21) and symmetry of graph shifts $\mathbf{S}_1, \dots, \mathbf{S}_d$, we can show that $\tilde{\mathbf{x}}$ is stationary and independent with the noise $\boldsymbol{\epsilon}$,

$$\mathbb{E}\tilde{\mathbf{x}} = \mathbf{0}, \ \mathbb{E}\tilde{\mathbf{x}}\boldsymbol{\epsilon}^T = \mathbf{0} \ \text{ and } \ \mathbb{E}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T = \widetilde{\mathbf{R}} = \tilde{r}(\mathbf{S}_1, \dots, \mathbf{S}_d), \tag{2.23}$$

and the output $\mathbf{y}$ is wide-band stationary,

$$\mathbb{E}\mathbf{y} = \mathbf{H}\mathbb{E}\tilde{\mathbf{x}} + \mathbb{E}\boldsymbol{\epsilon} + c\mathbf{H}\mathbf{1} = c\tau\mathbf{1} \tag{2.24}$$

and

$$\begin{aligned}\mathbb{E}(\mathbf{y} - \mathbb{E}\mathbf{y})(\mathbf{y} - \mathbb{E}\mathbf{y})^T &= \mathbf{H}(\mathbb{E}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T)\mathbf{H}^T + \mathbb{E}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T = \mathbf{H}\widetilde{\mathbf{R}}\mathbf{H}^T + \mathbf{G} \\ &= (h^2\tilde{r} + g)(\mathbf{S}_1, \dots, \mathbf{S}_d).\end{aligned} \tag{2.25}$$

For a reconstruction filter $\mathbf{W}$, we obtain from (2.1), (2.20) and (2.21) that

$$\mathbf{W}\mathbf{y} - \mathbf{x} = c(\mathbf{W}\mathbf{H}\mathbf{1} - \mathbf{1}) + (\mathbf{W}\mathbf{H} - \mathbf{I})\tilde{\mathbf{x}} + \mathbf{W}\boldsymbol{\epsilon}.$$

This implies that a reconstruction filter $\mathbf{W}$ of wide-band stationary signals is **unbiased** if and only if

$$\mathbf{W}\mathbf{1} = \tau^{-1}\mathbf{1}. \tag{2.26}$$

Therefore for an unbiased reconstruction filter $\mathbf{W}$ and a regularization matrix $\mathbf{K}$ satisfying (2.22), we have

$$\mathbf{W}\mathbf{y} - \mathbf{x} = (\mathbf{W}\mathbf{H} - \mathbf{I})\tilde{\mathbf{x}} + \mathbf{W}\boldsymbol{\epsilon}$$

and

$$\begin{aligned}\mathbf{y}^T\mathbf{W}^T\mathbf{K}\mathbf{W}\mathbf{y} &= (\mathbf{H}\tilde{\mathbf{x}} + \boldsymbol{\epsilon})^T\mathbf{W}^T\mathbf{K}\mathbf{W}(\mathbf{H}\tilde{\mathbf{x}} + \boldsymbol{\epsilon}) + \mathbf{1}^T\mathbf{K}\mathbf{1} \\ &\quad + \mathbf{1}^T\mathbf{K}\mathbf{W}(\mathbf{H}\tilde{\mathbf{x}} + \boldsymbol{\epsilon}) + (\mathbf{H}\tilde{\mathbf{x}} + \boldsymbol{\epsilon})^T\mathbf{W}^T\mathbf{K}\mathbf{1} \\ &= (\mathbf{H}\tilde{\mathbf{x}} + \boldsymbol{\epsilon})^T\mathbf{W}^T\mathbf{K}\mathbf{W}(\mathbf{H}\tilde{\mathbf{x}} + \boldsymbol{\epsilon}).\end{aligned}$$

Hence following the argument used in the proof of Theorem 2.1 with the signal $\mathbf{x}$ and polynomial r replaced by $\tilde{\mathbf{x}}$ and $\tilde{r}$ respectively, and applying (2.10), (2.20) and (2.23), we can show that the stochastic Wiener filter $\widetilde{\mathbf{W}}_{\text{mse}}$ in (2.28) is an optimal unbiased filter to reconstruct wide-band stationary signals.

Theorem 2.5. *Let the input $\mathbf{x}$, the noisy output $\mathbf{y}$ and the additive noise $\boldsymbol{\epsilon}$ be as in (2.21), (2.1c) and (2.1d) respectively, the covariance matrix $\mathbf{G}$ of the noise and the regularization matrix $\mathbf{K}$ satisfy (2.22), and let the filtering procedure associated with the filter $\mathbf{H}$ satisfy (2.1a) and (2.20). Assume that $\mathbf{H}\widetilde{\mathbf{R}}\mathbf{H}^T + \mathbf{G}$ and $\mathbf{P} + \mathbf{K}$ are strictly positive definite. Then*

$$F_{\text{mse},P,\mathbf{K}}(\mathbf{W}) \geq F_{\text{mse},P,\mathbf{K}}(\widetilde{\mathbf{W}}_{\text{mse}}) \tag{2.27}$$

hold for all unbiased reconstructing filters $\mathbf{W}$, where $F_{\text{mse},P,\mathbf{K}}(\mathbf{W})$ is the SMSE in (2.2) and

$$\widetilde{\mathbf{W}}_{\text{mse}} = (\mathbf{P} + \mathbf{K})^{-1}\mathbf{P}\widetilde{\mathbf{R}}\mathbf{H}^T(\mathbf{H}\widetilde{\mathbf{R}}\mathbf{H}^T + \mathbf{G})^{-1}. \tag{2.28}$$

Moreover, $\tilde{\mathbf{x}}_{\text{mse}} = \widetilde{\mathbf{W}}_{\text{mse}}\mathbf{y}$ is an unbiased estimator to the wide-band stationary signal $\mathbf{x}$.

Following the distributed algorithm used to implement the stochastic Wiener filtering procedure, the unbiased estimation $\tilde{\mathbf{x}}_{\text{mse}} = \widetilde{\mathbf{W}}_{\text{mse}}\mathbf{y}$ can be implemented at the vertex level with one-hop communication when

$$h^2(\mathbf{t})\tilde{r}(\mathbf{t}) + g(\mathbf{t}) > 0 \text{ for all } \mathbf{t} \in [\boldsymbol{\mu}, \boldsymbol{\nu}]. \tag{2.29}$$

Numerical demonstrations to denoise wide-band stationary signals are presented in Section 4.1.

Remark 2.6. The expectation $\mathbb{E}\mathbf{x} = c\mathbf{1}$ of wide-band stationary signals $\mathbf{x}$ in the definition (2.21) is often not known a priori in many engineering applications. For the case that it is given, we have

$$\mathbb{E}\mathbf{x}\mathbf{x}^T = c^2\mathbf{1}\mathbf{1}^T + \widetilde{\mathbf{R}}.$$

Consequently, we have the flexibility to dispense with the assumptions (2.20) and (2.22) concerning on the filter $\mathbf{H}$, the noise covariance matrix $\mathbf{G}$ and the regularization matrix $\mathbf{K}$. Thus, the Wiener filter to reconstruct wide-band stationary signals $\mathbf{x}$ with known c is the filter $\mathbf{W}_{\text{mse}}$ in (2.3) with $\mathbf{R}$ replaced by $c^2\mathbf{1}\mathbf{1}^T + \widetilde{\mathbf{R}}$, denoted by $\widetilde{\mathbf{W}}^*_{\text{mse}}$, i.e.,

$$F_{\text{mse},P,\mathbf{K}}(\mathbf{W}) \geq F_{\text{mse},P,\mathbf{K}}(\widetilde{\mathbf{W}}^*_{\text{mse}})$$

hold for all reconstruction filters $\mathbf{W}$. For the case that $c \neq 0$, $\widetilde{\mathbf{W}}^*_{\text{mse}}\mathbf{y}$ could be a biased estimate of the original signal $\mathbf{x}$, and the stochastic Wiener filtering procedure $\mathbf{y} \to \widetilde{\mathbf{W}}^*_{\text{mse}}\mathbf{y}$ cannot be implemented in a distributed manner with one-hop communication in general. If assumptions (2.20) and (2.22) hold, the Wiener filter $\widetilde{\mathbf{W}}^*_{\text{mse}}$ mentioned above will align with the unbiased reconstructing filter $\widetilde{\mathbf{W}}_{\text{mse}}$ in (2.28), and it can be implemented at the vertex level with one-hop communication.

3. Wiener filters for deterministic graph signals

In this section, we consider the scenario that the filtering procedure $\mathbf{x} \longmapsto \mathbf{y} = \mathbf{H}\mathbf{x}$ in (1.6) has the filter $\mathbf{H}$ given in (2.1a), its inputs $\mathbf{x} = (x(i))_{i\in V}$ are deterministic signals with their energy bounded by some $\delta_0 > 0$,

$$\|\mathbf{x}\|_2 \leq \delta_0, \tag{3.1}$$

and its outputs

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon} \tag{3.2}$$

are corrupted by some random noise $\boldsymbol{\epsilon}$ which has mean zero and covariance matrix $\mathbf{G} = \text{cov}(\boldsymbol{\epsilon})$ being a polynomial of graph shifts $\mathbf{S}_1, \dots, \mathbf{S}_d$, i.e.,

$$\mathbb{E}\boldsymbol{\epsilon} = \mathbf{0} \ \text{ and } \mathbf{G} = g(\mathbf{S}_1, \dots, \mathbf{S}_d) \tag{3.3}$$

for some multivariate polynomial g. For the above setting, we introduce the *worst-case mean squared error*, WMSE for abbreviation, of a reconstruction filter $\mathbf{W}$ by

$$F_{\text{wmse},P}(\mathbf{W}) = \sum_{i\in V} p(i) \max_{\|\mathbf{x}\|_2 \leq \delta_0} \mathbb{E}|(\mathbf{W}\mathbf{y})(i) - x(i)|^2, \tag{3.4}$$

where $p(i)_{i\in V}$ are the weights on the graph $\mathcal{G}$ [2,11]. In this section, we discuss the optimal reconstruction filter $\mathbf{W}_{\rm wmse}$ with respect to the WMSE $F_{{\rm wmse},P}$ in (3.4), and we propose a distributed algorithm to implement the worst-case Wiener filtering procedure (3.10) at the vertex level with one-hop communication.

First, we provide a **universal** solution to the minimization problem

$$\min_{\mathbf{W}} F_{{\rm wmse},P}(\mathbf{W}), \tag{3.5}$$

which is independent on the weight P, see Section 5.2 for the proof.

Theorem 3.1. *Let the filter $\mathbf{H}$, the input $\mathbf{x}$, the noisy output $\mathbf{y}$, the noise $\boldsymbol{\epsilon}$, and the WMSE $F_{{\rm wmse},P}$ be as in (2.1a), (3.1), (3.2), (3.3) and (3.4) respectively. Assume that $\delta_0^2\mathbf{H}\mathbf{H}^T+\mathbf{G}$ is strictly positive definite (or, equivalently, nonsingular). Then*

$$F_{{\rm wmse},P}(\mathbf{W}) \ge F_{{\rm wmse},P}(\mathbf{W}_{\rm wmse}) = \delta_0^2 - \delta_0^4 {\rm tr}\big((\delta_0^2\mathbf{H}\mathbf{H}^T+\mathbf{G})^{-1}\mathbf{H}\mathbf{P}\mathbf{H}^T\big) \tag{3.6}$$

hold for all reconstruction filters $\mathbf{W}$, where $\mathbf{P}$ is the diagonal matrix with diagonal entries $p(i), i\in V$, and

$$\mathbf{W}_{\rm wmse} = \delta_0^2\mathbf{H}^T\big(\delta_0^2\mathbf{H}\mathbf{H}^T+\mathbf{G}\big)^{-1}. \tag{3.7}$$

Moreover, the reconstruction filter $\mathbf{W}_{\rm wmse}$ is the unique solution of the minimization problem (3.5) if $\mathbf{P}$ is invertible, i.e., the weight $p(i)$ at every vertex $i\in V$ is positive.

We call the optimal reconstruction filter $\mathbf{W}_{\rm wmse}$ in (3.7) as the *worst-case Wiener filter*. Denote the order of the graph $\mathcal{G}$ by N. For the case that the probability measure P is the uniform probability measure P_U, we can simplify the estimate (3.6) as follows:

$$F_{{\rm wmse},P_U}(\mathbf{W}_{\rm wmse}) = \frac{\delta_0^2}{N}{\rm tr}\big((\delta_0^2\mathbf{H}\mathbf{H}^T+\mathbf{G})^{-1}\mathbf{G}\big), \tag{3.8}$$

cf. (2.7). If the random noises $\boldsymbol{\epsilon}$ are further assumed to be i.i.d. and have mean zero and variance $\sigma^2\mathbf{I}$, we can use singular values $\mu_i(\mathbf{H}), 1\le i\le N$, of the filter $\mathbf{H}$ to estimate the WMSE,

$$F_{{\rm wmse},P_U}(\mathbf{W}_{\rm wmse}) = \frac{\delta_0^2\sigma^2}{N}\sum_{i=1}^{N}\frac{1}{\delta_0^2\mu_i(\mathbf{H})^2+\sigma^2}. \tag{3.9}$$

Denote the reconstructed signal via the worst-case Wiener filter $\mathbf{W}_{\rm wmse}$ by

$$\mathbf{x}_{\rm wmse} = \mathbf{W}_{\rm wmse}\mathbf{y}, \tag{3.10}$$

where $\mathbf{y}$ is given in (3.2). By (3.7), the reconstructed signal $\mathbf{x}_{\rm wmse}$ can be obtained by the combination of an inverse filtering procedure

$$\mathbf{z} = \big(\delta_0^2\mathbf{H}\mathbf{H}^T+\mathbf{G}\big)^{-1}\mathbf{y} \tag{3.11a}$$

and a filtering procedure

$$\mathbf{x}_{\rm wmse} = \delta_0^2\mathbf{H}^T\mathbf{z}, \tag{3.11b}$$

where the noisy observation $\mathbf{y}$ is the input. As the graph shifts $\mathbf{S}_1,\dots,\mathbf{S}_d$ are symmetric and commutative, $\mathbf{H}$ is a polynomial graph filter in (2.1a) and (3.3) holds, we have that $\delta_0^2\mathbf{H}^T=\delta_0^2\mathbf{H}=\delta_0^2h(\mathbf{S}_1,\dots,\mathbf{S}_d)$ and

$$\delta_0^2\mathbf{H}\mathbf{H}^T+\mathbf{G} = \delta_0^2\mathbf{H}^2+\mathbf{G} = (\delta_0^2h^2+g)(\mathbf{S}_1,\dots,\mathbf{S}_d)$$

are polynomial filters of $\mathbf{S}_1,\dots,\mathbf{S}_d$. Therefore using [13, Algorithms 1 and 2], the filtering procedure (3.11b) can be implemented at the vertex level with one-hop communication. Here, we employ the iterative Chebyshev polynomial approximation algorithm in [13, Algorithm 4] to implement the inverse filtering procedure (3.11a) if the following requirement is met:

$\delta_0^2h^2(\mathbf{t})+g(\mathbf{t})>0$ for all $\mathbf{t}\in[\boldsymbol{\mu},\boldsymbol{\nu}]$.

Algorithm 2 Polynomial approximation algorithm to implement the worst-case Wiener filtering procedure $\mathbf{x}_{\rm wmse}=\mathbf{W}_{\rm wmse}\mathbf{y}$ at a vertex $i\in V$.

Inputs: Polynomial coefficients of polynomial filters $\mathbf{H},\mathbf{G}$ and $\mathbf{G}_M$ (Chebyshev polynomial approximation in [13, Eqn. (4.15)] to the inverse filter $(\delta_0^2\mathbf{H}\mathbf{H}^T+\mathbf{G})^{-1}$), entries $S_k(i,j), j\in\mathcal{N}_i$ in the i-th row of the shifts $\mathbf{S}_k, 1\le k\le d$, the value $y(i)$ of the input signal $\mathbf{y}=(y(i))_{i\in V}$ at the vertex i, and number L of iteration.

Pre-iteration: Find the polynomial coefficients of polynomial filter $\delta_0^2\mathbf{H}^2+\mathbf{G}$.

Initialization: $n=0$ and zero initial $x^{(0)}(i)=0$.

Iteration: Use [13, Algorithms 1 and 2] to implement the filtering procedures $\mathbf{e}^{(m)}=(\delta_0^2\mathbf{H}^2+\mathbf{G})\mathbf{x}^{(m-1)}-\mathbf{y}$ and $\mathbf{x}^{(m)}=\mathbf{x}^{(m-1)}-\mathbf{G}_M\mathbf{e}^{(m)}$ at the vertex i, with the output of the L-th iteration denoted by $x^{(L)}(i)$.

Post-iteration: Use [13, Algorithms 1 and 2] to implement the filtering procedure $\mathbf{x}_{\rm wmse}=\delta_0^2\mathbf{H}\mathbf{x}^{(L)}$ at the vertex i, with the output denoted by $x^{(L)}_{\rm wmse}(i)$.

Output: $x_{\rm wmse}(i)\approx x^{(L)}_{\rm wmse}(i)$, the approximate value of the output data of the Wiener filtering procedure $\mathbf{x}_{\rm wmse}=\mathbf{W}_{\rm wmse}\mathbf{y}$ at the vertex i.

Hence the worst-case Wiener filtering procedure (3.11) can be implemented at the vertex level with one-hop communication, see Algorithm 2 for the implementation at a vertex.

For a weight $P=(p(i))_{i\in V}$ on the graph $\mathcal{G}$ and a reconstruction filter $\mathbf{W}$,

$$\widetilde{F}_{{\rm wmse},P}(\mathbf{W}) = \max_{\|\mathbf{x}\|_2\le\delta_0}\sum_{i\in V}p(i)\mathbb{E}|(\mathbf{W}\mathbf{y})(i)-\mathbf{x}(i)|^2 \tag{3.12}$$

is another natural WMSE measurement, cf. (3.4). The above objective function is convex, however we are **unaware** of any explicit solution to the corresponding minimization problem, cf. Theorem 3.1. By (3.2) and (3.3), we obtain

$$\begin{aligned}\widetilde{F}_{{\rm wmse},P}(\mathbf{W}) &= \sup_{\|\mathbf{x}\|_2\le\delta_0}\mathbf{x}^T(\mathbf{H}^T\mathbf{W}^T-\mathbf{I})\mathbf{P}(\mathbf{W}\mathbf{H}-\mathbf{I})\mathbf{x}+{\rm tr}\big(\mathbf{P}\mathbf{W}(\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)\mathbf{W}^T\big)\\ &= \delta_0^2\lambda_{\max}\big((\mathbf{H}^T\mathbf{W}^T-\mathbf{I})\mathbf{P}(\mathbf{W}\mathbf{H}-\mathbf{I})\big)+{\rm tr}(\mathbf{P}\mathbf{W}\mathbf{G}\mathbf{W}^T)\\ &\le \delta_0^2{\rm tr}\big((\mathbf{H}^T\mathbf{W}^T-\mathbf{I})\mathbf{P}(\mathbf{W}\mathbf{H}-\mathbf{I})\big)+{\rm tr}(\mathbf{P}\mathbf{W}\mathbf{G}\mathbf{W}^T)\\ &= {\rm tr}\big(\mathbf{P}\big(\delta_0^2(\mathbf{W}\mathbf{H}-\mathbf{I})(\mathbf{H}^T\mathbf{W}^T-\mathbf{I})+\mathbf{W}\mathbf{G}\mathbf{W}^T\big)\big) = F_{{\rm wmse},P}(\mathbf{W}),\end{aligned}$$

where the inequality holds as the matrix $(\mathbf{H}^T\mathbf{W}^T-\mathbf{I})\mathbf{P}(\mathbf{W}\mathbf{H}-\mathbf{I})$ is positive semidefinite. Similarly, we have the following lower bound estimate,

$$\begin{aligned}\widetilde{F}_{{\rm wmse},P}(\mathbf{W}) &\ge \frac{\delta_0^2}{N}{\rm tr}\big((\mathbf{H}^T\mathbf{W}^T-\mathbf{I})\mathbf{P}(\mathbf{W}\mathbf{H}-\mathbf{I})\big)+{\rm tr}(\mathbf{P}\mathbf{W}\mathbf{G}\mathbf{W}^T)\\ &\ge \frac{F_{{\rm wmse},P}(\mathbf{W})}{N}.\end{aligned}$$

For the case that the weight P_U is uniform and the random noise vector $\boldsymbol{\epsilon}$ is i.i.d. with mean zero and variance $\sigma^2\mathbf{I}$, we get

$$\begin{aligned}\widetilde{F}_{{\rm wmse},P_U}(\mathbf{W}_{\rm wmse}) &= \frac{\delta_0^2\sigma^2}{N}\max_{1\le i\le N}\frac{\sigma^2}{(\delta_0^2\mu_i(\mathbf{H})^2+\sigma^2)^2}\\ &\quad+\frac{\delta_0^2\sigma^2}{N}\sum_{i=1}^{N}\frac{\delta_0^2\mu_i(\mathbf{H})^2}{(\delta_0^2\mu_i(\mathbf{H})^2+\sigma^2)^2},\end{aligned}$$

where $\mu_i(\mathbf{H}), 1\le i\le N$, are singular values of the filter $\mathbf{H}$, cf. (3.9) for the estimate for $F_{{\rm wmse},P_U}(\mathbf{W}_{\rm wmse})$.

4. Numerical demonstrations

Let $\mathcal{G}_N=(V_N,E_N), N\ge 2$, be random geometric graphs with vertices randomly deployed on $[0,1]^2$ and an undirected edge between two vertices if their physical distance is not larger than $\sqrt{2/N}$ [13,21,26]. In Sections 4.1 and 4.2, we consider denoising synthesis (wide-band) stationary signals and the real temperature data set collected at 32 weather stations in the region of Brest (France) in January 2014, via the Wiener procedures (2.3) and (2.11a) with/without regularization,

and we compare the denoising performances via the Tikhonov regularization method in (4.4) and (4.12). It is noted that the Wiener filtering procedures, whether with or without regularization, exhibit potentially superior performance in denoising synthesis (wide-band) stationary signals and real temperature data compared to the conventional Tikhonov regularization approach.

4.1. Denoising (wide-band) stationary signals on random geometric graphs

Let $\mathbf{L}$ be the Laplacian on the random geometric graph $\mathcal{G}_N$ with $N = 256$. In simulations of this subsection, we first consider stationary signals $\mathbf{x}$ on the random geometric graph $\mathcal{G}_{256}$ with correlation matrix $\mathbb{E}\mathbf{x}\mathbf{x}^T = \mathbf{I} + \mathbf{L}/2$, and noisy observations $\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon}$ being the inputs $\mathbf{x}$ corrupted by some additive noises $\boldsymbol{\epsilon}$ which are independent of the input signal $\mathbf{x}$ and whose entries are i.i.d. random variables with normal distribution $\mathcal{N}(0, \varepsilon)$ for some $\varepsilon > 0$, and we select the uniform probability measure $\mathbf{P}$ in the stochastic mean squared error (2.2). In other words, we consider the Wiener filtering procedure (2.8) in the scenario that

$$\mathbf{H} = \mathbf{I},\ \mathbf{R} = \mathbf{I} + \mathbf{L}/2,\ \mathbf{P} = N^{-1}\mathbf{I} \ \text{ and } \ \mathbf{G} = \varepsilon^2\mathbf{I}.$$

In this subsection, we use $\mathbf{K} = \varepsilon^2\mathbf{L}/(4N)$ as the regularization matrix to balance the fidelity and regularization terms in (2.2). Therefore

$$\mathbf{x}_{\mathrm{W0}} := \mathbf{W}^0_{\mathrm{mse}}\mathbf{y} = \mathbf{R}\big(\mathbf{R}+\mathbf{G}\big)^{-1}\mathbf{y} = (\mathbf{I}+\mathbf{L}/2)\big((1+\varepsilon^2)\mathbf{I}+\mathbf{L}/2\big)^{-1}\mathbf{y} \tag{4.1}$$

and

$$\mathbf{x}_{\mathrm{W}} := \mathbf{W}_{\mathrm{mse}}\mathbf{y} = (\mathbf{P}+\mathbf{K})^{-1}\mathbf{P}\mathbf{R}\big(\mathbf{R}+\mathbf{G}\big)^{-1}\mathbf{y} = (\mathbf{I}+\varepsilon^2\mathbf{L}/4)^{-1}\mathbf{x}_{\mathrm{W0}} \tag{4.2}$$

are signals reconstructed from the noisy observation $\mathbf{y}$ via the Wiener procedures (2.11a) and (2.3) without/with regularization taken into account respectively.

Define the input signal-to-noise ratio (ISNR) and the output signal-to-noise ratio (SNR) by

$$\mathrm{ISNR} = -20\log_{10}\frac{\|\boldsymbol{\epsilon}\|_2}{\|\mathbf{x}\|_2} \ \text{ and } \ \mathrm{SNR} = -20\log_{10}\frac{\|\widehat{\mathbf{x}}-\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \tag{4.3}$$

respectively, where $\widehat{\mathbf{x}}$ are either the reconstructed signal $\mathbf{x}_{\mathrm{W0}}$ via the Wiener procedure (2.11a) without regularization, or the reconstructed signal $\mathbf{x}_{\mathrm{W}}$ via the Wiener procedure (2.3) with regularization, or the reconstructed signal

$$\begin{aligned}\mathbf{x}_{\mathrm{Tik}} &= (\mathbf{P}+\mathbf{K})^{-1}\mathbf{P}\mathbf{y} = (\mathbf{I}+\varepsilon^2\mathbf{L}/4)^{-1}\mathbf{y}\\ &= \arg\min_{\mathbf{x}}\,(\mathbf{x}-\mathbf{y})^T\mathbf{P}(\mathbf{x}-\mathbf{y}) + \mathbf{x}^T\mathbf{K}\mathbf{x}\end{aligned} \tag{4.4}$$

via the Tikhonov regularization approach. Shown on the top left of Fig. 1 is a random signal on $\mathcal{G}_{256}$, while on the top middle and right are the performances of Wiener filtering procedure and the Tikhonov regularization approach. From the plot on top middle in Fig. 1, we observe that the Wiener procedures with regularization have the best performances on denoising stationary signals when the noise level is large, i.e. $\varepsilon > 1$ and the Wiener procedures without regularization have the best performances on denoising stationary signals when the noise level is small, i.e. $\varepsilon < 1$.

Next we consider denoising signals with certain regularity. Let $\mathbf{x}_{\mathrm{pp}}$ be the four-strip signal on the random geometric graph $\mathcal{G}_{256}$ that impose the polynomial $0.5 - 2c_x$ on the first and third diagonal strips and $0.5 + c_x^2 + c_y^2$ on the second and fourth strips respectively, where (c_x, c_y) are the coordinates of vertices [21, Fig. 2]. We do simulations on denoising the four-strip signal $\mathbf{x}_{\mathrm{pp}}$, i.e., we apply the same Tikhonov regularization and Wiener procedures with/without regularization except that stationary signals $\mathbf{x}$ are replaced by $\mathbf{x}_{\mathrm{pp}}$ and the correlation matrix $\mathbf{R}$ in (4.1) and (4.2) replaced by $\mathbf{R}_0 = \mathbf{x}_{\mathrm{pp}}\mathbf{x}_{\mathrm{pp}}^T$, see bottom middle plotted on Fig. 1. Let $\Pi(\mathbf{L})$ be the set of all polynomials of $\mathbf{L}$. Observe that the correlation matrix $\mathbf{R}_0$ does not commutate with graph shifts $\mathbf{L}$, and hence $\mathbf{R}_0 \notin \Pi(\mathbf{L})$. This indicates that $\mathbf{x}_{\mathrm{pp}}$, are not "stationary" signals on the graph $\mathcal{G}_N$, cf. [30].

Table 1
Stationarity level of four-strips graph signal $\mathbf{x}_{\mathrm{pp}}$ on random geometric graph over 100 trials and the data set of hourly temperature in the region of Brest (France).

Data set	Graph shift	Stationarity level
4-strips signal	$\mathbf{L}$	0.3335
USPS dataset	$\mathbf{L}$	0.7116
Temperature dataset	$\mathbf{L}_{\mathcal{T}} \otimes \mathbf{I}, \mathbf{I} \otimes \mathbf{L}$	0.8871

When dealing with real data, we have access to the covariance matrix, while the correlation matrix may not be a polynomial of graph shifts, thereby they are not "stationary" graph signals. Let

$$\mathbf{R}_P = \arg\min_{\mathbf{R}\in\Pi(\mathbf{L})} \|\mathbf{R}-\mathbf{R}_0\|_{\mathrm{F}} \tag{4.5}$$

be the best approximation to the correlation matrix $\mathbf{R}_0$ with $\|\mathbf{R}_P - \mathbf{R}_0\|_{\mathrm{F}} = 241.4$ in one implementation, see the argument used in the proof [13, Theorem A.4] for the construction of the polynomial filter $\mathbf{R}_P$. Now, we define the

$$\text{Stationarity Level} = \frac{\|\mathbf{R}_P\|_{\mathrm{F}}}{\|\mathbf{R}_0\|_{\mathrm{F}}}, \tag{4.6}$$

cf. [30]. In Table 1, we include the stationarity levels of the four-strips graph signal x_{pp} on $\mathcal{G}_{256}$ by take the average over 100 trials, the USPS dataset, and of temperature data sets in France used in Section 4.2.

In this section, we also perform the in-painting/de-noising experiment with the USPS dataset. For our experiments, we consider the same setup as the ones in [30]. That is every digit is considered as an independent realization of a graph stationary signal, and remove the mean. In the experiments, we use 20 digits for the graph construction and the stationarity level in (4.6) is 0.7116, and assume that the USPS dataset is corrupted by some random noises,

$$\widetilde{\mathbf{X}} = \mathbf{X} + \boldsymbol{\eta},$$

where the additive noises $\boldsymbol{\eta}$, are independent of the input signal $\mathbf{X}_d$, and they have mean zero and covariance matrix $\mathbf{G}_\eta = \eta^2(\mathbf{I})$ for some noise level parameter $\eta > 0$. Similar to the setup in Section 4.1, we consider using the Wiener filtering procedure (4.1) to inpainting the above noisy USPS dataset, where we use a mask covering 50% of the pixel and various amounts of noise. In Fig. 2, we compare the in-painting performances of Wiener filter with known correlation matrix $\mathbf{R}$, Wiener with the estimated matrix $\mathbf{R}_P$ in (4.5), and the Tikhonov approach and TV approach in [30]. To implement the Tikhonov approach and TV approach in [30], we use the code provided there. We observe that Wiener outperforms the other graph- based models for in-painting USPS dataset.

We will explore the Wiener filtering procedure (2.3) and (2.11a) with/without regularization for denoising the four-strips graph signal $\mathbf{x}_{\mathrm{pp}}$ on $\mathcal{G}_{256}$, utilizing the correlation matrix $\mathbf{R}_0$ and the approximation correlation matrix $\mathbf{R}_P$ in the polynomial form, shown at the bottom middle of Fig. 1. The performances of Wiener filtering procedure are also presented in Table 2, where the average Signal-to-Noise Ratios (SNRs) over 100 trials for the four-strips graph signal $\mathbf{x}_{\mathrm{pp}}$ are evaluated. SNR0 and SNRP represent the output SNRs in Wiener procedures without regularization in (2.11a) associated with the correlation matrix $\mathbf{R}_0$ and the approximated correlation matrix $\mathbf{R}_P$ respectively. SNRK and SNRKP correspond to the Wiener procedure with regularization in (2.3), associated with the correlation matrices $\mathbf{R}_0$ and $\mathbf{R}_P$ respectively. Additionally, SNRTik denotes the SNR obtained through the Tikhonov regularization approach. From the plot in the bottom middle of Fig. 1 and Table 2, it is evident that the Wiener filter without regularization, utilizing a known correlation matrix, exhibits excellent performance in denoising random graph signals. Furthermore, from the fourth and sixth columns of Table 2, we notice that the Wiener filtering procedures with the ap-

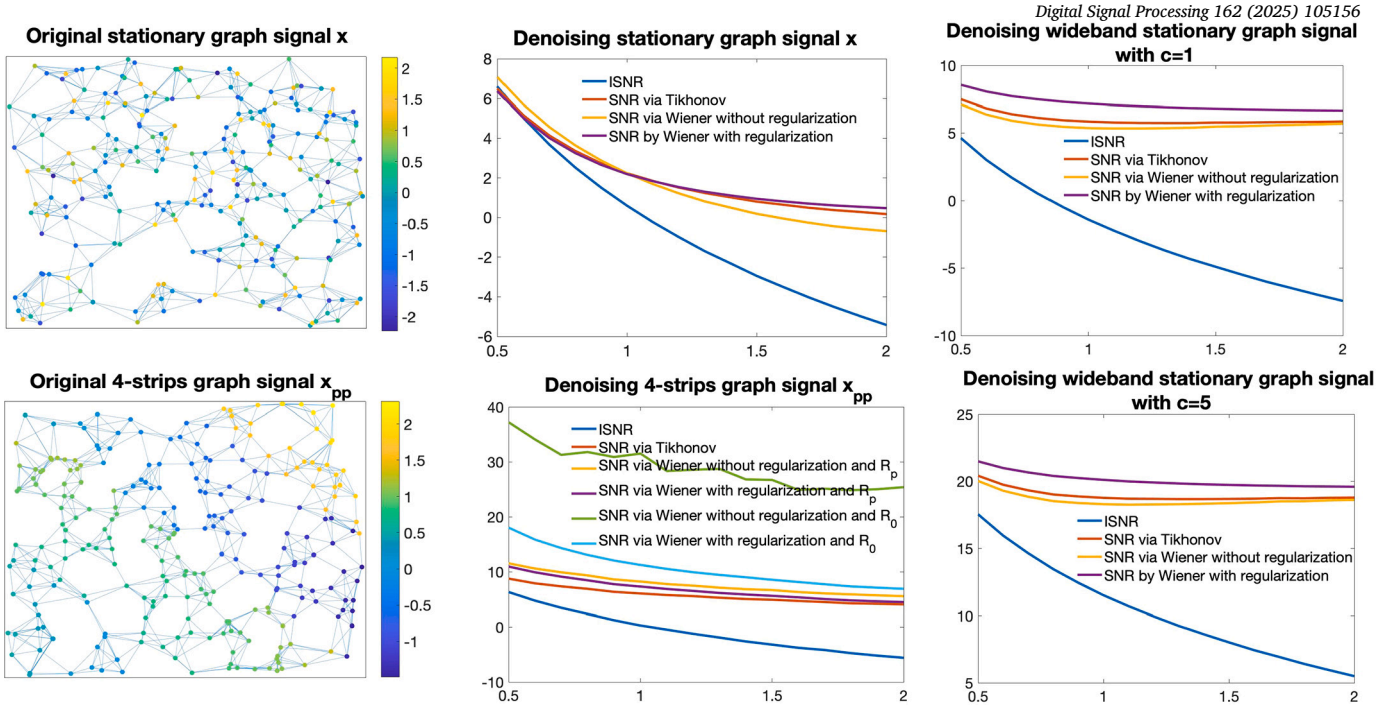

Fig. 1. Plotted on the left are the stationary signal $\mathbf{x}$ with correlation matrix $\mathbf{I}+\mathbf{L}/2$ (top left), the four-strip signal $\mathbf{x}_{pp}$ in [21] (bottom left) with $\|\mathbf{x}_{pp}\|_2 = 16.38$ and $\mathbf{x}_{pp}^T\mathbf{L}\mathbf{x}_{pp} = 285.87$. Plotted on the top middle/bottom middle/top right/bottom right are the averages of the input signal-to-noise ratio ISNR and output signal-to-noise ratio SNR of denoising stationary signals $\mathbf{x}$ (top middle), the four-strip signal $\mathbf{x}_{pp}$ (bottom middle), wide-band stationary signals with $c = 1$ (top right) and $c = 5$ (bottom right) on the random geometric graph $\mathcal{G}_{256}$ via the Wiener filtering procedures without/with regularization and Tikhonov regularization approach over 1000 trials for different noise levels $0.5 \le \varepsilon \le 2$.

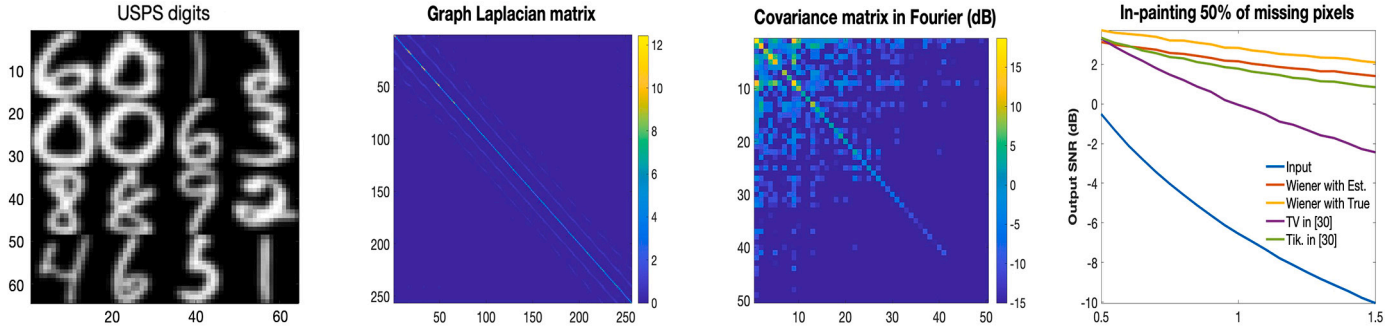

Fig. 2. From the left to right: Some digits of the USPS dataset, Laplacian matrix of the 10 nearest neighbors (patch) graph, spectral covariance matrix $\mathbf{U}^T\mathbf{R}\mathbf{U}$ for the first 50 graph frequencies, average of recovery errors over 500 trials for different noise levels $0.5 \le \varepsilon \le 1.5$. Wiener filter with known correlation matrix $\mathbf{R}$ and with the estimated correlation matrix $\mathbf{R}_P$ in (4.5) outperforms Tikhonov regularization approach and TV method in [30].

proximation correlation matrix $\mathbf{R}_P$ in polynomial form, with or without regularization, demonstrate satisfactory performance in denoising the noisy random stationary graph signal $\mathbf{x}_{pp}$. Interestingly, the stochastic Wiener filter in (2.11a), without regularization, surpasses its counterpart in (2.3) with the regularization matrix. From Table 2, it is evident that the Wiener filtering procedures, with the approximation correlation matrix $\mathbf{R}_P$ and with or without regularization, outperform the Tikhonov regularization approach in denoising the noisy random signal $\mathbf{x}_{pp}$. This suggests that the Wiener procedure without regularization may offer the **best** performance in denoising signals with certain regularity.

In this subsection, we finally consider denoising wide-band stationary signals. Let $\mathbf{x}$ be as in (2.21) on a random geometric graph $\mathcal{G}_{256}$ with

$$\mathbb{E}\mathbf{x} = c\mathbf{1} \quad \text{and} \quad \mathbb{E}(\mathbf{x}-\mathbb{E}\mathbf{x})(\mathbf{x}-\mathbb{E}\mathbf{x})^T = \mathbf{I}+\mathbf{L}/2,$$

where $c \neq 0$ is not necessarily to be given in advance. The observations $\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon}$ are the inputs $\mathbf{x}$ corrupted by some additive noises $\boldsymbol{\epsilon}$ which is independent of the input signal $\mathbf{x}$ and whose covariance matrix is $\mathbf{G} = \varepsilon^2\mathbf{L}$ for some $\varepsilon > 0$. We select the uniform probability measure $\mathbf{P}$ in the stochastic mean squared error. With the above scenario, we have

Table 2
Average SNRs of denoising four-strips graph signal $\mathbf{x}_{pp}$ on random geometric graph over 100 trials via the Wiener procedures (2.11a) with the correlation matrix $\mathbf{R}_0$ and with the approximated correlation matrix $\mathbf{R}_P$, via the Wiener procedures (2.3) with the correlation matrix $\mathbf{R}_0$ and with the approximated correlation matrix $\mathbf{R}_P$, and finally via the Tikhonov regularization approach.

ϵ	ISNR	SNR	SNRP	SNRK	SNRKP	SNRTik
0.5	6.4300	36.2787	11.6578	18.1801	11.0695	8.8684
0.6	4.7739	36.8671	10.6340	16.0175	9.9283	7.9353
0.8	2.2342	31.1045	9.2365	13.1347	8.4280	6.7976
1.0	0.3434	28.3604	8.3268	11.3660	7.4358	6.2035
1.2	-1.2585	28.6795	7.4793	10.0266	6.5486	5.5442
1.6	-3.7719	25.4907	6.3014	8.2448	5.3556	4.7422
2.0	-5.6573	24.6062	5.6178	6.9541	4.5010	4.0602

$$\mathbf{H}=\mathbf{I},\ \widetilde{\mathbf{R}}=\mathbf{I}+\mathbf{L}/2,\ \mathbf{P}=N^{-1}\mathbf{I}\ \text{ and }\ \mathbf{G}=\varepsilon^2\mathbf{L}$$

in the Wiener filtering procedure (2.8). From the simulation results presented in the top/bottom right of Fig. 1, we see that the Wiener procedure with regularization has the best performance on denoising than the Wiener procedure without regularization and Tikhonov regularization approach do.

4.2. Denoising temperature data sets in France

In this subsection, we consider denoising the data set of hourly temperature measured in Celsius collected at 32 weather stations in the region of Brest (France) in January 2014 [30,35]. The temperature data set is published by French national meteorological service and it is of size $32\times 24\times 31$.

We represent the data set by vectors $\mathbf{X}_d^{\text{org}}=[(\mathbf{x}_d(t_0))^T,\dots,(\mathbf{x}_d(t_{23}))^T]^T$ of dimension $N=768=24\times 32$, where for each $1\le d\le 31$, the column vectors $\mathbf{x}_d(t_i), 0\le i\le 23$, is the regional temperature at t_i-th hour of d-th day in January 2014. As $\mathbf{X}_1^{\text{org}},\dots,\mathbf{X}_{31}^{\text{org}}$ can be considered as 31 samples of the temperate data set in January 2014, we can pre-process the temperature data set by eliminating the average temperature $\mathbf{X}_{\text{avg}}^{\text{org}}=\sum_{d=1}^{31}\mathbf{X}_d^{\text{org}}/31$ in January 2014, and set

$$\mathbf{X}_d=\mathbf{X}_d^{\text{org}}-\mathbf{X}_{\text{avg}}^{\text{org}},\ 1\le d\le 31.$$

We model vectors $\mathbf{X}_d, 1\le d\le 31$, as graph signals residing on the Cartesian product $\mathcal{T}\square\mathcal{S}$ of the undirected circulant graph $\mathcal{T}$ with 24 vertices and the undirected weather station graph $\mathcal{S}$ with 32 locations constructed by the 5 nearest neighboring stations in physical distances. Denote Laplacian matrices on the graph $\mathcal{S}$ and $\mathcal{T}$ by $\mathbf{L}_{\mathcal{S}}$ and $\mathbf{L}_{\mathcal{T}}$ respectively, and set $\mathbf{S}_1=\mathbf{L}_{\mathcal{T}}\otimes\mathbf{I}$ and $\mathbf{S}_2=\mathbf{I}\otimes\mathbf{L}_{\mathcal{S}}$. By the property of Kronecker product, one may verify that $\mathbf{S}_1$ and $\mathbf{S}_2$ are commutative graph shifts on the product graph $\mathcal{T}\square\mathcal{S}$ [13].

In our simulations, we assume that the temperature data set are corrupted by some random noises,

$$\widetilde{\mathbf{X}}_d=\mathbf{X}_d+\boldsymbol{\eta}_d,\ d=1,\dots,31, \tag{4.7}$$

where the additive noises $\boldsymbol{\eta}_d, 1\le i\le 31$, are independent of the input signal $\mathbf{X}_d$, and they have mean zero and covariance matrix $\mathbf{G}_\eta=\eta^2(\mathbf{I}+(\mathbf{S}_1+\mathbf{S}_2)/2)$ for some noise level parameter $\eta>0$. Similar to the setup in Section 4.1, we consider using the Wiener filtering procedure (2.8) and the Tikhonov regularization approach (4.12) to denoise the above noisy temperature data set in the scenario that

$$\mathbf{H}=\mathbf{I},\ \mathbf{P}=N^{-1}\mathbf{I}\ \text{ and }\ \mathbf{G}_\eta=\eta^2(\mathbf{I}+(\mathbf{S}_1+\mathbf{S}_2)/2).$$

Shown in the first and second columns of Table 3 are the noise levels η and averages of the corresponding ISNR in (4.3) with $\boldsymbol{\epsilon}$ and $\mathbf{x}$ replaced by $\boldsymbol{\eta}_d$ and $\mathbf{X}_d$ over 100 trials for every $1\le d\le 31$.

As $\mathbf{X}_1,\dots,\mathbf{X}_{31}$ can be thought as 31 samples of the pre-processed temperate data set in the region of Brest (France), we may estimate their correlation (covariance) matrix by

$$\mathbf{R}_0=\frac{\mathbf{X}_1\mathbf{X}_1^T+\cdots+\mathbf{X}_{31}\mathbf{X}_{31}^T}{31},$$

which has its Frobenius norm $\|\mathbf{R}_0\|_{\text{F}}=3533.8$ and all eigenvalues in the range $[0,976.4427]\cup\{3388.9\}$. Now we apply the stochastic Wiener filter (2.3) with the correlation matrix $\mathbf{R}$ replaced by $\mathbf{R}_0$ without the regularization matrix to denoise the noisy weather data set $\widetilde{\mathbf{X}}_d, 1\le d\le 31$. Denote the output of the above filtering procedure by

$$\mathbf{X}_d^0=\mathbf{R}_0(\mathbf{R}_0+\mathbf{G}_\eta)^{-1}\widetilde{\mathbf{X}}_d, 1\le d\le 31.$$

Shown in the third column of Table 3 is the performance of the above denoising procedure in the term of average SNRs over 100 trials for every $1\le d\le 31$, where SNR0 is the SNR in (4.3) with $\hat{\mathbf{x}}$ and $\mathbf{x}$ replaced by $\mathbf{X}_d^0$ and $\mathbf{X}_d, 1\le d\le 31$.

Table 3
Average SNRs of denoising the noisy observation $\widetilde{\mathbf{X}}_d$ over 100 trials for every $1\le d\le 31$ via the Wiener procedures (2.11a) with the correlation matrix $\mathbf{R}_0$ and with the approximated correlation matrix $\mathbf{R}_P$, via the Wiener procedures (2.3) with the approximated correlation matrix $\mathbf{R}_P$, and finally via the Tikhonov regularization approach.

η	ISNR	SNR0	SNRP	SNRK	SNRTik
0.1	10.1355	27.2839	15.2134	14.9590	13.7221
0.2	7.1127	24.3854	13.8253	13.3387	12.4387
0.3	5.3646	22.8508	13.0647	12.4099	11.7821
0.4	4.1111	21.6392	12.5557	11.7545	11.3508
0.6	2.3372	20.0777	11.9061	10.9130	10.7766
1	0.1269	18.3237	11.1009	9.9151	10.0404
1.4	-1.3347	17.1772	10.6292	9.3127	9.5541

Let $\Pi(\mathbf{S}_1,\mathbf{S}_2)$ be the set of all polynomials of $\mathbf{S}_1$ and $\mathbf{S}_2$. Observe that the correlation matrix $\mathbf{R}_0$ does not commutate with graph shifts $\mathbf{S}_1$ and $\mathbf{S}_2$, and hence $\mathbf{R}_0\notin\Pi(\mathbf{S}_1,\mathbf{S}_2)$. This indicates that $\mathbf{X}_d, 1\le d\le 31$, are not "stationary" signals on the product graph $\mathcal{T}\square\mathcal{S}$, cf. [30]. Let

$$\mathbf{R}_P=\arg\min_{\mathbf{R}\in\Pi(\mathbf{S}_1,\mathbf{S}_2)}\|\mathbf{R}-\mathbf{R}_0\|_{\text{F}} \tag{4.8}$$

be the best approximation to the correlation matrix $\mathbf{R}_0$ with $\|\mathbf{R}_P-\mathbf{R}_0\|_{\text{F}}=1677.4$. Next we will explore the Wiener filtering procedure with/without regularization on denoising the noisy weather data set $\{\widetilde{\mathbf{X}}_d, 1\le d\le 31\}$ by using the approximation filter $\mathbf{R}_P$ of the correlation matrix $\mathbf{R}_0$, which is of the polynomial form. Denote the output applying the stochastic Wiener filter (2.3) with the correlation matrix $\mathbf{R}$ replaced by $\mathbf{R}_P$ (4.8) without the regularization matrix by

$$\mathbf{X}_d^P=\mathbf{R}_P(\mathbf{R}_P+\mathbf{G}_\eta)^{-1}\widetilde{\mathbf{X}}_d, 1\le d\le 31. \tag{4.9}$$

The temperatures at neighboring observation stations and consecutive hours have strong correlation, which implies that the pre-processed hourly weather data sets $\mathbf{X}_d, 1\le d\le 31$, have certain regularity. In particular, the average energy $\sum_{d=1}^{31}\|\mathbf{X}_d\|_2/31$ of the data set is 67.55, and the average regularity measurements $\sum_{d=1}^{31}(\mathbf{X}_d^T\mathbf{S}_1\mathbf{X}_d)^{1/2}/31$ in the temporal domain and $\sum_{d=1}^{31}(\mathbf{X}_d^T\mathbf{S}_2\mathbf{X}_d)^{1/2}/31$ in the spatial domain are 14.4209 and 22.6823 respectively. This inspires us to consider the Wiener filtering procedure with regularization to denoise the hourly temperature data set. In our simulations, we use

$$\mathbf{K}_\eta=\eta^2\tilde{\alpha}\mathbf{S}_1+\eta^2\tilde{\beta}\mathbf{S}_2 \tag{4.10}$$

as the regularization matrix to balance the fidelity and regularization in (2.2), where $\tilde{\alpha}=(\mathbf{X}_d^T\mathbf{S}_1\mathbf{X}_d)^{-1}$ and $\tilde{\beta}=(\mathbf{X}_d^T\mathbf{S}_2\mathbf{X}_d)^{-1}$ are regularization factors in spatial and temporal directions. Denote the output of the stochastic Wiener filter (2.3) with the correlation matrix $\mathbf{R}$ replaced by $\mathbf{R}_P$ (4.8) and the regularization matrix $\mathbf{K}$ given in (4.10) by

$$\mathbf{X}_d^K=(\mathbf{I}+N\mathbf{K}_\eta)^{-1}\mathbf{R}_P(\mathbf{R}_P+\mathbf{G}_\eta)^{-1}\widetilde{\mathbf{X}}_d. \tag{4.11}$$

Shown in the fourth and fifth columns of Table 3 are the average SNRs of the above denoising procedures over 100 trials for every $1\le d\le 31$, where SNRP and SNRK are the SNR in (4.3) with the input $\mathbf{x}$ replaced by $\mathbf{X}_d$ and $\hat{\mathbf{x}}$ replaced by $\mathbf{X}_d^P$ in (4.9) and $\mathbf{X}_d^K$ in (4.11), $1\le d\le 31$, respectively.

Finally we compare the performance of the proposed Wiener filtering procedures with the Tikhonov regularization approach,

$$\begin{aligned}\mathbf{X}_d^{\text{Tik}}&=\arg\min_{\mathbf{Z}}(\mathbf{Z}-\widetilde{\mathbf{X}}_d)^T\mathbf{P}(\mathbf{Z}-\widetilde{\mathbf{X}}_d)+\mathbf{Z}^T\mathbf{K}_\eta\mathbf{Z}\\&=(\mathbf{I}+N\mathbf{K}_\eta)^{-1}\widetilde{\mathbf{X}}_d,\ 1\le d\le 31,\end{aligned} \tag{4.12}$$

where the regularization matrix $\mathbf{K}_\eta$ is given in (4.10). Shown in the last column of Table 3 is the average SNRs of the denoising procedure via the Tikhonov regularization approach over 100 trials for every $1\le d\le 31$,

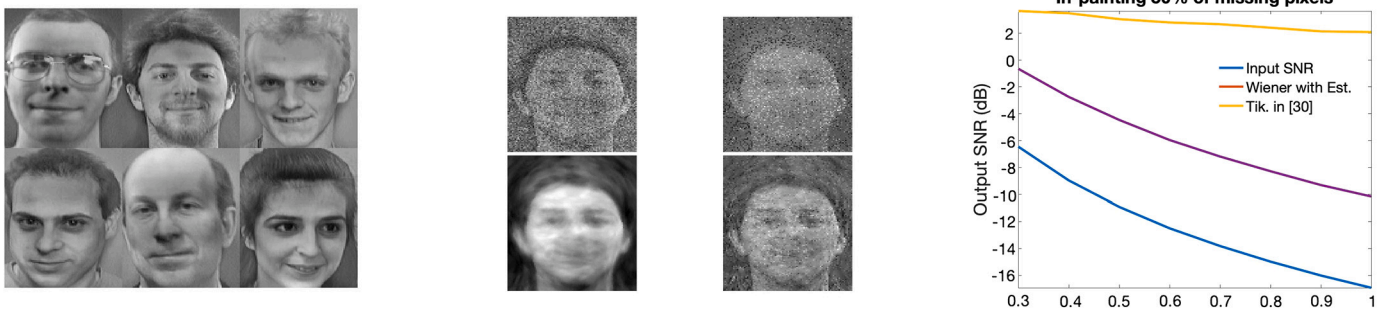

Fig. 3. On the left, we present samples from the ORL dataset. In the middle, arranged from top left to bottom right, we show a noisy sample image with a noise level of $\varepsilon = 0.3$ (its input signal-to-noise ratio (ISNR) is -2.6541), its 50% measurement, its denoised image through the Wiener filter (its SNR is 4.0414), and the denoised image through the Tikhonov regularization approach (its SNR is 2.4233). On the right, we plot the averages of the ISNR, SNR for denoising the ORL dataset using the Wiener filtering procedure (without regularization and with the estimated correlation matrix) and the Tikhonov regularization approach over 50 trials for different noise levels ranging from $0.3 \le \varepsilon \le 1$.

where SNRTik is the SNR in (4.3) with $\mathbf{x}$ and $\hat{\mathbf{x}}$ replaced by $\mathbf{X}_d$ and $\mathbf{X}_d^{\mathrm{Tik}}$, $1 \le d \le 31$.

From the third column on Table 3, we observe that the Wiener filtering procedure with known correlation matrix has **impressive performance** on denoising random signals, particularly with the hourly weather data set in the region of Brest (France). We also notice from the fourth and fifth columns of Table 3 that the Wiener filtering procedures with the approximation correlation matrix $\mathbf{R}_P$ in (4.8) and with/without regularization $\mathbf{K}_\eta$ in (4.10) demonstrate satisfactory performances in denoising the noisy temperature datasets $\widetilde{\mathbf{X}}_d$, $1 \le d \le 31$, and the stochastic Wiener filter in (2.11a) without the regularization outperforms the one in (2.3) with the regularization matrix $\mathbf{K}_\eta$ being chosen as in (4.10). Comparing the SNR data from fifth and sixth columns of Table 3, we see that the denoising procedure via the Tikhonov regularization approach exhibits a slightly better performance compared to the Wiener filtering procedure associated with $\mathbf{R}_P$ in (4.8) and $\mathbf{K}_\eta$ in (4.10) does.

4.3. Denoising ORL faces dataset

In this section, we conduct experiments on the ORL faces dataset to demonstrate the performance of the Wiener filtering procedure. Each image in the ORL dataset consists of $112 \times 92 = 10304$ pixels, and its stationarity level is 0.7170 (see the left plot of Fig. 3). We apply the Wiener filter procedure (without regularization in (2.11a) but with the estimated correlation matrix $\mathbf{R}_P$ in (4.5)) and the Tikhonov regularization approach in [30] to the ORL dataset. The right plot of Fig. 3 presents the average ISNR and SNR values over 50 trials for denoising the ORL dataset using the Wiener filtering procedure without regularization and with the estimated correlation matrix, compared to the Tikhonov regularization approach. Consequently, we see that the Wiener procedure, without regularization in (2.11a) and with the approximated correlation matrix, outperforms the Tikhonov regularization approach in denoising the noisy image signal.

5. Proofs

In this section, we collect the proofs of Theorems 2.1 and 3.1.

5.1. Proof of Theorem 2.1

By (2.1b), (2.1c) and (2.1d), we have

$$\mathbb{E}\mathbf{y}\mathbf{y}^T = \mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G} \quad \text{and} \quad \mathbb{E}\mathbf{y}\mathbf{x}^T = \mathbf{H}\mathbf{R}. \tag{5.1}$$

By (2.1b), (2.2) and (5.1), we obtain

$$\begin{aligned} F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W}) &= \mathrm{tr}\left(\mathbf{P}\mathbb{E}((\mathbf{W}\mathbf{y}-\mathbf{x})(\mathbf{W}\mathbf{y}-\mathbf{x})^T)\right) + \mathrm{tr}\mathbf{W}^T\mathbf{K}\mathbf{W}\mathbb{E}(\mathbf{y}\mathbf{y}^T) \\ &= \mathrm{tr}\left(\mathbf{W}^T(\mathbf{P}+\mathbf{K})\mathbf{W}(\mathbf{H}\mathbf{R}\mathbf{H}^T+\mathbf{G})\right) + \mathrm{tr}(\mathbf{P}\mathbf{R}) \\ &\quad -\mathrm{tr}(\mathbf{H}\mathbf{R}\mathbf{P}\mathbf{W}) - \mathrm{tr}(\mathbf{W}^T\mathbf{P}\mathbf{R}\mathbf{H}^T). \end{aligned} \tag{5.2}$$

Substituting $\mathbf{W}$ in (5.2) by $\mathbf{W}_{\mathrm{mse}}$ proves (2.5).

By (2.3) and (5.2), we obtain

$$\begin{aligned} F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W}) &= F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W}_{\mathrm{mse}}) + \mathrm{tr}\left(\mathbf{V}^T(\mathbf{P}+\mathbf{K})\mathbf{V}(\mathbf{H}\mathbf{R}\mathbf{H}^T+\mathbf{G})\right) \\ &\quad +\mathrm{tr}\left(\mathbf{V}^T(\mathbf{P}+\mathbf{K})\mathbf{W}_{\mathrm{mse}}(\mathbf{H}\mathbf{R}\mathbf{H}^T+\mathbf{G}) - \mathbf{V}^T\mathbf{P}\mathbf{R}\mathbf{H}^T\right) \\ &\quad +\mathrm{tr}\left(\mathbf{W}_{\mathrm{mse}}^T(\mathbf{P}+\mathbf{K})\mathbf{V}(\mathbf{H}\mathbf{R}\mathbf{H}^T+\mathbf{G}) - \mathbf{H}\mathbf{R}\mathbf{P}\mathbf{V}\right) \\ &= F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W}_{\mathrm{mse}}) \\ &\quad +\mathrm{tr}\left((\mathbf{H}\mathbf{R}\mathbf{H}^T+\mathbf{G})^{1/2}\mathbf{V}^T(\mathbf{P}+\mathbf{K})\mathbf{V}(\mathbf{H}\mathbf{R}\mathbf{H}^T+\mathbf{G})^{1/2}\right) \\ &\ge F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W}_{\mathrm{mse}}), \end{aligned} \tag{5.3}$$

where $\mathbf{V} = \mathbf{W} - \mathbf{W}_{\mathrm{mse}}$, the first and second equality follows from (5.2) and (2.3) respectively, and the inequality holds as $(\mathbf{H}\mathbf{R}\mathbf{H}^T + \mathbf{G})^{1/2}\mathbf{V}^T(\mathbf{P}+\mathbf{K})\mathbf{V}(\mathbf{H}\mathbf{R}\mathbf{H}^T+\mathbf{G})^{1/2}$ are positive semidefinite for all matrices $\mathbf{V}$. This proves that $\mathbf{W}_{\mathrm{mse}}$ is a minimizer to the minimization problem $\min_{\mathbf{W}} F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W})$.

The conclusion that $\mathbf{W}_{\mathrm{mse}}$ is a unique minimizer to the minimization problem

$$\min_{\mathbf{W}} F_{\mathrm{mse},P,\mathbf{K}}(\mathbf{W})$$

follows from (5.3) and the assumptions that $\mathbf{P}+\mathbf{K}$ and $\mathbf{H}\mathbf{R}\mathbf{H}^T+\mathbf{G}$ are strictly positive definite.

5.2. Proof of Theorem 3.1

Define the worst-case mean squared error of a reconstruction vector $\mathbf{w}$ with respect to a given unit vector $\mathbf{u}$ by

$$f_{\mathrm{wmse},\mathbf{u}}(\mathbf{w}) = \max_{\|\mathbf{x}\|_2 \le \delta_0} \mathbb{E}|\mathbf{w}^T\mathbf{y} - \mathbf{u}^T\mathbf{x}|^2 \tag{5.4}$$

and set

$$\mathbf{w}_{\mathrm{wmse},\mathbf{u}} = \mathbf{W}_{\mathrm{wmse}}^T\mathbf{u}. \tag{5.5}$$

By direct computation, we have

$$F_{\mathrm{wmse},P}(\mathbf{W}) = \sum_{i\in V} p(i) f_{\mathrm{wmse},\mathbf{e}_i}(\mathbf{W}^T\mathbf{e}_i), \tag{5.6}$$

where $\mathbf{e}_i, i \in V$, are delta signals taking value one at vertex i and zero at all other vertices. Then it suffices to show that $\mathbf{w}_{\mathrm{wmse},\mathbf{u}}$ is the optimal reconstructing vector with respect to the measurement $f_{\mathrm{wmse},\mathbf{u}}(\mathbf{w})$, i.e.,

$$\mathbf{w}_{\mathrm{wmse},\mathbf{u}} = \arg\min_{\mathbf{w}} f_{\mathrm{wmse},\mathbf{u}}(\mathbf{w}). \tag{5.7}$$

By (3.2), (3.3) and the assumption $\|\mathbf{u}\|_2 = 1$, we have

$$\begin{aligned} f_{\text{wmse},\mathbf{u}}(\mathbf{w}) &= \max_{\|\mathbf{x}\|_2 \le \delta_0} \mathbb{E}|(\mathbf{w}^T\mathbf{H} - \mathbf{u}^T)\mathbf{x} + \mathbf{w}^T\boldsymbol{\epsilon}|^2 \\ &= \max_{\|\mathbf{x}\|_2 \le \delta_0} \left|(\mathbf{w}^T\mathbf{H} - \mathbf{u}^T)\mathbf{x}\right|^2 + \mathbb{E}|\mathbf{w}^T\boldsymbol{\epsilon}|^2 \\ &= \delta_0^2(\mathbf{w}^T\mathbf{H} - \mathbf{u}^T)(\mathbf{H}^T\mathbf{w} - \mathbf{u}) + \mathbf{w}^T\mathbf{G}\mathbf{w} \\ &= \mathbf{w}^T\left(\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G}\right)\mathbf{w} - 2\delta_0^2\mathbf{w}^T\mathbf{H}\mathbf{u} + \delta_0^2. \end{aligned}$$

By (3.7) and (5.5), we have

$$\begin{aligned} f_{\text{wmse},\mathbf{u}}(\mathbf{w}) &= f_{\text{wmse},\mathbf{u}}(\mathbf{w}_{\text{wmse},\mathbf{u}}) + \mathbf{v}^T\left(\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G}\right)\mathbf{v} \\ &\quad + 2\mathbf{v}^T\left(\left(\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G}\right)\mathbf{w}_{\text{wmse},\mathbf{u}} - \delta_0^2\mathbf{H}\mathbf{u}\right) \\ &= f_{\text{wmse},\mathbf{u}}(\mathbf{w}_{\text{wmse},\mathbf{u}}) + \mathbf{v}^T\left(\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G}\right)\mathbf{v} \ge f_{\text{wmse},\mathbf{u}}(\mathbf{w}_{\text{wmse},\mathbf{u}}), \end{aligned} \tag{5.8}$$

where $\mathbf{v} = \mathbf{w} - \mathbf{w}_{\text{wmse},\mathbf{u}}$ and the last inequality holds as $\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G}$ is strictly positive definite. This proves (5.7) and hence that $\mathbf{W}_{\text{wmse}}$ is a minimizer of the minimization problem (3.5), i.e., the inequality in (3.6) holds.

By (5.5), (5.6) and (5.7), we have

$$\begin{aligned} F_{\text{wmse},P}(\mathbf{W}_{\text{wmse}}) &= \sum_{i\in V} p(i) f_i(\mathbf{w}_{\text{wmse},\mathbf{e}_i}) \\ &= \sum_{i\in V} p(i)\big(-\delta_0^4\mathbf{e}_i^T\mathbf{H}^T(\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G})^{-1}\mathbf{H}\mathbf{e}_i + \delta_0^2\big) \\ &= \delta_0^2 - \delta_0^4\text{tr}\left(\mathbf{P}\mathbf{H}^T(\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G})^{-1}\mathbf{H}\right) \\ &= \delta_0^2 - \delta_0^4\text{tr}\left((\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G})^{-1}\mathbf{H}\mathbf{P}\mathbf{H}^T\right). \end{aligned}$$

This proves the equality in (3.6) and hence completes the proof of the conclusion (3.6).

The uniqueness of the minimization problem (3.5) follows from (5.6) and (5.8), and the strictly positive definiteness of the matrices $\mathbf{P}$ and $\delta_0^2\mathbf{H}\mathbf{H}^T + \mathbf{G}$.

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.

Acknowledgment

The authors would like to thank the reviewers for their constructive comments for the improvement. This work is partially supported by National Key R&D Program of China (No. 2024YFA1013703), National Nature Science Foundation of China (12171490), Guangdong Basic and Applied Basic Research Foundation (2022A1515011060), and Fundamental Research Funds for the Central Universities, Sun Yat-sen University (24lgqb019).

Data availability

Data will be made available on request.

References

[1] M.A. Abu-Hashem, M. Shehab, M.K.Y. Shambour, M.S. Daoud, L. Abualigah, Improved Black Widow Optimization: an investigation into enhancing cloud task scheduling efficiency, Sustain. Comput. Inf. Syst. 41 (Jan. 2024) 100949.

[2] N. Bi, M.Z. Nashed, Q. Sun, Reconstructing signals with finite rate of innovation from noisy samples, Acta Appl. Math. 107 (1) (July 2009) 339–372.

[3] S. Biswas, A. Shaikh, A.E.-S. Ezugwu, J. Greeff, S. Mirjalili, U.K. Bera, L. Abualigah, Enhanced prairie dog optimization with Levy flight and dynamic opposition-based learning for global optimization and engineering design problems, Neural Comput. Appl. 36 (Mar. 2024) 11137–11170.

[4] M.M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: going beyond Euclidean data, IEEE Signal Process. Mag. 34 (4) (2017) 18–42.

[5] C. Cheng, N. Emirov, Q. Sun, Preconditioned gradient descent algorithm for inverse filtering on spatially distributed networks, IEEE Signal Process. Lett. 27 (Oct. 2020) 1834–1838.

[6] C. Cheng, J. Jiang, N. Emirov, Q. Sun, Iterative Chebyshev polynomial algorithm for signal denoising on graphs, in: Proceeding 13th Int. Conf. on SampTA, Bordeaux, France, July 2019, pp. 1–5.

[7] C. Cheng, Y. Jiang, Q. Sun, Spatially distributed sampling and reconstruction, Appl. Comput. Harmon. Anal. 47 (1) (July 2019) 109–148.

[8] C. Chong, S. Kumar, Sensor networks: evolution, opportunities, and challenges, Proc. IEEE 91 (Aug. 2003) 1247–1256.

[9] M. Coutino, E. Isufi, G. Leus, Advances in distributed graph filtering, IEEE Trans. Signal Process. 67 (9) (May 2019) 2320–2333.

[10] X. Dong, D. Thanou, L. Toni, M. Bronstein, P. Frossard, Graph signal processing for machine learning: a review and new perspectives, IEEE Signal Process. Mag. 37 (6) (2020) 117–127.

[11] Y. Eldar, M. Unser, Nonideal sampling and interpolation from noisy observations in shift-invariant spaces, IEEE Trans. Signal Process. 54 (7) (June 2006) 2636–2651.

[12] N. Emirov, G. Song, Q. Sun, A divide-and-conquer algorithm for distributed optimization on networks, Appl. Comput. Harmon. Anal. 70 (May 2024) 101623.

[13] N. Emirov, C. Cheng, J. Jiang, Q. Sun, Polynomial graph filter of multiple shifts and distributed implementation of inverse filtering, Sampl. Theory Signal Process. Data Anal. 20 (2022) 2.

[14] B. Girault, Stationary graph signals using an isometric graph translation, in: Proc. 23rd Eur. Signal Process. Conf, 2015, pp. 1516–1520.

[15] A. Gavili, X. Zhang, On the shift operator, graph frequency, and optimal filtering in graph signal processing, IEEE Trans. Signal Process. 65 (23) (Dec. 2017) 6303–6318.

[16] J. Hara, Y. Tanaka, Y.C. Eldar, Graph signal sampling under stochastic priors, IEEE Trans. Signal Process. 71 (Apr. 2023) 1421–1434.

[17] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 2012.

[18] E. Isufi, A. Loukas, A. Simonetto, G. Leus, Autoregressive moving average graph filtering, IEEE Trans. Signal Process. 65 (2) (Jan. 2017) 274–288.

[19] E. Isufi, F. Gama, D.I. Shuman, S. Segarra, Graph filters for signal processing and machine learning on graphs, IEEE Trans. Signal Process. (Jan. 2024) 1–32.

[20] X. Jian, W.P. Tay, Wide-sense stationarity in generalized graph signal processing, IEEE Trans. Signal Process. 70 (June 2022) 3414–3428.

[21] J. Jiang, C. Cheng, Q. Sun, Nonsubsampled graph filter banks: theory and distributed algorithms, IEEE Trans. Signal Process. 67 (15) (Aug. 2019) 3938–3953.

[22] J. Jiang, D.B. Tay, Q. Sun, S. Ouyang, Design of nonsubsampled graph filter banks via lifting schemes, IEEE Signal Process. Lett. 27 (Feb. 2020) 441–445.

[23] K. Lu, A. Ortega, D. Mukherjee, Y. Chen, Efficient rate-distortion approximation and transform type selection using Laplacian operators, in: 2018 Picture Coding Symposium (PCS), San Francisco, CA, June 2018, pp. 76–80.

[24] G. Mao, B. Fidan, B.D.O. Anderson, Wireless sensor network localization techniques, Comput. Netw. 51 (10) (July 2007) 2529–2553.

[25] N. Motee, Q. Sun, Sparsity and spatial localization measures for spatially distributed systems, SIAM J. Control Optim. 55 (1) (Jan. 2017) 200–235.

[26] P. Nathanael, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, D.K. Hammond, GSPBOX: a toolbox for signal processing on graphs, arXiv:1408.5781, Aug. 2014.

[27] M. Onuki, S. Ono, M. Yamagishi, Y. Tanaka, Graph signal denoising via trilateral filter on graph spectral domain, IEEE Trans. Signal Inf. Process. Netw. 2 (2) (June 2016) 137–148.

[28] A. Ortega, P. Frossard, J. Kovačević, J.M.F. Moura, P. Vandergheynst, Graph signal processing: overview, challenges, and applications, Proc. IEEE 106 (5) (May 2018) 808–828.

[29] E. Osaba, E. Villar-Rodriguez, J. Del Ser, A.J. Nebro, D. Molina, A. LaTorre, P.N. Suganthan, C.A. Coello Coello, F. Herrera, A tutorial on the design, experimentation and application of metaheuristic algorithms to real-world optimization problems, Swarm Evol. Comput. 64 (July 2021) 100888.

[30] N. Perraudin, P. Vandergheynst, Stationary signal processing on graphs, IEEE Trans. Signal Process. 65 (13) (July 2017) 3462–3477.

[31] K. Rajwar, K. Deep, S. Das, An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges, Artif. Intell. Rev. 56 (Apr. 2023) 13187–13257.

[32] T. Salehnia, A. Seyfollahi, S. Raziani, A. Noori, A. Ghaffari, A.R. Alsoud, L. Abualigah, An optimal task scheduling method in IoT-Fog-Cloud network using multi-objective moth-flame algorithm, Multimed. Tools Appl. 83 (2024) 34351–34372.

[33] A. Sandryhaila, J.M.F. Moura, Discrete signal processing on graphs: frequency analysis, IEEE Trans. Signal Process. 62 (12) (June 2014) 3042–3054.

[34] A. Sandryhaila, J.M.F. Moura, Discrete signal processing on graphs, IEEE Trans. Signal Process. 61 (7) (Apr. 2013) 1644–1656.

[35] S. Segarra, A.G. Marques, G. Leus, A. Ribeiro, Stationary graph processes: parametric power spectral estimation, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, Mar. 2017, pp. 4099–4103.

[36] S. Segarra, A.G. Marques, A. Ribeiro, Optimal graph-filter design and applications to distributed linear network operators, IEEE Trans. Signal Process. 65 (15) (Aug. 2017) 4117–4131.

[37] X. Shi, H. Feng, M. Zhai, T. Yang, B. Hu, Infinite impulse response graph filters in wireless sensor networks, IEEE Signal Process. Lett. 22 (8) (Aug. 2015) 1113–1117.

[38] 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) (May 2013) 83–98.

[39] D.I. Shuman, P. Vandergheynst, D. Kressner, P. Frossard, Distributed signal processing via Chebyshev polynomial approximation, IEEE Trans. Signal Inf. Process. Netw. 4 (4) (Dec. 2018) 736–751.

[40] L. Stanković, M. Daković, E. Sejdić, Introduction to graph signal processing, in: Vertex-Frequency Analysis of Graph Signals, Springer, 2019, pp. 3–108.

[41] W. Waheed, D.B.H. Tay, Graph polynomial filter for signal denoising, IET Signal Process. 12 (3) (Apr. 2018) 301–309.

[42] J. Xiao, X. Pan, J. Liu, J. Wang, P. Zhang, L. Abualigah, Load balancing strategy for SDN multi-controller clusters based on load prediction, J. Supercomput. 80 (Sep. 2023) 5136–5162.

[43] A.C. Yagan, M.T. Ozgen, Spectral graph based vertex-frequency Wiener filtering for image and graph signal denoising, IEEE Trans. Signal Inf. Process. Netw. 6 (Feb. 2020) 226–240.

[44] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Comput. Netw. 52 (12) (Aug. 2008) 2292–2330.