Shift-invariant spaces, bandlimited spaces and reproducing kernel spaces with shift-invariant kernels on undirected finite graphs

Seok-Young Chung and Qiyu Sun

Abstract—**In this paper, we introduce the concept of graph shift-invariant spaces (GSISs) of graph signals on undirected finite graphs and investigate their bandlimiting, reproducing kernel, and sampling properties. Bandlimited spaces are the most widely adopted models for representing graph signals. In this paper, under some technical conditions on the graph shift, we show that every GSIS is a graph bandlimited space (GBLS), and every GBLS is a principal GSIS. Functions in a reproducing kernel Hilbert space with shift-invariant kernel on graphs could be learnt with significantly low computational cost. In this paper, we demonstrate that every GSIS is a reproducing kernel Hilbert space with a shift-invariant kernel.**

Based on the nested Krylov structure of GSISs, we propose a finite-step sampling and reconstruction algorithm in the spatial domain. Its effectiveness is demonstrated on well-localized signals over cycle graphs and the flight delay dataset of the 50 busiest U.S. airports. On this flight delay dataset, the GSIS approach surpasses bandlimited spaces and RKHS models with diffusion and random walk kernels in both model suitability, data fitting capability and practical interpretability.

I. Introduction

A shift-invariant space (SIS) $\mathcal{H}$ of functions on the real line is a linear space invariant under integer shifts, i.e., $T_k f := f(\cdot - k) \in \mathcal{H}$ for all $f \in \mathcal{H}$ and $k \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of all integers. It has been widely used in approximation theory, wavelet analysis, sampling theory, Gabor analysis and many other mathematical and engineering fields ([1], [2], [3], [4], [5], [6], [7]). Bandlimited spaces and principal shift-invariant spaces are two typical examples of SISs.

Graph signal processing (GSP) offers an opportunity to represent, process, analyze, and visualize graph signals and network data ([8], [9], [10], [11], [12], [13], [14], [15], [16]). Similar to the one-order delay in classical signal processing, the concept of graph shifts has been introduced in GSP. Graph shifts are usually selected to have specific features, and designed to capture the topology of the underlying graph; see Section II-A. In this paper, we introduce the concept of graph shift-invariant spaces (GSISs) on undirected finite graphs, namely linear spaces of graph signals that remain invariant under some graph shift $\mathbf{S}$ (see (III.1)), and we study their various properties on bandlimiting, kernel reproducing and sampling. The GSIS is defined in analogy to the SIS on the real line, with the integer shifts $T_k, k \in \mathbb{Z}$, on the line replaced by the graph shift operations $\mathbf{S}^l, l \in \mathbb{N}$, where $\mathbb{N}$ is the set of all natural numbers. Our motivation stems from the observation that the graph-Fourier basis often fails to exhibit clear spatial patterns on many real-world graphs. This limitation reduces the descriptive power of purely spectral approaches. A natural question, therefore, is whether one can develop a spatial perspective for graph signals that complements and enriches the traditional bandlimiting framework in GSP. The GSISs introduced in this paper provide such a perspective: they emphasize spatial structure directly through graph-shift operations, and their underlying nested-Krylov-subspaces structure offers greater flexibility and practical interpretability.

Graph bandlimited spaces (GBLSs) are the most commonly used models for graph signals in GSP. The bandlimited framework enables sparse representation, fast reconstruction and efficient processing of graph signals that exhibit spatial regularity or smooth spectrum, and it forms a foundation for modern graph-based signal models and algorithms across networked systems, imaging, and data science ([17], [18], [19], [20], [21], [22], [23]). One may verify that every GBLS is invariant under the graph shift $\mathbf{S}$. The **first main contribution** of this paper is to demonstrate that, under the distinct eigenvalue conditions on the graph shift $\mathbf{S}$ (see Assumption II.1), every GSIS is a GBLS and every GBLS is a principal graph shift-invariant space (PGSIS); see Theorems III.1 and III.2. Therefore the terminologies on bandlimitedness, shift-invariance and principal-shift-invariance of a

Chung is with Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA; and Sun is with School of Data, Mathematical and Statistical Sciences, University of Central Florida, Orlando, Florida 32816, USA. Emails: sychung@msu.edu; qiyu.sun@ucf.edu.

linear space of graph signals are essentially the same in the undirected finite graph setting, with the first one illustrating its bandlimiting property in the spectral domain, the second emphasizing its shift-invariance in the spatial domain, and the third highlighting the spatial-frequency localization for its generator.

Every GBLS is a reproducing kernel Hilbert space (RKHS) ([17], [18], [24], [25]). For efficient learning of functions in a RKHS on an undirected graph, the kernels are usually selected to be shift-invariant; see (IV.1). Graph RKHSs with shift-invariant kernel (SIGRKHSs) have their inner products being defined by a generalized dot product in the graph Fourier domain, see Theorem IV.2. Common selection of such shift-invariant kernels includes diffusion kernels, regularization kernels, random walk kernels, spline kernels and their combinations ([26], [27], [28], [29], [30], [31], [33], [32], [34], [35], [36], [37], [38], [39]). A SIGRKHS is clearly invariant under the graph shift $\mathbf{S}$ and hence a GSIS. The **second main contribution** of this paper is to show that every GSIS embedded with the standard Euclidean inner product is a SIGRKHS; see Theorem IV.1.

Sampling and reconstruction in GBLSs is a well-studied topic in GSP. This framework provides characterizations of admissible sampling sets, computationally efficient reconstruction formulas, robustness guarantees in the presence of noise, and guidelines for designing sampling strategies ([17], [18], [19], [20], [21], [22], [40], [41]). In this paper, we study sampling and reconstruction in PGSISs. Krylov subspace method is a widely used iterative technique in numerical linear algebra, designed to solve large linear systems or compute eigenvalues without forming dense matrix factorizations. Motivated by the observation that every PGSIS possesses a nested Krylov subspace structure (see (V.9)), our **third main contribution** is to establish a sampling theorem for PGSISs and to develop a finite-step iterative algorithm for graph signal reconstruction without performing the eigendecomposition for the graph shift; see Theorem V.1, Corollaries V.2 and V.3, and Algorithm V.1. Our numerical simulations in Section VI show that the proposed nested-Krylov-subspace-based reconstruction Algorithm V.1 demonstrates robust and promising performance in recovering well-localized signals in a GSIS on cycle graphs, as well as in the flight-delay dataset of the 50 busiest airports in the United States. We also notice that the proposed reconstruction algorithm is self-adaptive and achieves accurate localized signal recovery with substantially lower computational overhead than the conventional bandlimiting method, further demonstrating the practical utility of the GSIS approach.

SISs and RKHSs on the real line are suitable for modeling signals with smooth spectrum ([3], [5], [7], [42], [43], [44]). Flight delays usually start from few airports and their ripple effects happen subsequently throughout the entire airport network. Our numerical simulations indicate that the on-time performance data of the 50 busiest airports in the USA can be modeled as graph signals living **more suitably** in a GSIS with the generators appropriately chosen than in a bandlimited space and RKHS with shift-invariant kernel, and moreover, the GSIS model effectively captures the ripple effect originating from a few major airports and propagating throughout the entire network; see Section VI-B. This suggests that the GSIS may offer flexibility and practical interpretability for modeling signals and datasets with strong localization pattern, such as those arising in transportation delays and diffusion phenomena in networked systems.

The paper is organized as follows. In Section II, we recall some preliminaries on graph shifts, polynomial filters, and graph Fourier transform. In Section III, we introduce GSISs, GBLSs and GSISs generated by multiple graph signals, and show that they are essentially the same; see Theorem III.1. In Section III-B, based on the graph uncertainty principle, we show that a PGSIS generated by a localized graph signal has large dimension and a PGSIS with small dimension is generated by some signal with large support; see Proposition III.3. In Section IV, we introduce the concept of shift-invariant kernel, discuss the inner product structure of SIGRKHSs, and show that every GSIS embedded with standard Euclidean inner product is a SIGRKHS; see (IV.1), and Theorems IV.1 and IV.2. In Section V, we consider sampling and reconstruction of signals in a finitely-generated GSIS; see Theorem V.1, Corollary V.2 and Algorithm V.1. Performance of Algorithm V.1 for signal reconstruction is presented in Section VI. All proofs are collected in Section VII. The conclusions and future study are summarized in Section VIII.

Notation: We denote the standard inner product and p-norm on the Euclidean space $\mathbb{R}^N$ by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_p, 1 \le p \le \infty$, respectively. We use $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{H}}$ to represent the inner product and norm on a Hilbert space $\mathcal{H}$. For a matrix $\mathbf{A}$, we denote its transpose by $\mathbf{A}^T$. As usual, we denote the set of nonnegative integers by $\mathbb{Z}_+$, the set of natural numbers by $\mathbb{N}$, the cardinality of a set W by $\#W$, the support of a vector $\mathbf{x} \in \mathbb{R}^N$ by $\operatorname{supp} \mathbf{x}$ and the dimension of a linear space $\mathcal{H}$ by $\dim \mathcal{H}$.

II. Preliminaries

Polynomial filters have been widely used in GSP, and graph shifts are building blocks for polynomial filters ([8], [9], [10], [11], [15], [37], [45]). In this section, we first recall some preliminaries on graph shifts and polynomial filters. Graph Fourier transform (GFT) decomposes graph signals into different frequency components. In this section we recall the conventional definition of GFT on undirected graphs ([9], [12], [14], [15], [39], [46], [47]).

A. Graph shifts

Let $\mathcal{G} = (V, E)$ be an undirected graph of order $N \ge 1$. We say that $\mathbf{S} = [S(i,j)]_{i,j \in V}$ is a *graph shift* if $S(i,j) = 0$ except that either $i = j$ or $(i,j) \in E$. Illustrative examples of graph shifts are the adjacency matrix $\mathbf{A}$, the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$, the symmetrically normalized Laplacian matrix $\mathbf{L}^{\mathrm{sym}} := \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$, and their variants, where $\mathbf{D}$ is the degree matrix.

Under additional real-valued and symmetric assumptions on the graph shift $\mathbf{S}$, we have the following eigendecomposition:

$$\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \tag{II.1}$$

where $\mathbf{U}$ is an orthogonal matrix and the diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}\,[\lambda(n)]_{1 \le n \le N}$ contains the eigenvalues of the graph shift $\mathbf{S}$ as its diagonal entries, ordered so that $|\lambda(1)| \le |\lambda(2)| \le \cdots \le |\lambda(N)|$ [12], [14]. To ensure meaningful spectral behavior, we make the following assumption on the graph shift $\mathbf{S}$.

Assumption II.1. *The graph shift* $\mathbf{S}$ *is real-valued and symmetric, and its eigenvalues* $\lambda(n), 1 \le n \le N$, *are assumed to be distinct.*

The distinct eigenvalue problem for the graph Laplacian, our illustrative example of graph shifts, is a classical topic in spectral graph theory [48]. It is known that Assumption II.1 is satisfied for the Laplacian on a graph of order N if it has diameter $N - 1$ [49], and with high probability $1 - o(1)$ if it is an Erdos–Renyi random graph $G(N, 1/2)$ [50].

B. Polynomial filters

We say that a filter $\mathbf{H}$ is a *polynomial filter* of the graph shift $\mathbf{S}$ if there exists a polynomial $h(t) = \sum_{l \in \mathbb{Z}_+} h_l t^l$ such that

$$\mathbf{H} = h(\mathbf{S}) = \sum_{l \in \mathbb{Z}_+} h_l \mathbf{S}^l, \tag{II.2}$$

where the sum is taken on a finite subset of $\mathbb{Z}_+$. A significant advantage is that the filtering procedure associated with a polynomial filter can be implemented at the vertex level in which each vertex is equipped with a one-hop communication subsystem. For a polynomial filter $\mathbf{H}$, one may verify that it *commutes* with the graph shift $\mathbf{S}$, i.e., $\mathbf{HS} = \mathbf{SH}$. Under Assumption II.1 on the graph shift $\mathbf{S}$, the converse also holds; see [8], [9], [10], [11], [37].

Proposition II.2. *Let* $\mathcal{G}$ *be an undirected finite graph,* $\mathbf{S}$ *be a graph shift satisfying Assumption II.1, and* $\mathbf{U}$ *be the orthogonal matrix in* (II.1). *Then a filter* $\mathbf{H}$ *commutes with the graph shift* $\mathbf{S}$ *if and only if* $\mathbf{H} = \mathbf{U}\mathbf{\Theta}\mathbf{U}^T$ *for some diagonal matrix* $\mathbf{\Theta}$ *if and only if* $\mathbf{H}$ *is a polynomial filter.*

We remark that the above equivalence between terminologies polynomial filters and shift-invariant filters no longer holds once the distinct-eigenvalue assumption on the graph shift $\mathbf{S}$ is removed [51].

C. Graph Fourier transform

Let $\mathbf{U} := [\mathbf{u}_1, \ldots, \mathbf{u}_N]$ be the orthogonal matrix in (II.1). We define the *graph Fourier transform* (GFT) $\widehat{\mathbf{x}} := \mathcal{F}\mathbf{x}$ of a graph signal $\mathbf{x}$ by

$$\widehat{\mathbf{x}} = \mathbf{U}^T\mathbf{x} = [\mathbf{u}_1^T\mathbf{x}, \ldots, \mathbf{u}_N^T\mathbf{x}]^T. \tag{II.3}$$

Analogous to the classical discrete Fourier transform, we may interpret the eigenvalues of the graph shift $\mathbf{S}$ as frequencies to the above GFT, and the corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$ to form its graph Fourier basis ([9], [12], [14], [15], [39], [46], [47], [52], [53]). For the GFT in (II.3), we obtain from (II.1) that

$$\widehat{\mathbf{S}\mathbf{x}} = \mathbf{\Lambda}\widehat{\mathbf{x}}, \tag{II.4}$$

where $\mathbf{\Lambda}$ is the diagonal matrix in (II.1).

III. Graph shift-invariant spaces and bandlimited spaces

In Section III-A, we investigate the bandlimiting and finite-generating properties of a GSIS; see Theorems III.1 and III.2. Since every GBLS is generated by a graph signal of sinc type and every GSIS is a GBLS, in Section III-B, we show that a GSIS generated by a well-localized graph signal does not have small dimension; see Proposition III.3.

A. Bandlimiting and finite-generating properties of graph shift-invariant spaces

Let $\mathcal{G}$ be an undirected graph of order $N \ge 1$, and $\mathbf{S}$ be a graph shift on $\mathcal{G}$. In analogy with the SIS on the real line, we say that a linear space $\mathcal{H}$ of graph signals on $\mathcal{G}$ is *shift-invariant* if it is closed under the action of graph shift operation $\mathbf{S}$, i.e.,

$$\mathbf{S}\mathbf{x} \in \mathcal{H} \quad \text{for all } \mathbf{x} \in \mathcal{H}. \tag{III.1}$$

Applying the graph shift $\mathbf{S}$ repeatedly and using the linear space property, we see that

$$\mathbf{Hx} \in \mathcal{H} \text{ for all } \mathbf{x} \in \mathcal{H} \text{ and polynomial filters } \mathbf{H}. \quad \text{(III.2)}$$

Let Φ be a nonempty set of nonzero graph signals on the graph $\mathcal{G}$. We say that a graph shift-invariant space (GSIS) $\mathcal{H}$ is *generated by* Φ if it is the minimal shift-invariant space containing Φ, i.e.,

$$\mathcal{H} = \mathcal{H}(\Phi) := \mathrm{span}\{\mathbf{S}^l \phi \mid l \in \mathbb{Z}_+,\ \phi \in \Phi\} \quad \text{(III.3)}$$

and that $\mathcal{H}$ is *principal* if it is generated by a single graph signal. As $\mathcal{H}(\Phi)$ is a linear subspace of $\mathbb{R}^N$, we may assume that Φ has finite cardinality, i.e., $\#\Phi < \infty$. By the classical Cayley-Hamilton theorem, $\mathbf{S}^N$ is the linear combination of $\mathbf{S}^m, 0 \le m \le N-1$. Therefore the GSIS generated by Φ is given by

$$\mathcal{H}(\Phi) = \mathrm{span}\{\mathbf{S}^m \phi \mid 0 \le m \le N-1, \phi \in \Phi\}. \quad \text{(III.4)}$$

We say that a linear space $\mathcal{H}$ of graph signals on $\mathcal{G}$ is *bandlimited* if

$$\mathcal{H} = \mathcal{B}_\Omega \quad \text{(III.5)}$$

for some $\Omega \subset \{1, \ldots, N\}$, where

$$\mathcal{B}_\Omega = \{\mathbf{x} \mid \mathrm{supp}\ \widehat{\mathbf{x}} \subset \Omega\}. \quad \text{(III.6)}$$

The graph bandlimited spaces (GBLSs) $\mathcal{B}_\Omega$, also known as Paley-Wiener spaces, have been widely used in GSP; see ([17], [18], [19], [20], [21], [22]) and references therein. In the classical real-line setting, a bandlimited space is a principal shift-invariant space and a principal shift-invariant space is shift-invariant, while the converse does **not** hold in general. In the following theorem, we show that the terminologies on bandlimitedness, shift-invariance, principal-shift-invariance are essentially the same in the undirected finite graph setting.

Theorem III.1. *Let $\mathcal{G}$ be an undirected finite graph, and $\mathbf{S}$ be a graph shift satisfying Assumption II.1, and $\mathcal{H}$ be a linear space of graph signals on the graph $\mathcal{G}$. The following statements are equivalent to each other:*

(i) *$\mathcal{H} = \mathcal{B}_\Omega$ for some $\Omega \subset \{1, \ldots, N\}$.*
(ii) *$\mathcal{H}$ is shift-invariant.*
(iii) *$\mathcal{H}$ is a finitely-generated graph shift-invariant space (FGSIS).*
(iv) *$\mathcal{H}$ is a principal graph shift-invariant space (PGSIS).*

We divide the proof into the following steps: (iv)$\Longrightarrow$(iii)$\Longrightarrow$(ii) $\Longrightarrow$(i) $\Longrightarrow$(iv); see Section VII-A for the detailed proof. In the proof of the implication (ii) $\Longrightarrow$(i), we show that a GSIS $\mathcal{H}$ is the bandlimited space $\mathcal{B}_\Omega$ with

$$\Omega = \cup_{\mathbf{x} \in \mathcal{H}} \mathrm{supp}\ \widehat{\mathbf{x}} \ \text{ and } \ \#\Omega = \dim \mathcal{H}, \quad \text{(III.7)}$$

where the second conclusion on the cardinality of the supporting set Ω holds as the bandlimited space $\mathcal{B}_\Omega$ has dimension $\#\Omega$.

For the case that $\mathcal{H}$ is a GSIS generated by Φ, we obtain from (II.4) that for any $m \in \mathbb{Z}_+$ and $\phi \in \Phi$, the Fourier transform of $\mathbf{S}^m \phi$ has its support contained in the supporting set of $\widehat{\phi}$, i.e.,

$$\mathrm{supp}\ \mathcal{F}(\mathbf{S}^m \phi) = \mathrm{supp}\ \mathbf{\Lambda}^m \widehat{\phi} \subset \mathrm{supp}\ \widehat{\phi}.$$

Therefore the set Ω in the bandlimited space $\mathcal{H} = \mathcal{B}_\Omega$ is completely determined by the generator Φ,

$$\Omega = \cup_{\phi \in \Phi} \mathrm{supp}\ \widehat{\phi}. \quad \text{(III.8)}$$

The implication (i) $\Longrightarrow$(iv) follows from the following theorem, see Section VII-B for the proof.

Theorem III.2. *Let $\mathcal{G}$ be an undirected finite graph, $\mathbf{S}$ be a graph shift satisfying Assumption II.1, and Ω be a nonempty subset of $\{1, \ldots, N\}$. Then there exists a graph signal $\phi_0 \in \mathcal{B}_\Omega$ satisfying*

$$\widehat{\phi}_0(n) \ne 0 \ \text{ if and only if } \ n \in \Omega, \quad \text{(III.9)}$$

such that

$$\mathcal{B}_\Omega = \mathrm{span}\{\mathbf{S}^m \phi_0 \mid 0 \le m \le \#\Omega - 1\}. \quad \text{(III.10)}$$

We remark that the requirement (III.9) for the generator ϕ_0 is also a necessary condition for (III.10) to hold. Let χ_Ω be the characteristic function on the set Ω whose n-th component takes value one for every $n \in \Omega$ and value zero for any $n \notin \Omega$. Similar to the sinc function for the classical bandlimited space on the real line, our illustrative example of the generator ϕ_0 in (III.10) for the bandlimited space $\mathcal{B}_\Omega$ is the graph signal $\mathcal{F}^{-1}\chi_\Omega$, the inverse graph Fourier transform of the characteristic function χ_Ω on the set Ω.

As a consequence of Theorems III.1 and III.2, every GSIS $\mathcal{H}$ is generated by finite shifts of some generator ϕ_0, i.e.,

$$\mathcal{H} = \mathrm{span}\{\mathbf{S}^m \phi_0 \mid 0 \le m \le M-1\} \quad \text{(III.11)}$$

holds for all $M \ge \dim \mathcal{H}$.

B. Uncertainty principle for principal graph shift-invariant spaces

The generator ϕ_0 of a given GSIS $\mathcal{H}$ selected in (III.9) is more like the classical sinc function in a bandlimited space, typically exhibiting large spatial support when the bandlimiting set Ω has small cardinality (or equivalently when the GSIS $\mathcal{H}$ has low dimension). As it is desirable to identify generators of a GSIS with localized support, a natural question is whether a GSIS can admit a generator that has localized support and yields a low-dimensional space. In this subsection, we address this question by establishing an uncertainty principle that quantifies the trade-off between the spatial support of the generator of a PGSIS and the dimension of the space; see Proposition III.3. The reader may refer to [54], [55], [56], [57], [58], [59] and references therein for various uncertainty principles on graphs.

For the orthogonal matrix $\mathbf{U} = [u_n(i)]_{1\le n\le N, i\in V}$ in (II.1), define

$$\|\mathbf{U}\|_\infty = \sup_{1\le n\le N, i\in V} |u_n(i)|$$

and

$$\|\mathbf{U}\|_\infty^* = \sup_{W\subset V, \Omega\subset\{1,\ldots,N\}} \Big\{ (\#W\#\Omega)^{-1/2} \,\Big|\, \sum_{i\in W, n\in\Omega} |u_n(i)|^2 \ge 1 \Big\}. \quad \text{(III.12)}$$

Then the following bound estimate for $\|\mathbf{U}\|_\infty^*$ holds:

$$N^{-1/2} \le \|\mathbf{U}\|_\infty^* \le \|\mathbf{U}\|_\infty \quad \text{(III.13)}$$

[54]. Here the first inequality is true as $\sum_{i\in W}\sum_{n\in\Omega}|u_1(i)|^2 = 1$ for the sets $W = V$ and $\Omega = \{1\}$, and the second inequality follows from

$$(\#W\#\Omega)^{-1} \le \frac{1}{\#W\#\Omega}\sum_{i\in W, n\in\Omega}|u_n(i)|^2 \le \|\mathbf{U}\|_\infty^2$$

for any pair of sets $W\subset V$ and $\Omega\subset\{1,\ldots,N\}$ satisfying $\sum_{i\in W, n\in\Omega}|u_n(i)|^2 \ge 1$. By the uncertainty principle in [54],

$$\#(\text{supp }\mathbf{x}) \times \#(\text{supp }\widehat{\mathbf{x}}) \ge \left(\|\mathbf{U}\|_\infty^*\right)^{-2} \quad \text{(III.14)}$$

hold for all nonzero graph signals $\mathbf{x}$.

For a nonzero graph signal ϕ_0, we obtain from (III.8) that the PGSIS generated by ϕ_0 is a GBLS $\mathcal{B}_\Omega$ with $\Omega = \text{supp }\widehat{\phi_0}$. Then by (III.14) and the above observation, we have the following uncertainty principle.

Proposition III.3. *Let the graph $\mathcal{G}$ and the graph shift $\mathbf{S}$ be as in Theorem III.2. Let ϕ_0 be a nonzero graph signal with its support denoted by W, and denote the PGSIS generated by ϕ_0 by $\mathcal{H}(\phi_0)$. Then*

$$\#W \times \dim\mathcal{H}(\phi_0) \ge (\|\mathbf{U}\|_\infty^*)^{-2} \quad \text{(III.15)}$$

where $\|\mathbf{U}\|_\infty^*$ is given in (III.12).

We finish this subsection with a remark on the optimality of the lower bound estimate (III.15) for the PGSIS generated by a localized signal on cycle graphs.

Remark III.4. Let $\mathcal{C}(N) = (V_N, E_N)$ be the cycle graph that has the vertex set $V_N = \{0, 1, \ldots N-1\}$ and the edge set $E_N = \{(i, i\pm 1 \mod N), i\in V_N\}$. Define $\mathbf{S} = (S(i,j))_{i,j\in V_N}$ with $S(i,j) = 1$ if $j = i$, $S(i,j) = -1/2$ if $i - j = \pm 1 \mod N$, and $S(i,j) = 0$ otherwise. One may verify that $\mathbf{S}$ is a symmetric graph shift. Denote the orthogonal matrix to diagonalize the graph shift $\mathbf{S}$ on the cycle graph $\mathcal{C}(N)$ by $\mathbf{U}_N$. For the above orthogonal matrix, we have

$$N^{-1/2} \le \|\mathbf{U}_N\|_\infty^* \le \|\mathbf{U}_N\|_\infty \le 2^{1/2}N^{-1/2}, \quad \text{(III.16)}$$

see [52, Section 6.1]. This indicates that the order $N^{-1/2}$ in the lower bound estimate in (III.13) is optimal for the orthogonal matrix $\mathbf{U}_N$.

Let ϕ_0 be the delta signal supported at vertex $\lfloor N/2\rfloor + 1$, where $\lfloor t\rfloor$ is the integral part of a real number t. One may verify that the GSIS space $\mathcal{H}(\phi_0)$ contains all graph signals $\mathbf{x} = [x(0), \ldots, x(N-1)]^T$ satisfying the following symmetry property:

$$x(\lfloor N/2\rfloor - i) = x(\lfloor N/2\rfloor + i),\ 1\le i\le N-1-\lfloor N/2\rfloor,$$

and it has dimension $\lfloor N/2\rfloor + 1$. Therefore, for the GSIS $\mathcal{H}(\phi_0)$ on the cycle graph $\mathcal{C}(N)$, the estimate in (III.15) becomes **accurate** when N is odd, and the difference between the left and right hand sides in (III.15) is at most one for even N.

IV. Shift-invariant graph reproducing kernel Hilbert spaces

Let $\mathbf{S}$ be a graph shift on an undirected graph $\mathcal{G}$ of order N. We say that a reproducing kernel Hilbert space $\mathcal{H}$ of graph signals on $\mathcal{G}$ is *shift-invariant* if it has a shift-invariant reproducing kernel $\mathbf{K}$, i.e.,

$$\mathbf{SK} = \mathbf{KS}. \quad \text{(IV.1)}$$

Shift-invariant graph RKHSs (SIGRKHSs) have been widely appreciated in machine learning, and their popularity can be ascribed to their simplicity to represent signals, flexibility to select kernels, and efficiency to learn signals with low computational costs. By Proposition II.2, a shift-invariant reproducing kernel $\mathbf{K}$ is a **polynomial filter** of the graph shift $\mathbf{S}$. Common selection of shift-invariant kernels include the diffusion kernels $\exp(\sigma^2\mathbf{L}^{\rm sym}/2)$ with $\sigma > 0$, the p-step random walk kernels $(a\mathbf{I} - \mathbf{L}^{\rm sym})^{-p}$ with $a > 2$ and $p \ge 1$, the Laplacian regularization kernels $\mathbf{I} + \sigma^2\mathbf{L}^{\rm sym}$ with $\sigma > 0$, the spline kernels $((\mathbf{L}^{\rm sym})^\dagger)^\alpha$ and their variants, where $(\mathbf{L}^{\rm sym})^\dagger$ is pseudo-inverse of the symmetrically normalized Laplacian $\mathbf{L}^{\rm sym}$ on the graph $\mathcal{G}$ ([9], [26], [27], [28], [29], [31], [33], [34], [35], [36], [37], [38]).

For a SIGRKHS $\mathcal{H}$ with a kernel $\mathbf{K}$, we can express any element in $\mathcal{H}$ as a linear combination of the columns of $\mathbf{K}$, i.e.,

$$\mathcal{H} = \left\{\mathbf{Kc},\ \mathbf{c}\in\mathbb{R}^N\right\}. \quad \text{(IV.2)}$$

The SIGRKHS with a spline kernel $((\mathbf{L}^{\rm sym})^\dagger)^p, p \ge 1$, is known as graph spline space in [36]. For the case that all eigenvalues of the symmetrically normalized Laplacian $\mathbf{L}^{\rm sym}$ are distinct (i.e., Assumption II.1 holds), one may verify that the SIGRKHS associated with the diffusion kernels, the p-step random walk kernels, and the Laplacian regularization kernels are the **whole** Euclidean space $\mathbb{R}^N$ with **different** inner products embedded.

By (IV.1) and (IV.2), we observe that a SIGRKHS $\mathcal{H}$ is a GSIS. In the following theorem, we show that the converse holds too, see Section VII-C for the proof.

Theorem IV.1. *Let $\mathcal{G}$ be an undirected finite graph, and $\mathbf{S}$ be a graph shift satisfying Assumption II.1. Then a GSIS $\mathcal{H}$ embedded with the standard Euclidean inner product is a SIGRKHS.*

From the proof of Theorem IV.1, we observe that the standard inner product $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^T\mathbf{y}$ of two graph signals $\mathbf{x}$ and $\mathbf{y} \in \mathcal{H}$ can be represented in the Fourier domain as follows:

$$\langle \mathbf{x}, \mathbf{y}\rangle = \widehat{\mathbf{x}}^T \mathbf{\Lambda}_\Omega \widehat{\mathbf{y}}, \tag{IV.3}$$

where $\Omega \subset \{1, \ldots, N\}$ is given in Theorem III.1 and $\mathbf{\Lambda}_\Omega = \mathrm{diag}\big[\chi_\Omega(1), \ldots, \chi_\Omega(N)\big]$ is the diagonal matrix with $\chi_\Omega(n) = 1$ for $n \in \Omega$ and $\chi_\Omega(n) = 0$ otherwise. In the following theorem, we show that the inner product of any SIGRKHS is a generalized dot product in the graph Fourier domain, see Section VII-D for the proof.

Theorem IV.2. *Let $\mathcal{G}$ be an undirected finite graph, and $\mathbf{S}$ be a graph shift satisfying Assumption II.1. Then a SIGRKHS $\mathcal{H}$ has its inner product defined by*

$$\langle \mathbf{x}, \mathbf{y}\rangle_{\mathcal{H}} = \widehat{\mathbf{x}}^T \mathbf{B} \widehat{\mathbf{y}}, \quad \mathbf{x}, \mathbf{y} \in \mathcal{H} \tag{IV.4}$$

for some diagonal matrix $\mathbf{B}$ with nonnegative diagonal entries.

Let $\mathbf{K}$ be the shift-invariant kernel of a SIGRKHS $\mathcal{H}$, $\Omega \subset \{1, \ldots, N\}$ be as in Theorem III.1 and $\mathbf{U}$ be the orthogonal matrix in (II.1). By (IV.1) and Proposition II.2, we observe that $\mathbf{K} = \mathbf{U}\mathbf{\Lambda}_{\mathbf{K}}\mathbf{U}^T$ for some diagonal matrix $\mathbf{\Lambda}_{\mathbf{K}}$ with its n-th entries taking positive value for $n \in \Omega$ and zero value otherwise. From the proof of Theorem IV.2, we observe that the pseudo-inverse $\mathbf{\Lambda}_{\mathbf{K}}^\dagger$ of the diagonal matrix $\mathbf{\Lambda}_{\mathbf{K}}$ can be used in (IV.4) to define the inner product of the SIGRKHS $\mathcal{H}$. In particular, one may verify that a matrix $\mathbf{B}$ with nonnegative diagonal entries can be used in (IV.4) if and only if $\mathbf{B} - \mathbf{\Lambda}_{\mathbf{K}}^\dagger$ has n-th diagonal entries taking zero value for all $n \in \Omega$.

Given a diagonal matrix $\mathbf{B}$ with nonnegative diagonal entries, one may verify that the bandlimited space $\mathcal{B}_{\Omega_{\mathbf{B}}} = \big\{\mathbf{x} \,|\, \mathrm{supp}\ \widehat{\mathbf{x}} \subset \Omega_{\mathbf{B}}\big\}$ embedded with the inner product in (IV.4) is a RKHS with the shift-invariant kernel $\mathbf{U}\mathbf{B}^\dagger\mathbf{U}^T$, where $\Omega_{\mathbf{B}}$ is the supporting set of diagonal entries of the diagonal matrix $\mathbf{B}$ and $\mathbf{B}^\dagger$ is the pseudo-inverse of the matrix $\mathbf{B}$.

Given the reproducing kernel space $\mathcal{H}$ with the inner product defined by (IV.4), we see that the evaluation functional is uniformly bounded, i.e.,

$$|x(i)| \le \|\mathbf{x}\|_2 = \|\widehat{\mathbf{x}}\|_2 \le \Big(\min_{b(n)\neq 0} b(n)\Big)^{-1/2} \|\mathbf{x}\|_{\mathcal{H}}, \ i \in V, \tag{IV.5}$$

hold for all graph signals $\mathbf{x} = [x(i)]_{i\in V} \in H$, where $\mathbf{B} = \mathrm{diag}[b(1), \ldots, b(N)]$ is the diagonal matrix in (IV.4).

V. Sampling and reconstruction in a finitely-generated graph shift-invariant space

Let $\mathbf{S}$ be a graph shift on an undirected graph $\mathcal{G}$ of order N, $\mathcal{H}(\Phi)$ be the GSIS generated by $\Phi := \{\phi_0, \ldots, \phi_{R-1}\}$. In this section, we consider the following linear sampling and reconstruction problem in the GSIS $\mathcal{H}(\Phi)$,

$$S: \ \mathbf{x} \longmapsto \mathbf{A}\mathbf{x}, \tag{V.1}$$

where $\mathbf{A} = [a(s,i)]_{s\in W, i\in V}$ is the sampling matrix, and we propose a Krylov-based algorithm for signal recovery.

An illustrative example of the sampling scheme (V.1) is the ideal sampling on the subset $W \subset V$, in which the sampling matrix is given by

$$\mathbf{A}_W = [a_W(i,j)]_{i\in W, j\in V} \tag{V.2}$$

where $a_W(i,j) = 0$ except that $a_W(i,i) = 1, i \in W$. The above sampling scheme is also known as *subset sampling* ([17], [18], [19], [20], [21], [22]). Another illustrative example of the sampling scheme (V.1) is *dynamic sampling* (also known as *aggregation sampling*),

$$S_{\mathbf{D}, i_0}: \ \mathbf{x} \longmapsto [\mathbf{x}(i_0), (\mathbf{D}\mathbf{x})(i_0), \ldots, (\mathbf{D}^{K-1}\mathbf{x})(i_0)] \tag{V.3}$$

where $i_0 \in V, K \ge 1$, $\mathbf{D}$ is the state matrix and $\mathbf{x}(i_0)$ is i_0-th component of a graph signal $\mathbf{x}$ ([22], [60], [61], [62], [40]).

For the sampling procedure in (V.1), we have the following characterization on its injectivity.

Theorem V.1. *Let $\mathcal{G}$ be an undirected finite graph of order N, $\mathbf{S}$ be a graph shift satisfying Assumption II.1, $\mathcal{H}(\Phi)$ be the GSIS generated by $\Phi := \{\phi_0, \ldots, \phi_{R-1}\}$, and set the frame matrix $\mathbf{F}_N = [\mathbf{S}^m\phi]_{0\le m\le N-1, \phi\in\Phi}$. Then the sampling procedure* (V.1) *with sampling matrix $\mathbf{A}$ is one-to-one if and only if $\mathbf{F}_N$ and $\mathbf{A}\mathbf{F}_N$ have the same rank.*

The characterization in the above theorem regarding admissible sampling procedure S is well established in the existing literature. For completeness of this paper, we restate it in our setting and provide a sketch of the proof; see Section VII-E for details.

Let $\Omega \subset \{1, \ldots, N\}$ and $\phi_0 \in \mathcal{B}_\Omega$ be a nonzero signal and its GFT $\widehat{\phi_0}$ has its n-th component taking nonzero value for all $n \in \Omega$. Then the GSIS $\mathcal{H}(\phi_0)$ generated by ϕ_0 is identical to the bandlimited space $\mathcal{B}_\Omega$. Combining the above observation with Theorem V.1, we obtain the following result on the uniqueness for ideal sampling on bandlimited spaces, which has been established in the literature; see [17], [18], [19], [21], [22] and references therein for some historical remarks on the subset sampling on (graph) bandlimited spaces.

Corollary V.2. *Let $\mathcal{G}$ be an undirected graph of order N, $\mathbf{S}$ be a graph shift satisfying Assumption II.1, and $\mathbf{U} = [u_n(i)]_{1\le n\le N, i\in V}$ be the orthogonal matrix in* (II.1). *Then for any $\Omega \subset \{1, \ldots, N\}$ and $W \subset V$, the ideal sampling of the bandlimited space $\mathcal{B}_\Omega$ on the set W is one-to-one if and only if the submatrix $[u_n(i)]_{n\in\Omega, i\in W}$ of the orthogonal matrix $\mathbf{U}$ has rank $\#\Omega$.*

For the dynamic sampling (V.3) with the state matrix $\mathbf{D}$ being commutative with the graph shift $\mathbf{S}$, one may verify that the corresponding sampling matrix is given by

$$\mathbf{A}_{\mathbf{D},i_0} = [(\lambda_{\mathbf{D}}(n))^k]_{0\le k\le K-1, 1\le n\le N} \times \mathrm{diag}[u_1(i_0), \ldots, u_N(i_0)]\mathbf{U},$$

where $\mathrm{diag}[\lambda_{\mathbf{D}}(1), \ldots, \lambda_{\mathbf{D}}(N)] = \mathbf{U}^T\mathbf{D}\mathbf{U}$, and $[u_1(i_0), \ldots, u_N(i_0)]$ is the i_0-th row of the orthogonal matrix. This together with Theorem V.1 yields the following result on the uniqueness of dynamic sampling scheme (V.3) on a GSIS; see [40], [41], [62] and references for dynamical sampling.

Corollary V.3. *Let $\mathcal{G}$ be an undirected graph of order N, $\mathbf{S}$ be a graph shift satisfying Assumption II.1, and $\mathcal{H}(\Phi)$ be the GSIS generated by Φ. Consider the dynamic sampling scheme $S_{\mathbf{D},i_0}$ on the GSIS $\mathcal{H}(\Phi)$ at location $i_0 \in V$ with the state matrix $\mathbf{D}$ commutative with graph shift. Then the sampling procedure in* (V.3) *is one-to-one if and only if $K \ge \#\Omega$, $\lambda_A(n), n \in \Omega$ are distinct, and $u_n(i_0) \ne 0$ for all $n \in \Omega$, where $\Omega \subset \{1, \ldots, N\}$ is given in Theorem III.1.*

Next consider signal reconstruction associated with the sampling scheme (V.1) on the GSIS $\mathcal{H}(\Phi)$, under the assumption that the sampling scheme (V.1) is one-to-one and the given observation data is the sampling of some signal $\mathbf{x}_0 \in \mathcal{H}(\Phi)$ corrupted by some random/deterministic noise $\boldsymbol{\epsilon}$,

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \boldsymbol{\epsilon}. \tag{V.4}$$

Let $\Omega \subset \{1, \ldots, N\}$ be the set in Theorem III.1 so that $\mathcal{H}(\Phi) = \mathcal{B}_\Omega$ and set $\mathbf{U}_\Omega = [\mathbf{u}_n]_{n\in\Omega}$. Considering the GSIS $\mathcal{H}(\Phi)$ as a bandlimited space, we may use

$$\mathbf{x}_0^\sharp = \mathbf{U}_\Omega(\mathbf{U}_\Omega^T\mathbf{A}^T\mathbf{A}\mathbf{U}_\Omega)^{-1}\mathbf{U}_\Omega^T\mathbf{A}^T\mathbf{y}, \tag{V.5}$$

a solution of the minimization problem

$$\mathbf{x}_0^\sharp = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{H}(\Phi)} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2, \tag{V.6}$$

as the reconstructed signal in $\mathcal{H}(\Phi)$. We remark that $\mathbf{U}_\Omega^T\mathbf{A}^T\mathbf{A}\mathbf{U}_\Omega$ is invertible by the assumption that the sampling scheme (V.1) is one-to-one and its characterization in Theorem V.1. For the numerical implementation of the reconstruction algorithm (V.5), we observe that $\mathbf{U}_\Omega^T\mathbf{A}^T\mathbf{y}$ is essentially the restriction of the Fourier transform of $\mathbf{A}^T\mathbf{y}$ onto the set Ω, and hence we may consider the reconstruction formula (V.5) for the sampling scheme (V.1) as an implementation in the graph Fourier domain. Applying the conventional reconstruction formula (V.5) in the bandlimited framework requires three computationally expensive steps: (i) computing the eigendecomposition (II.1) of the graph shift $\mathbf{S}$; (ii) identifying the support set Ω of the bandlimited space associated with the GSIS space $\mathcal{H}(\Phi)$; and (iii) inverting the matrix $\mathbf{U}_\Omega^T\mathbf{A}^T\mathbf{A}\mathbf{U}_\Omega$. Together, the three steps incur a total computational cost of $O(N^3)$. Here we say that $A = O(B)$ for quantities A and B if $|A/B|$ is bounded by an absolute constant. Moreover, the formula becomes inapplicable if the matrix $\mathbf{U}_\Omega^T\mathbf{A}^T\mathbf{A}\mathbf{U}_\Omega$ is poorly conditioned, leading to numerical instability.

For $0 \le n \le N$, define

$$\mathcal{H}_n(\Phi) = \mathrm{span}\{\mathbf{S}^m\phi \mid 0 \le m \le n, \phi \in \Phi\}, \tag{V.7}$$

and

$$\langle \mathbf{x}_1, \mathbf{x}_2\rangle_{\mathbf{A}} = \mathbf{x}_1^T\mathbf{A}^T\mathbf{A}\mathbf{x}_2 \quad \text{and} \quad \|\mathbf{x}_1\|_{\mathbf{A}} = \sqrt{\langle \mathbf{x}_1, \mathbf{x}_1\rangle_{\mathbf{A}}} \tag{V.8}$$

for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$. We observe that the GSIS $\mathcal{H}(\Phi)$ generated by Φ could be approximated successively by nested Krylov subspaces,

$$\mathrm{span}\{\Phi\} = \mathcal{H}_0(\Phi) \subset \mathcal{H}_1(\Phi) \subset \cdots \subset \mathcal{H}_{N-1}(\Phi) = \mathcal{H}(\Phi). \tag{V.9}$$

Furthermore, due to the one-to-one property of the sampling scheme (V.1), Theorem V.1 ensures that the sampling-induced inner product in (V.8) is a valid inner product on the full GSIS space $\mathcal{H}(\Phi)$, as well as each nested Krylov subspace $\mathcal{H}_n(\Phi), 0 \le n \le N-1$. Based on the above two observations, we propose the following iterative algorithm with finite steps to recover the signal $\mathbf{x}_0^\sharp$ in (V.5); see Algorithm V.1.

(i) Initial Step: We apply the Gram-Schmidt process to construct an orthonormal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_{d_0}\}$

Algorithm V.1 Reconstruction algorithm for signals in a FGSIS

Input: N (order of the underlying graph $\mathcal{G}$), $\mathbf{S}$ (the graph shift), $\phi_0, \ldots, \phi_{R-1}$ (nonzero generators of the FGSIS), $\mathbf{A}$ (the sampling matrix), $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \boldsymbol{\epsilon}$ (the noisy observation of some unknown signal $\mathbf{x}_0 \in \mathcal{H}(\phi_0, \ldots, \phi_{R-1})$), and sampling error threshold δ.

Initial: Perform Gram-Schmidt process to $\phi_0, \ldots, \phi_{R-1}$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathbf{A}}$, and obtain an orthonormal basis $\mathbf{W}_0$. Set $d_0 = \#\mathbf{W}_0 = \dim \mathcal{H}_0(\Phi)$, $\mathbf{W}_0 = \{\mathbf{w}_1, \ldots, \mathbf{w}_{d_0}\}$ and $\widetilde{\mathbf{W}}_0 = \mathbf{W}_0$. Define $\tilde{\mathbf{x}}_0 = \sum_{m=1}^{d_0} \langle \mathbf{y}, \mathbf{A}\mathbf{w}_m \rangle \mathbf{w}_m$ and $\mathbf{e}_0 = \mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}_0$.

If $\|\mathbf{e}_0\|_2 \le \delta$, set $\mathbf{x}_{\text{out}} = \tilde{\mathbf{x}}_0$, $\mathbf{e}_{\text{final}} = \mathbf{e}_0$, $D = 0$, and stop, else do the following iteration.

Iteration:

for $n = 1 : N-1$

a) Set $\mathbf{v}_{m'} = \mathbf{S}\mathbf{w}_{m'} - \sum_{m=1}^{d_{n-1}} \langle \mathbf{S}\mathbf{w}_{m'}, \mathbf{w}_m \rangle_{\mathbf{A}} \mathbf{w}_m$ for all $\mathbf{w}_{m'} \in \widetilde{\mathbf{W}}_{n-1}$.

b) If $\mathbf{v}_{m'} = 0$ for all $\mathbf{w}_{m'} \in \widetilde{\mathbf{W}}_{n-1}$, then set $\mathbf{x}_{\text{out}} = \tilde{\mathbf{x}}_{n-1}, \mathbf{e}_{\text{final}} = \mathbf{e}_{n-1}$, $D = n-1$ and stop.

c) Else perform Gram-Schmidt process to $\{\mathbf{v}_{m'} \mid \mathbf{w}_{m'} \in \widetilde{\mathbf{W}}_{n-1}\}$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathbf{A}}$, and obtain an orthonormal basis $\widetilde{\mathbf{W}}_n$.

d) Set $d_n = d_{n-1} + \#\widetilde{\mathbf{W}}_n$, write $\widetilde{\mathbf{W}}_n = \{\mathbf{w}_{d_{n-1}+1}, \ldots, \mathbf{w}_{d_n}\}$, and update $\mathbf{W}_n = \mathbf{W}_{n-1} \cup \widetilde{\mathbf{W}}_n$.

e) Set $\mathbf{z}_n = \sum_{m=d_{n-1}+1}^{d_n} \langle \mathbf{e}_{n-1}, \mathbf{A}\mathbf{w}_m \rangle \mathbf{w}_m$, $\tilde{\mathbf{x}}_n = \tilde{\mathbf{x}}_{n-1} + \mathbf{z}_n$ and $\mathbf{e}_n = \mathbf{e}_{n-1} - \mathbf{A}\mathbf{z}_n$.

f) If $\|\mathbf{e}_n\|_2 \le \delta$, set $\mathbf{x}_{\text{out}} = \tilde{\mathbf{x}}_n, \mathbf{e}_{\text{final}} = \mathbf{e}_n$, $D = n$ and stop, else continue the loop.

end

Output: The reconstruction signal $\mathbf{x}_{\text{out}}$, the sampling error $\|\mathbf{e}_{\text{final}}\|_2$, and the Krylov subspace level D to represent the reconstructed signal $\mathbf{x}_{\text{out}}$.

for the Krylov subspace $\mathcal{H}_0(\Phi)$, which is spanned by $\phi_0, \ldots, \phi_{R-1}$. The reconstructed signal $\tilde{\mathbf{x}}_0$ is the optimal approximation within $\mathcal{H}_0(\Phi)$ that best matches the noisy sampling data $\mathbf{y}$, i.e., it solves the least-squares problem $\tilde{\mathbf{x}}_0 = \arg\min_{\mathbf{x} \in \mathcal{H}_0(\Phi)} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$.

(ii) Step a) of the n-th iteration: Denote the projection operator onto the Krylov space $\mathcal{H}_n(\Phi)$ by $\mathbf{P}_n, 0 \le n \le N-1$. The vectors $\mathbf{v}_{m'} = \mathbf{S}\mathbf{w}_{m'} - \mathbf{P}_{n-1}\mathbf{S}\mathbf{w}_{m'}$ with $\mathbf{w}_{m'} \in \widetilde{\mathbf{W}}_{n-1}$, are constructed to form a spanning set for the orthogonal complement $\mathcal{H}_n(\Phi) \cap \mathcal{H}_{n-1}(\Phi)^\perp$ of the Krylov space $\mathcal{H}_{n-1}(\Phi)$ at level $n-1$ in the Krylov space $\mathcal{H}_n(\Phi)$ at level n.

(iii) Step b) of the n-th iteration: The condition in step b) asserts that $\mathcal{H}_n(\Phi) \cap \mathcal{H}_{n-1}(\Phi)^\perp = \{0\}$, which implies that the Krylov subspaces have stabilized in the sense that

$$\mathcal{H}_{n-1}(\Phi) = \mathcal{H}_m(\Phi) = \mathcal{H}(\Phi), \ n \le m \le N-1.$$

Thus the solution $\mathbf{x}_{n-1}$ in the $(n-1)$-th iteration solves the original least-squares problem (V.6).

(iv) Steps c) and d) of the n-th iteration: The condition for steps c) and d) is that $\mathcal{H}_n(\Phi) \cap \mathcal{H}_{n-1}(\Phi)^\perp \ne \{0\}$. In step c) and d), we perform the Gram-Schmidt procedure to obtain an orthonormal basis $\{\mathbf{w}_{d_{n-1}+1}, \ldots, \mathbf{w}_{d_n}\}$ for the orthogonal complement $\mathcal{H}_n(\Phi) \cap \mathcal{H}_{n-1}(\Phi)^\perp$. As a consequence, we obtain that $d_n = \dim \mathcal{H}_n(\Phi)$ and $W_m = \{\mathbf{w}_m, 1 \le m \le d_n\}$ is an orthonormal basis of $\mathcal{H}_n(\Phi)$.

(v) Step e) of the n-th iteration: The correction term $\mathbf{z}_n$ and the updated signal $\tilde{\mathbf{x}}_n$ are obtained as solutions to the following least-squares problems:

$$\mathbf{z}_n := \underset{\mathbf{w} \in \mathcal{H}_n(\Phi) \cap \mathcal{H}_{n-1}(\Phi)^\perp}{\operatorname{argmin}} \|\mathbf{e}_{n-1} - \mathbf{A}\mathbf{w}\|_2$$

and

$$\tilde{\mathbf{x}}_n := \underset{\mathbf{x} \in \mathcal{H}_n(\Phi)}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2,$$

and the residue term $\mathbf{e}_n$ is given by $\mathbf{e}_n := \mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}_n$.

As a consequence of the above observations, the output $\mathbf{x}_{\text{out}} \in \mathcal{H}(\Phi)$ of the Algorithm V.1 is either a graph signal to match the noisy sampling data $\mathbf{y}$ within the threshold δ in the sense that $\|\mathbf{y} - \mathbf{A}\mathbf{x}_{\text{out}}\|_2 \le \delta$, when the iteration terminates at step f) or during the initial step; or the graph signal $\mathbf{x}_0^\sharp$ in (V.5) achieves the minimal approximation error over the GSIS $\mathcal{H}(\Phi)$,

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}_{\text{out}}\|_2 = \min_{\mathbf{x} \in \mathcal{H}(\Phi)} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2,$$

when the iteration terminates at step b).

Comparing with the reconstruction formula (V.5) for the sampling scheme (V.1), we may consider the proposed Algorithm V.1 as a reconstruction scheme in the **spatial** domain. It can be implemented without prior knowledge of the support set Ω. A signal residing in the GSIS $\mathcal{H}(\Phi)$ will be reconstructed with an error tolerance δ, which is typically chosen to reflect both the sampling noise and the signal's deviation from the GSIS $\mathcal{H}(\Phi)$. Moreover, for spatially localized signals, as discussed in Section VI-A, the iterative process is **self-adaptive** and could terminate early, in which case the reconstructed signal involves only a few nonzero coefficients corresponding to graph shift operations of the generators. This efficiency stems from the fact that, in such scenarios, Krylov subspace approximations yield accurate reconstructions with minimal computational overhead. Consider the scenario where the graph shift operation $\mathbf{S}$ and the sampling operation $\mathbf{A}$ each have computational cost $O(N)$, which holds when the underlying graph has bounded degree and the sampling matrix is spatially localized. Under the above assumptions, if the Krylov-based algorithm V.1 terminates at early step n_0, its total cost for signal recovery is $O(m_0^2 N)$, where $m_0 = \dim \mathcal{H}_{n_0}(\Phi) \le n_0 \# \Phi$. Therefore when the original signal is well approximated by the Krylov spaces (hence m_0 is much smaller than the graph order N), the computational cost for signal recovery is far lower than the $O(N^3)$ cost of the direct spectral approach in (V.5).

VI. Numerical Simulations

In this section, we demonstrate that GSIS may provide a natural and efficient framework for modeling datasets with strong localization pattern that can be well-approximated by Krylov subspaces of the GSIS. In addition, we evaluate the performance of Algorithm V.1 in consideration of the sampling and reconstruction problem (V.1) for damped cosine wave signals in a GSIS on the cycle graph from their noisy samples and for flight delay dataset of the 50 busiest airports in the USA. Also in this section we compare the performance on model suitability and data fitting capability for the GSIS, bandlimiting and RKHS models. For the cycle graph simulations in Section VI-A, we aim to illustrate the performance of the proposed Krylov-subspace-based algorithm to reconstruct signals in a GSIS, while for the flight-delay experiment in Section VI-B, we focus on model suitability and data fitting across the national air traffic network.

A. GSIS signal reconstruction on cycle graphs

Let the cycle graph $\mathcal{C}(N)$ and the graph shift $\mathbf{S}$ be as in Remark III.4. Take the delta signal ϕ_0 at the vertex $\lfloor N/2 \rfloor$ and consider the GSIS $\mathcal{H}(\phi_0)$ generated by ϕ_0. We observe that the GSIS $\mathcal{H}(\phi_0)$ admits the successive Krylov subspace approximation,

$$\mathrm{span}\{\phi_0\} = \mathcal{H}_0(\phi_0) \subset \cdots \subset \mathcal{H}_{N-1}(\phi_0) = \mathcal{H}(\phi_0),$$

where Krylov subspaces $\mathcal{H}_n(\phi_0), n \ge 0$, are given by

$$\mathcal{H}_n(\phi_0) = \mathrm{span}\{\mathbf{S}^m \phi_0 \mid 0 \le m \le n\}. \tag{VI.1}$$

On the cycle graph $\mathcal{C}(N)$, one may verify that the Krylov subspaces $\mathcal{H}_n(\phi_0)$ in (VI.1) have dimension $n+1$ for $0 \le n \le \lfloor N/2 \rfloor$, and $\mathcal{H}_n(\phi_0) = \mathcal{H}(\phi_0) \ne \mathbb{R}^N$ for all $n \ge \lfloor N/2 \rfloor$.

In this subsection, we consider sampling and reconstruction of the damped cosine wave signals $\mathbf{x}_0 = [x_0(i)]_{0 \le i \le N-1}$ on the cycle graph $\mathcal{C}(N)$, where

$$x_0(i) = A e^{-\lambda |i - \lfloor N/2 \rfloor|} \cos\left(\omega |i - \lfloor N/2 \rfloor|\right),\ 0 \le i \le N-1, \tag{VI.2}$$

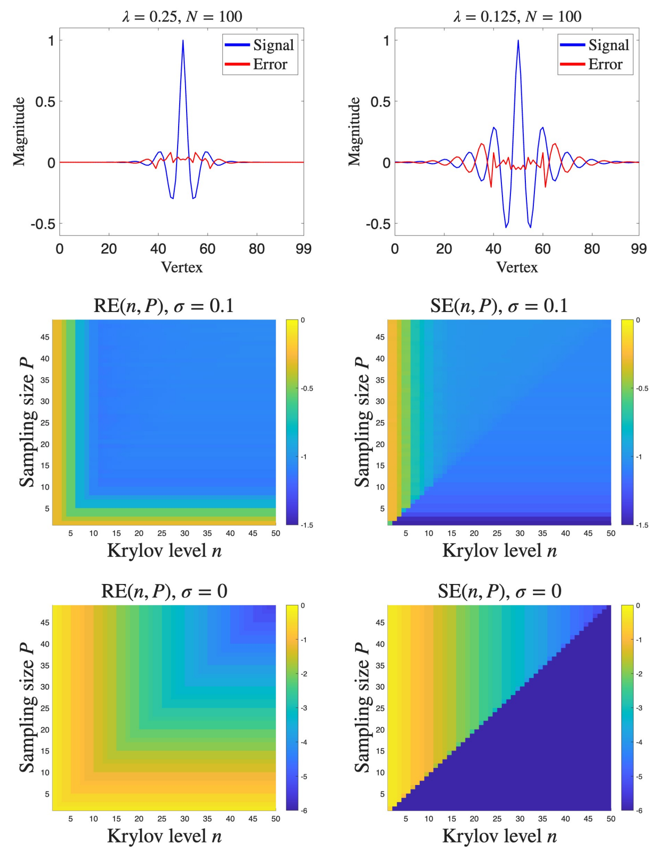

Fig. 1. Plotted on the top are the damped cosine wave signals $\mathbf{x}_0$ (in blue) and the reconstruction errors $\mathbf{x}_{n,W} - \mathbf{x}_0$ (in red), where $n = 10, N = 100, P = \lfloor N/6 \rfloor = 16, \sigma = 0.1$, and $(A, \lambda, \omega) = (1, 1/4, 2\pi/10)$ (top left) and $(1, 1/8, 2\pi/10)$ (top right). The relative maximal reconstruction error $\mathrm{RE}(10, 16)$ and the relative maximal sampling error $\mathrm{SE}(10, 16)$ are $(0.0799, 0.0876)$ (top left) and $(0.2046, 0.2379)$ (top right) respectively. Presented at the middle and bottom rows are the average of the relative maximal signal reconstruction error $\mathrm{RE}(n, P)$ (left) and the relative maximal sampling error $\mathrm{SE}(n, P)$ (right) over $M = 100$ trials, where $1 \le n \le 50, 1 \le P \le 45$, $\mathbf{x}_0$ is the damped cosine wave signal with $(A, \lambda, \omega) = (1, 1/4, 2\pi/10)$ and the noise level $\sigma = 0.1$ (the middle row) and $\sigma = 0$ (the bottom row).

A is the amplitude, λ is the decay constant, and ω is the angular frequency. The above wave signals belong to the GSIS space $\mathcal{H}(\phi_0)$ and are well-localized and symmetric around the vertex $\lfloor N/2 \rfloor$; see the top plots of Figure 1 in blue, where $N = 100$, $\omega = 2\pi/10$, and $\lambda = 0.25$ (left) and 0.125 (right). Moreover, one may verify that maximal relative approximation errors from the Krylov subspaces $\mathcal{H}_n(\phi_0)$ have exponential decay,

$$E_n := \inf_{\mathbf{x}_n \in \mathcal{H}_n(\phi_0)} \frac{\|\mathbf{x}_0 - \mathbf{x}_n\|_\infty}{\|\mathbf{x}_0\|_\infty} \le e^{-(n+1)\lambda},\ 1 \le n \le \lfloor N/2 \rfloor. \tag{VI.3}$$

In this subsection, we consider the subset sampling scheme (V.2) with the sampling set W being symmetric around $\lfloor N/2 \rfloor$, and the observation data $\mathbf{y}$ is corrupted by some uniform random noises,

$$\mathbf{y}(i) = \mathbf{x}_0(i) + \boldsymbol{\epsilon}(i), \ i \in W_P,$$

where $W_P = \{\lfloor N/2 \rfloor - P, \dots, \lfloor N/2 \rfloor + P\}$ for some $P \geq 1$, and $\boldsymbol{\epsilon}(i), i \in W_P$, are i.i.d. random variables with uniform distribution on $[-\sigma, \sigma]$ for some noise level $\sigma > 0$.

In our simulations, we use average of the relative maximal signal reconstruction error in the logarithmic scale

$$\mathrm{RE}(n, P) = \log_{10}\left(\frac{\|\mathbf{x}_{n,W_P} - \mathbf{x}_0\|_\infty}{\|\mathbf{x}_0\|_\infty} + 10^{-6}\right),$$

and the relative maximal sampling error in the logarithmic scale,

$$\mathrm{SE}(n, P) = \log_{10}\left(\frac{\|\mathbf{A}_{W_P}(\mathbf{x}_{n,W_P} - \mathbf{x}_0)\|_\infty}{\|\mathbf{A}_{W_P}\mathbf{x}_0\|_\infty} + 10^{-6}\right)$$

over M trials to measure the performance of the Algorithm V.1, where $M \geq 1$, $\mathbf{A}_{W_P}$ is the sampling matrix, and $\mathbf{x}_{n,W_P} \in \mathcal{H}_n(\phi_0), n \geq 0$, are signals reconstructed by Algorithm V.1 in its n-th iteration.

Presented in Figure 1 are the performances of Algorithm V.1 to reconstruct the damped cosine wave signals in (VI.2) from their noiseless and noisy samples. For the noiseless scenario illustrated in the bottom row of Figure 1, we observe the following: 1) The relative maximal signal reconstruction error and the relative maximal sampling error measured without logarithmic scaling decrease as the Krylov level n increases and become small once the Krylov subspace is sufficiently rich to represent the sampled data on W_P; 2) For any fixed sampling size $1 \leq P \leq \lfloor N/2 \rfloor$, those errors become small once the Krylov level reaches the sampling radius, namely $n \geq P$; and 3) For any fixed Krylov level $n \leq \lfloor N/2 \rfloor$, increasing the sampling size beyond $P = n$ does not change the reconstructed signal and only affects the sampling error through the exponentially small tail of the damped cosine signal. These observations are consistent with the Krylov structure in (VI.1). Starting from the delta signal, the subspace $\mathcal{H}_n(\phi_0)$ expands locally one layer at a time on the cycle graph. Hence, for any given sampling size P, the Krylov subspace of order $n_0 = P$, instead of the whole GSIS, is used in Algorithm V.1. This choice enables effective signal recovery with low computational cost while preserving accuracy; see the middle and bottom rows of Figure 1.

Our numerical results also show that the maximal relative signal reconstruction error $\|\mathbf{x}_{n_0,W_P} - \mathbf{x}_0\|_\infty / \|\mathbf{x}_0\|_\infty$ is proportional to the best maximal approximation error E_n in (VI.3). The possible reasons we believe are that signals in the Krylov subspace $\mathcal{H}_n(\phi_0)$ are supported in $[\lfloor N/2 \rfloor - n, \lfloor N/2 \rfloor + n]$ and they can be fully recovered from their samples on vertices in that interval.

For the noisy scenario shown in the middle row of Figure 1, we observe that the relative maximal signal reconstruction error $\mathrm{RE}(n, P)$ and the relative maximal sampling error $\mathrm{SE}(n, P)$ have the same monotonic property about the Krylov level n and the sampling size P, however they do not improve significantly when $P \geq 13$ and $n \geq 10$. The possible reason is that the damped cosine wave function $\mathbf{x}_0 = [x_0(i)]_{0 \leq i \leq N-1}$ is well-localized around $\lfloor N/2 \rfloor$ with $\sup_{|i - \lfloor N/2 \rfloor| \geq 13} |x_0(i)| = 0.0244 < 0.1 = \sigma$ and its noisy samples at vertices i with $|i - \lfloor N/2 \rfloor| \geq 13$ are essentially dominated by random noise at level $\sigma = 0.1$.

B. Flight-delay dataset and graph shift-invariant spaces

Flight delays are inevitable especially for frequent flyers. The three most common causes of flight delays are extreme weather conditions, air traffic congestion, and operational issues by airlines. These factors can trigger widespread delays at key airports, particularly within specific regions or airline networks, and often initiate ripple effects that propagate throughout the broader air transport system. In this subsection, we demonstrate that GSIS and its Krylov subspace approximation provide a natural and efficient framework for modeling flight delay performance and for capturing the ripple effect phenomenon across the top 50 busiest airports among the 323 highest-traffic airports in the United States [35], [63]. Our justification for selecting the 50 busiest airports is that those airports generate observable ripple effects on flight delay, while the remaining airports (most of them are regional airports) show only limited delay propagation across the national air traffic network.

In our simulations, we take the dataset on July, August, and September in 2014 and 2015 for total 184 days; see the top left plot of Figure 2. We model the flight delay dataset (measured in minutes) as a family of graph signals $\mathbf{x}_d = [x_d(j)]_{j \in A}, 1 \leq d \leq 184$, on the undirected graph $\mathcal{F} = (A, D)$, where d indicates the date in the dataset, the vertex set A has vertices indicating the top 50 airports and the edge set D contains all airport pairs so that the number of mutual flights between them exceeds 100 in the three months (roughly one direct flight between them per day). The underlying graph is designed to ensure that our analysis focuses on the busiest hubs, which are likely to have a greater influence on overall air traffic patterns and network dynamics. In the simulations, we use the corresponding symmetrically normalized graph Laplacian $\mathbf{L}^{\mathrm{sym}}$ as the graph shift in the definition of our proposed GSIS, bandlimiting and RKHS models.

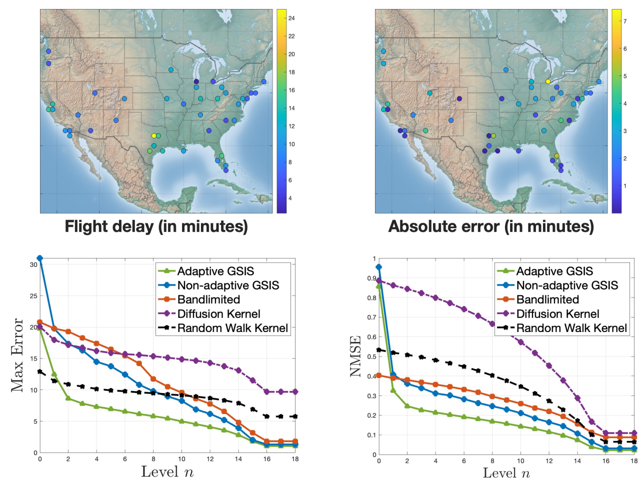

Fig. 2. Plotted on the top are the flight delay data $\mathbf{x}_d$ (top left) and absolute error data $|\mathbf{x}_{\mathcal{H};d,n} - \mathbf{x}_d|$ (top right) across the top 50 US airports on 29 August 2014, except Honolulu International airport (HNL) and Luis Munoz Marin International airport (SJU), where $d = 60$ and $\mathbf{x}_{K;d,n}$ is reconstructed from Algorithm V.1 in the noiseless environment with adaptive GSIS model and Krylov level $n = 2$. Shown on the bottom are the average performance over 184 days: maximal error in minutes (bottom left) and normalized mean square error (bottom right), comparing five modeling approaches via adaptive GSIS (in green), non-adaptive GSIS (in blue), bandlimited space (in red), and RKHS with diffusion kernel $\mathbf{K}_G = \exp(\mathbf{L}^{\rm sym}/2)$ (in purple) and with 2-step random walk kernel $\mathbf{K}_R = (4\mathbf{I} - \mathbf{L}^{\rm sym})^{-2}$ (in black). The sampling set W in this figure contains the whole 50 airports, and the maximal error in the top right plot is $\|\mathbf{x}_{\mathcal{H};d,n} - \mathbf{x}_d\|_\infty = 7.4253$.

In our simulations, we consider the sampling procedure over a subset W of size $10 \le m \le 50$ that are randomly selected from the 50 airports, and denote the corresponding sampling matrix in (V.1) by $\mathbf{A}_W$. We conduct simulations under the full sampling scenario (i.e., $m = 50$ for which $\mathbf{A}_W = \mathbf{I}$) to demonstrate that GSIS, when equipped with appropriately selected generators, provides a more effective modeling framework for the flight delay dataset than both the bandlimiting and RKHS approaches; see Figure 2 and its accompanying explanation. We also do experiments under the random sampling scenario (i.e., $m = 10, 20, 30, 40$) to evaluate the performance of the proposed Algorithm V.1, and to compare its data fitting capabilities against those of bandlimited and RKHS-based signal reconstruction methods; see Figure 3 and its accompanying explanation.

In our simulations, we consider nonadaptive GSIS

$$\mathcal{H} = {\rm span}\{(\mathbf{L}^{\rm sym})^n \delta_i \mid n \ge 0, 1 \le i \le 3\} \tag{VI.4}$$

and adaptive GSISs

$$\mathcal{H}_d = {\rm span}\{(\mathbf{L}^{\rm sym})^n \delta_{i,d} \mid n \ge 0, 1 \le i \le 3\},\ \ 1 \le d \le 184, \tag{VI.5}$$

to model graph signals $\mathbf{x}_d, 1 \le d \le 184$, where δ_i, $1 \le i \le 3$, are delta signals at vertices representing airports with the i-th highest average delay across the sampled dataset $W \subset V$, and similarly, $\delta_{i,d}$, $1 \le i \le 3$, are delta signals centered at vertices representing the airports with the i-th highest delays on each day $d, 1 \le d \le 184$. Denote the Krylov subspaces of the nonadaptive GSIS $\mathcal{H}$ in (VI.4) and the adaptive GSIS $\mathcal{H}_d$ in (VI.5) at level $0 \le n \le 49$ by $\mathcal{H}_n$ and $\mathcal{H}_{d,n}, 1 \le d \le 184$, respectively. In our simulations, we assess the model suitability and data fitting performance of GSISs for the flight delay dataset using two error metrics: the maximal error in minutes

$$\mathrm{ME}_{\mathcal{H},n} = \frac{1}{184}\mathbb{E}_W\left(\sum_{d=1}^{184} \|\mathbf{A}_W(\mathbf{x}_{\mathcal{H};d,n} - \mathbf{x}_d)\|_\infty\right) \tag{VI.6}$$

and the normalized mean square error (NMSE)

$$\mathrm{SE}_{\mathcal{H},n} = \left[\frac{1}{184}\mathbb{E}_W\left(\sum_{d=1}^{184} \frac{\|\mathbf{A}_W(\mathbf{x}_{\mathcal{H};d,n} - \mathbf{x}_d)\|_2^2}{\|\mathbf{A}_W \mathbf{x}_d\|_2^2}\right)\right]^{1/2}. \tag{VI.7}$$

Here $\mathbf{x}_{\mathcal{H};d,n}$ is the signal in $\mathcal{H}_{d,n}$ (respectively $\mathcal{H}_n$) reconstructed from Algorithm V.1 in the noiseless environment from the sampled data of $\mathbf{x}_d$ on W, and the expectation $\mathbb{E}_W$ is taken over randomly-selected sampling sets W of fixed size m over 100 trials.

Following the eigendecomposition (II.1), we write $\mathbf{L}^{\rm sym} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ and define GFT as in (II.3), where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{50}]^T$ is an orthogonal matrix and $\mathbf{\Lambda} = {\rm diag}[\lambda(1), \ldots, \lambda(50)]$ is a diagonal matrix satisfying $0 \le \lambda(1) \le \lambda(2) \le \cdots \le \lambda(50)$. Let $\mathcal{B}_{d,n}, n \ge 0$, be the subspaces spanned by the first d_n components $\mathbf{u}_1, \ldots, \mathbf{u}_{d_n}$ corresponding to the lowest frequencies and have dimension matching that of the corresponding Krylov subspace $\mathcal{H}_{d,n}$, i.e., $d_n = \dim \mathcal{B}_{d,n} = \dim \mathcal{H}_{d,n}$. Similar to (V.9), we have the following successive approximation property in bandlimiting setting:

$$\mathcal{B}_{d,0} \subset \mathcal{B}_{d,1} \subset \cdots \subset \mathcal{B}_{d,49} \subset \mathbb{R}^{50}. \tag{VI.8}$$

Similar to the error measurement in (VI.6) and (VI.7) in the GSIS model, we assess the model suitability and data fitting performance of bandlimited spaces using the maximal error in minutes

$$\mathrm{ME}_{\mathcal{B},n} = \frac{1}{184}\mathbb{E}_W\left(\sum_{d=1}^{184} \|\mathbf{A}_W(\mathbf{x}_{\mathcal{B};d,n} - \mathbf{x}_d)\|_\infty\right), \tag{VI.9}$$

and the NMSE

$$\mathrm{SE}_{\mathcal{B},n} = \left[\frac{1}{184}\mathbb{E}_W\left(\sum_{d=1}^{184} \frac{\|\mathbf{A}_W(\mathbf{x}_{\mathcal{B};d,n} - \mathbf{x}_d)\|_2^2}{\|\mathbf{A}_W \mathbf{x}_d\|_2^2}\right)\right]^{1/2}, \tag{VI.10}$$

where, as in Algorithm V.1, the reconstructed signal $\mathbf{x}_{\mathcal{B};d,n}$ in the bandlimited model is the solution of the following ridge regression problem:

$$\mathbf{x}_{\mathcal{B};d,n} = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{B}_{d,n}} \|\mathbf{x}_d - \mathbf{A}_W\mathbf{x}\|_2^2. \tag{VI.11}$$

Let $\mathcal{H}$ be a RKHS on the graph $\mathcal{F}$ with kernel $\mathbf{K} = (K(j,j'))_{j,j'\in A}$. We apply the kernel ridge regression framework in [24], [35], [38] for signal reconstruction from its sampling data $\mathbf{A}_W\mathbf{x}_d$,

$$\mathbf{x}_{K;d} = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{H}} \|\mathbf{A}_W(\mathbf{x}_d - \mathbf{x})\|_2^2 + \gamma\|\mathbf{x}\|_{\mathcal{H}}. \tag{VI.12}$$

In our simulations, we consider two families of kernels: the diffusion kernel $\mathbf{K} = \exp(\sigma^2\mathbf{L}^{\mathrm{sym}}/2)$ with $\sigma = 1$ and the p-step random walk kernel $\mathbf{K} = (a\mathbf{I} - \mathbf{L}^{\mathrm{sym}})^{-p}$ with $a = 4$ and $p = 2$, and set the regularization parameter to $\gamma = 0.05$. By the representer theorem, there exist coefficients $c_d(j), j \in W \subset A$, such that

$$\mathbf{x}_{\mathbf{K};d} = \sum_{j\in W} c_d(j)\mathbf{K}(\cdot, j). \tag{VI.13}$$

Following the error measures in (VI.6) and (VI.7), we assess the model suitability and data fitting performance of RKHS using the maximal error in minutes

$$\mathrm{ME}_{\mathbf{K},n} = \frac{1}{184}\mathbb{E}_W\left(\sum_{d=1}^{184} \|\mathbf{A}_W(\mathbf{x}_{\mathbf{K};d,n} - \mathbf{x}_d)\|_\infty\right), \tag{VI.14}$$

and the NMSE

$$\mathrm{SE}_{\mathbf{K},n} = \left[\frac{1}{184}\mathbb{E}_W\left(\sum_{d=1}^{184} \frac{\|\mathbf{A}_W(\mathbf{x}_{\mathbf{K};d,n} - \mathbf{x}_d)\|_2^2}{\|\mathbf{A}_W\mathbf{x}_d\|_2^2}\right)\right]^{1/2}, \tag{VI.15}$$

where the reconstructed signal $\mathbf{x}_{\mathbf{K};d,n}$ at level n is the best d_n-term approximation to the solution $\mathbf{x}_{d,W}$ of the kernel ridge regression (VI.12), i.e.,

$$\mathbf{x}_{\mathbf{K};d,n} = \sum_{j\in W_n} c_d(j)\mathbf{K}(\cdot, j),$$

where W_n contains sampling indices corresponding to the d_n-largest-magnitude coefficients $c_d(j), j \in W$.

Plotted at the bottom left and right of Figure 2 are the average maximal error and NMSE for the GSIS, bandlimiting, and RKHS models over 184 days respectively, where the sampling set W contains the whole 50 airports (i.e., $m = 50$) in which the sampling matrix $\mathbf{A}_W$ becomes the identity matrix $\mathbf{I}$. For $n \geq 16$, we observe that both the maximal error and NMSE stabilize across all models. The GSIS and bandlimiting models maintain maximal errors below two minutes—well within the U.S. Department of Transportation's 15-minute "on-time" threshold. In contrast, the RKHS model yields higher maximal errors: over five minutes for the 2-step random walk kernel and around ten minutes for the diffusion kernel. This discrepancy may stem from the lack of sparsity in the kernel ridge regression solution (VI.12), combined with the subspace dimension $d_n = 49$ on most days and 50 on few days for $n \geq 16$. Additional simulations show that the maximal error drops to near zero on days when $d_n = 50$, indicating that RKHS models with diffusion and random-walk kernels span the whole signal domain $\mathbb{R}^{50}$. These findings suggest that GSIS, bandlimited space, and RKHS may essentially coincide, even though GSIS and bandlimiting models yield sparser representation.

The selection of generators within a GSIS framework offers both model flexibility and practical interpretability. As shown in Figure 2 the maximal Krylov approximation error, at level $n = 2$ for the adaptive GSIS models using delta generators centered at up to three major hub airports with the highest delays, is less than 9 minutes across the national air traffic network, which is below the "on-time" threshold of 15 minutes. This outcome aligns with strong localization patterns of the flight-delay data set and also with operational intuition: delay ripple effects typically propagate over at most **two** subsequent flight legs, after which the aircraft either exits the network or realigns with its scheduled trajectory. Given our choice of graph shift and the generators in the adaptive GSIS, we conclude that the dominant sources of flight delay are the **top three** delay airports and their propagated **ripple effects** across the network.

Bandlimited spaces are well-suited for analyzing global smoothness and regularity in GSP. Graph reproducing kernel spaces with shift-invariant kernels offer a flexible and rigorous analytic framework for sampling and learning. However, they may lack the practical interpretability and representational efficiency offered by GSISs in some real-world datasets. As illustrated in the bottom plots of Figure 2, bandlimiting and RKHS-based approximations require a larger number of frequency components or kernel components to achieve comparable performance, which does not align with the ripple-effect dynamics typically observed in flight delay data. In contrast, GSIS models yield more compact representations with **meaningful** interpretations. Moreover, GSISs with adaptively selected generators demonstrate improved approximation quality, highlighting their efficiency to capture the structure of flight delay dynamics. A plausible explanation is that the flight delay data is more spatially localized and well-structured than concentrated in low-frequency components in the graph Fourier domain. This alignment makes GSIS models particularly advantageous for broader applications requiring localized modeling.

We also conduct a performance comparison of the GSIS, bandlimiting, and RKHS models for sampling and reconstruction of the flight delay dataset. Figure 3 presents the data fitting NMSE for each model, based on sampling sets W of size $m = 10, 20, 30, 40$ that are randomly selected from the total 50 airports, c.f. the bottom right plot of Figure 2 with $m = 50$. We observe that, the sampled data are well-fitted by the GSIS, bandlimiting and RKHS models across varying sample sizes, particularly once performance stabilizes. We also observe slower convergence for larger sampling sizes m, which may be attributed to the increased data volume required to be fitted. This, in turn, necessitates higher levels n for accurate reconstruction. Comparing GSIS, bandlimiting and RKHS models across all sampling scenarios, our proposed adaptive GSIS (in green) consistently outperforms all competing methods. This is particularly evident at low Krylov levels, where the adaptive method achieves a much faster error decay. This result underscores the importance of our adaptive approach, where selecting generators based on daily signal characteristics (the three busiest airports per day) captures the signal structure far more efficiently.

We remark that the sampling procedure defined in (V.1) is not one-to-one for the flight delay dataset if the sampling set W has size $m \le 40$, and hence sampling data fitting and signal reconstruction are distinct problems, c.f. the top right plot of Figure 1. As shown in Figure 3, the GSIS models exhibit strong performance in fitting the sampled data; however our simulations reveal that this does not necessarily translate to accurate signal reconstruction. This suggests that the limitations of these models to recover the original localized signal approximately from incomplete samples.

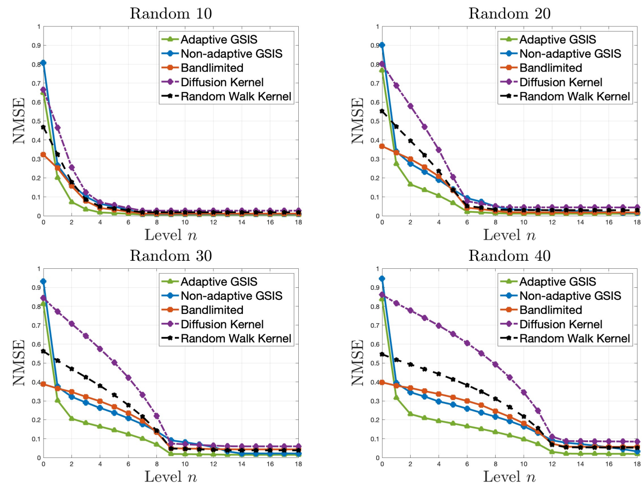

Fig. 3. Average normalized mean square error (NMSE) performance over 184 days of five modeling approaches with 100 trials: adaptive GSIS (in green), non-adaptive GSIS (in blue), bandlimited space (in red), and RKHS with diffusion kernel $\mathbf{K}_G = \exp(\mathbf{L}^{\rm sym}/2)$ (in purple) and with 2-step random walk kernel $\mathbf{K}_R = (4\mathbf{I} - \mathbf{L}^{\rm sym})^{-2}$ (in black). Here the random sampling set W has size $m = 10$ (top left), 20 (top right), 30 (bottom left) and 40 (bottom right).

VII. Proofs

In this section, we collect the proofs of Theorems III.1, III.2, IV.1, IV.2 and V.1.

A. Proof of Theorem III.1

The implications (iv)$\Longrightarrow$(iii)$\Longrightarrow$ (ii) follow from the definition of a FGSIS, and the implication (i)$\Longrightarrow$ (iv) holds by Theorem III.2 with its proof given in Section VII-B. Then it remains to prove the implication (ii)$\Longrightarrow$(i).

For $1 \le n \le N$, let $\mathbf{e}_n$ be the unit vector with zero entries except taking value one at n-th entries, and $\mathbf{E}_n$ be the diagonal matrix with $\mathbf{e}_n$ as the diagonal vector. By Assumption II.1, there exist polynomials $p_n, 1 \le n \le N$, of degree at most $N-1$ such that

$$p_n(\mathbf{S}) = \mathbf{U}\mathbf{E}_n\mathbf{U}^T,\ 1 \le n \le N, \tag{VII.1}$$

where $\mathbf{U}$ is the orthogonal matrix in (II.1). In particular, the polynomials $p_n, 1 \le n \le N$, are chosen to satisfy the interpolation property

$$p_n(\lambda(m)) = \delta_{nm},\ 1 \le m, n \le N, \tag{VII.2}$$

where δ_{nm} is the standard Kronecker symbol and $\lambda(m)$ are eigenvalues of the graph shift $\mathbf{S}$.

For $1 \le n \le N$, let $\mathbf{u}_n$ be the n-th column of the orthogonal matrix $\mathbf{U}$ in (II.1), and set

$$G_n = p_n(\mathbf{S})\mathcal{H},\ 1 \le n \le N. \tag{VII.3}$$

This together with the shift-invariance assumption on $\mathcal{H}$ implies that

$$G_n = \{\widehat{\mathbf{x}}(n)\mathbf{u}_n \mid \mathbf{x} \in \mathcal{H}\} \subset \mathcal{H},\ 1 \le n \le N, \tag{VII.4}$$

where $\widehat{\mathbf{x}}(n) = \mathbf{e}_n^T\widehat{\mathbf{x}}$ is the n-th component of the Fourier transform $\widehat{\mathbf{x}}$ of a signal $\mathbf{x} \in \mathcal{H}$. Therefore for any $1 \le n \le N$, either G_n is trivial or it has dimension one, i.e.,

$$\text{either}\quad G_n = \{0\}\quad \text{or}\quad G_n = \text{span}\{\mathbf{u}_n\}. \tag{VII.5}$$

Let $\Omega = \{n \mid G_n \ne \{0\}\}$, and set

$$\mathcal{B}_\Omega = \{\mathbf{x} \mid \text{supp}\ \widehat{\mathbf{x}} \subset \Omega\} = \bigoplus_{n\in\Omega} G_n. \tag{VII.6}$$

Since $G_n \subset \mathcal{H}$ for all $1 \le n \le N$, we have

$$\mathcal{B}_\Omega \subset \mathcal{H} \tag{VII.7}$$

by (VII.5) and (VII.6). On the other hand, for any $\mathbf{x} \in \mathcal{H}$, we obtained from (VII.3) and the definition of polynomials $p_n, 1 \le n \le N$, that

$$\mathbf{x} = \sum_{n=1}^{N} p_n(\mathbf{S})\mathbf{x} = \sum_{n\in\Omega} p_n(\mathbf{S})\mathbf{x} \in \mathcal{B}_\Omega.$$

This together with (VII.7) proves that $\mathcal{H} = \mathcal{B}_\Omega$. $\square$

B. Proof of Theorem III.2

Let $\lambda(n), 1 \le n \le N$, be the eigenvalues of the graph shift $\mathbf{S}$ and ϕ_0 be as in (III.9). Then it remains to verify that

$$\mathcal{B}_\Omega = \mathrm{span}\{\mathbf{S}^m \phi_0 \mid 0 \le m \le \#\Omega - 1\}. \tag{VII.8}$$

By (II.4) and (III.9), we see that the Fourier transforms of $\mathbf{S}^m\phi_0,\ 0 \le m \le \#\Omega - 1$, are supported on Ω and hence

$$\mathrm{span}\{\mathbf{S}^m \phi_0 \mid 0 \le m \le \#\Omega - 1\} \subset \mathcal{B}_\Omega. \tag{VII.9}$$

Let $q_n, n \in \Omega$, be univariate polynomials of degree at most $\#\Omega - 1$ that satisfy the interpolation property

$$q_n(\lambda(n')) = \delta_{nn'}, \quad n, n' \in \Omega, \tag{VII.10}$$

cf. (VII.2). The existence of such a polynomial follows from the distinct assumption on $\lambda(n'), n' \in \Omega$. Therefore

$$q_n(\mathbf{S}) = \mathbf{U} q_n(\mathbf{\Lambda}) \mathbf{U}^T \tag{VII.11}$$

with the diagonal matrix $q_n(\mathbf{\Lambda})$ having diagonal entries $q_n(\lambda(n')), 1 \le n' \le N$.

Take arbitrary bandlimited signal $\mathbf{x} \in \mathcal{B}_\Omega$. By (VII.10) and (VII.11), we have

$$\widehat{\mathbf{x}} = \sum_{n\in\Omega} \frac{\widehat{\mathbf{x}}(n)}{\widehat{\phi_0}(n)} q_n(\mathbf{\Lambda}) \widehat{\phi_0}.$$

Therefore taking the inverse Fourier transform $\mathcal{F}^{-1}$ at both sides yields

$$\mathbf{x} = \sum_{n\in\Omega} \frac{\widehat{\mathbf{x}}(n)}{\widehat{\phi_0}(n)} q_n(\mathbf{S}) \phi_0.$$

This together with the degree property for the polynomials $q_n, n \in \Omega$ proves that

$$\mathcal{B}_\Omega \subset \mathrm{span}\{\mathbf{S}^m \phi_0 \mid 0 \le m \le \#\Omega - 1\}. \tag{VII.12}$$

Combining (VII.9) and (VII.12) proves (VII.8) and hence completes the proof of Theorem III.2. $\square$

C. Proof of Theorem IV.1

By Theorem III.1, there exists $\Omega \subset \{1, \dots, N\}$ such that (III.5) holds. Define the kernel $\mathbf{K}$ by

$$\mathbf{K} = \mathbf{U} \mathbf{\Lambda}_\Omega \mathbf{U}^T, \tag{VII.13}$$

where $\mathbf{U}$ is the orthogonal matrix in (II.1) and $\mathbf{\Lambda}_\Omega$ is the diagonal matrix in (IV.3). Then we obtain from (II.1) and (VII.13) that the kernel $\mathbf{K}$ in (VII.13) is shift-invariant.

Let $\boldsymbol{\delta}_j, j \in V$, be delta signals taking value zero at all vertices except value one at the vertex j. By (III.5) and (VII.13), the linear space $\mathcal{H}$ is spanned by $\mathbf{K}\boldsymbol{\delta}_j, j \in V$, and satisfies the following reproducing kernel property, $x(j) = \langle \mathbf{x}, \mathbf{K}\boldsymbol{\delta}_j \rangle$, where the standard Euclidean inner product on $\mathbb{R}^N$ is used for its linear subspace $\mathcal{H} \subset \mathbb{R}^N$. This completes the proof of the conclusion that $\mathcal{H}$ is a RKHS with the shift-invariant kernel $\mathbf{K}$. $\square$

D. Proof of Theorem IV.2

Let $\mathbf{K}$ be the shift-invariant kernel of the RKHS $\mathcal{H}$, and let $\mathbf{x}, \mathbf{y}$ be two arbitrary elements in $\mathcal{H}$. By (IV.2) and the definition of the inner product on the RKHS $\mathcal{H}$,

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{H}} = \mathbf{c}^T \mathbf{K} \mathbf{d} \tag{VII.14}$$

for some $\mathbf{c}$ and $\mathbf{d} \in \mathbb{R}^N$ with $\mathbf{x} = \mathbf{K}\mathbf{c}$ and $\mathbf{y} = \mathbf{K}\mathbf{d}$.

By Assumption II.1 and Proposition II.2, there exists a diagonal matrix $\mathbf{\Lambda}$ for the shift-invariant kernel $\mathbf{K}$ such that $\mathbf{K} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U}$ is the orthogonal matrix in (II.1). Denote the pseudo-inverse of the diagonal matrix $\mathbf{\Lambda}$ by $\mathbf{\Lambda}^\dagger$. Then it follows from (VII.14) that $\widehat{\mathbf{x}} = \mathbf{\Lambda}\widehat{\mathbf{c}}, \widehat{\mathbf{y}} = \mathbf{\Lambda}\widehat{\mathbf{d}}$ and

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{H}} = \widehat{\mathbf{c}}^T \mathbf{\Lambda} \widehat{\mathbf{d}} = \widehat{\mathbf{c}}^T \mathbf{\Lambda}\mathbf{\Lambda}^\dagger \mathbf{\Lambda} \widehat{\mathbf{d}} = \widehat{\mathbf{x}}^T \mathbf{\Lambda}^\dagger \widehat{\mathbf{y}}.$$

As $\mathbf{K}$ is a reproducing kernel, it is represented by a positive semidefinite matrix, which implies that $\mathbf{\Lambda}^\dagger$ has nonnegative entries. Therefore (IV.4) holds and the proof is completed. $\square$

E. Proof of Theorem V.1

By (III.11), the columns of $\mathbf{F}_N$ form a frame for $\mathcal{H}(\Phi)$. Hence its rank is the same as the dimension of the GSIS $\mathcal{H}(\Phi)$ and also the cardinality of the supporting set Ω of the corresponding bandlimited space in Theorem III.1, i.e.,

$$\mathrm{rank}(\mathbf{F}_N) = \dim \mathcal{H}(\Phi) = \#\Omega. \tag{VII.15}$$

This together with the observation that dimension of the range space of the sampling operator S in (V.1) is the same as the rank of the matrix $\mathbf{A}\mathbf{F}_N$ proves the desired equivalence in Theorem V.1. $\square$

VIII. Conclusions and Future Study

Bandlimited spaces efficiently represent graph signals with spatial regularity or smooth spectral behavior, but the graph-Fourier basis often lacks clear spatial patterns on many real-world graphs and thus limits the descriptive power of purely spectral approaches. Graph reproducing kernel spaces with shift-invariant kernels offer a flexible analytic framework for sampling and learning. The GSIS introduced in this paper offers a powerful and interpretable approach to model localized signals on undirected finite graphs. Under suitable assumptions on the graph shift, we show that bandlimited spaces, shift-invariant RKHS, and GSIS are essentially equivalent from the perspective of functional analysis.

By leveraging the nested structure of Krylov subspaces, GSIS provides interpretable and compact representations of localized graph signals. On the flight delay dataset, it outperforms bandlimited spaces and RKHS models based on diffusion and random walk kernels in both model suitability and data fitting capability. This suggests that GSIS provides a natural and efficient framework for modeling datasets characterized by short-range diffusion dynamics and sparsity in both spatial and spectral domains.

The performance of the Krylov-based reconstruction algorithm V.1 depends on the appropriate selection of the graph shift $\mathbf{S}$, the generator Φ of the GSIS $\mathcal{H}(\Phi)$, and the original signal. For its optimal performance, the underlying graph should possess a modest degree (and thus it is sparse), the original signals should be well approximated by the low-level Krylov subspaces of the GSIS $\mathcal{H}(\Phi)$, and the sampling procedure should be numerically stable. We plan to investigate these aspects further in our future work.

Assumption II.1 plays an essential role in our development; however, it does not hold when the graph shift has repeated eigenvalues, as is the case for the Laplacian of a cycle graph and for many real-world graphs. The multiple–graph-shift framework provides substantially greater modeling flexibility than single-graph-shift setting in this paper, making it particularly effective for capturing directional flows, such as traffic, diffusion, or propagation, through more accurate and context-aware representations; see [37, Appendix A], [53], and references therein. In the forthcoming paper [51], we will address the challenges associated with removing the spectral simplicity assumption and extending the framework to the multiple–graph-shift setting.

Although the current GSIS formulation targets static graphs, we believe that it could be applicable for dynamical settings where shift and generators evolve over time. With appropriate design to capture temporal dynamics, GSIS can offer meaningful insights into localized time-varying signals, which we aim to explore in future work.

Acknowledgment: The authors thank the associate editor and anonymous reviewers for their constructive comments and insightful suggestions, which have significantly contributed to the improvement of this manuscript. In preparing the revision of this paper, we made use of Microsoft Copilot to assist with phrasing refinements and stylistic improvements.

REFERENCES

[1] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
[2] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.
[3] M. Unser, "Sampling-50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, 2000.
[4] K. Grochenig, *Foundations of Time-Frequency Analysis*, Birkhauser, Boston, 2001.
[5] A. Aldroubi and K. Grochenig, "Nonuniform sampling and reconstruction in shift-invariant spaces," *SIAM Review*, vol. 43, no. 4, pp. 585–620, 2001.
[6] A. Aldroubi, Q. Sun, and W. S. Tang, "p-frames and shift invariant subspaces of L^p," *J. Fourier Anal. Appl.*, vol. 7, pp. 1–21, 2001.
[7] A. Aldroubi, Q. Sun, and W.-S. Tang, "Convolution, average sampling, and a Calderon resolution of the identity for shift-invariant spaces," *J. Fourier Anal. Appl.*, vol. 11, pp. 215–244, 2005.
[8] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, pp. 1644–1656, 2013.
[9] 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, pp. 83–98, 2013.
[10] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, pp. 3042–3054, 2014.
[11] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.
[12] L. Stankovic, M. Dakovic, and E. Sejdic, "Introduction to graph signal processing," *Vertex-Frequency Analysis of Graph Signals*, Springer Cham, pp. 3–108, 2019.
[13] X. Dong, D. Thanou, L. Toni, M. Bronstein, and P. Frossard, "Graph signal processing for machine learning: A review and new perspectives," *IEEE Signal Process. Mag.*, vol. 37, pp. 117–127, 2020.
[14] A. Ortega, *Introduction to Graph Signal Processing*, Cambridge University Press, 2022.
[15] E. Isufi, F. Gama, D. I. Shuman, and S. Segarra, "Graph filters for signal processing and machine learning on graphs," *IEEE Trans. Signal Process.*, vol. 72, pp. 4745–4781, 2024.
[16] G. Leus, A. G. Marques, J. M. F. Moura, A. Ortega and D. I. Shuman, "Graph signal processing: History, development, impact, and outlook," *IEEE Signal Process. Mag.*, vol. 40, pp. 49-60, 2023.
[17] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.
[18] I. Pesenson, "Variational splines and Paley–Wiener spaces on combinatorial graphs," *Constr. Approx.*, vol. 29, no. 1, pp. 1–21, 2009.
[19] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, 2015.
[20] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, 2016.
[21] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Appl. Comput. Harmon. Anal.*, vol. 44, no. 2, pp. 446–475, 2018.
[22] C. Huang, Q. Zhang, J. Huang, and L. Yang, "Reconstruction of bandlimited graph signals from measurements," *Digit. Signal Process.*, vol. 101, article no. 102728, 13 pp., 2020.
[23] E. Isufi, P. Banelli, P. Di. Lorenzo, and G. Leus, "Observing and tracking bandlimited graph processes from sampled measurements," *Signal Process.*, vol. 177, article no. 107749, 2020.

[24] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, Massachusetts, 2002.

[25] G. Nikolentzos, G. Siglidis, and M. Vazirgiannis, “Graph kernels: a survey,” *J. Artif. Intell. Res.*, vol. 72, pp. 943–1027, 2021.

[26] R. S. Kondor and J. Lafferty, “Diffusion kernels on graphs and other discrete structures,” *Proc. ICML*, pp. 315–322, 2002.

[27] A. J. Smola and R. Kondor, “Kernels and regularization on graphs,” *Learning Theory and Kernel Machines*, Lecture Notes in Computer Science, vol. 2777, Springer, Berlin, Heidelberg, 15 pp., 2003.

[28] D. Zhou and B. Scholkopf, “A regularization framework for learning from graph data,” *Proc. ICML Workshop Statist. Relational Learn. Connections Other Fields*, Banff, Canada, pp. 132–137, 2004.

[29] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold regularization: a geometric framework for learning from labeled and unlabeled examples,” *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, 2006.

[30] H. Zhang, Y. Xu, and J. Zhang, “Reproducing kernel Banach spaces for machine learning,” *J. Mach. Learn. Res.*, vol. 10, pp. 2741–2775, 2009.

[31] M. Seto, S. Suda, and T. Taniguchi, “Gram matrices of reproducing kernel Hilbert spaces over graphs,” *Linear Algebra Appl.*, vol. 445, pp. 56–68, 2014.

[32] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, 2014.

[33] P. A. Forero, K. Rajawat, and G. B. Giannakis, “Prediction of partially observed dynamical processes over networks via dictionary learning,” *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3305–3320, 2014.

[34] M. S. Kotzagiannidis and P. L. Dragotti, “Sampling and reconstruction of sparse signals on circulant graphs – an introduction to graph-FRI”, *Appl. Comput. Harmon. Anal.*, vol. 47, no. 3, pp. 539–565, 2019.

[35] D. Romero, M. Ma, and G. B. Giannakis, “Kernel-based reconstruction of graph signals,” *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 764–778, 2017.

[36] J. P. Ward, F. J. Narcowich, and J. D. Ward, “Interpolating splines on graphs for data science applications,” *Appl. Comput. Harmon. Anal.*, vol. 49, no. 2, pp. 540–557, 2020.

[37] N. Emirov, C. Cheng, J. Jiang, and Q. Sun, “Polynomial graph filters of multiple shifts and distributed implementation of inverse filtering,” *Sampl. Theory Signal Process. Data Anal.*, vol. 20, article no. 2, 2022.

[38] X. Jian and W. P. Tay, “Kernel ridge regression for generalized graph signal processing,” *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, pp. 1–5, 2023.

[39] S.-Y. Chung and Q. Sun, “Barron space for graph convolutional neural networks,” *J. Fourier Anal. Appl.*, vol. 32, article no. 52, 2026.

[40] L. Huang, D. Needell, and S. Tang, “Robust estimation of smooth graph signals from randomized space–time samples,” *Inf. Inference*, vol. 13, no. 2, article no. iaae012, 2024.

[41] L. Huang, D. Li, S. Tang, and Q. Yao, “Random space-time sampling and reconstruction of sparse bandlimited graph diffusion field,” *J. Fourier Anal. Appl.*, vol. 31, article no. 43, 2025.

[42] M. Z. Nashed and Q. Sun, “Sampling and reconstruction of signals in a reproducing kernel subspace of $L^p(\mathbb{R}^d)$,” *J. Funct. Anal.*, vol. 258, pp. 2422–2452, 2010.

[43] M. Nagahama, K. Yamada, Y. Tanaka, S. H. Chan, and Y. C. Eldar, “Graph signal restoration using nested deep algorithm unrolling,” *IEEE Trans. Signal Process.*, vol. 70, pp. 3296–3311, 2022.

[44] J. H. Giraldo, A. Mahmood, B. Garcia-Garcia, D. Thanou, and T. Bouwmans, “Reconstruction of time-varying graph signals via Sobolev smoothness,” *IEEE Trans. Signal Inf. Process. Netw.*. vol. 8, pp. 201–214, 2022.

[45] J. Jiang, C. Cheng, and Q. Sun, “Nonsubsampled graph filter banks: theory and distributed algorithms,” *IEEE Trans. Signal Process.*, vol. 67, pp. 3938–3953, 2019.

[46] F. R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, No. 92. Providence, RI, Amer. Math. Soc., 1997.

[47] B. Ricaud, P. Borgnat, N. Tremblay, P. Goncalves, and P. Vandergheynst, “Fourier could be a data scientist: from graph Fourier transform to signal processing on graphs,” *C. R. Phys.*, vol. 20, pp. 474–488, 2019.

[48] F. Harary and A. J. Schwenk, “Which graphs have integral spectra?” In *Graphs and Combinatorics*, edited by R. Bari and F. Harary, Lecture Notes in Math., vol. 406, Springer-Verlag, Berlin, pp. 45–51, 1974.

[49] R. A. Brualdi, *The Mutually Beneficial Relationship of Graphs and Matrices*, vol. 115, American Mathematical Soc., 2011.

[50] T. Tao and V. Vu, “Random matrices have simple spectrum,” *Combinatorica*, vol. 37, pp. 539–553, 2017.

[51] Y. Chen, S.-Y. Chung, and Q. Sun, “Shift-invariant spaces on finite undirected graphs,” arXiv:2602.19234, 55 pp., 2026.

[52] Y. Chen, C. Cheng, and Q. Sun, “Graph Fourier transform based on singular value decomposition of directed Laplacian,” *Sampl. Theory Signal Process. Data Anal.*, vol. 21, article no. 24, 28 pp., 2023.

[53] C. Cheng, Y. Chen, Y. J. Lee, and Q. Sun, “SVD-based graph Fourier transforms on directed product graphs,” *IEEE Trans. Signal Inf. Process. Netw.*, vol. 9, pp. 531–541, 2023.

[54] E. Rebrova and P. Salanevich, “On graph uncertainty principle and eigenvector delocalization,” *2023 International Conference on Sampling Theory and Applications (SampTA)*, pp. 1–5, 2023.

[55] D. L. Donoho and P. B. Stark, “Uncertainty principles and signal recovery,” *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, 1989.

[56] A. Agaskar and Y. M. Lu, “A spectral graph uncertainty principle,” *IEEE Trans. Inf. Theory*, vol. 59, pp. 4338-4356, 2013.

[57] B. Ricaud and B. Torresani, “A survey of uncertainty principles and some signal processing applications,” *Adv. Comput. Math.*, vol. 40, pp. 629–650, 2014.

[58] M. Tsitsvero, S. Barbarossa and P. Di Lorenzo, “Signals on graphs: uncertainty principle and sampling,” *IEEE Transactions on Signal Processing*, vol. 64, pp. 4845-4860, 2016.

[59] O. Teke and P. P. Vaidyanathan. “Uncertainty principles and sparse eigenvectors of graphs,” *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5406–5420, 2017.

[60] A. Aldroubi, J. Davis, and I. Krishtal, “Dynamical sampling: Time–space trade-off,” *Appl. Comput. Harmon. Anal.*, vol. 34, no. 3, pp. 495–503, 2013.

[61] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, “Sampling of graph signals with successive local aggregations,” *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1832–1843, 2016.

[62] A. Aldroubi, C. Cabrelli, U. Molter, and S. Tang, “Dynamical sampling,” *Appl. Comput. Harmon. Anal.*, vol. 42, no. 3, pp. 378–401, 2017.

[63] Airline On-Time Performance Data collected by Bureau of Transportation Statistics, USA. https://www.transtats.bts.gov/