Doctoral Dissertation

Graph signal sampling under arbitrary signal priors
(任意の先験情報に基づくグラフ信号のサンプリング)

Junya Hara

July 2024

Graduate School of Engineering,
Osaka University

Abstract

This dissertation describes the developments and results of graph signal sampling under arbitrary signal priors, based on research papers published within the Graduate School of Bio-Applications and Systems Engineering, Department of Bio-Functions and Systems Science at Tokyo University of Agriculture and Technology, and Graduate School of Engineering, Division of Electrical, Electronic, and Infocommunications Engineering at Osaka University. Contributions of this dissertation include proposing a series of graph signal sampling methods tailored to various signal models, including deterministic, random, and their mixture models. The effectiveness of these proposed methods is validated via recovery experiments with synthetic and real-world datasets. The dissertation is organized into six chapters, detailed as follows.

Chapter 1 serves as the introduction, setting the stage for the dissertation by discussing the limitations of conventional signal processing and the unique challenges posed by sampling of graph signals. It outlines the necessity for a flexible sampling framework to address these issues effectively.

Chapter 2 reviews fundamental concepts in sampling theory, from Shannon's theorem to its generalization required for graph signals. This chapter lays the groundwork by defining graph signals and introducing the concept of generalized sampling theory (GST) which is adapted for the graph domain.

Chapter 3 addresses the challenges of applying generalized graph sampling theory (GGST) to random graph signals, which are non-shift-invariant and conform to graph-wide sense stationarity (GWSS). This chapter extends the concept of wide sense stationarity from time-domain to graph signals, incorporating a correction transform to improve the accuracy of non-ideal measurements. The framework supports arbitrary sampling methods and demonstrates its effectiveness through experiments with synthetic and real datasets.

Chapter 4 presents a sampling set selection (SSS) method that operates under arbitrary signal priors, moving beyond the conventional bandlimited assumptions. This method utilizes the direct sum condition between sampling and reconstruction subspaces, ensuring optimal recovery of graph signals. The effictiveness of this approach is validated through computational experiments using various graph signal models.

Chapter 5 expands SSS method to address sensor placement problem on graphs (SPPG)

with sensors of diverse specifications. It proposes a solution based on difference-of-convex (DC) optimization, balancing the maximization of coverage area with the minimization of sensing budget. Experimental results demonstrate the effectiveness of this approach.

Chapter 6 extends the GGST framework to multi-channel systems, addressing the complexities posed by full-band graph signals. This chapter not only clarifies the relationships among existing analysis methods but also demonstrates the framework's effectiveness through the multi-channel sampling (MCS) of full-band graph signals.

Chapter 7 summarizes the dissertation and offers comprehensive directions for future research.

Contents

List of Figures

List of Tables

List of Notations

We list major notations in the following table.

Symbol	Definition
calligraphic letter	set of indices, e.g., $\mathcal{X}$ and $\mathcal{Y}$.
bold-type upper/lower-case letter	vector or matrix.
$[\cdot]_{n,k}$	(n,k)-th element in the matrix.
$[\cdot]_{n,\cdot}[n]$	n-th element in the vector.
$\langle \cdot, \cdot \rangle$	inner product between two vectors.
$\mathrm{tr}(\cdot)$	trace of a matrix.
$\mathrm{diag}(\cdot)$	diagonal element of a matrix.
$\sigma(\cdot)$	singular-/eigen-value of a matrix. $\sigma_{\min}(\cdot)$ and $\sigma_{\max}(\cdot)$ are the minimum and maximum value of $\sigma(\cdot)$.
$(\circ)$	Hadamard (elementwise) product.
$(\otimes)$	Kronecker product.
$\Vert \cdot \Vert_p$	ℓ_p norm of a vector.
$\Vert \cdot \Vert_2$	ℓ_2 spectral norm of a matrix.
$\Vert \cdot \Vert_F$	Frobenius norm of a matrix.
$\mathbf{A}_{\mathcal{X}}$	submatrix of $\mathbf{A}$ indexed by $\mathcal{X}$ and $\mathcal{X}$.
$\mathbf{A}_{\mathcal{X}\mathcal{Y}}$	submatrix of $\mathbf{A}$ indexed by $\mathcal{X}$ and $\mathcal{Y}$.
$\mathbf{A}_{xy}$	(x,y) element of $\mathbf{A}$.
$\mathbf{A}^\top / \mathbf{A}^*$	transpose/adjoint of matrix.
$\mathcal{M}^c$	complement set of $\mathcal{M}$.
$\vert\mathcal{M}\vert$	cardinality of the set $\mathcal{M}$.
$\mathcal{V}$	sets of nodes.
$\mathcal{E}$	sets of edges.
$\mathcal{G}$	weighted undirected graph, i.e., $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.
N	number of nodes, i.e., $N = \vert\mathcal{V}\vert$ (unless otherwise specified).
$\boldsymbol{\mathcal{W}}$	weighted adjacency matrix. Its (m,n)-element $[\boldsymbol{\mathcal{W}}]_{mn} \geq 0$ is the edge weight between the mth and nth nodes; $[\boldsymbol{\mathcal{W}}]_{mn} = 0$ for unconnected nodes.
$\mathbf{D}$	degree matrix, i.e., $\mathbf{D} = \mathrm{diag}\,(d_0, d_1, \ldots, d_{N-1})$, where $d_m = \sum_n [\boldsymbol{\mathcal{W}}]_{mn}$ is the mth diagonal element.
$\mathbf{L}$	graph Laplacian $\mathbf{L} \coloneqq \mathbf{D} - \boldsymbol{\mathcal{W}}$. We use $\mathbf{L}$ as a graph variation operator.
$\boldsymbol{x}$	graph signal. $\boldsymbol{x} \in \mathbb{R}^N$ is defined as a mapping from the node set to the set of real numbers, i.e., $\boldsymbol{x} : \mathcal{V} \to \mathbb{R}$. We can easily extend to the complex number version.
$\mathbf{U}$	graph Fourier transform (GFT) matrix. It is obtained by the eigenvalue decomposition of the graph Laplacian $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ with the eigenvalue matrix $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_0, \lambda_1, \ldots, \lambda_{N-1})$.
λ_i	ith graph frequency.
$\hat{\boldsymbol{x}}$	GFT of $\boldsymbol{x}$. It is defined as $\hat{x}(\lambda_i) = \langle \boldsymbol{u}_i, \boldsymbol{x} \rangle = \sum_{n=0}^{N-1} u_i[n] x[n]$.
$\mathcal{B}$	subset of the bandwidth, where B is the bandwidth.

Acronyms

Acronym	Definition
BGFB	Bipartite graph filter bank
BIBO	Bounded-input-bounded-output
BL	Bandlimited
BP	Bandpass
CTFT	Continuous-time Fourier transform (or simply Fourier transform)
DC	Difference-of-convex
DS	Direct sum
DTFT	Discrete-time Fourier transform
ER	Erdős-Rényi
FIR	Finite impluse response
GFB	Graph filter bank
GFT	Graph Fourier transform
GGST	Generalized graph sampling theory
GSP	Graph signal processing
GST	Generalized sampling theory
GWSS	Graph wide sense sationary, graph wide sense sationarity
ICTFT	Inverse of CTFT
IDTFT	Inverse of DTFT
IIR	Infinite impluse response
LS	Least squares
LSI	linear shift-invariant, linear shift-invariance (or linear time-invariance)
MCS	Multi-channel sampling
MMSE	Minimum mean squared error
MSE	Mean squared error
MX	Minimax
NLI	Node-localization invariant
PDS	Primal-dual splitting
PGS	Periodic graph spectrum
PR	Perfect reconstruction
PSD	Power spectral density
PWC	Piecewise constant
PWS	Piecewise smooth
RIP	Restricted isometry property
RKHS	Reproducing kernel Hilbert space
RS	Random sensor
SI	Shift-invariant, Shift-invariance
SPPG	Sensor placement problem on graphs
SSS	Sampling set selection
UBP	Union of bandpass
WSS	Wide sense sationary, wide sense sationarity

Chapter 1

Introduction

Sampling has played a crucial role in signal processing. Sensors are ubiquitous and capable of collecting large volumes of diverse information in real-time. This information includes multimedia (images, audio, and videos), 3D point clouds, and environmental and behavioral data [1–4], to name a few. Although modern sensors can capture signals with significantly high temporal/spatial resolutions, most instances are wasted in practice due to storage constraints and the limited scale of analysis techniques. Therefore, the data acquisition in sensors must incorporate a sampling process for the spatial and temporal instances to extract informative (significant) ones from the sensing environment.

Spatial sampling presents more challenges than temporal sampling because of the irregular distribution of sensors in space. This irregularity leads to the absence of a unified sampling paradigm for spatial data. We focus on spatial domain sampling in this dissertation.

Networks are powerful tools for analyzing signals with a spatial structure, where each node in a network carries an attribute value and edges represent the inter-node relationships. These signals are mathematically modeled as signals whose domain is given by nodes on the graph, i.e., *graph signals*, and are present in various applications, including sensor, social, infrastructure, neuroscientific, and biological networks [5–7]. Graph signal processing (GSP) has been developed to provide efficient analysis tools for graph signals [8].

In recent decades, the diversification and globalization of infocommunication have significantly increased the demands for high-quality and high-quantity services, industries, and businesses. This has led to the extensive distribution of GSP applications utilizing versatile sensors [5, 9]. Consequently, several emerging challenges in GSP have arisen beyond conventional signal communication applications [8]. For instance, there is often insufficient domain knowledge of GSP applications and thus further analysis of real data is required to accurately understand various measuring environments. Additionally, the scale of current GSP techniques, which are typically applicable to networks of thousands of nodes, has not yet

met the demands of large-scale applications which involve millions of nodes. This disparity hinders practical applications. Since sampling tackles both the analysis and reduction of signal components, it addresses these challenges effectively. In this scenario, sampling graph signals is a key focus within GSP [10–15].

A critical distinction from standard signal processing is that GSP systems are generally not shift-invariant (SI), due to the irregular connectivity of the nodes. While SI sampling can be equivalently expressed in time and Fourier domains [16], graph signal sampling does not generally have a similar relationship between the nodal and graph Fourier domains. This discrepancy leads to a divergence among various sampling approaches [10–15, 17].

Node-wise sampling is a principal approach in graph signal sampling [10–14]. It involves selecting a subset of nodes, referred to as sampling set selection (SSS). A challenge of SSS is that the optimal sampling set varies depending on the graph and the sampling criterion; SSS is considered *non-uniform sampling in space*, in contrast to the uniform intervals of SI sampling. The sampling criterion is designed based on the requirements of downstream applications [18, 19].

Sensor networks are a compelling application of SSS. Since monitoring all nodes might be infeasible due to various constraints, one may need to select sensor positions from the (large) sensor network, whose nodes and edges represent the candidates for sensor positions and the inter-node communication of sensors, respectively. Thus, the significance of nodes can be evaluated based on the estimation error of the unobserved nodes, reconstructed from the observed ones. In this context, SSS is also considered as a *sensor placement problem* on the graph (SPPG) [20–22].

The least squares (LS) criterion is representative for SPPG [23]. It calculates the average mean squared error (MSE) in *sampled* signals. However, LS recovery can cause significant errors in *non-sampled* signals, since it only focuses on minimizing the estimation error in *sampled* signals [24]. To address this limitation, SPPG requires an alternative criterion that leverages any available information about the non-sampled signals to achieve the best possible performance.

A prior domain, where the original signal is likely to reside, can be specified by a *signal prior*. For instance, the frequency band and historical statistics of objective signals are often used for identifying these domains [25, 26]. There are three widely-studied priors in the literature [15]: subspace, smoothness, and stochastic priors. In essence, they are categorized as constraints on a basis function, energy bound, and covariance of the original signal, respectively. By utilizing signal priors, sampling theory can achieve greater flexibility in practice and higher accuracy.

In summary, the aforementioned challenges in applications boil down to a fundamental problem: Sampling theory on graphs under arbitrary signal priors, which we abbreviate as generalized graph sampling theory (GGST).

Accordingly, this dissertation establishes a sampling paradigm for graph signals by ad-

dressing three challenges:

1. We construct a GGST, providing a general framework for sampling that allows for the best possible recovery under arbitrary graph signal priors.

2. We develop an SPPG based on the GGST framework and then extend it to more challenging scenarios where sensors have multiple specifications.

3. We expand SSS from a single-channel system to a multi-channel one, broadening the available signal models beyond those in SPPGs to include a mixture of multiple signal priors.

We describe the above three contributions from Chapter 3 onwards. The overview of Chapters is visualized in Fig. 1.1.

While this dissertation focuses on SPPG, it is merely one of the possible applications of SSS and encompasses several other applications, such as bottleneck detection in traffic networks [27] and leak detection in water supply networks [28]. Throughout this dissertation, we use the term SPPG, differentiating it from SSS when addressing specific constraints related to sensors.

The remainder of this dissertation is organized as follows: Chapter 2 reviews the basics of sampling theory for both time domain and graph signals. We begin with an overview of the classical Shannon's theorem and discuss its limitations due to the bandlimited assumption. To address these limitations, we introduce the concept of *generalized sampling theory* (GST), which aims to ensure the best possible recovery under arbitrary signal priors [16]. This framework comprises sampling, correction, and reconstruction transforms. A key innovation of GST is the incorporation of a correction transform into the standard sampling theory. The correction transform is specifically designed to minimize the worst-case estimation error, compensating for the non-ideal reconstruction transform. Because GST operates within a general framework in Hilbert spaces, it accommodates not only SI signals but also graph signals [15]. Finally, we lay out the mathematical definition of graph signals and adapt GST to the graph setting to derive GGST.

Chapter 3 considers a GGST framework for random graph signals, which adhere to graph wide sense stationarity (GWSS) [26, 29]. In fact, random graph signals are ill-posed due to their non-SI nature, compared to the deterministic counterparts. As a result, GGST for random graph signals is challenging and yet requires further insights and arguments. We extend the concept of wide sense stationarity (WSS) from time-domain signals to the graph domain. In this framework, a correction transform is placed between the sampling and reconstruction transforms to compensate for non-ideal measurements, aiming to minimize the MSE between the original and reconstructed signals. The proposed framework allows for arbitrary sampling methods, i.e., sampling in the nodal or graph frequency domain. The

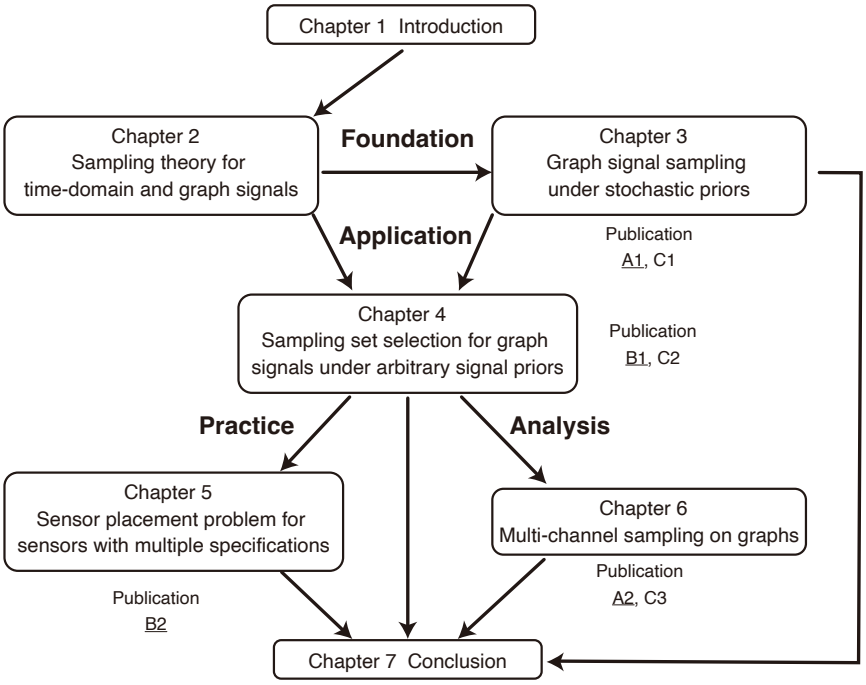


Figure 1.1: Overview of this dissertation.

effectiveness of the proposed method is validated through recovery experiments with synthetic and real datasets.

Chapter 4 proposes an SSS method that accommodates arbitrary signal priors, moving beyond the typical bandlimited assumption [10–12,14]. This method is applicable to arbitrary linear graph signal models and is founded on a GGST framework, considered in Chapters 2 and 3. Our approach is uniquely built upon the direct sum condition between sampling and reconstruction subspaces [15], which is crucial for the perfect recovery of sampled signals. Additionally, we develop a computationally-efficient algorithm for the proposed SSS, utilizing the Neumann series approximation. The effectiveness of our method has been proven through experiments with various graph signal models.

Chapter 5 expands the scope of SSS in Chapter 4 by addressing the challenges in the context of SPPG. It explores an SPPG method with varying specifications on sensors, which is a key concern in environmental monitoring across networks like water supply, electricity, and gas pipelines [30]. Unlike most studies that assume uniform sensor specifications, our approach takes into account sensors with diverse specifications. We propose a solution based on difference-of-convex (DC) optimization. Sensor placements are selected such that they simultaneously maximize the overall coverage area and minimize the total sensing budget. Our experimental results demonstrate the effectiveness of our proposed method.

Chapter 6 develops multi-channel sampling (MCS) for graph signals. Unlike the traditional focus on bandlimited graph signals, this work addresses *full-band* graph signals often encountered in applications [31]. These signals are generally represented by a mixture of multiple signals following different generation models. This necessitates MCS for effective analysis, diverging from the existing single-channel sampling practices. We propose an MCS framework built upon GGST, introduced in Chapters 2 and 3, and design an SSS method tailored for MCS based on Chapter 4. Experimental results demonstrate the effectiveness of the proposed method in recovering full-band graph signals.

Chapter 7 provides a summary of the proposed sampling paradigm for graph signals and concludes the dissertation.

Chapter 2

Sampling theory for time-domain and graph signals

2.1 Introduction

Sampling theory has profoundly influenced modern signal processing tasks since being established by Shannon [32]. Understanding sampling theory is useful for navigating through classical signal processing literature. Its principle, i.e., extracting an informative subset from a pool of signals, addresses a crucial challenge in numerous signal processing applications, such as compressed sensing [33], key-frame detection for videos [34], anomaly detection [35], and feature extraction in machine learning [36], to name a few.

While the elegance of Shannon's paradigm attracts many researchers, it assumes that signals are ideally bandlimited, which can be challenging in practice. This assumption has motivated numerous studies tailored to several challenging scenarios. Generalized sampling theory (GST) is one of the promising alternatives [16]. It allows for the best possible recovery of arbitrary signals from non-ideally sampled signals, given appropriate signal priors. This framework embraces arbitrary linear signals, thus extending its applications beyond time-domain signals to include signals in various domains. Graph signals have been an emerging signal model in signal processing applications [8], whose domain is given by nodes in a network. This extension motivates us to develop a GST for graph signals.

In this chapter, we start by reviewing fundamental properties of linear shift-invariant (LSI) systems, Fourier analysis, and random processes. We revisit the Shannon-Nyquist sampling theorem from a practical perspective. To overcome its limitations, we then introduce the well-studied GST in Hilbert spaces. Lastly, we derive a GST framework on graphs based on the preceding framework.

2.2 Classical signal processing tools

In this section, we review the basic signal processing tools. First, we refer to a LSI system to introduce Fourier transforms of LSI signals. Next, we consider the stationarity of LSI signals.

GSP is not a LSI system, but an analog of LSI is introduced, which plays a fundamental role in GSP, including spectral analysis, filter designs, sampling, and restoration of graph signals. Therefore, revisiting an LSI system is beneficial as a good starting point for understanding and entering the field of GSP.

Linear shift-invariant system

We quickly review a LSI system, which plays a key role in signal processing. It yields many desired properties. The classical sampling theory is built on these properties.

Throughout this dissertation, we assume that both of continuous and discrete linear systems are bounded. That is, they are defined over L_p and ℓ_p spaces ($1 \leq p < \infty$), respectively. For an arbitrary function $x(t)$, the vector space L_p satisfies the following condition.

$$\|x(t)\|_p = \left(\int_{-\infty}^{\infty} |x(t)|^p dt \right)^{1/p} < \infty. \tag{2.1}$$

For an arbitrary sequence $x[n]$, the vector space ℓ_p satisfies the following condition.

$$\|x[n]\|_p = \left(\sum_{n \in \mathbb{Z}} |x[n]|^p \right)^{1/p} < \infty. \tag{2.2}$$

Continuous-time system

We introduce several important terminologies of conventional signal processing:

Linearity Let $x(t)$ and $y(t)$ be the input and output of a system $\mathcal{T}$, i.e., $y(t) = \mathcal{T}\{x(t)\}$. Then, $\mathcal{T}$ is linear if and only if

$$\mathcal{T}\{ax_1(t) + bx_2(t)\} = ay_1(t) + by_2(t), \tag{2.3}$$

where $a, b \in \mathbb{R}$. In particular, if the delay between $x(t)$ and $y(t)$ is invariant, i.e., $y(t-\tau) = \mathcal{T}\{x(t-\tau)\}$, $\mathcal{T}$ is a linear shift-invariant (LSI) system.

Dirac delta The Dirac delta $\delta(t)$ is defined as

$$\int_{-\infty}^{\infty} x(t)\delta(t) = x(0), \tag{2.4}$$

where $x(t)$ is an arbitrary test function that has a compact support and infinitely differentiable. We summarize properties of $\delta(t)$ as follows:

$$\int_{-\infty}^{\infty} \delta(t)dt = 1, \tag{2.5a}$$

$$\int_{-\infty}^{\infty} x(t)\delta(t-t_0)dt = x(t_0) \tag{2.5b}$$

$$x(t)\delta(t) = x(0)\delta(t) \tag{2.5c}$$

We will frequently use the Dirac function to express sampling process.

Convolution The convolution between two functions $x(t), y(t)$ is defined as

$$x(t) * y(t) = \int_{-\infty}^{\infty} x(\tau)y(t-\tau)d\tau. \tag{2.6}$$

Impulse response The basis property of a LSI system is that it is characterized by the impulse response $h(t)$. The impulse response $h(t)$ is given by the output of the system when the input is $\delta(t)$, i.e., $h(t) = \mathcal{T}\{\delta(t)\}$. By using (2.5b), $y(t)$ can be expressed by

$$\begin{aligned} y(t) = \mathcal{T}\{x(t)\} &= \mathcal{T}\left\{\int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau\right\} \\ &= \int_{-\infty}^{\infty} x(\tau)\mathcal{T}\left\{\delta(t-\tau)\right\}d\tau. \end{aligned}$$

Since $\mathcal{T}$ is a LSI system, we have $\mathcal{T}\{\delta(t-\tau)\} = h(t-\tau)$. Therefore, the LSI system is characterized by

$$\mathcal{T}\{x(t)\} = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)dt = x(t) * h(t). \tag{2.7}$$

For a bounded input $x(t)$, the system is *stable* if the output of the LSI system $y(t)$ is bounded. This is referred to as the bounded-input-bounded-output (BIBO) stability. By using (2.7), the BIBO condition leads to

$$|y(t)| \leq \int_{-\infty}^{\infty} |x(t-\tau)||h(\tau)|d\tau \leq \max_{t\in\mathbb{R}} |x(t)| \int_{-\infty}^{\infty} |h(\tau)|d\tau. \tag{2.8}$$

Thus, supposing that $x(t)$ is bounded, the BIBO condition implies that $h(t)$ is absolutely integrable, i.e., $h(t) \in L_1$.

Continuous-time Fourier transform (CTFT)

Here, suppose that all functions that we treat are absolutely integrable[1]. The CTFT of $x(t) \in L_1$ is defined as

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt. \tag{2.9}$$

Its inverse (ICTFT) is given by

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega. \tag{2.10}$$

We describe important properties of CTFT below. We denote the Fourier transform pair by $x(t) \leftrightarrow X(\omega)$.

- **Convolution and product**

  The convolution between $x(t), y(t)$ can be expressed by the product between the corresponding CTFTs $X(\omega), Y(\omega)$ in Fourier domain, i.e.,

  $$x(t) * y(t) \leftrightarrow X(\omega) Y(\omega). \tag{2.11}$$

  Similarly, the dual relationship is commutable, i.e.,

  $$x(t) y(t) \leftrightarrow \frac{1}{2\pi} X(\omega) * Y(\omega). \tag{2.12}$$

- **Symmetry**

  The CTFT of a real function, $X(\omega)$, is Hermitian symmetric, i.e.,

  $$X(\omega) = \overline{X(-\omega)}. \tag{2.13}$$

- **Shift and scaling**

  The time-shift corresponds to the frequency-modulation in Fourier domain, i.e.,

  $$x(t - t_0) \leftrightarrow X(\omega) e^{-j\omega t_0}. \tag{2.14}$$

[1]In principle, if a signal $f(t)$ is absolutely integrable, i.e., $f(t) \in L_1$, the CTFT of $f(t)$ is defined. Exceptionally, if $f(t)$ is square-integrable, i.e., $f(t) \in L_2$, the CTFT of $f(t)$ could be defined. For example, $\delta(t)$, $1/t$ and $\mathrm{sinc}(t)$ are all in L_2 but not in L_1. The existence of a Fourier transform depends on either $f(t) \in L_1$ or $f(t) \in L_2$, which are not necessarily simultaneously satisfied.

The dual relationship is commutable, i.e.,

$$x(t)e^{j\omega_0 t} \leftrightarrow X(\omega - \omega_0). \tag{2.15}$$

The time-scaling of $x(t)$ is reciprocal to the frequency-scaling of $X(\omega)$, i.e.,

$$x(at) \leftrightarrow \frac{1}{|a|} X\left(\frac{\omega}{a}\right), \tag{2.16}$$

where a is a scalar. In particular, considering $a = -1$, it is clear that $x(-t) \leftrightarrow X(-\omega)$. This relationship is known as the time-reversal symmetry.

- **Correlation**

  The correlation between real functions $x(t), y(t)$ is defined as

$$r_{xy}(t) = \int_{-\infty}^{\infty} x(\tau)y(t+\tau)d\tau. \tag{2.17}$$

  In terms of the convolution, this can be expressed by $r_{xy}(t) = x(-t) * y(t)$. Further, with (2.11) and (2.13), the CTFT of $r_{xy}(t)$, $R_{XY}(\omega)$, can be expressed by

$$R_{XY}(\omega) = \overline{X(\omega)}Y(\omega). \tag{2.18}$$

- **Energy conservation**

  According to Parseval's theorem, we have the following equality:

$$\int_{-\infty}^{\infty} \overline{x(t)}y(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{X(\omega)}Y(\omega)d\omega. \tag{2.19}$$

  Replacing with $x(t) = y(t)$, (2.19) becomes

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(\omega)|^2 d\omega. \tag{2.20}$$

  Therefore, the signal energy is conserved in the CTFT.

Discrete-time system

Basically, a discrete-time system parallels properties of a corresponding continuous-time system. The discrete-time system can typically be obtained by sampling of the continuous-time system. In particular, $x[n]$ is often assumed to be samples of $x(t)$ at uniform period $t = nT, n \in \mathbb{Z}$.

Suppose that $x(t)$ is a LSI system. For its uniformly-sampled system $x[n]$, the shift invariance holds at every intervals of T. That is, when the input delayed m samples, the output of the system $\mathcal{T}$ is also delayed m samples, i.e., $y[n-m] = \mathcal{T}\{x[n-m]\}$ in terms of $y[n] = \mathcal{T}\{x[n]\}$. Here, we introduce a few terminologies of the discrete-time system.

Kronecker delta The Kronecker delta is defined as

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0. \end{cases} \tag{2.21}$$

We can rewrite an arbitrary sequence $x[n]$ with the use of $\delta[n]$ as

$$x[n] = \sum_{m \in \mathbb{Z}} x[m]\delta[n-m]. \tag{2.22}$$

Convolution The convolution between two sequences $x[n], y[n]$ is defined as

$$x[n] * y[n] = \sum_{m \in \mathbb{Z}} x[m]y[n-m]. \tag{2.23}$$

Discrete-time impulse response In the same manner as the continuous-time system, the discrete-time LSI system is also characterized by the impulse response $h[n]$. The impulse response $h[n]$ is given by the output of the system when $\delta[n]$ is the input, i.e., $h[n] = \mathcal{T}\{\delta[n]\}$. By using (2.22), the output $\mathcal{T}$ results in

$$\begin{aligned} \mathcal{T}\{x[n]\} &= \sum_{m \in \mathbb{Z}} x[m]\mathcal{T}\{\delta[n-m]\} \\ &= \sum_{m \in \mathbb{Z}} x[m]h[n-m] = x[n] * h[n] \end{aligned} \tag{2.24}$$

where the second equality is followed by the fact that $\mathcal{T}$ is a LSI system, i.e., $\mathcal{T}\{\delta[n-m]\} = h[n-m]$.

Importantly, if $h[n]$ has a compact support, (2.24) is expressed by the finite-length filtering. Such a filter $h[n]$ is referred to as a finite impulse response (FIR) filter. In the same manner as the continuous-time system, the stability can be evaluated by impulse response $h[n]$. The BIBO stability is guaranteed if $h[n]$ is absolutely summable, i.e., $h[n] \in \ell_1$.

Discrete-time Fourier transform (DTFT)

The DTFT of $x[n] \in \ell_1$, $X(e^{j\omega})$, is defined as

$$X(e^{j\omega}) = \sum_{n \in \mathbb{Z}} x[n]e^{-j\omega n}. \tag{2.25}$$

Note that $X(e^{j\omega})$ is 2π-periodic. Correspondingly, the inverse DTFT (IDTFT) is given by

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega n}d\omega. \tag{2.26}$$

Since most of properties of DTFT is inherited from those in CTFT, we omit them in this dissertation for simplicity. We highlight a few important properties below.

- **Convolution and product**

  The convolution between $x[n], y[n]$ can be expressed by the product between the corresponding DTFTs $X(e^{j\omega}), Y(e^{j\omega})$ in Fourier domain, i.e.,

$$x[n] * y[n] \leftrightarrow X(e^{j\omega})Y(e^{j\omega}). \tag{2.27}$$

- **Discrete correlation**

  The correlation between real sequences $x[n], y[n]$ is defined as

$$r_{xy}[n] = \sum_{m\in\mathbb{Z}} x[m]y[n+m]. \tag{2.28}$$

  In terms of the convolution, $r_{ry}[n]$ can be expressed by $r_{xy}[n] = x[n]*y[-n]$. In addition, with their DTFTs $X(e^{j\omega}), Y(e^{j\omega})$, the DTFT of $r_{xy}[n]$, $R_{XY}(e^{j\omega})$, can be expressed by

$$R_{XY}(e^{j\omega}) = X(e^{j\omega})\overline{Y(e^{j\omega})}. \tag{2.29}$$

- **Energy conservation**

  Parseval's theorem also holds for the discrete-time system as

$$\sum_{n\in\mathbb{Z}} \overline{x[n]}y[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{X(e^{j\omega})}Y(e^{j\omega})d\omega. \tag{2.30}$$

  Considering $x[n] = y[n]$, the signal energy is also conserved in the DTFT, i.e.,

$$\sum_{n\in\mathbb{Z}} |x[n]|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(e^{j\omega})|d\omega. \tag{2.31}$$

Random process

We review the stationarity of continuous/discrete-time signals. Stationarity is characterized by the second order statistics. The second order statistics are widely used in several applications in signal processing.

Continuous-time random process

First, we consider a continuous-time signal $x(t)$. Its stationarity is defined as follows:

Definition 1 (Wide sense stationary for continuous-time signals)**.** *Let $x(t)$ and $\gamma_x(t)$ be a stochastic signal in the time domain and its autocovariance function, respectively. The signal $x(t)$ is a wide sense stationary process if and only if the following two conditions are satisfied:*

1. $\mathbb{E}[x(t)] = \mu_x = \text{const},$ (2.32)
2. $\mathbb{E}[(x(t) - \mu_x)(x(\tau) - \mu_x)] = \gamma_x(t - \tau).$ (2.33)

In addition, a wide sense stationary (WSS) process is characterized in Fourier domain, stated by Wiener-Khinchin theorem [37]. If $x(t)$ is a WSS process, the power spectral density (PSD) function coincides with the CTFT of the autocovariance function of $x(t)$, $\gamma_x(t) \in L_1$, i.e.,

$$\Gamma_x(\omega) = \int_{-\infty}^{\infty} \gamma_x(t) e^{-j\omega t} d\omega. \tag{2.34}$$

Discrete-time random process

Next, we consider a discrete-time signal $x[n]$. Its stationarity is defined as follows:

Definition 2 (Wide sense stationary for discrete-time signals)**.** *Let $x[n]$ and $\gamma_x[n]$ be a stochastic signal in the time domain and its autocovariance function, respectively. The signal $x[n]$ is a wide sense stationary process if and only if the following two conditions are satisfied:*

1. $\mathbb{E}[x[n]] = \mu_x = \text{const},$ (2.35)
2. $\mathbb{E}[(x[n] - \mu_x)(x[m] - \mu_x)] = \gamma_x[n - m].$ (2.36)

In the same manner as the above, if $x[n]$ is a WSS process, the power spectral density (PSD) function coincides with the DTFT of the autocovariance function of $x[n]$, $\gamma_x[n] \in \ell_1$, i.e.,

$$\Gamma_x(e^{j\omega}) = \sum_{n \in \mathbb{Z}} \gamma_x[n] e^{-jn\omega}. \tag{2.37}$$

We have thus far reviewed several fundamental components of classical signal processing. Based on them, we will examine classical sampling theory in the next section.

2.3 Classical sampling theory

Although the standard sampling theory has been well-studied, many books omit its detailed derivation. In this section, we review Shannon-Whittaker sampling theorem and address several limitations inherent in its formulation.

Sampling in shift-invariant spaces

Let $s(-t) \in L_1$ be some sampling filter and let T be a sampling interval. A sampled signal is given by

$$c[n] = \langle x(t - nT), s(-t) \rangle = x(t) * s(-t)|_{t=nT}. \tag{2.38}$$

Under presense of noise $\eta[n]$, the sampled signal can be obtained by $y[n] = c[n] + \eta[n]$. In Fourier domain, the sampled signal is obtained by the following theorem:

Theorem 1 (Sampling in Fourier domain)**.** *Let $x(t) \in L_1$ be a continuous time signal with the CTFT $X(\omega)$, $c[n] \in \ell_1$ be the sequence obtained by sampling in* (2.38)*. Then, the DTFT of $c[n]$ is given by*

$$C(e^{j\omega}) = \frac{1}{T} \sum_{k \in \mathbb{Z}} X\left(\frac{\omega - 2\pi k}{T}\right) \overline{S\left(\frac{\omega - 2\pi k}{T}\right)}. \tag{2.39}$$

Proof. Let $x_s(t) = x(t) * s(-t)$ denotes the filtered signal. A sampled signal is given by $c[n] = x_s(nT)$. We can can rewrite $c[n]$ in terms of a real number $n' \in \mathbb{R}$ instead of $n \in \mathbb{Z}$ as follows:

$$c[n] = x_s(n'T) \cdot \sum_{\tau \in \mathbb{Z}} \delta((\tau - n')T) = \begin{cases} x_s(n'T) & \text{if } n'T \text{ is an integer} \\ 0 & \text{otherwise.} \end{cases} \tag{2.40}$$

The CTFT of $x_s(n'T)$ with respect to n' is given by

$$\mathcal{F}\{x_s(n'T)\} = \frac{1}{T} X_s\left(\frac{\omega}{T}\right), \tag{2.41}$$

where $X_s(\omega)$ is the CTFT of $x_s(t)$. The CTFT of $\sum_{\tau \in \mathbb{Z}} \delta((\tau - n')T)$ with respect to n' is given by

$$\mathcal{F}\left\{\sum_{\tau \in \mathbb{Z}} \delta((\tau - n')T)\right\} = 2\pi \sum_{k \in \mathbb{Z}} \delta\left(\frac{2\pi k - \omega}{T}\right). \tag{2.42}$$

Thus, by using the convolution property of Fourier domain, i.e., $x(t) \cdot y(t) \leftrightarrow \frac{1}{2\pi} X(\omega) * Y(\omega)$, (2.40) is equivalent to

$$\begin{aligned} \frac{1}{2\pi} \frac{1}{T} X_s\left(\frac{\omega}{T}\right) * 2\pi \sum_{k \in \mathbb{Z}} \delta\left(\frac{2\pi k - \omega}{T}\right) &= \frac{1}{T} \sum_{k \in \mathbb{Z}} X_s\left(\frac{\omega - 2\pi k}{T}\right) \\ &= \frac{1}{T} \sum_{k \in \mathbb{Z}} X\left(\frac{\omega - 2\pi k}{T}\right) \overline{S\left(\frac{\omega - 2\pi k}{T}\right)}. \end{aligned} \tag{2.43}$$

This completes the proof. □

Let $a(t)$ be a reconstruction filter. The sampled signal (2.38) can be reconstructed by

$$\tilde{x}(t) = \sum_{n\in\mathbb{Z}} c[n]w(t-nT) = a(t) * \left(\sum_{n\in\mathbb{Z}} c[n]\delta(t-nT)\right). \tag{2.44}$$

Let $c_0(t)$ be the zero-inserted signal with $c[n]$, i.e.,

$$c_0(t) = \sum_{n\in\mathbb{Z}} c[n]\cdot\delta(t-nT) = \begin{cases} c[t/T] & \text{if } t/T \text{ is an integer} \\ 0 & \text{otherwise.} \end{cases} \tag{2.45}$$

By the scaling property of DTFT, $x[n/T] \leftrightarrow X(e^{j\omega T})$, the CTFT of $c_0(t)$ results in

$$\mathcal{F}\{c_0(t)\} = C(e^{j\omega T}). \tag{2.46}$$

Finally, the reconstructed signal is given by

$$\widetilde{X}(\omega) = C(e^{j\omega T})A(\omega). \tag{2.47}$$

Note that $C(e^{j\omega T})$ in (2.47) is $2\pi/T$-periodic while $C(e^{j\omega})$ in (2.39) is 2π-periodic.

We now define a feasible signal class given by (2.44). Given the reconstruction process in (2.44) with the basis $\{a(t-nT)\}_{n\in\mathbb{Z}}$, it spans the subspace characterized as follows:

Definition 3 (Shift invariant space)**.** *A shift invariant (SI) space is the space of signals that can be expressed as linear combinations of shifts of a given generator, i.e.,*

$$\mathcal{A} \coloneqq \left\{ x(t) \mid x(t) = \sum_{n\in\mathbb{Z}} d[n]a(t-nT), \textit{for some } a \in \ell_2 \right\}, \tag{2.48}$$

where $d[n]$ is an expansion coefficient and $a(t)$ is the generator.

Note that the notion of (SI) subspaces plays a key role in generalized sampling theory.

Bandlimited sampling

We now derive the well-known Shannon-Whittaker theorem. Consider a bandlimited signal $x_{\mathrm{BL}}(t)$, whose CTFT is given by

$$X_{\mathrm{BL}}(\omega) = \begin{cases} X(\omega) & |\omega| < \omega_{\max} \\ 0 & |\omega| \geq \omega_{\max}, \end{cases} \tag{2.49}$$

where $\omega_{\max}$ is the highest frequency of $x_{\mathrm{BL}}(t)$ and $X(\omega)$ is the spectrum of the original signal $x(t)$. The DTFT of the sampled signal is given by

$$C(e^{j\omega}) = \frac{1}{T}\sum_{k\in\mathbb{Z}} X_{\mathrm{BL}}\left(\frac{\omega-2\pi k}{T}\right). \tag{2.50}$$

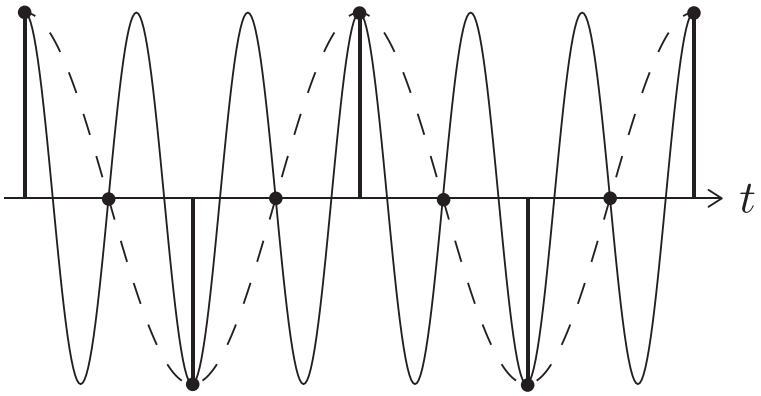


Figure 2.1: Aliasing.

Aliasing According to (2.50), the overlap from the spectrum shifted by $2\pi k$ distorts itself. It is the well-known aliasing phenomenon caused by sampling. Its toy example is shown in Fig. 2.1. In order to maintain the original spectrum for $\omega \in [-\pi, \pi]$ after sampling, the aliasing component is required to be zero, i.e., $X_s\left(\frac{\omega \pm 2\pi}{T}\right) = 0$. Therefore, the following inequality is required:

$$\frac{|\omega \pm 2\pi|}{T} \geq \omega_{\max}. \tag{2.51}$$

By solving with respect to T, we have

$$T \leq \frac{|\omega \pm 2\pi|}{\omega_{\max}}. \tag{2.52}$$

Since the inequality holds for all $\omega \in [-\pi, \pi]$, T is also bounded in the case of $|\omega \pm 2\pi| = \pi$, i.e.,

$$T \leq \frac{\pi}{\omega_{\max}} \leq \frac{|\omega \pm 2\pi|}{\omega_{\max}}. \tag{2.53}$$

In terms of $f_s = 1/T$ and $f_{\max} = \omega_{\max}/2\pi$, (2.53) is translated as

$$f_s \geq 2f_{\max}. \tag{2.54}$$

This statement is well-known as Shannon-Nyquist sampling theorem. When the equality holds in (2.53), f_s is referred to as the Nyquist rate.

Ideal low-pass filter We consider recovery for a ideally-sampled signal in (2.54). In terms of the cutoff frequency $2f_{\max} \leq f_c \leq f_s$, the CTFT of the reconstruction filter can be designed by

$$A(\omega) = T \operatorname{rect}_{2\pi f_c}(\omega), \tag{2.55}$$

where

$$\operatorname{rect}_{2\pi f_c}(\omega) := \begin{cases} 1 & |\omega| < 2\pi f_c \\ 0 & |\omega| \geq 2\pi f_c. \end{cases} \tag{2.56}$$

That is, (2.56) is referred to as the ideal low-pass filter. We will mention why it is to be ideal later.

The reconstruction filter is obtained by

$$\begin{aligned} a(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} A(\omega) e^{j\omega t} d\omega \\ &= \frac{T}{2\pi} \int_{-2\pi f_c}^{2\pi f_c} e^{j\omega t} d\omega \\ &= 2T f_c \frac{\sin 2\pi f_c t}{2\pi f_c t} \\ &= 2\frac{f_c}{f_s} \operatorname{sinc}(2f_c t), \end{aligned} \tag{2.57}$$

where $\operatorname{sinc}(ax) = \sin(a\pi x)/(a\pi x)$. As a result, the bandlimited signal is recovered by

$$x_{\mathrm{BL}}(t) = 2\frac{f_c}{f_s} \sum_{n \in \mathbb{Z}} c[n] \operatorname{sinc}(2f_c(t - nT)). \tag{2.58}$$

This formula is stated by the Whittaker sampling theorem.

In the following, we show important properties of the basis $\{\operatorname{sinc}(2f_c(t - nT))\}_{n \in \mathbb{Z}}$.

Orthogonality of sinc function For simplicity, we denote $2\pi f_c$ by b. In Fourier domain, the inner product of an arbitrary set of the basis function in $\{\operatorname{sinc}(\frac{b}{\pi}(t - nT))\}_{n \in \mathbb{Z}}$ results in

$$\begin{aligned} \int_{-\infty}^{\infty} \operatorname{sinc}\left(\frac{b}{\pi}(t - nT)\right) \operatorname{sinc}\left(\frac{b}{\pi}(t - mT)\right) dt &= \frac{1}{2\pi}\frac{\pi^2}{b^2} \int_{-\infty}^{\infty} \operatorname{rect}_b(\omega) e^{-j\omega(n-m)T} \\ &= \frac{1}{2\pi}\frac{\pi^2}{b^2} \int_{-b}^{b} e^{-j\omega(n-m)T} \\ &= \begin{cases} \frac{\pi}{b} \operatorname{sinc}\left(\frac{b(n-m)T}{\pi}\right) & n \neq m \\ \frac{\pi}{b} & n = m. \end{cases} \end{aligned} \tag{2.59}$$

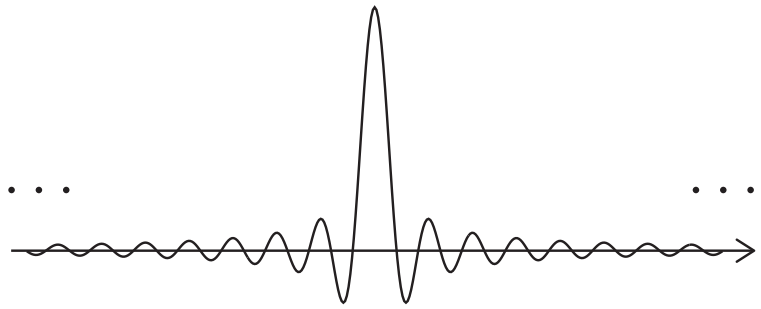


Figure 2.2: Sinc function.

If T is a multiple of $\pi/b = 1/(2f_c)$, $\{\mathrm{sinc}(2f_c(t-nT))\}_{n\in\mathbb{Z}}$ forms the orthogonal basis, i.e.,

$$\int_{-\infty}^{\infty} \mathrm{sinc}\left(\frac{b}{\pi}(t-nT)\right)\mathrm{sinc}\left(\frac{b}{\pi}(t-mT)\right)dt = \frac{\pi}{b}\delta_{n,m}, \tag{2.60}$$

where $\delta_{n,m}$ is the Kronecker delta. Similarly, it is also clear that $\{\frac{1}{\sqrt{T}}\mathrm{sinc}(2f_c(t-nT))\}_{n\in\mathbb{Z}}$ with $T = 1/(2f_c)$ forms the orthonormal basis.

Implementation of the ideal low-pass filter The sinc function $\{\mathrm{sinc}(2f_c(t-nT))\}_{n\in\mathbb{Z}}$ is hard to efficiently approximate due to the slow decay, which is proportional to $1/t$. We plot sinc function in Fig. 2.2. Furthermore, it is a typical non-causal infinite-impulse-response (IIR) filter and thus it causes the difficulty of the implementation[2]. This is why the sinc function is referred to as the *ideal* low-pass filter. The limitation of the Shannon's paradigm motivates studies of generalized sampling theory [16,38], which aim recovery from an arbitrary basis beyond bandlimited signals.

Stability of the ideal low-pass filter Let $w(t) = \mathrm{sinc}(2f_c(t))$ be a reconstruction kernel. The sinc function has a finite energy , i.e., $\{w(t-nT)\}_{n\in\mathbb{Z}} \in L_2$. This implies that the reconstruction process can be stable in the energy sense. To see this, suppose that sampled signals $c[n]$ are corrupted by noise, i.e., $y[n] = c[n] + \eta[n]$. Then, reconsturcted error can be expressed by $\eta(t) = \sum_n \eta[n]w(t-nT)$. The reconstruction process via $w(t)$ is stable in the energy sense if the following inequality is satisfied [16, 39]

$$\langle \eta(t), \eta(t)\rangle = \int_{-\infty}^{\infty} \eta^2(t)dt \le C\sum_{n\in\mathbb{Z}} \eta^2[n]. \tag{2.61}$$

[2]The non-causality of a filter becomes a issue for time domain signals. In contrast, the non-causality does not matter for spatial domain signals like images.

By utilizing the orthogonality of sinc funciton,

$$\langle \eta(t), \eta(t) \rangle = \left\langle \sum_{n\in\mathbb{Z}} \eta[n] w(t-nT), \sum_{m\in\mathbb{Z}} \eta[m] w(t-mT) \right\rangle = T^2 \sum_{n\in\mathbb{Z}} \eta^2[n]. \tag{2.62}$$

Therefore, sinc function is stable by viewing $T^2 = C$.

On the other hand, the sinc function is not integrable, i.e., $\{w(t-nT)\}_{n\in\mathbb{Z}} \notin L_1$. In other words, the reconstruction process cannot be BIBO stable. In order to guarantee the BIBO stability, one needs to satisfy

$$\left| \sum_{n\in\mathbb{Z}} c[n] w(t_0 - nT) \right| \leq \sum_{n\in\mathbb{Z}} |c[n]| \cdot \sum_{n\in\mathbb{Z}} |w(t_0 - nT)| < \infty, \tag{2.63}$$

for all instances t_0. Suppose that $c[n]$ is absolutely summable. The BIBO condition is necessary to hold

$$\sup_{t_0\in[0,T]} \sum_{n\in\mathbb{Z}} |w(t_0 - nT)| < \infty. \tag{2.64}$$

Note that the sequence sampled from the sinc function is in general not absolutely summable, i.e., $\{w(t_0 - nT)\} \notin \ell_1$ for $t_0 \in [0, T]$, due to the above reason. Therefore, the reconstruction process via the sinc function is not BIBO stable.

As observed so far, Shannon's theorem has several limitations due to the ideal low-pass filter, which is non-smooth and discontinuous. In the next section, we will redefine sampling within the context of a general vector space to lay the groundwork for introducing generalized sampling theory.

2.4 Sampling in general vector spaces

In this section, we generalize sampling in SI spaces to Hilbert spaces. First, we review the basics of vector spaces. Next, we redefine sampling as a *linear transform* within this context and formulate a generalized sampling framework.

Basics of vector space

We introduce the notion of Riesz and biorthogonal bases, which play a significant role in generalized sampling theory.

Riesz basis

In this dissertation, the signal $x(t)$ is assumed to be expressed by the linear combination of vectors with an appropriate basis $\{a_n\}$, i.e., $x = \sum_n d[n]a_n$. Here, to guarantee the unique decomposition of x of the form $x = \sum_n d[n]a_n$, we consider that x lies in a Hilbert space $\mathcal{H}$ where $\{a_n\}$ forms a well-conditioned basis for $\mathcal{H}$. To this aim, the following condition is desired.

Definition 4 (Riesz basis condtion)**.** *A sequence $\{a_n \in \mathcal{H}, n \in \mathcal{I}\}$ forms a Riesz basis for $\mathcal{H}$ if it is complete and there exist constants $\alpha > 0$ and $\beta < \infty$ such that*

$$\alpha \sum_{n\in\mathcal{I}} |d[n]|^2 \leq \left\| \sum_{n\in\mathcal{I}} d[n]a_n \right\|^2 \leq \beta \sum_{n\in\mathcal{I}} |d[n]|^2, \tag{2.65}$$

for all $d \in \ell_2$.

Along with a Riesz basis, we may use the following operator.

Definition 5 (Set transformation)**.** *Let $\{a_n : n \in \mathcal{I}\}$ be a countable set vectors in a Hilbert space $\mathcal{H}$. The set transformation $A : \ell_2 \to \mathcal{H}$ corresponding to these vectors is defined by $Ad = \sum_{n\in\mathcal{I}} d[n]a_n$ for any $d[n] \in \ell_2$.*

Throughout this dissertation, we denote the adjoint of A by $A^* : \mathcal{H} \mapsto \ell_2$. By definition of the adjoint, i.e., $\langle Ad, y\rangle = \langle d, A^*y\rangle$, we have

$$\begin{aligned} \langle Ad, y\rangle &= \langle \sum_{n\in\mathcal{I}} d[n]a_n, y\rangle \\ &= \sum_{n\in\mathcal{I}} d^*[n] \langle a_n, y\rangle \\ &= \langle d, A^*y\rangle . \end{aligned} \tag{2.66}$$

Therefore, we can conclude that $A^*y = \langle a_n, y\rangle$.

In terms of the set transformation A, (2.65) is simplified to

$$\alpha \langle d, d\rangle \leq \langle d, A^*Ad\rangle \leq \beta \langle d, d\rangle . \tag{2.67}$$

We summarize the merit of using the Riesz basis below.

Remark 1. If the corresponding set vectors of A form a Riesz basis for $\mathcal{H}$, A^*A is bounded, which implies that A^*A is invertible. This property is useful for constructing sampling theory in Hilbert spaces. We will see this in the following section.

We introduce another important properties of a Riesz basis below.

SI expression In the SI setting, we have the following expression:

$$\begin{aligned} A^*Ad &= \left[a(t) * \sum_{n\in\mathbb{Z}} d[n]a(t-nT) \right]_{t=nT} \\ &\leftrightarrow \frac{1}{T} D(e^{j\omega}) \sum_{k\in\mathbb{Z}} \left| A\left(\frac{\omega - 2\pi k}{T}\right)\right|^2 \\ &\leftrightarrow R_{AA}(e^{j\omega}) \cdot D(e^{j\omega}), \end{aligned} \tag{2.68}$$

where $\leftrightarrow$ denotes the equivalent expression in Fourier domain and

$$R_{AA}(e^{j\omega}) \coloneqq \sum_{k\in\mathbb{Z}} \left| A\left(\frac{\omega - 2\pi k}{T}\right)\right|^2. \tag{2.69}$$

Therefore, we can recast the Riesz condition in Definition 4 in the SI space.

Theorem 2 (Riesz condition in Fourier domain)**.** *The signals* $\{a(t-nT)\}$ *form a Riesz basis for a SI space* $\mathcal{H}$ *if and only if there exists* $\alpha > 0$, $\beta < \infty$ *such that* $\alpha \le R_{AA}(e^{j\omega}) \le \beta$ *almost everywhere*[3]*,* $R_{AA}(e^{j\omega})$ *is defined by* (2.69)*.*

Proof. Applying (2.68) to (2.67), the proof is completed. □

Stability of a Riesz basis The reconstruction process through the Riesz basis is stable in the energy sense. Here, we also assume SI sampling for simplicity. Let $\{a(t-nT)\}_{n\in\mathbb{Z}}$ be a Riesz basis for SI spaces $\mathcal{H}$. Let $y[n] = c[n] + \eta[n]$ denote the noisy sample. Then, reconstructed error is given by $\eta(t) = \sum_n \eta[n]a(t-nT)$. Again, the sample is stable in the energy sense if

$$\langle \eta(t), \eta(t) \rangle = \int_{-\infty}^{\infty} \eta^2(t)dt \le C \sum_{n\in\mathbb{Z}} \eta^2[n] \tag{2.70}$$

By comparing with (2.65), it is clear that (2.70) accords with the upper bound in the Riesz condition.

Biorthogonal basis

We next introduce a biorthogonal (Riesz) basis in Hilbert spaces. We start by the definition of biorthogonal basis in the following:

[3]Roughly speaking, the term *almost everywhere* means that everywhere except for a discontinuity.

Definition 6 (Biorthogonal basis)**.** *Let $\{a_n, n \in \mathcal{I}\}$ be a Riesz basis for $\mathcal{H}$. Then, there exists an unique vector set $\{\tilde{a}, n \in \mathcal{I}\}$, namely, biorthogonal basis, such that*

$$\langle \tilde{a}_n, a_m \rangle = \delta_{n,m}. \tag{2.71}$$

In terms of the set transformation, we have the following corollary.

Corollary 1. *Let A and $\widetilde{A}$ be set transformations corresponding to $\{a_n\}$ and $\{\tilde{a}_n\}$, respectively. Then, Definition 6 can be expressed by*

$$\widetilde{A}^* A = I_{\ell_2}, \tag{2.72}$$

I_{ℓ_2} is the identitiy transformation for ℓ_2.

In the following, we describe two explicit designs of the biorthogonal basis.

Orthonormal basis Denoting the range space of the set transformation A by $\mathcal{A} \subset \mathcal{H}$, we would like to seek the biorthogonal basis on $\mathcal{A}$. By the Riesz condition in Definition 4, A^*A and $(A^*A)^{-1}$ are bounded. Hence, we can choose the set transformation corresponding to the biorthogonal basis $\{\tilde{a}_n\}$ as

$$\widetilde{A} = A(A^*A)^{-1}. \tag{2.73}$$

It is easy to verify that $\widetilde{A}$ satisfies (2.72).

We next show the SI counterpart of (2.73). Recall that the Riesz condition is given by Theorem 2 in the SI setting. For an arbitary Riesz basis $\{a(t - nT)\}$ with the CTFT $A(\omega)$, the biorthogonal basis $\{\tilde{a}(t - nT)\}$ with the CTFT $\widetilde{A}(\omega)$ can be obtained by

$$\widetilde{A}(\omega) = \frac{1}{R_{AA}(e^{j\omega T})} A(\omega), \tag{2.74}$$

where $R_{AA}(e^{j\omega})$ is defined by (2.69).

Oblique biorthogonal-basis In general, we may seek the biorthogonal basis for $\mathcal{A}$ on the different subspace $\mathcal{B}$ such that $\mathcal{A} \cup \mathcal{B} = \mathcal{H}$. Since this problem involves two subspaces, we exploit the following theorem.

Theorem 3. *Let the vectors $\{a_n, n \in \mathcal{I}\}$ form a Riesz basis for a subspace $\mathcal{A}$ of a Hilbert space $\mathcal{H}$, and let the vectors $\{b_n, n \in \mathcal{I}\}$ form a Riesz basis for a subspace $\mathcal{B}$ of $\mathcal{H}$. Let $A : \ell_2 \to \mathcal{H}$ and $B : \ell_2 \to \mathcal{H}$ denote the set transformations corresponding to the vectors $\{a_n, n \in \mathcal{I}\}$ and $\{b_n, n \in \mathcal{I}\}$, respectively. Then, B^*A is invertible if and only if $\mathcal{A} \oplus \mathcal{B}^\perp = \mathcal{H}$.*

The proof can be found in [16, 40].

With Theorem 3, we can express an arbitrary obiorthogonal basis as

$$\widetilde{A} = B(A^*B)^{-1}. \tag{2.75}$$

It is easily verifiable that $\widetilde{A}$ satifies (2.72). We can view (2.72) as the space case of (2.75) where $B = A$.

In order to show the SI counterpart of (2.75), we utilize the following corollary.

Corollary 2. *Let $\mathcal{A}, \mathcal{S}$ be SI spaces with bases $\{a(t - nT)\}, \{s(t - nT)\}$. Let $A(\omega), S(\omega)$ be the CTFTs of $a(t), s(t)$, respectively. Then, $L_2 = \mathcal{A} \oplus \mathcal{S}^\perp$ if and only if there exists a constant $\alpha > 0$ such that*

$$|R_{SA}(e^{j\omega})| > \alpha, \quad \textit{almost everywhere.} \tag{2.76}$$

Supposing that Corollary 2 is satisfied, we can rewrite (2.75) as

$$\widetilde{A}(\omega) = \frac{1}{R_{AS}(e^{j\omega T})} A(\omega). \tag{2.77}$$

We are now prepared to reformulate sampling within general vector spaces. In the following, we introduce a generalized sampling framework.

Generalized sampling framework

Let s_n be an arbitrary Riesz basis and $S : \ell_2 \to \mathcal{H}$ be its set transformation. Suppose that $\{s_n\}$ spans a sampling subspace $\mathcal{S} \subset \mathcal{H}$. Using the set transformation notation, the sequence of sampled signals $c \in \ell_2$ can be expressed by $x = S^*x$ and its element $c[n]$ is given by

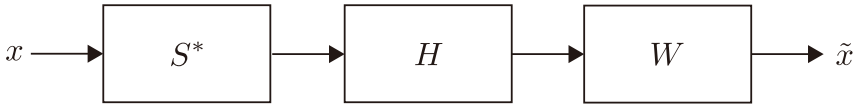
$$c[n] = \langle s_n, x \rangle, \tag{2.78}$$

where $S^* : \mathcal{H} \to \ell_2$ is the adjoint operator of S, i.e., $\langle S^*a, b \rangle = \langle a, Sb \rangle$ for arbitrary $a \in \mathcal{H}$ and $b \in \ell_2$.

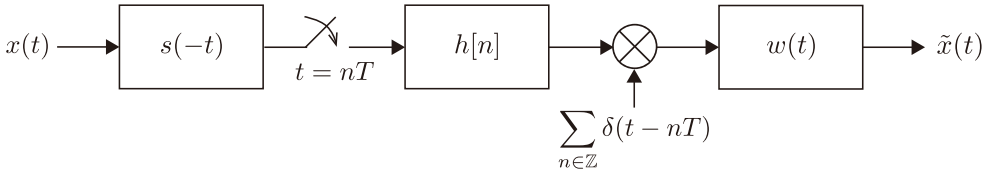
Let w_n be an arbitrary Riesz basis and $W : \ell \to \mathcal{H}$ be its set transformation. Suppose that $\{w_n\}$ spans a reconstruction subspace $\mathcal{W}$. Correspondingly, the reconstructed signal is given by

$$\tilde{x} = \sum_n w_n \tilde{d} = W\tilde{d}, \tag{2.79}$$

where $\tilde{d} \in \ell_2$ is the sequence of appropriate expansion coefficients whose element is $d[n]$. In general, $\{s_n\}$ and $\{w_n\}$ may differ due to some practical limitations. Therefore, we would like



(a) Hilbert space expression



(b) SI space expression

Figure 2.3: Generalized sampling framework.

to seek the best $\tilde{d}$ with arbitrary vectors $\{s_n\}, \{w_n\}$. Here, we insert the following processing prior to the reconstruction

$$\tilde{d} = Hc, \tag{2.80}$$

where $H : \ell_2 \to \ell_2$ is referred to as a correction transform. As a result, generalized sampling aims to design H so that $\tilde{x}$ is close to x in some sense. We depict the framework in Fig. 2.3. In the following section, we will describe the optimal design of H in detail.

Indeed, generalized sampling encompasses a wider class of signals. We introduce promising examples of signal subspaces below.

Sampling in multiple SI-spaces Since the generalized sampling framework does not specify a single-variable sample, we can naturally extend the standard single-channel sampling to multi-channel sampling. Let $\{s_\ell(t - nT)\}_{\ell=1,\dots,M,\, n\in\mathbb{Z}}$ be a Riesz basis for a sampling space $\mathcal{S}$, where M is the number of channels. The sample in the ℓth channel is given by

$$c_\ell[n] = \langle s_\ell(t - nT), x \rangle \,. \tag{2.81}$$

In this setting, x is assumed to be composed of a linear combination of multi-generators. Let $\{w_\ell(t - nT)\}_{\ell=1,\dots,M,\, n\in\mathbb{Z}}$ be a Riesz basis for a reconstruction space $\mathcal{W}$. Then, the reconstructed signal is given by

$$\tilde{x}[n] = \sum_{n\in\mathbb{Z}} \sum_{\ell=1}^{M} \tilde{d}_\ell[n] w_\ell(t - nT), \tag{2.82}$$

where $\tilde{d}_\ell$ is an appropriate expansion coefficient.

Sampling in reproducing kernel Hilbert spaces (RKHSs) Although the generalized sampling framework, in principle, allows for an arbitrary linear sampling process, we can possibly extend it to a non-linear sampling process. We consider a signal $x \in \mathcal{H}$ defined in an arbitrary RKHS $\Omega \subset \mathcal{H}$. Let $\{\phi_{t_n n}\}$ be a Riesz basis for $\mathcal{S} \subset \mathcal{H}$. The sampled signal is defined as

$$c(t_n) = \langle x(t), \phi_{t_n} \rangle , \quad x \in \Omega. \tag{2.83}$$

Let k be the reproducing kernel induced by the inner product of $\{\phi_{t_n n}\}$, where $\{k(\cdot, x_n)\}$ forms Riesz basis for Ω. Then, the reconstructed signal is given by [41]

$$\tilde{x}(t) = \sum_{n \in \mathbb{Z}} \tilde{d}(t_n) k(t, t_n), \tag{2.84}$$

where $\tilde{d}$ is an appropriate expansion coefficient.

In (2.84), we did not specify concrete designs for the recovery transform. In the following section, we summarize representative designs commonly used in practice.

Design of recovery

We describe the setup of recovery and criteria for recovery in detail. We consider two scenarios in recovery, where reconstruction process is constrained a priori due to practical limitations [16, 38]:

Unconstrained recovery Without consideration of practical limitations, our problem is to directly recover the sampled signal. In this setting, there are no constraints in the reconstruction (interpolation) process. Since the recovery process is formulated by $\tilde{x} = WHc$, we need to optimize W jointly with H based on the appropriate criterion. If the signal subspace is known, denoted by $\mathcal{A}$, we can perfectly recover the original signals. Even when $\mathcal{A}$ is not known but other priors are available, we can still estimate the reliable signal subspace and recover the appropriate signal.

Predefined recovery If the unconstrained recovery is difficult to implement, we may resort to the predefined interpolation method, which is easy to implement. In this setting, the reconstruction process is constrained by the imposed kernel. Therefore, we only need to design H to compensate for a non-ideal behavior of S as well as W. The predefined solution differs from unconstrained one in that the signal subspace is fixed by the predefined kernel, denoted by $\mathcal{W}$. This implies that we can never achieve perfect recovery. Instead, we seek the best possible signal on $\mathcal{W}$ with the appropriate criterion.

In principle, the reconstruction error $\|\tilde{x} - x\|$ should be minimized. However, we can never directly access to the original signal to minimize it. Therefore, $\|\tilde{x} - x\|$ is replaced by another appropriate error. In what follows, we introduce two criteria widely-studied in the literature of signal processing [16].

Least squares (LS) criterion Recalling that the sequence of sampled signals is $c = S^*x$, the LS criterion seeks the signal that minimizes the error-in-sample, i.e.,

$$\tilde{x} = \underset{\substack{x \in \mathcal{R}, \\ \mathcal{R} \in \{\mathcal{Q}, \mathcal{W}\}}}{\arg\min} \|S^*x - c\|^2, \tag{2.85}$$

where $\mathcal{Q}$ and $\mathcal{W}$ are the sets of signals under the given prior and reconstruction constraint, respectively. Because of the relatively simple design, the LS criterion has been widely used in the inverse problem. However, it is also well-known that the LS solution is sensitive to noise in samples and leads to a large error. Another drawback of the LS criterion is that it can only incorporate either $\mathcal{Q}$ or $\mathcal{W}$ into the solution, as long as $\mathcal{Q} \neq \mathcal{W}$. This is because the freedom in (2.85) is only on x and either $\mathcal{Q}$ or $\mathcal{W}$ can impose on x. Due to this limitation, the LS solution in the constrained case may generally lead to poor performance.

Minimax (MX) criterion The MX criterion corresponds to the reconstruction that minimizes the worst-case error, i.e.,

$$\tilde{x} = \underset{\tilde{x} \in \mathcal{W}}{\arg\min} \max_{x \in \mathcal{Q},\ S^*x = c} \|\tilde{x} - x\|^2. \tag{2.86}$$

Since (2.86) handles the reconstruction error $\|\tilde{x} - x\|$, we expect that the MX recovery results in superior performance than the LS recovery in many cases. Furthermore, there are two freedom in (2.86), x and $\tilde{x}$. This implies that the MX criterion can incorporate the signal prior $\mathcal{Q}$ jointly with the reconstruction constraint $\mathcal{W}$ into the solution. As a result, the MX solution can find the better compromise than the LS solution in the predefined setting, where $\tilde{x}$ is close to $\mathcal{Q}$ in $\mathcal{W}$.

We have thus far not specified concrete signal priors. In the next section, we introduce two well-known signal priors and derive optimal recovery transforms tailored to each prior.

2.5 Generalized sampling theory

In this section, we formulate recovery problems based on the aforementioned designs and provide the optimal correction filter designs. We focus on recovery under two well-known priors. For simplicity, we omit derivations of all problem herein. Their proofs can be found in [16]. We will briefly introduce those solutions below.

Subspace priors

When a signal is known to lie in the subspace $\mathcal{A}$ of Hilbert space $\mathcal{H}$, i.e., $\mathcal{A} \subset \mathcal{H}$, it is possible to recover arbitrary signals in $\mathcal{A}$ from their sample, which are not necessarily ideally sampled. This is achieved by selecting the appropriate criterion, including the LS and MX criteria, in both cases of the unconstrained and predefined reconstruction. While we focus on recovery in the SI space to highlight its elegant expression, all results herein hold in arbitrary Hilbert spaces.

Unconstrained recovery

Technically, the solution in (2.85) is generally not unique and there may be numerous solutions. Alternatively, we seek the minimum norm solution in (2.85) as follows:

$$\tilde{x} = \underset{x \in \mathcal{Q}, S^*x = c}{\arg\min} \|x\|^2. \tag{2.87}$$

Solving (2.87) with $\mathcal{Q} = \mathcal{A}$, the optimal solution is given by

$$H_{\text{ULS}} = (S^*A)^\dagger, \tag{2.88}$$

where † denotes the Moore-Penrose inverse and the associated reconstruction operator is $W = A$.

In fact, the MX solution in the unconstrained case eventually coincides with (2.88) (see [16]). Particularly, given $\mathcal{W} = \mathcal{Q} = \mathcal{A}$, (2.86) is reduced to (2.87).

Suppose that $\mathcal{S}$ and $\mathcal{A}$ span $\mathcal{H}$ and intersect at the origin, called the direct sum (DS) condition, i.e., $\mathcal{H} = \mathcal{A} \oplus \mathcal{S}^\perp$. Then, the recovery transform based on the LS and MX criteria in the unconstrained case is given by

$$\tilde{x} = A(S^*A)^{-1}c. \tag{2.89}$$

Recall that $A(S^*A)^{-1}$ corresponds to the (oblique) biorthogonal basis in $\mathcal{A}$. Therefore, when the DS condition is satisfied, the sampled signal is perfectly recovered from an arbitrary sampling operator.

Predefined recovery

In this setting, the reconstruction operator is constrained by W a priori. Solving (2.85) with $\mathcal{Q} = \mathcal{W}$, the optimal solution is given by

$$H_{\text{PLS}} = (S^*W)^\dagger. \tag{2.90}$$

Note that this solution ignores the signal prior $\mathcal{A}$.

Next, solving (2.86) with $\mathcal{Q} = \mathcal{A}$, the optimal solution is given by

$$H_{\text{PMX}} = (W^*W)^{-1}W^*A(S^*A)^\dagger. \tag{2.91}$$

SI expression All solutions introduced in this section have the same form as follows:

$$H = \begin{cases} \frac{R_{WQ}(e^{j\omega})}{R_{WW}(e^{j\omega})R_{SQ}(e^{j\omega})} & R_{SQ}(e^{j\omega}) \neq 0 \\ 0 & R_{SQ}(e^{j\omega}) = 1 \end{cases}, \tag{2.92}$$

where $R_{WQ}(e^{j\omega})$ is defined in (2.69), and $W(\omega)$ and $Q(\omega)$ denote filters corresponding to $\mathcal{W}$ and $\mathcal{Q} = \mathcal{A}$, respectively.

Smoothness priors

We have shown that an arbitrarily sampled signal can perfectly be recovered under the DS condition in the previous section. In this section, we tackle with a milder problem, in which a signal is smooth in a norm sense. In this setting, the signal subspace is explicitly not confined, but the signal norm weighted by an appropriate measuring operator V, i.e., $\|Vx\|$, is bounded. Typically, V is designed by a sharpning (high-pass) filter. Although perfect recovery is no longer possible here, we can still achieve the best possible recovery based on the smoothness prior. Similar to the subspace prior, we utilize the LS and MX criteria in both of the unconstrained and predefined reconstruction filters. Suppose that V is bounded. We consider the following signal class

$$\mathcal{V} = \{x \mid \|Vx\| \leq \rho\}. \tag{2.93}$$

Unconstrained recovery

In the same scenario as the subspace prior, (2.85) may have numerous solutions under the smoothness prior. Alternatively, we seek the smoothest solution in (2.85) by solving

$$\tilde{x} = \underset{x \in \mathcal{Q},\ S^*x = c}{\arg\min} \|Vx\|^2. \tag{2.94}$$

With $\mathcal{Q} = \mathcal{V}$, the optimal solution in (2.94) is given by

$$H_{\text{ULS}} = (S^*\widetilde{W})^{-1}, \tag{2.95}$$

where $\widetilde{W} = (V^*V)^{-1}S$. The optimal reconstruction operator is $\widetilde{W}$.

Like recovery under the subspace prior, the MX solution under the smoothness prior coincides with (2.95). Particularly, given $\mathcal{W} = \mathcal{H}$ and $\mathcal{Q} = \mathcal{V}$, (2.86) is reduced to (2.94).

Predefined recovery

Recall that, in principle, (2.85) cannot simultaneously take into account a signal prior $\mathcal{Q}$ and reconstruction constraint $\mathcal{W}$. Nevertheless, the smoothness prior $\mathcal{V}$ can be merged with $\mathcal{W}$, i.e.,

$$\mathcal{V}' = \{x \mid x \in \mathcal{W}, \|Vx\| \leq \rho\}. \tag{2.96}$$

Since $\mathcal{V}'$ is the convex set, we can incorporate two priors, namely $\mathcal{V}'$, into the LS solution. Solving (2.85) with $\mathcal{Q} = \mathcal{V}'$, the optimal solution is given by

$$H_{\text{PLS}} = (W^*V^*VW)^{-1}W^*S(S^*\widehat{W})^{-1}, \tag{2.97}$$

where $\widehat{W} = W(W^*V^*VW)^{-1}W^*S$.

Solving (2.86) with $\mathcal{Q} = \mathcal{V}$, the optimal solution is given by

$$H_{\text{PMX}} = (W^*W)^{-1}W^*\widetilde{W}(S^*\widetilde{W})^{-1}, \tag{2.98}$$

where $\widetilde{W} = (V^*V)^{-1}S$.

Subspace interpretation The underlying signal subspace is explicitly not given under the smoothness prior. Nevertheless, the recovered signal with (2.95) lies in the subspace determined by $\widetilde{\mathcal{W}} = \mathcal{R}(\widetilde{W})$. In other words, $\widetilde{\mathcal{W}}$ can be viewed as the best possible signal subspace for $\mathcal{V}$. Similarly, $\widehat{\mathcal{W}} = \mathcal{R}(\widehat{W})$ is the best possible signal subspace for $\mathcal{V}'$. In this sense, for the sampling operator S^*, $\widetilde{W}(S^*\widetilde{W})^{-1}$ and $\widehat{W}(S^*\widehat{W})^{-1}$ are associated by oblique biorthogonal bases in $\widetilde{\mathcal{W}}$ and $\widehat{\mathcal{W}}$, respectively.

SI expression Interestingly, all solutions in this section also have the same form as (2.92). As mentioned the above, $\widetilde{\mathcal{W}}$ and $\widehat{\mathcal{W}}$ can be viewed as optimal subspaces for $\mathcal{V}$ and $\mathcal{V}'$, respectively. Let $\widetilde{W}(\omega)$ and $\widehat{W}(\omega)$ denote filters corresponding to $\widetilde{\mathcal{W}}$ and $\widehat{\mathcal{W}}$, respectively. Then, $Q(\omega)$ in (2.92) is replaced by either $\widetilde{W}(\omega)$ or $\widehat{W}(\omega)$. Throughout this dissertation, $Q(\omega)$, is called a *prior filter*, which is replaceable with an arbitrary filter depending on the signal prior.

In the above regard, we can view (2.92) as a recovery transform tailored to arbitrary signal priors, specifically within the context of subspace and smoothness priors. Interestingly, we will see that this perspective plays a key role throughout this dissertation.

We have thoroughly reviewed the groundwork of sampling in SI spaces. In the next section, we transition to reviewing the graph space, which serves as an analog to time-domain signals in some aspects but is unique in others.

2.6 Basics of graph signal processing

In this section, we review the spectral analysis of graph signals, which plays a key role in GSP. We also introduce the graph counterpart of the SI system, which is advantageous for developing GSP applications.

Graphs and graph signals

We consider a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ represent sets of nodes and edges, respectively. The number of nodes is $N = |\mathcal{V}|$ unless otherwise specified. The adjacency matrix of $\mathcal{G}$ is denoted by $\boldsymbol{\mathcal{W}}$ where its (m, n)-element $[\boldsymbol{\mathcal{W}}]_{mn} \geq 0$ is the edge weight between the mth and nth nodes; $[\boldsymbol{\mathcal{W}}]_{mn} = 0$ for unconnected nodes. The degree matrix $\mathbf{D}$ is defined as $\mathbf{D} = \mathrm{diag}\,(d_0, d_1, \ldots, d_{N-1})$, where $d_m = \sum_n [\boldsymbol{\mathcal{W}}]_{mn}$ is the mth diagonal element. A graph signal $\boldsymbol{x} \in \mathbb{R}^N$ is defined as a mapping from the node set to the set of real numbers, i.e., $x[n] : \mathcal{V} \to \mathbb{R}$.

Graph variation operator In this dissertation, we refer to graph Laplacian $\mathbf{L} \coloneqq \mathbf{D} - \boldsymbol{\mathcal{W}}$ as a *graph variation operator*. In fact, $\mathbf{L}$ plays a significant role in GSP to capture the signal variation on the graph. Here, the Laplacian quadratic form is defined as

$$\boldsymbol{x}^{\mathsf{T}}\mathbf{L}\boldsymbol{x} = \sum_{(n,m)\in\mathcal{E}} [\boldsymbol{\mathcal{W}}]_{mn}(x[n] - x[m])^2 \tag{2.99}$$

$$= \sum_{i=0}^{N-1} \lambda_i \hat{x}(\lambda_i)^2. \tag{2.100}$$

Since $\mathbf{L}$ is positive semidefinite, $\boldsymbol{x}^{\mathsf{T}}\mathbf{L}\boldsymbol{x} \geq 0$ always holds. If the signal differences $|x[n] - x[m]|$ are large among some nodes n and m, $\boldsymbol{x}^{\mathsf{T}}\mathbf{L}\boldsymbol{x}$ may become large, i.e., graph signals vary greatly. In contrast, when $\boldsymbol{x}^{\mathsf{T}}\mathbf{L}\boldsymbol{x}$ is small, graph signals seem to be smooth on the graph $\mathcal{G}$. Therefore, $\mathbf{L}$ reflects the smoothness on $\mathcal{G}$.

Graph Fourier transform

The graph Fourier transform (GFT) of $\boldsymbol{x}$ is defined as

$$\hat{x}(\lambda_i) = \langle \boldsymbol{u}_i, \boldsymbol{x} \rangle = \sum_{n=0}^{N-1} u_i[n]x[n], \tag{2.101}$$

where $\boldsymbol{u}_i$ is the ith column of a unitary matrix $\mathbf{U}$ and it is obtained by the eigendecomposition of the graph Laplacian $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathsf{T}}$ with the eigenvalue matrix $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_0, \lambda_1, \ldots, \lambda_{N-1})$.

Correspondingly, the inverse GFT is given by

$$x[n] = \sum_{i=0}^{N-1} \hat{x}(\lambda_i) u_i[n]. \tag{2.102}$$

Without of loss of generality, we suppose that $\{\lambda_i\}_{i=1,\dots,N}$ are arranged in ascending order, i.e., $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1}$. We refer to λ_i as a *graph frequency*.

Most of regular properties of Fourier transform, such as convolution, modulation, shift, and scaling, do not generally hold in the graph setting. In what follows, we introduce a few properties of GFT.

Energy conservation Since $\mathbf{U}$ is a unitary matrix, the GFT does not change a signal energy, i.e.,

$$\sum_{n=0}^{N-1} |x[n]|^2 = \sum_{n=0}^{N-1} |\hat{x}(\lambda)|^2. \tag{2.103}$$

Since GSP is generally not SI, there may be fewer beneficial links between vertex and frequency domains compared to standard signal processing. In the following, we demonstrate this distinction and present its desirable analog.

Node-localized system

Several favorable properties in SI spaces are absent in the graph setting due to its non-SI nature. Still, the node-localization invariant (NLI) system, which is a counterpart of the SI system, is widely-used in GSP [42].

Localization There is no regular shift in GSP due to the irregular incidence of nodes. To develop a counterpart of shift, we introduce the notion of *localization*: The signal variation is characterized by filtering whose response is localized at several nodes. The localization operator is defined as [42]

$$\mathcal{L}_m\{x[n]\} \coloneqq \sum_{i=0}^{N-1} \hat{x}(\lambda_i) u_i[n] u_i[m]. \tag{2.104}$$

In the vector form, (2.104) is given by

$$\boldsymbol{\mathcal{L}}_m \boldsymbol{x} = \mathbf{U}\mathrm{diag}(\mathbf{U}^\mathsf{T}\boldsymbol{\delta}_m)\mathbf{U}^\mathsf{T}\boldsymbol{x}, \tag{2.105}$$

where $\boldsymbol{\delta}_m$ is the Kronecker delta centered at the node m.

Generalized convolution A convolution (filtering) is a fundamental processing in signal processing. Similar to the time domain signal, we would like to build the following dual relationship.

$$x[n] \star y[n] \leftrightarrow \hat{x}(\lambda_i)\hat{y}(\lambda_i), \tag{2.106}$$

where generalized convolution is defined as [42]

$$\begin{aligned} x[n] \star y[n] &\coloneqq \sum_{\ell=0}^{N-1} \hat{x}(\lambda_\ell)\hat{y}(\lambda_\ell)u_\ell[n] && (2.107)\\ &= \sum_{m=0}^{N-1} x[n] \sum_{\ell=0}^{N-1} \hat{y}(\lambda_\ell)u_\ell[n]u_\ell[m] \\ &= \sum_{m=0}^{N-1} x[n]\mathcal{L}_n\{y[m]\}. && (2.108) \end{aligned}$$

In the vector form, (2.108) is given by

$$\boldsymbol{x} \star \boldsymbol{y} = \mathbf{U}\mathrm{diag}(\hat{\boldsymbol{y}})\mathbf{U}^\mathsf{T}\boldsymbol{x}. \tag{2.109}$$

Node-localization invariance As aforementioned, the significant difference between the standard signal processing and GSP is that a GSP is generally not SI due to irregularly-distributed node degrees. Nevertheless, the node localization parallels SI in some respects. Let $y[n] = \mathcal{T}\{x[n]\}$ be the output of a system. $\mathcal{T}$ is a NLI system if and only if

$$\mathcal{T} \circ \mathcal{L}_m\{x[n]\} = \mathcal{L}_m\{y[n]\} \quad \text{for all } m \in \mathcal{V} \tag{2.110}$$

where $f \circ g$ denotes applying g and following f.

Recall that, in the matrix form, the localization operator is given by $\boldsymbol{\mathcal{L}}_m = \mathbf{U}\mathrm{diag}(\mathbf{U}^\mathsf{T}\boldsymbol{\delta}_m)\mathbf{U}^\mathsf{T}$. A system $\boldsymbol{\mathcal{T}}$ is NLI, i.e., $\boldsymbol{\mathcal{T}}\boldsymbol{\mathcal{L}}_m = \boldsymbol{\mathcal{L}}_m\boldsymbol{\mathcal{T}}$ holds, if $\boldsymbol{\mathcal{L}}_m$ and $\boldsymbol{\mathcal{T}}$ are simultaneously diagonalizable by $\mathbf{U}$. In particular, $\boldsymbol{\mathcal{T}}$ is a NLI system if $\boldsymbol{\mathcal{T}}$ is a polynomial of $\mathbf{L}$, i.e., $\boldsymbol{\mathcal{T}} = \sum_{\ell=0}^{L} \alpha_\ell(\mathbf{L})^\ell$, which is a sufficient condition to be diagonalizable by $\mathbf{U}$.

Impulse response By utilizing (2.108), we can characterize the NLI system with the impulse response. A NLI system in (2.110) can be rewritten by

$$\begin{aligned}
\mathcal{T} \circ \mathcal{L}_m\{x[n]\} &= \mathcal{T} \circ \mathcal{L}_m \left\{ \sum_{\ell=0}^{N-1} x[\ell]\delta_\ell[n] \right\} \\
&= \sum_{\ell=0}^{N-1} x[\ell]\mathcal{T} \circ \mathcal{L}_m \left\{\delta_\ell[n]\right\} \\
&= \sum_{\ell=0}^{N-1} x[\ell]\mathcal{T} \circ \mathcal{L}_\ell \left\{\delta_m[n]\right\} \\
&= \sum_{\ell=0}^{N-1} x[\ell]\mathcal{L}_\ell \left\{h_m[n]\right\} \\
&= x[n] \star h_m[n], \qquad (2.111)
\end{aligned}$$

where $\mathcal{T}\{\delta_m[n]\} = h_m[n]$ for all m. The second equality is the linearity. In the third equality, instances m and ℓ are commutative by definition of $\mathcal{L}_m$. The forth equality is followed by the NLI condition in (2.110).

Although (2.111) parallels SI properties in (2.24) and (2.7), the significant difference is that the NLI system generally has N unique impulse responses $\{h_m[n]\}_{m\in\mathcal{V}}$ for each centered nodes. Therefore, we cannot express the NLI system with a single impulse response in contrast to a SI system.

We often encounter NLI systems in various studies on GSP [25, 26, 42]. In the next section, we will introduce one of the widely-studied signal generation models, bandlimited graph signals, and conventional sampling approaches for this model."

2.7 Sampling framework of graph signals

While we assume sampling with uniform intervals in SI space, sampling of graph signals is not unique due to the irregular connectivity of nodes. Therefore, we need to design a sampling strategy as well as a recovery problem in contrast to the SI setting. In this section, we introduce conventional sampling approaches for graph signals.

Bandlimited sampling

There are many attempts to build efficient sampling strategies for graph signals [10–12, 14, 20, 43]. They mainly differ from deterministic/random approaches and their cost functions.

On the other hand, most of studies assume that a graph signal is bandlimited. Here, we assume that a graph signal lies in a following subspace:

$$\mathcal{A}_{\mathrm{BL}} \coloneqq \left\{ \boldsymbol{x} \,\middle|\, \boldsymbol{x} = \mathbf{U}_{\mathcal{B}}\boldsymbol{d}, \text{ for some } \boldsymbol{d} \in \mathbb{R}^{|\mathcal{B}|} \right\}, \tag{2.112}$$

where $\mathcal{B} \in \{1, \ldots, K\}$ is a subset of the low-passband with the bandwidth K.

Suppose that sampled signal is given by

$$\boldsymbol{c} = \mathbf{I}_{\mathcal{M}\mathcal{V}}\boldsymbol{x} + \boldsymbol{\eta}, \tag{2.113}$$

where $\mathcal{M} \subset \mathcal{V}$, $\mathbf{I}_{\mathcal{X}\mathcal{Y}}$ denotes the submatrix that retains rows in $\mathcal{X}$ and columns in $\mathcal{Y}$ and $\boldsymbol{\eta}$ is noise. In the following, we introduce two representative conventional sampling approaches.

Deterministic sampling

In the deterministic approach, the sampling problem is generally formulated as follows:

$$\mathcal{M}^* = \operatorname*{arg\,max}_{\mathcal{M} \subset \mathcal{V}} f(\mathcal{V}), \tag{2.114}$$

where $f : \mathcal{V} \to \mathbb{R}$ is a cost function. However, since this is a combinatorial optimization, which is NP-hard. Alternatively, the greedy method is typically utilized by replacing it with

$$y^* = \operatorname*{arg\,max}_{y \in \mathcal{M}^c} f(\mathcal{M} \cup \{y\}) - f(\mathcal{M}). \tag{2.115}$$

To design the cost function f, we need to the quality of the sampling set. Typically, the quality of sampling depends on the recovery performance. We describe examples of the design of f below.

Error-covariance based design The norm of reconstruction error leads to

$$\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\| = \|\mathbf{U}_{\mathcal{B}}(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}})^{\dagger}\boldsymbol{\eta}\| \leq \|\mathbf{U}_{\mathcal{B}}\|_2 \|(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}})^{\dagger}\|_2 \|\boldsymbol{\eta}\|, \tag{2.116}$$

where $\|\cdot\|$ and $\|\cdot\|_2$ are the ℓ_2 norm and spectral norm, respectively.

Since $\mathbf{U}_{\mathcal{B}}$ and $\boldsymbol{e}$ do not depend on the sampling set $\mathcal{M}$, (2.116) is essentially bounded by the following inequality

$$0 < \alpha \leq \sigma_{\min}(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}}). \tag{2.117}$$

Note that (2.116) can be arbitrarily large without (2.117). Therefore, $\sigma_{\min}(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}})$ should be maximized to select a robust sampling set.

In practice, the spectral norm may not be preferred due to the computational intractability. Alternatively, other cost functions are widely used in the literature of experimental designs [20, 44, 45]. $\sigma_{\min}(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}})$ is referred to as the E-experimental design. The D-experimental design is given by $\det(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}}) = \prod_{i=1}^{\min(|\mathcal{M}|,|\mathcal{B}|)} \sigma_i(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}})$. The T-experimental design is given by $\mathrm{tr}(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}}) = \sum_{i=1}^{\min(|\mathcal{M}|,|\mathcal{B}|)} \sigma_i(\mathbf{U}_{\mathcal{M}\mathcal{B}}^{\mathsf{T}}\mathbf{U}_{\mathcal{M}\mathcal{B}})$. For interested readers, please see [20] and references therein.

Random sampling

Let $\boldsymbol{p} \in \mathbb{R}_+^N$ be a probability distribution on $\mathcal{V}$, i.e., its components satisfy $\sum_{n=1}^N p[n] = 1$, $p[0], \ldots, p[N-1] \geq 0$. $\boldsymbol{p}$ is used as a sampling distribution in random sampling approaches [14, 46, 47]. Here, we define the following matrix:

$$\mathbf{P} \coloneqq \mathrm{diag}(\boldsymbol{p}). \tag{2.118}$$

We denote the sampling set by $\Omega \coloneqq \{\omega_1, \ldots, \omega_K\}$, whose elements independently drawn from $\mathcal{V}$ according to $\boldsymbol{p}$.

Here, we exploit the following theorem:

Theorem 4 (Restricted isometry property (RIP))**.** *Let Ω be a sampling set independently drawn from $\mathcal{V}$ conforming to the probabilistic distribution $\boldsymbol{p}$. For any $\delta \in (0, 1)$, $\mathbf{I}_{\Omega\mathcal{V}}$ satisfies a RIP property if and only if*

$$(1-\delta)\|\boldsymbol{x}\|^2 \leq \frac{1}{m}\|\mathbf{I}_{\Omega\mathcal{V}}\mathbf{P}^{-1/2}\boldsymbol{x}\|^2 \leq (1+\delta)\|\boldsymbol{x}\|^2, \tag{2.119}$$

for all $\boldsymbol{x} \in \mathcal{A}_{BL}$ and some constant m.

If the RIP condition is satisfied, $\boldsymbol{p}$ is optimal in that $\boldsymbol{x} \in \mathcal{A}_{\mathrm{BL}}$ can have a chance to be perfectly recovered.

Local graph coherence The local graph coherence [14, 42] is a widely-accepted quantity for $\boldsymbol{p}$ under the bandlimited assumption $\mathcal{A}_{\mathrm{BL}}$. Let $\|\mathbf{U}_{\mathcal{B}}^{\mathsf{T}}\boldsymbol{\delta}_m\|$ denotes the local graph coherence at node m. This quantity represents the importance of of $\mathbf{U}_{\mathcal{B}}^{\mathsf{T}}\boldsymbol{\delta}_i$ in rows in $\mathbf{U}_{\mathcal{B}}^{\mathsf{T}}$: $\mathbf{U}_{\mathcal{B}}^{\mathsf{T}}\boldsymbol{\delta}_i$ captures features of $\boldsymbol{x} \in \mathcal{A}_{\mathrm{BL}}$ better as its score is higher. In fact, $\boldsymbol{p}_i \propto \|\mathbf{U}_{\mathcal{B}}^{\mathsf{T}}\boldsymbol{\delta}_i\|$ satisfies that the RIP condition with the sufficient number of samples [14].

While bandlimited graph signals are widely accepted in the literature, in practice, graph signals may not be ideally bandlimited. To address this issue, we apply generalized sampling theory, as discussed in Sec. 2.5, to the graph setting below.

2.8 Generalized graph sampling theory

In this section, we briefly review our generalized sampling on graphs. Based on the previous framework in Hilbert space, we formulate recovery problems for LS/MX criteria under three priors. First of all, we define a nodal domain sampling operator as follows:

Definition 7 (Nodal domain sampling)**.** *Let $\mathbf{I}_{\mathcal{M}\mathcal{V}} \in \{0, 1\}^{K \times N}$ be the submatrix of the identity matrix indexed by $\mathcal{M} \subset \mathcal{V}$ ($|\mathcal{M}| = K$) and $\mathcal{V}$. The sampling operator is defined by*

$$\mathbf{S}^{\mathsf{T}} \coloneqq \mathbf{I}_{\mathcal{M}\mathcal{V}}\mathbf{G}, \tag{2.120}$$

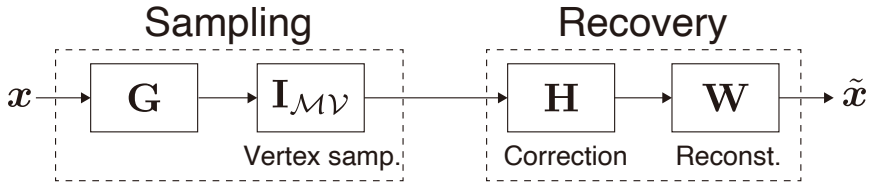


Figure 2.4: Generalized sampling framework. The dotted left and right boxes are sampling and recovery phases, respectively.

where $\mathbf{G} \in \mathbb{R}^{N \times N}$ *is an arbitrary graph filter. A sampled graph signal is thus given by* $\boldsymbol{y} = \mathbf{S}^{\mathsf{T}}\boldsymbol{x}$.

In the graph setting, the reconstruction operator depends on the chosen sampling operator $\mathbf{S}^{\mathsf{T}}$ because $\mathbf{I}_{\mathcal{MV}}$ is not unique in general.

The sampling and recovery framework based on generalized sampling is illustrated in Fig. 2.4. The input signal $\boldsymbol{x}$ is sampled by $\mathbf{S}^{\mathsf{T}}$, and the sample is corrected by a *correction operator* $\mathbf{H} \in \mathbb{R}^{K \times K}$. After correction, the corrected samples are reconstructed by a reconstruction operator $\mathbf{W} \in \mathbb{R}^{N \times K}$. We assume that $\mathbf{W}^{\mathsf{T}}\mathbf{W}$ is invertible in the same manner as the Riesz condition in Definition 4. This framework explains many existing graph signal sampling, including sampling for bandlimited signals [15].

As a result, the reconstructed graph signal is represented as follows:

$$\tilde{\boldsymbol{x}} = \mathbf{WH}\boldsymbol{y} = \mathbf{WHS}^{\mathsf{T}}\boldsymbol{x}. \tag{2.121}$$

In this scenario, the recovery problem turns out to be seeking the best possible $\mathbf{H}$ based on signal priors.

In the following, we introduce representative signal models and recovery methods corresponding to the models (i.e., designs of $\mathbf{H}$ and $\mathbf{W}$).

Subspace prior

Thanks to the aforementioned generalized sampling framework in Hilbert spaces, we can utilize all results therein. Here, suppose that the generation subspace $\mathcal{A}$ spanned by $\mathbf{A} \in \mathbb{R}^{N \times M}$ $(M \leq N)$ is known, i.e.,

$$\mathcal{A} = \left\{ \boldsymbol{x} \in \mathbb{R}^N : \ \boldsymbol{x} = \mathbf{A}\boldsymbol{d} \right\}, \tag{2.122}$$

where $\boldsymbol{d}$ is an expansion coefficient. We summarize recovery solutions for each strategy below.

Predefined LS solution:

Recovery with the LS criterion in the predefined case is given by

$$\tilde{\boldsymbol{x}} = \mathbf{W}(\mathbf{S}^{\mathsf{T}}\mathbf{W})^{\dagger}\boldsymbol{y}. \tag{2.123}$$

Predefined MX solution:

Alternatively, graph signal recovery with the MX criterion in the predefined case results in

$$\tilde{\boldsymbol{x}} = \mathbf{W}(\mathbf{W}^{\mathsf{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathsf{T}}\mathbf{A}(\mathbf{S}^{\mathsf{T}}\mathbf{A})^{\dagger}\boldsymbol{y}. \tag{2.124}$$

Unconstrained LS/MX solution:

Regardless of LS and MX, recovery in the unconstrained case is reduced to

$$\tilde{\boldsymbol{x}} = \mathbf{A}(\mathbf{S}^{\mathsf{T}}\mathbf{A})^{\dagger}\boldsymbol{y}. \tag{2.125}$$

Recall that, $Q(\omega)$ in (2.92), which depending on a signal prior, is called the prior filter in SI sampling. This also passes down to the graph setting, where $Q(\omega)$ is replaced with a matrix $\mathbf{Q} \in \mathbb{R}^{N\times M}$ $(M \leq N)$. The prior operator is given by $\mathbf{Q} = \mathbf{A}$. The DS condition is identical to that of the predefined MX solution. In addition, if the DS condition is satisfied for $\mathbb{R}^N = \mathcal{A} \oplus \mathcal{S}^{\perp}$, i.e., $(\mathbf{S}^{\mathsf{T}}\mathbf{A})$ is invertible, it leads to perfect reconstruction.

Graph signals encompass rich variations of signal subspaces because of the irregular structure of networks. In the following, we introduce examples of graph signal subspaces.

Graph signal subspaces One of the well-studied generator is the bandlimited model which can be characterized as

$$\boldsymbol{x} = \sum_{i=0}^{K-1} d[i]\boldsymbol{u}_i = \mathbf{U}_{\mathcal{VB}}\boldsymbol{d}, \tag{2.126}$$

where $\mathbf{U}_{\mathcal{VB}} \in \mathbb{R}^{N\times K}$ is the submatrix of $\mathbf{U}$ whose rows are extracted within $\mathcal{B} = \{0, \ldots, K-1\}$. In this case, $\mathbf{A} = \mathbf{U}_{\mathcal{VB}}$.

Periodic graph spectrum (PGS) model [48] assumes the periodicity of the graph spectrum as follows:

$$\boldsymbol{x} = \mathbf{U}A(\mathbf{\Lambda})\mathbf{D}_{\text{samp}}^{\mathsf{T}}\boldsymbol{d}, \tag{2.127}$$

where $A(\mathbf{\Lambda}) := \text{diag}(A(\lambda_0), \ldots, A(\lambda_{N-1}))$ is a graph spectral response of the generator and $\mathbf{D}_{\text{samp}}^{\mathsf{T}}$ is the matrix for the GFT domain upsampling [17].

Piecewise constant model [49] is a node domain signal model and is defined as follows.

$$\boldsymbol{x} = \sum_{i=1}^{K} d_i \mathbf{1}_{\mathcal{T}_i} = [\mathbf{1}_{\mathcal{T}_1}, \ldots, \mathbf{1}_{\mathcal{T}_K}]\boldsymbol{d}, \tag{2.128}$$

where $[\mathbf{1}_{\mathcal{T}_i}]_n = 1$ when the node n is in the ith cluster $\mathcal{T}_i$ and 0 otherwise [49].

Smoothness prior

In the same manner as the subspace prior, all results herein are also directly derived from the aforementioned Hilbert space sampling. Here, suppose that the signal energy that is measured by the smoothness measuring function (i.e., high-pass filter) $\mathbf{V}$ is bounded by $\sigma \in \mathbb{R}_+$, i.e., $\mathcal{V} = \left\{\boldsymbol{x} \in \mathbb{R}^N : \ \|\mathbf{V}\boldsymbol{x}\| \leq \sigma\right\}$. For simplicity, $\mathbf{V}$ is assumed to be invertible. We summarize recovery solutions for each strategy below.

Predefined LS solution:

The LS solution in the predefined case results in

$$\tilde{\boldsymbol{x}} = \widehat{\mathbf{W}}(\mathbf{S}^\mathsf{T}\widehat{\mathbf{W}})^\dagger \boldsymbol{y}, \tag{2.129}$$

where $\widehat{\mathbf{W}} = \mathbf{W}(\mathbf{W}^\mathsf{T}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{W})^{-1}\mathbf{W}^\mathsf{T}\mathbf{S}$. Denoting $\widehat{\mathcal{W}}$ as the range space of $\widehat{\mathbf{W}}$, the DS condition can be written by $\mathbb{R}^N = \widehat{\mathcal{W}} \oplus \mathcal{S}^\perp$. While this solution is the predefined case, we can also view the prior operator as $\mathbf{Q} = \widehat{\mathbf{W}}$ in another perspective.

Predefined MX solution:

The MX solution in the predefined case leads to

$$\tilde{\boldsymbol{x}} = \mathbf{W}(\mathbf{W}^\mathsf{T}\mathbf{W})^{-1}\mathbf{W}^\mathsf{T}\widetilde{\mathbf{W}}(\mathbf{S}^\mathsf{T}\widetilde{\mathbf{W}})^\dagger \boldsymbol{y}, \tag{2.130}$$

where $\widetilde{\mathbf{W}} = (\mathbf{V}^\mathsf{T}\mathbf{V})^{-1}\mathbf{S}$. Denoting the range space of $\widetilde{\mathbf{W}}$ by $\widetilde{\mathcal{W}}$, the DS condition can be written by $\mathbb{R}^N = \widetilde{\mathcal{W}} \oplus \mathcal{S}^\perp$.

Unconstrained LS/MX solution:

The LS and MX solutions in the unconstrained case coincides with each other, i.e.,

$$\tilde{\boldsymbol{x}} = \widetilde{\mathbf{W}}(\mathbf{S}^\mathsf{T}\widetilde{\mathbf{W}})^{-1}\boldsymbol{y}. \tag{2.131}$$

The prior operator is give by $\mathbf{Q} = \widetilde{\mathbf{W}}$. The DS condition is identical to that of the predefined MX solution.

Since $\mathbf{V}$ can be viewed as a graph variation operator, namely high-pass filter, its spectral representation is advantageous to implement itself. We describe the design of $\mathbf{V}$ below.

Smoothness measuring function on a graph There are other several graph variation operators in the literature [50]. In terms of $\mathbf{V}$, the spectral response of an arbitrary graph variation operator can generally be expressed by

$$\hat{V}(\lambda_i) = \sum_{\ell=0}^{L} \alpha_\ell (\lambda_i)^{\ell/2}, \tag{2.132}$$

where $\{\alpha_\ell\}$ is a coefficient and L is the order of a polynomial. It is easy to check that $\mathbf{V} = \mathbf{L}^{1/2}$ is a special case.

2.9 Conclusion

This chapter delves into the fundamentals of sampling theory for both time-domain and graph signals. It starts with a critical look at Shannon's theorem, highlighting its limitation with the assumption that signals must be bandlimited. To overcome these constraints, the chapter introduces generalized sampling theory (GST), designed for optimal recovery of signals under any signal priors. This framework is built on three core processes: sampling, correction, and reconstruction. The correction transform aims to reduce worst-case estimation errors, making the theory applicable to a broader range of signals. The chapter concludes by formally defining graph signals and extending GST to the graph counterpart, i.e., generalized graph sampling theory.

Chapter 3

Graph signal sampling under stochastic priors

3.1 Introduction

Graph signal processing (GSP) has presented various analysis tools since being developed [8, 51]. Recent interest in GSP is to extend classical signal processing theory to the graph setting [48,52–55]. One of the main differences between standard signal processing and GSP is that GSP systems are not shift-invariant (SI) in general. This leads to the challenge that GSP systems may have different definitions in the node and spectral (graph frequency) domains: They do not coincide in general. Such examples include sampling [10–12, 14, 15, 17, 20, 56, 57], translation [29, 42, 58–60], and filtering [61].

While graph signal sampling has been extensively studied, most works focus on building parallels of the Shannon-Nyquist theorem [32, 62] and its generalizations [16, 63] to the graph setting. Therefore, sampling of bandlimited graph signals has been a main interest of existing studies [10–12, 57]. Other graph signal subspaces are also useful for practical applications: For example, piecewise smooth graph signals and periodic graph spectrum signals have been considered [31, 49, 64, 65].

Since sampling of deterministic signals has been well studied in the literature of sampling in Hilbert space [16, 40, 63, 66], we can immediately derive its GSP counterpart. In contrast, sampling of *random graph signals* has not been considered so far because existing (generalized) sampling methods for graph signals have mainly focused on deterministic signal models [10–12, 14, 15, 20, 48, 56, 57, 67].

In this chapter, we consider a graph signal sampling framework for random graph signals. Random graph signals are modeled by graph wide sense stationarity (GWSS), which is a counterpart of wide sense stationarity (WSS) of standard signals. WSS is characterized by the statistical moments, i.e., mean and covariance, that are invariant to shift. Graph

signals do not lie in SI spaces in general. Hence, existing definitions of GWSS are based on the spectral characteristics of signals [26, 29, 68], which are extensions of the power spectral density (PSD) of the standard WSS. However, existing definitions of GWSS require additional assumptions that are not necessary for WSS. Therefore, we first define GWSS as a natural extension of WSS based on signal modulation. With this definition, we can appropriately analyze GWSS signals with their covariance, which conforms to a GWSS process.

Subsequently, we develop a generalized sampling framework for graph signals under stochastic priors. We assume that the covariance of graph signals is known and satisfies the GWSS conditions. Our framework parallels generalized sampling of SI signals [69]. It consists of sampling, correction, and reconstruction transforms. The correction transform is inserted between the sampling and reconstruction transforms to compensate for non-ideal measurements. We derive the correction transform that minimizes the mean-squared error (MSE). Our framework can be applied to any sampling method that is linear. In other words, both node and graph frequency domain sampling methods [10–12, 17, 20, 57] are applicable without changing the framework.

Existing works of (generalized) sampling theory on graphs [48, 67] have been studied only for deterministic signal models. In particular, deterministic models are categorized into subspace and smoothness priors [15]. Since these priors constrain the basis function and energy bound of original signals, respectively, sampling under these priors may be weak in the face of unexpected disturbances and slight changes in the input. In contrast, our generalized sampling approach is designed for random graph signals, as described in Table 3.1. Our framework, utilizing stochastic priors, is expected to be robust against undesirable randomness in graph signals. To the best of our knowledge, graph signals can almost always be identified by one of the three priors as long as they are linear [16]. Therefore, this work complements the existing body of generalized sampling theory on graphs.

Interestingly, our solution parallels that in the SI setting [69] when sampling is performed in the graph frequency domain. Moreover, we show that existing signal recovery methods under different priors [48] are special cases of our framework. Experiments for synthetic signals validate that our proposed recovery is effective for stochastic graph signals with sampling in both node and spectral domains.

The remainder of this chapter is organized as follows. Section 3.2 reviews generalized sampling for time-domain signals. For WSS signals, we introduce the standard Wiener filter. In Section 3.3, a generalized sampling framework for stochastic graph signals is introduced. In Section 3.4, we define GWSS as an extension of WSS for time domain signals. Section 3.5 derives graph Wiener filters for recovery of a GWSS process based on minimization of the MSE between the original and reconstructed graph signals. In the special case of graph frequency domain sampling, the graph frequency response parallels that of sampling in SI spaces. In Section 3.6, we discuss the relationship between the proposed Wiener filter and existing generalized graph signal sampling. Signal recovery experiments for synthetic and

Table 3.1: Comparison of generalized graph signal sampling.

Signal priors	SI spaces	Graph subspaces
Subspace	[16]	[48, 67]
Smootheness	[16]	[48, 67]
Stochatic	[69]	**This work**

real-world signals are demonstrated in Section 3.7. Section 3.8 concludes the chapter.

For time-domain signals, $x(t)$ and $x[n]$ denote a continuous-time signal and discrete-time signal, respectively. The continuous-time Fourier transform (CTFT) of a signal $x(t) \in L_2$ is denoted by $X(\omega)$ and the discrete-time Fourier transform (DTFT) of a sequence $x[n] \in \ell_2$ is denoted by $X(e^{j\omega})$.

3.2 Generalized sampling in SI spaces

Generalized sampling for stochastic standard signals is reviewed in this section [69]. Detailed derivations and discussions can be found in [16, 69]. The Wiener filter introduced in this section parallels that in our graph signal sampling framework.

Sampling and recovery framework for time-domain signals

We first review standard generalized sampling for time-domain signals [16]. The framework is depicted in Fig. 3.1.

Let $x(t)$ be a continuous-time signal and $\eta[n]$ be stationary noise. We consider samples $c[n]$ at $t = nT$ of a filtered signal of $x(t)$, $c(t) = x(t) * s(-t)$, where $s(-t)$ denotes a sampling filter. The DTFT of the samples, $C(e^{j\omega})$, can be written as

$$C(e^{j\omega}) = \frac{1}{T} \sum_{k\in\mathbb{Z}} S^* \left(\frac{\omega - 2\pi k}{T} \right) X \left(\frac{\omega - 2\pi k}{T} \right), \tag{3.1}$$

where $S(\omega)$ is the CTFT of $s(t) \in L_2$. The measured samples are corrupted by noise and are given by

$$y[n] = \langle s(t - nT), x(t) \rangle + \eta[n]. \tag{3.2}$$

The reconstructed signal is constrained to lie in a SI space $\mathcal{W}$, spanned by the shifts of a reconstruction kernel $w(t) \in L_2$, i.e.,

$$\tilde{x}(t) = \sum_{n\in\mathbb{Z}} d[n] w(t - nT), \tag{3.3}$$

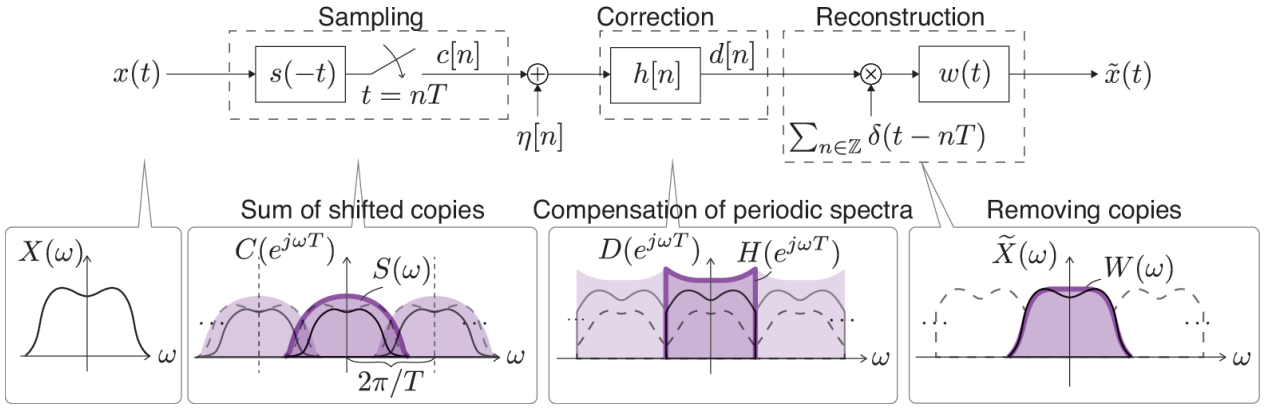


Figure 3.1: Generalized sampling framework in SI spaces and its counterpart in the Fourier domain. Dashed-lined and solid-lined areas in the bottom represent input and filtered spectra, respectively. Colored areas represent spectral responses of each filter.

where $d[n] \in \ell_2$ are unknown expansion coefficients, which yield the best possible recovery $\tilde{x}(t)$. To obtain an appropriate $d[n]$, we apply a digital correction filter $h[n]$ to $y[n]$, i.e., $d[n] = (h * y)[n]$. The CTFT of $\tilde{x}(t)$ can be expressed as

$$\widetilde{X}(\omega) = D(e^{j\omega T})W(\omega), \tag{3.4}$$

where $D(e^{j\omega})$ is the DTFT of $d[n]$ and $W(\omega)$ is the CTFT of $w(t)$.

In practice, sampling and reconstruction filters, i.e., $s(t)$ and $w(t)$, may be predefined prior to sampling based on technical requirements. For generalized sampling, instead, we assume that we can design the correction filter $h[n]$ freely. The best correction filter $h[n]$ is designed such that the input signal $x(t)$ and the reconstructed signal $\tilde{x}(t)$ are close enough in some metric. This is a widely-accepted sampling framework for time-domain signals and we follow it here for the graph setting [24, 69].

Wiener filter

Next, suppose that $x(t)$ is a zero-mean WSS process with known power spectral density (PSD) $\Gamma_x(\omega)$ and $\eta[n]$ is a zero-mean WSS noise process with known PSD $\Gamma_\eta(e^{j\omega})$, independent of $x(t)$. Detailed definitions of WSS are given in Appendix 3.A.

The optimal correction filter proposed in [69] is obtained by solving the following problem:

$$\min_{h[n]} \mathbb{E}\left\{|\tilde{x}(t) - P_{\mathcal{W}}x(t)|^2\right\}, \tag{3.5}$$

where $P_{\mathcal{W}}$ is an orthogonal projection onto the reconstruction subspace $\mathcal{W}$. Since the recovered signal $\tilde{x}(t)$ is constrained by (3.3) to lie in $\mathcal{W}$, we want it to best approximate the

orthogonal projection of $x(t)$ onto that same space $\mathcal{W}$ [69]. The frequency response of the solution of (3.5) is given by [69]:

$$H_{\mathrm{PRE}}(e^{j\omega}) = \frac{\sum_{k\in\mathbb{Z}} \Gamma_x(\frac{\omega}{T}+\frac{2\pi k}{T}) S(\frac{\omega}{T}+\frac{2\pi k}{T}) W^*(\frac{\omega}{T}+\frac{2\pi k}{T})}{R_W(e^{j\omega})\left(\sum_{k\in\mathbb{Z}} \Gamma_x(\frac{\omega}{T}+\frac{2\pi k}{T}) \left|S(\frac{\omega}{T}+\frac{2\pi k}{T})\right|^2 + \Gamma_\eta(e^{j\omega})\right)}, \tag{3.6}$$

where $R_W(e^{j\omega}) = \sum_{k\in\mathbb{Z}} |W(\frac{\omega}{T}+\frac{2\pi k}{T})|^2$.

When the reconstruction filter is unconstrained, we can optimize the correction filter jointly with the reconstruction filter. The optimal filter in the unconstrained setting is a special case of (3.6), where $w(t)$ is optimally chosen. In order to obtain the unconstrained solution, we thus solve (3.5) with respect to $w(t)$ by substituting (3.6) for (3.5) [16]. The resulting optimal correction filter in the unconstrained case is given by

$$H_{\mathrm{UNC}}(e^{j\omega}) = \frac{1}{\sum_{k\in\mathbb{Z}} \Gamma_x(\frac{\omega}{T}+\frac{2\pi k}{T}) \left|S(\frac{\omega}{T}+\frac{2\pi k}{T})\right|^2 + \Gamma_\eta(e^{j\omega})}, \tag{3.7}$$

with the reconstruction filter $W(\omega) = \Gamma_x(\omega)S(\omega)$.

In Section 3.3, we introduce a sampling framework for GWSS as a counterpart to what we described in this section.

3.3 Sampling of graph signals

We next introduce the sampling and recovery framework used throughout the dissertation. Then, two representative sampling methods for graph signals are reviewed.

Sampling and recovery framework for graph signals

We begin by reviewing generalized graph signal sampling [48, 70], illustrated in Fig. 3.2. The main differences among various methods stem from the sampling domain and signal generation model.

Let $\boldsymbol{c} \in \mathbb{C}^K (K \leq N)$ be the sampled graph signal. When $\mathbf{S}^*$ is the sampling transformation, $\boldsymbol{c} = \mathbf{S}^*\boldsymbol{x}$, we would like to recover $\boldsymbol{x}$ from the noisy samples $\boldsymbol{y} = \boldsymbol{c} + \boldsymbol{\eta}$ using a reconstruction transformation $\mathbf{W} \in \mathbb{C}^{N\times K}$. The recovered signal is then

$$\tilde{\boldsymbol{x}} = \mathbf{WH}\boldsymbol{y} = \mathbf{WH}(\boldsymbol{c}+\boldsymbol{\eta}) = \mathbf{WH}(\mathbf{S}^*\boldsymbol{x}+\boldsymbol{\eta}), \tag{3.8}$$

where $\mathbf{H} \in \mathbb{C}^{K\times K}$ is some transformation that compensates for non-ideal measurements. It is an analog of the generalized sampling framework introduced in Section 3.2: We use three

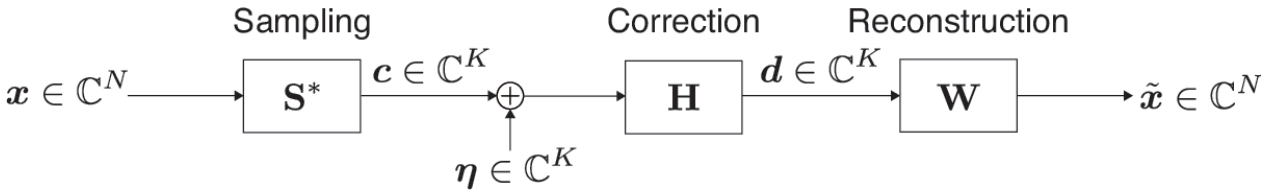


Figure 3.2: Generalized sampling framework for graph signals.

filters $\mathbf{S}$, $\mathbf{H}$, and $\mathbf{W}$. In the following, we seek the best $\mathbf{H}$ so that $\tilde{\boldsymbol{x}}$ is close to $\boldsymbol{x}$ in some sense under appropriate signal assumptions [24, 69].

Sampling methods

Here, we introduce representative graph signal sampling operators in the node and graph frequency domains.

Sampling in the nodal domain

Node domain sampling is expressed as the selection of samples on a node subset. Therefore, it corresponds to non-uniform sampling in the time domain.

Sampling in the nodal domain is defined as follows:

Definition 8 (Sampling in the nodal domain)**.** *Let $\boldsymbol{x} \in \mathbb{C}^N$ be the original graph signal and $\mathbf{G} \in \mathbb{C}^{N \times N}$ be an arbitrary graph filter. In addition, let $\mathbf{I}_{\mathcal{MV}} \in \{0,1\}^{K \times N}$ be a submatrix of the identity matrix $\mathbf{I}_N$ extracting $K = |\mathcal{M}|$ rows corresponding to the sampling set $\mathcal{M} \subset \mathcal{V}$. The sampled graph signal $\boldsymbol{c} \in \mathbb{C}^K$ is given as follows:*

$$\boldsymbol{c} = \mathbf{I}_{\mathcal{MV}} \mathbf{G} \boldsymbol{x}. \tag{3.9}$$

The sampling matrix is then expressed as $\mathbf{S}^* = \mathbf{I}_{\mathcal{M}} \mathbf{G}$. In contrast to time domain signals, $\mathbf{I}_{\mathcal{MV}}$ may depend on the graph since the local connectivity of the graph is often irregular. This implies that there may exist a "best" node set for graph signal sampling described in [15].

Sampling in the GFT domain

When sampling SI signals, the spectrum folding phenomenon occurs [16]. Graph frequency domain sampling utilizes the behavior in the graph spectrum.

Formally, graph frequency domain sampling, which is a counterpart of (3.1), is defined as follows:

Definition 9 (Sampling in the graph frequency domain [17]). *Let $\hat{\boldsymbol{x}} \in \mathbb{C}^N$ be the original signal in the graph frequency domain, i.e., $\hat{\boldsymbol{x}} = \mathbf{U}^* \boldsymbol{x}$, and let $S(\lambda_i)$ be an arbitrary sampling filter defined in the graph frequency domain. For any sampling ratio $M \in \mathbb{Z}^1$, the sampled graph signal in the graph frequency domain is given by $\hat{\boldsymbol{c}} \in \mathbb{C}^K$, where $K = N/M$, and*

$$\hat{c}(\lambda_i) = \sum_{l=0}^{M-1} S(\lambda_{i+Kl}) \hat{x}(\lambda_{i+Kl}). \tag{3.10}$$

In matrix form, the sampled graph signal is represented as $\hat{\boldsymbol{c}} = \mathbf{D}_{\text{samp}} S(\boldsymbol{\Lambda}) \hat{\boldsymbol{x}}$, where $\mathbf{D}_{\text{samp}} = [\mathbf{I}_K \ \mathbf{I}_K \ \cdots] \in \mathbb{C}^{K \times N}$. Then, we define the sampling matrix $\mathbf{S}^*$ as [48]

$$\mathbf{S}^* = \mathbf{U}_{\text{reduced}} \mathbf{D}_{\text{samp}} S(\boldsymbol{\Lambda}) \mathbf{U}^*, \tag{3.11}$$

where $\mathbf{U}_{\text{reduced}} \in \mathbb{C}^{K \times K}$ is an arbitrary unitary matrix, which may correspond to the GFT for a reduced-size graph. Theoretically, we can use any unitary matrix for $\mathbf{U}_{\text{reduced}}$. One choice which has been studied in multiscale transforms for graph signals is graph reduction [71].

Here, we define reconstruction in the graph frequency domain, which is the counterpart of (3.4), as follows:

$$[\mathbf{U}^* \tilde{\boldsymbol{x}}](\lambda_i) = \hat{d}(\lambda_{i \bmod K}) W(\lambda_i), \tag{3.12}$$

where $\hat{\boldsymbol{d}} \in \mathbb{C}^K$ is a vector composed of expansion coefficients and $W(\lambda_i)$ is the reconstruction filter kernel defined in the graph frequency domain. Correspondingly, the reconstruction matrix is represented as

$$\mathbf{W} = \mathbf{U} W(\boldsymbol{\Lambda}) \mathbf{D}_{\text{samp}}^* \mathbf{U}_{\text{reduced}}^*. \tag{3.13}$$

The reconstruction in (3.12) is performed by replicating the original spectrum in the same manner as standard signal processing [17].

In the following, we propose a definition of GWSS based on the classical WSS, in which the covariance is invariant to modulation.

3.4 Graph wide sense stationarity

In this section, we consider the stationarity of random graph signals. We begin by defining GWSS as the invariance of statistical moments with modulation which extends the classical WSS definition. We then derive the properties of our GWSS. For clarity, the classical definition of WSS is detailed in Appendix 3.A.

[1] M is assumed to be a divisor of N for simplicity.

Definition of GWSS

Several definitions of GWSS have been proposed [26, 29, 72]. A definition of GWSS based on the WSS by shift (see Collorary 3 in Appendix 3.A) is shown in Appendix 3.B.

In this dissertation, we define GWSS based on the WSS by modulation (see Collorary 4 in Appendix 3.A) as follows:

Definition 10 (Graph wide sense stationary by modulation ($\mathrm{GWSS_M}$))**.** *Let $\boldsymbol{x}$ be a graph signal on a graph[2] $\mathcal{G}$. Then, $\boldsymbol{x}$ is a graph wide sense stationary process if and only if the following two conditions are satisfied for all m:*

1. $\mathbb{E}\left[(\boldsymbol{x} \star \boldsymbol{\delta}_n)_m\right] = \mu_x = \mathrm{const},$ (3.14)
2. $\mathbb{E}\left[(\boldsymbol{x} \star \boldsymbol{\delta}_n - \mu_x \mathbf{1})_m (\boldsymbol{x} \star \boldsymbol{\delta}_k - \mu_x \mathbf{1})^*_m\right] = [\boldsymbol{\Gamma}_x]_{n,k}.$ (3.15)

The operator $\cdot \star \boldsymbol{\delta}_k$ is defined as follows:

$$\boldsymbol{x} \star \boldsymbol{\delta}_k \coloneqq \mathbf{M}\mathrm{diag}(u_0[k], u_1[k], \ldots)\mathbf{U}^* \boldsymbol{x}, \tag{3.16}$$

where $\mathbf{M} = [e^{j2\pi \cdot 0/N}\mathbf{1},\ e^{j2\pi \cdot 1/N}\mathbf{1},\ \cdots]$.

Note that $\cdot \star \boldsymbol{\delta}_k$ corresponds to the standard sinusoidal modulation, since $(\boldsymbol{x} \star \boldsymbol{\delta}_k)_n = (\mathbf{U}\exp(j\boldsymbol{\Omega})\mathbf{U}^*\boldsymbol{x})_k$ where $\boldsymbol{\Omega} = \mathrm{diag}(0, 2\pi/N, \ldots, 2\pi(N-1)/N)$. Therefore, we refer to $\cdot \star \boldsymbol{\delta}_k$ as a modulation operator on a graph.

Properties of GWSS

Next, we study the properties of Definition 10 for GWSS used in our generalized sampling, as a counterpart of Corollary 4 in Appendix 3.A.

In Definition 10, the condition in (3.14) is identical to

$$\mathbb{E}[\boldsymbol{x} \star \boldsymbol{\delta}_n] = \mathbf{M}\mathrm{diag}(u_0[n], u_1[n], \ldots)\mathbf{U}^*\mathbb{E}[\boldsymbol{x}]. \tag{3.17}$$

Since $\mathbf{M}\mathrm{diag}(1, 0, \ldots, 0) = [\mathbf{1}\,\mathbf{0}\,\cdots\,\mathbf{0}]$, the equality in (3.17) holds if and only if $\mathbb{E}[\boldsymbol{x}] = \mu_x \boldsymbol{u}_0 = \mu_x \mathbf{1}$. Therefore, $\cdot \star \boldsymbol{\delta}_n$ does not change the mean of graph signals. The condition in (3.15) implies that the covariance $\boldsymbol{\Gamma}_x$ has to be diagonalizable by $\mathbf{U}$. We show this fact in the following lemma.

Lemma 1. *Let $\boldsymbol{x}$ and $\boldsymbol{\Gamma}_x$ be a stochastic signal on $\mathcal{G}$ and its covariance matrix, respectively, by Definition 10. Then, $\boldsymbol{\Gamma}_x$ is diagonalizable by $\mathbf{U}$.*

[2]We suppose that $\mathcal{G}$ is connected for simplicity. Nevertheless, we can extend our definition of GWSS to the non-connected case.

Proof. The LHS of (3.15) is expressed as

$$\begin{aligned}
&\mathbb{E}[(\boldsymbol{x} \star \boldsymbol{\delta}_n - \mu_x \mathbf{1})_m (\boldsymbol{x} \star \boldsymbol{\delta}_k - \mu_x \mathbf{1})_m^*] \\
&= [\mathbf{M}\mathrm{diag}(u_0[n], u_1[n], \ldots)\widehat{\boldsymbol{\Gamma}}_x \mathrm{diag}(u_0^*[k], u_1^*[k], \ldots)\mathbf{M}^*]_{m,m} \\
&= \boldsymbol{\delta}_m^* \mathbf{M}\mathrm{diag}(\mathbf{U}^*\boldsymbol{\delta}_n)\widehat{\boldsymbol{\Gamma}}_x \mathrm{diag}(\mathbf{U}^*\boldsymbol{\delta}_k)\mathbf{M}^*\boldsymbol{\delta}_m \\
&= \boldsymbol{\delta}_n^* \mathbf{U}\mathrm{diag}(\mathbf{M}^*\boldsymbol{\delta}_m)\widehat{\boldsymbol{\Gamma}}_x \mathrm{diag}(\mathbf{M}^*\boldsymbol{\delta}_m)\mathbf{U}^*\boldsymbol{\delta}_k \\
&= [\mathbf{U}(\widehat{\boldsymbol{\Gamma}}_x \circ \mathbf{M}^*\boldsymbol{\delta}_m\boldsymbol{\delta}_m^*\mathbf{M})\mathbf{U}^*]_{n,k} \\
&= [\mathbf{U}(\widehat{\boldsymbol{\Gamma}}_x \circ \boldsymbol{\Xi})\mathbf{U}^*]_{n,k}, \qquad (3.18)
\end{aligned}$$

where $\widehat{\boldsymbol{\Gamma}}_x = \mathbb{E}[\mathbf{U}^*(\boldsymbol{x} - \mu_x\mathbf{1})(\boldsymbol{x} - \mu_x\mathbf{1})^*\mathbf{U}]$ and $[\boldsymbol{\Xi}]_{i,l} = [\mathbf{M}^*\boldsymbol{\delta}_m\boldsymbol{\delta}_m^*\mathbf{M}]_{i,l} = \exp(-j2\pi(i-l)/N)$. Since $[\boldsymbol{\Xi}]_{i,l} = 1$ for $i = l$ and $[\boldsymbol{\Xi}]_{i,l} \neq 1$ otherwise, the condition satisfying (3.18), i.e., $\widehat{\boldsymbol{\Gamma}}_x = \widehat{\boldsymbol{\Gamma}}_x \circ \boldsymbol{\Xi}$, holds if and only if $\widehat{\boldsymbol{\Gamma}}_x$ is diagonal. Therefore, $\boldsymbol{\Gamma}_x$ is diagonalizable by $\mathbf{U}$, which completes the proof. □

One of the important properties of Definition 10 is that $\boldsymbol{\Gamma}_x$ is diagonalizable by $\mathbf{U}$, as stated in Lemma 1. We refer to $\widehat{\Gamma}_x(\boldsymbol{\Lambda}) \coloneqq \mathbf{U}^*\boldsymbol{\Gamma}_x\mathbf{U}$ as the graph PSD. This property is preferred in GSP since it parallels the Wiener-Khinchin relation in the time domain [73] and it is advantageous for exploiting spectral tools in GSP.

Recall that the autocovariance function is invariant to shift in the standard WSS [73]. Since a graph operator is often used as a shift operator in the graph setting, the invariance with shift for GWSS can be translated to

$$\boldsymbol{\Gamma}_x\mathbf{L} = \mathbf{L}\boldsymbol{\Gamma}_x. \qquad (3.19)$$

This property is also used in [74,75]. Note that $\boldsymbol{\Gamma}_x$ is diagonalizable by $\mathbf{U}$ if and only if (3.19) is satisfied [76]. In fact, the alternative definitions of GWSS require additional assumptions to satisfy (3.19). For example, [29] requires that all eigenvalues of $\mathbf{L}$ are distinct. However, we sometimes encounter graphs whose graph operator has an eigenvalue with multiplicity greater than one [26]. Moreover, [26, 72] require that the covariance $\boldsymbol{\Gamma}_x$ is expressed by a polynomial in $\mathbf{L}$, while this assumption may not be true in general, including the multiple eigenvalue case. On the other hand, our GWSS definition does not require such assumptions, but it still satisfies (3.19). As a result, Definition 10 can be regarded as a more natural extension of the classical WSS than existing GWSSs. The relationship among the GWSS definitions is discussed in detail in Appendix 3.B. Note that our generalized sampling is applicable under different definitions of GWSS with slight modifications.

We develop a Wiener filter for our generalized sampling framework. We consider two scenarios of reconstruction design: predefined and unconstrained reconstruction. We demonstrate the similarity of the resulting filters to the classical ones.

3.5 Graph Wiener filter: Recovery for stochastic graph signals

We now consider the design of the correction filter. It is optimized such that the reconstructed graph signal is close to the original one in MSE. We show that the resulting Wiener filter parallels that in the SI setting when graph spectral sampling (Definition 9) is performed.

Graph Wiener filter

Suppose that $\boldsymbol{x}$ is a zero-mean GWSS process with a known covariance $\boldsymbol{\Gamma}_x$ and $\boldsymbol{\eta}$ is a zero-mean GWSS noise process with a known covariance $\boldsymbol{\Gamma}_\eta$, independent of each other. For simplicity, we suppose that columns of $\mathbf{W}$ and $\mathbf{S}$ satisfy the Riesz condition [16], i.e., $\mathbf{W}^*\mathbf{W}$ and $\mathbf{S}^*\mathbf{S}$ are invertible.

We now formulate the design the correction filter. Unlike the formulation (3.5) in the time domain, we can directly minimize the expectation of the normed error in the graph setting because the subspace of graph signals is finite-dimensional.

We consider minimizing the MSE between the original and reconstructed graph signals by solving

$$\min_{\mathbf{H}} \ \mathbb{E}[\,\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2\,]. \tag{3.20}$$

The optimal correction filter is obtained in the following Theorem.

Theorem 5. *Suppose that the input signal $\boldsymbol{x}$ and noise $\boldsymbol{w}$ are zero-mean GWSS processes with covariances $\boldsymbol{\Gamma}_x$ and $\boldsymbol{\Gamma}_w$, respectively, which are uncorrelated with each other. Then, the solution of* (3.20) *is given by*

$$\mathbf{H}_{\mathrm{PRE}} = (\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*\boldsymbol{\Gamma}_x\mathbf{S}(\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S} + \boldsymbol{\Gamma}_\eta)^{-1}. \tag{3.21}$$

Proof. The MSE $\varepsilon_{\mathrm{MSE}} = \mathbb{E}[\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2]$ is given by

$$\begin{aligned}
\varepsilon_{\mathrm{MSE}} =& \mathbb{E}[\mathrm{tr}\,(\mathbf{W}\mathbf{H}\mathbf{S}^*\boldsymbol{x}\boldsymbol{x}^*\mathbf{S}\mathbf{H}^*\mathbf{W}^*)] \\
&+ \mathbb{E}[\mathrm{tr}\,(-\boldsymbol{x}\boldsymbol{x}^*\mathbf{S}\mathbf{H}^*\mathbf{W}^* - \boldsymbol{x}\boldsymbol{x}^*\mathbf{W}\mathbf{H}\mathbf{S}^* + \boldsymbol{x}\boldsymbol{x}^*)] \\
&+ \mathbb{E}[\mathrm{tr}\,(\mathbf{W}\mathbf{H}\boldsymbol{\eta}\boldsymbol{\eta}^*\mathbf{H}^*\mathbf{W}^* - \boldsymbol{x}\boldsymbol{\eta}^*\mathbf{H}^*\mathbf{W}^*)] \\
&+ \mathbb{E}[\mathrm{tr}\,(-\mathbf{W}\mathbf{H}\boldsymbol{\eta}\boldsymbol{x}^* + \boldsymbol{\eta}\boldsymbol{\eta}^*)] \\
=& \,\mathrm{tr}\,(\mathbf{W}\mathbf{H}\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S}\mathbf{H}^*\mathbf{W}^* - \boldsymbol{\Gamma}_x\mathbf{S}\mathbf{H}^*\mathbf{W}^*) \\
&+ \mathrm{tr}\,(-\boldsymbol{\Gamma}_x\mathbf{W}\mathbf{H}\mathbf{S}^* + \boldsymbol{\Gamma}_x) \\
&+ \mathrm{tr}\,(\mathbf{W}\mathbf{H}\boldsymbol{\Gamma}_\eta\mathbf{H}^*\mathbf{W}^* + \boldsymbol{\Gamma}_\eta) \\
=& \,\mathrm{tr}\,(\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S}\mathbf{H}^*\mathbf{W}^*\mathbf{W}\mathbf{H} - 2\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{W}\mathbf{H} + \boldsymbol{\Gamma}_x) \\
&+ \mathrm{tr}\,(\boldsymbol{\Gamma}_\eta\mathbf{H}^*\mathbf{W}^*\mathbf{W}\mathbf{H} + \boldsymbol{\Gamma}_\eta)\,,
\end{aligned} \tag{3.22}$$

where the second equality holds since expectation and trace are interchangeable and the third equality follows by the cyclic property of trace.

The infinitesimal perturbation of $\varepsilon_{\mathrm{MSE}}$ with respect to $\mathbf{H}$ results in

$$\begin{aligned}\mathrm{d}\varepsilon_{\mathrm{MSE}} =& \operatorname{tr}\left(2\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S}\mathbf{H}^*\mathbf{W}^*\mathbf{W}\mathrm{d}\mathbf{H}\right) \\ &+ \operatorname{tr}\left(-2\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{W}\mathrm{d}\mathbf{H} + 2\boldsymbol{\Gamma}_\eta\mathbf{H}^*\mathbf{W}^*\mathbf{W}\mathrm{d}\mathbf{H}\right). \end{aligned} \tag{3.23}$$

Here, $\mathrm{d}\varepsilon_{\mathrm{MSE}}$ is defined by

$$\mathrm{d}\varepsilon_{\mathrm{MSE}} = \sum_{n,m}\left[\frac{\partial\varepsilon_{\mathrm{MSE}}}{\partial\mathbf{H}} \circ \mathrm{d}\mathbf{H}\right]_{n,m} = \operatorname{tr}\left(\left(\frac{\partial\varepsilon_{\mathrm{MSE}}}{\partial\mathbf{H}}\right)^* \mathrm{d}\mathbf{H}\right). \tag{3.24}$$

As a result, the derivative of $\varepsilon_{\mathrm{MSE}}$ in (3.22) with respect to $\mathbf{H}$ results in

$$\frac{\partial\varepsilon_{\mathrm{MSE}}}{\partial\mathbf{H}} = \mathbf{W}^*\mathbf{W}\mathbf{H}(\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S} + \boldsymbol{\Gamma}_\eta) - \mathbf{W}^*\boldsymbol{\Gamma}_x\mathbf{S}. \tag{3.25}$$

Setting (3.25) to zero, we obtain the optimal $\mathbf{H}$ of (3.21). □

In Theorem 5, we consider the predefined case. Next, we move on to the unconstrained case: The reconstruction filter $\mathbf{W}$ can also be freely chosen along with $\mathbf{H}$. The optimal solution is obtained by solving

$$\min_{\mathbf{W},\,\mathbf{H}} \mathbb{E}\left[\,\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2\,\right]. \tag{3.26}$$

The optimal correction filter is given in the following Theorem.

Theorem 6. *Suppose that the input signal $\boldsymbol{x}$ and noise $\boldsymbol{w}$ are zero-mean GWSS processes with covariances $\boldsymbol{\Gamma}_x$ and $\boldsymbol{\Gamma}_\eta$, respectively, which are uncorrelated with each other. Then, the solution of* (3.26) *is given by*

$$\mathbf{H}_{\mathrm{UNC}} = (\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S} + \boldsymbol{\Gamma}_\eta)^{-1}, \quad \mathbf{W}_{\mathrm{UNC}} = \boldsymbol{\Gamma}_x\mathbf{S}. \tag{3.27}$$

Proof. Since (3.21) is optimal for an arbitrary $\mathbf{W}$, we plug it into (3.22) and have

$$\begin{aligned}\varepsilon_{\mathrm{MSE}} =& \operatorname{tr}(\mathbf{P}_{\mathcal{W}}\mathbf{Z}\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S}\mathbf{Z}^*\mathbf{P}_{\mathcal{W}} - 2\boldsymbol{\Gamma}_x\mathbf{S}\mathbf{Z}^*\mathbf{P}_{\mathcal{W}} + \boldsymbol{\Gamma}_x) \\ &+ \operatorname{tr}(\mathbf{P}_{\mathcal{W}}\mathbf{Z}\boldsymbol{\Gamma}_\eta\mathbf{Z}^*\mathbf{P}_{\mathcal{W}} + \boldsymbol{\Gamma}_\eta), \end{aligned} \tag{3.28}$$

where $\mathbf{P}_{\mathcal{W}} = \mathbf{W}(\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*$ and $\mathbf{Z} = \boldsymbol{\Gamma}_x\mathbf{S}(\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S} + \boldsymbol{\Gamma}_\eta)^{-1}$. In the same manner as (3.25), the derivative of $\varepsilon_{\mathrm{MSE}}$ with respect to $\mathbf{W}$ is derived from (3.28) as

$$\begin{aligned}\frac{\partial\varepsilon_{\mathrm{MSE}}}{\partial\mathbf{W}} =& 2\mathbf{P}_{\mathcal{W}}\mathbf{Z}\mathbf{S}^*\boldsymbol{\Gamma}_x\mathbf{S}\mathbf{Z}^*\mathbf{W}(\mathbf{W}^*\mathbf{W})^{-1} \\ &- 2\boldsymbol{\Gamma}_x\mathbf{S}\mathbf{Z}^*\mathbf{W}(\mathbf{W}^*\mathbf{W})^{-1} \\ &+ 2\mathbf{P}_{\mathcal{W}}\mathbf{Z}\boldsymbol{\Gamma}_\eta\mathbf{Z}^*\mathbf{W}(\mathbf{W}^*\mathbf{W})^{-1}. \end{aligned} \tag{3.29}$$

By setting (3.29) to zero, we have

$$(\mathbf{P}_{\mathcal{W}}\mathbf{\Gamma}_x\mathbf{S} - \mathbf{\Gamma}_x\mathbf{S})\mathbf{Z}^*\mathbf{W}(\mathbf{W}^*\mathbf{W})^{-1} = \mathbf{0}. \tag{3.30}$$

Therefore, $\mathbf{P}_{\mathcal{W}}$ is necessary to be the identity mapping over $\mathcal{R}(\mathbf{\Gamma}_x\mathbf{S})$, leading to $\mathcal{W} = \mathcal{R}(\mathbf{\Gamma}_x\mathbf{S})$ where $\mathcal{R}(\cdot)$ is the range space of a matrix. Graph signal recovery with (3.21) can be expressed as

$$\mathbf{WH} = \mathbf{P}_{\mathcal{W}}\mathbf{\Gamma}_x\mathbf{S}(\mathbf{S}^*\mathbf{\Gamma}_x\mathbf{S} + \mathbf{\Gamma}_\eta)^{-1}. \tag{3.31}$$

Since $\mathcal{W} = \mathcal{R}(\mathbf{\Gamma}_x\mathbf{S})$, we can choose $\mathbf{W} = \mathbf{\Gamma}_x\mathbf{S}$ and obtain $\mathbf{H} = (\mathbf{S}^*\mathbf{\Gamma}_x\mathbf{S} + \mathbf{\Gamma}_\eta)^{-1}$. This completes the proof. □

We note that graph signal restoration based on Bayesian estimation in [57, Proposition 1] coincides with the proposed recovery transformation in Theorem 2. This implies that the proposed recovery transformation in the unconstrained case can also be viewed as a Bayesian estimation. The recovery transformation in Theorem 1 covers both of the unconstrained and predefined cases: This situation is not studied in [57].

Additionally, the Wiener filter presented in [74, Theorem 5] appears similar to the one presented in Theorem 2, however, it requires the bandlimited assumption. In contrast, our Wiener filter can be applied to full-band graph signals.

So far, we considered a general solution for stochastic recovery. By choosing graph frequency domain sampling, we can show that the resulting recovery parallels that in SI space, studied in [16, 69].

Special cases for GFT domain sampling

Suppose that the graph signal and noise conform to zero-mean GWSS processes with PSD $\widehat{\Gamma}_x(\lambda)$ and $\widehat{\Gamma}_\eta(\lambda)$, respectively. We assume that $\mathbf{S}^*$ and $\mathbf{W}$ are defined in the graph frequency domain by (3.11) and (3.13), respectively. In this setting, (3.21) is diagonalizable by $\mathbf{U}_{\text{reduced}}$, i.e., it has a graph frequency response

$$H_{\text{PRE}}(\lambda_i) = \frac{\sum_l \Gamma_x(\lambda_{i+Kl})W^*(\lambda_{i+Kl})S(\lambda_{i+Kl})}{R_W(\lambda_i)(\sum_l \widehat{\Gamma}_x(\lambda_{i+Kl})|S(\lambda_{i+Kl})|^2 + \widehat{\Gamma}_\eta(\lambda_i))}, \tag{3.32}$$

where

$$R_W(\lambda_i) \coloneqq \sum_{l=0}^{M-1} W^*(\lambda_{i+Kl})W(\lambda_{i+Kl}). \tag{3.33}$$

Similarly, the graph spectral responses of (3.27) are given by

$$H_{\mathrm{UNC}}(\lambda_i) = \frac{1}{\sum_l \Gamma_x(\lambda_{i+Kl})|S(\lambda_{i+Kl})|^2 + \Gamma_\eta(\lambda_i)}, \tag{3.34}$$

with reconstruction filter $W(\lambda_i) = \Gamma_x(\lambda_i)S(\lambda_i)$. The correction filters (3.32) and (3.34) inherit frequency responses of the Wiener filter in standard sampling (3.6) and (3.7), respectively.

In the following, we discuss the relationship among the proposed graph Wiener filter and existing generalized graph signal sampling.

3.6 Relationship to other priors

In this section, we consider the relationship between the proposed graph Wiener filter and graph signal recovery under subspace and smoothness priors [48]. For simplicity, we only consider the predefined case. The unconstrained case can be derived in a similar fashion.

Subspace prior

Under the subspace prior, we assume the following graph signal model [48].

$$\boldsymbol{x} = \mathbf{A}\boldsymbol{d}, \tag{3.35}$$

where $\mathbf{A} \in \mathbb{C}^{N\times K}$ is a known generator and $\boldsymbol{d} \in \mathbb{C}^K$ is expansion coefficients. Suppose that $\boldsymbol{d}$ is a random vector and $\boldsymbol{\Sigma}_d = \mathbb{E}[\boldsymbol{d}\boldsymbol{d}^*]$ is the covariance of $\boldsymbol{d}$. The covariance of $\boldsymbol{x}$ is then written as

$$\boldsymbol{\Sigma}_x = \mathbb{E}[\mathbf{A}\boldsymbol{d}\boldsymbol{d}^*\mathbf{A}^*] = \mathbf{A}\mathbb{E}[\boldsymbol{d}\boldsymbol{d}^*]\mathbf{A}^* = \mathbf{A}\boldsymbol{\Sigma}_d\mathbf{A}^*. \tag{3.36}$$

By substituting (3.36) to (3.21), we have the following correction filter:

$$\mathbf{H}_{\mathrm{MX,SUB}} = (\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*\mathbf{A}\boldsymbol{\Sigma}_d\mathbf{A}^*\mathbf{S}(\mathbf{S}^*\mathbf{A}\boldsymbol{\Sigma}_d\mathbf{A}^*\mathbf{S} + \boldsymbol{\Gamma}_\eta)^{-1}. \tag{3.37}$$

For a subspace prior, if $\mathbf{S}^*\mathbf{A}$ is invertible and there is no noise, (3.37) reduces to

$$\begin{aligned}\mathbf{H}_{\mathrm{MX,SUB}} =& (\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*\mathbf{A}\boldsymbol{\Sigma}_d\mathbf{A}^*\mathbf{S}(\mathbf{S}^*\mathbf{A}\boldsymbol{\Sigma}_d\mathbf{A}^*\mathbf{S})^{-1} \\ =& (\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*\mathbf{A}\boldsymbol{\Sigma}_d\mathbf{A}^*\mathbf{S}(\mathbf{A}^*\mathbf{S})^{-1}\boldsymbol{\Sigma}_d^{-1}(\mathbf{S}^*\mathbf{A})^{-1} \\ =& (\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*\mathbf{A}(\mathbf{S}^*\mathbf{A})^{-1},\end{aligned} \tag{3.38}$$

where we assume that $\boldsymbol{\Sigma}_d$ is invertible. As observed, $\mathbf{H}_{\mathrm{MX,SUB}}$ does not depend on $\boldsymbol{\Sigma}_d$. Interestingly, the solution in (3.38) coincides with the minimax solution for signal recovery under subspace prior [48]. As a result, we can view the minimax recovery under the subspace prior (3.38) as a special case of (3.21) with random expansion coefficients and known $\mathbf{A}$.

Smoothness prior

Next, we consider a smoothness prior [48]:

$$\|\mathbf{V}\boldsymbol{x}\|^2 \leq \rho^2, \tag{3.39}$$

where $\mathbf{V} = \mathbf{U}V(\mathbf{\Lambda})\mathbf{U}^*$ is a smoothness measuring function, i.e., graph high-pass filter, and $\rho > 0$ is a constant. We assume that $\mathbf{V}$ is bounded below on ℓ_2, i.e., $\mathbf{V}^*\mathbf{V}$ is invertible.

To derive the solution of (3.20) subject to the constraint in (3.39), the problem may be written as minimization of the Lagrangian

$$L(\mathbf{H}, \xi) = \mathbb{E}[\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2] + \xi(\mathbb{E}[\|\mathbf{V}\boldsymbol{x}\|^2] - \rho^2), \tag{3.40}$$

where ξ is a constant. The optimal solution of (3.40) satisfies the following identity [23]:

$$\lambda(\mathbb{E}[\|\mathbf{V}\boldsymbol{x}\|^2] - \rho^2) = \xi(\|\mathbf{V}\mathbf{\Sigma}_x\mathbf{V}^*\|_F^2 - \rho^2) = 0. \tag{3.41}$$

The second equality of (3.41) is satisfied with either $\xi = 0$ or $\|\mathbf{V}\mathbf{\Sigma}_x\mathbf{V}\|_F^2 = \rho^2$. Therefore, the nontrivial solution of (3.40) is obtained with $\|\mathbf{V}\mathbf{\Sigma}_x\mathbf{V}\|_F^2 = \rho^2$. Since $\|\mathbf{V}(\mathbf{V}^*\mathbf{V})^{-1}\mathbf{V}^*\|_F^2 = \|\mathbf{I}\|_F^2 = N$, $\mathbf{\Sigma}_x$ can be written as

$$\mathbf{\Sigma}_x = \frac{\rho^2}{N}(\mathbf{V}^*\mathbf{V})^{-1}. \tag{3.42}$$

Since the derivative of $L(\mathbf{H}, \xi)$ with respect to $\mathbf{H}$ coincides with (3.22), we immediately obtain the following solution:

$$\mathbf{H}_{\text{MX,SMO}} = (\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*(\mathbf{V}^*\mathbf{V})^{-1}\mathbf{S}(\mathbf{S}^*(\mathbf{V}^*\mathbf{V})^{-1}\mathbf{S})^{-1}, \tag{3.43}$$

where we also assume the noiseless case ($\mathbf{\Gamma}_\eta = \mathbf{0}$ in (3.22)) as in the subspace prior.

In fact, (3.43) coincides with the minimax solution under the smoothness prior [48]. Similar to (3.38), we may view minimax recovery under the smoothness prior (3.43) as a special case of (3.21).

3.7 Signal recovery experiments

In this section, we validate the effectiveness of the proposed method via signal recovery experiments. We demonstrate that the approach described in Section 3.5 reduces the reconstruction error in comparison with existing recovery techniques under the stochastic setting.

Synthetic graph signals

In this subsection, we perform signal recovery experiments for synthetic graph signals.

Sampling and recovery setting

We perform signal recovery experiments of synthetic graph signals. The recovery framework is illustrated in Fig. 3.2.

Here, we consider the following three graphs with $N = 256$:

- Random sensor graph[3];
- Erdős-Rényi (ER) graph with edge connection probability $p = 0.3$; and
- 2D grid graph.

We use the following functions throughout the experiments.

- PSD function:

$$\widehat{\Gamma}_x(\lambda_i) = \exp\left(-\left\{\frac{2\lambda_i - \lambda_{\max}}{\sqrt{\lambda_{\max}}}\right\}^2\right). \tag{3.44}$$

- Sampling function #1 (for full-band sampling):

$$S(\lambda_i) = \begin{cases} 1 & \lambda_{\max}/\lambda_i \leq 2 \\ 2 - 2\lambda_i/(\lambda_{\max}) & \text{otherwise.} \end{cases} \tag{3.45}$$

- Sampling function #2 (for bandlimited sampling):

$$S(\lambda_i) = \begin{cases} 1 & i \in [0, K-1], \\ 0 & \text{otherwise.} \end{cases} \tag{3.46}$$

- Reconstruction function (for constrained recovery):

$$W(\lambda_i) = \cos\left(\frac{\pi}{2} \cdot \frac{\lambda_i}{\lambda_{\max}}\right). \tag{3.47}$$

- Smoothness function (for recovery with smoothness prior by [48]):

$$V(\lambda_i) = \frac{\lambda_i}{\lambda_{\max}} + \varepsilon \tag{3.48}$$

  where we set to $\varepsilon = 0.1$.

[3]Random sensor graphs are implemented by k nearest neighbor graphs, whose nodes are randomly distributed in 2-D space $[0, 1] \times [0, 1]$ [77].

Their spectral responses are shown in Fig. 3.3.

A stochastic graph signal is generated by $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_x)$, where $\mathbf{\Gamma}_x$ satisfies the GWSS conditions described in Section 3.4, i.e., $\mathbf{\Gamma}_x = \mathbf{U}\widehat{\Gamma}_x(\mathbf{\Lambda})\mathbf{U}^*$. As mentioned in Appendix 3.B, existing definitions of GWSS coincide when $\mathbf{\Gamma}_x$ is diagonalizable by $\mathbf{U}$. Therefore, this setting simultaneously satisfies Definition 10 and the other existing definitions of GWSS [26, 29, 72].

Under the presence of noise, we suppose that noise conforms to i.i.d. zero-mean Gaussian distribution with $\sigma^2 = 0.3$, i.e., $\eta \sim \mathcal{N}(0, 0.3)$ and thus $\Gamma_\eta(\lambda_i) = 0.3$ for all i. Graph signals are sampled with sampling ratio $N/K = 4$.

We perform experiments for two cases. 1) The sampling function $S(\lambda_i)$ is full-band as in (3.45). 2) $S(\lambda_i)$ is bandlimited as in (3.46). In addition, two sampling domains, i.e., samplings in node and graph frequency domains, are considered. For node domain sampling, sampled nodes are selected randomly. For the predefined recovery, we use the given kernel $W(\lambda_i)$ in (3.47).

We calculate the average MSE in 20 independent runs and compare with existing graph signal interpolation methods, MKVV [78], MKVD [79], NLPD [10], GSOD [80] and NLPI [81], and recovery methods under smoothness priors proposed in [48]. Smoothness is measured by the energy of high frequency graph spectra with the measuring function $V(\lambda_i)$ in (3.48). We also show the result of the bandlimited reconstruction [17], i.e., reconstruction with the sampling filter #2 without correction.

Since there has been no prior work on graph signal sampling with stochastic prior, we also compare with graph signal recovery with smoothness prior [48] as a benchmark.

Results

Table 3.2 summarizes the average MSEs with the standard deviations (in decibels). Examples of recovered signals via full-band sampling are visualized in Fig. 3.4.

From Table 3.2, it is observed that the proposed methods present the best MSEs in most cases regardless of the sampling domain. Sometimes, the MSEs for bandlimited sampling are close to the proposed method, especially for the ER graph. This is because of the eigenvalue distribution: The ER graph may produce eigenvalues sparsely distributed in low graph frequencies, i.e., $\lambda < \lambda_{\max}/2$ where we consider $\mathbf{L}$ as a graph operator, and many eigenvalues exist in high graph frequencies ($\lambda \geq \lambda_{\max}/2$). This implies that the cutoff frequency at the Kth eigenvalue can be relatively high, resulting in wide bandwidth. This also results in the fact that the bandlimited sampling filter in (3.46) may pass the spectra in the middle graph frequencies ($\lambda \in [\lambda_{\max}/4, 3\lambda_{\max}/4]$); This behavior improves the reconstruction MSEs. Note that the proposed method presents stable recovery performances for all the graph and signal models considered.

The recovery with smoothness prior increases MSEs especially for node domain sampling because the PSD function $\widehat{\Gamma}_x(\lambda_i)$ does not represent smooth signals in this setting. For the

Table 3.2: Average MSEs of the reconstructed signals for synthetic data on 20 independent realizations (in decibels). We also show the standard deviations along with the MSEs. ER, BL, FB, VS and SS denote an Erdős-Rényi graph, bandlimited, full-band, node domain sampling and spectral domain sampling, respectively. UNC, PRE, LS and MX denote unconstrained, predefined, least-squares and minimax criteria, respectively.

Prior				Stochastic		Smoothness [48]			BL	MKVV	MKVD	NLPD	GSOD	NLPI
Reconst./ Criterion				UNC	PRE	UNC	PRE LS	PRE MX	[17]	[78]	[79]	[10]	[80]	[81]
Sensor	Noisy	BL	VS	**-8.87**±1.18	-8.86±1.19	443±14.4	208±11.5	442±14.4	-8.69±1.16	136±23.9	68.9±30.4	119±23.6	18.9±0.10	-6.79±1.15
			SS	**-8.88**±1.99	**-8.88**±1.99	-8.13±1.99	-6.99±1.98	-8.63±1.99	-8.19±1.99	-	-	-	-	-
		FB	VS	**-11.9**±1.15	-10.7±1.15	-4.73±1.44	-9.82±1.21	-9.59±1.21	-	143±18.8	203±22.7	162±17.8	18.9±0.11	-2.10±2.22
			SS	**-14.1**±1.99	-10.9±1.99	-5.05±1.99	-10.8±1.99	-10.6±1.99	-	-	-	-	-	-
	Clean	BL	VS	**-8.99**±1.20	-8.90±1.19	126±12.5	44.2±8.94	125±12.6	-8.90±1.19	97.0±20.4	-5.93±1.02	-6.94±1.22	18.7±0.09	-8.91±1.19
			SS	**-8.99**±1.99	-8.95±1.99	**-8.99**±1.99	-8.88±1.99	-8.95±1.99	-8.98±1.99	-	-	-	-	-
		FB	VS	**-13.5**±1.35	-11.5±1.23	-6.66±1.38	-11.4±1.23	-11.4±1.23	-	132±19.7	176±27.6	152±17.1	18.8±0.11	-4.01±2.13
			SS	**-14.7**±1.99	-11.2±1.99	-5.51±1.99	-11.1±1.99	-11.0±1.99	-	-	-	-	-	-
ER	Noisy	BL	VS	**-18.6**±1.40	-16.6±1.44	209±19.2	198±18.5	209±19.2	-17.3±1.43	-14.8±1.46	30.1±22.8	142±12.2	18.0±0.32	-16.5±1.40
			SS	-24.2±1.99	-18.2±1.99	-24.2±1.99	**-24.3**±1.99	-12.6±1.99	-18.2±1.99	-	-	-	-	-
		FB	VS	**-18.8**±1.53	-17.2±1.46	-14.3±1.50	-14.1±1.43	-14.1±1.43	-	-14.8±1.46	53.1±14.1	147±11.8	18.1±0.31	-16.8±1.42
			SS	**-25.9**±2.00	-19.5±2.00	-21.3±1.99	-19.5±1.99	-19.5±1.99	-	-	-	-	-	-
	Clean	BL	VS	**-26.7**±1.64	-17.5±1.51	62.4±15.9	66.1±14.0	62.1±15.9	-17.6±1.47	-14.8±1.46	-4.95±22.8	137±12.7	17.8±0.33	-16.9±1.46
			SS	**-26.7**±2.00	-18.7±1.99	**-26.7**±2.00	-13.9±2.00	-18.7±1.99	-26.5±2.00	-	-	-	-	-
		FB	VS	**-25.2**±1.91	-18.8±1.61	-19.9±1.63	-18.8±1.61	-18.8±1.61	-	-14.8±1.46	14.4±20.2	141±12.2	18.0±0.32	-17.7±1.54
			SS	**-28.2**±2.01	-20.1±2.00	-22.1±2.00	-20.1±2.00	-20.1±2.00	-	-	-	-	-	-
2D grid	Noisy	BL	VS	**-8.25**±1.16	-8.24±1.16	311±18.1	204±13.8	310±18.1	-8.11±1.15	-8.07±1.15	-2.85±17.1	132±23.5	18.9±0.09	-5.96±1.14
			SS	**-8.31**±1.99	-8.24±1.99	-7.57±1.99	-6.13±1.98	-7.85±1.99	-7.67±1.99	-	-	-	-	-
		FB	VS	**-12.4**±1.15	-10.6±1.17	-5.58±1.35	-10.5±1.16	-10.3±1.15	-	-8.12±1.15	66.2±18.7	156±18.4	18.9±0.10	-2.34±1.88
			SS	**-13.7**±1.99	-10.7±1.99	-5.75±1.99	-10.6±1.99	-10.4±1.99	-	-	-	-	-	-
	Clean	BL	VS	**-8.51**±1.18	-8.33±1.18	72.0±17.3	36.0±10.4	71.2±17.2	-8.32±1.17	-8.23±1.16	-6.38±1.08	-6.41±1.62	18.7±0.11	-8.32±1.17
			SS	**-8.51**±1.99	-8.39±1.99	**-8.51**±1.99	-8.25±1.99	-8.39±1.99	-8.48±1.99	-	-	-	-	-
		FB	VS	**-14.0**±1.29	-11.3±1.24	-6.45±1.37	-11.2±1.24	-11.2±1.24	-	-8.13±1.15	61.3±20.6	153±18.3	18.9±0.11	-3.29±1.82
			SS	**-14.5**±2.00	-11.0±1.99	-6.06±1.99	-10.9±1.99	-10.8±1.99	-	-	-	-	-	-

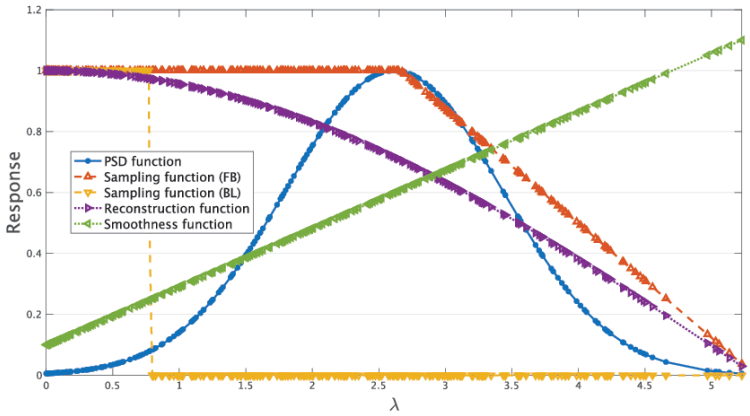


Figure 3.3: Graph frequency responses of several functions used for experiments.

noiseless cases, the proposed methods demonstrate MSE improvements by 3–5 dB on average compared to the noisy case. The unconstrained solution of the proposed stochastic recovery results in the best performance for almost all cases.

Real-world data

We also perform signal recovery experiments on real-world data.

Sampling and recovery setting

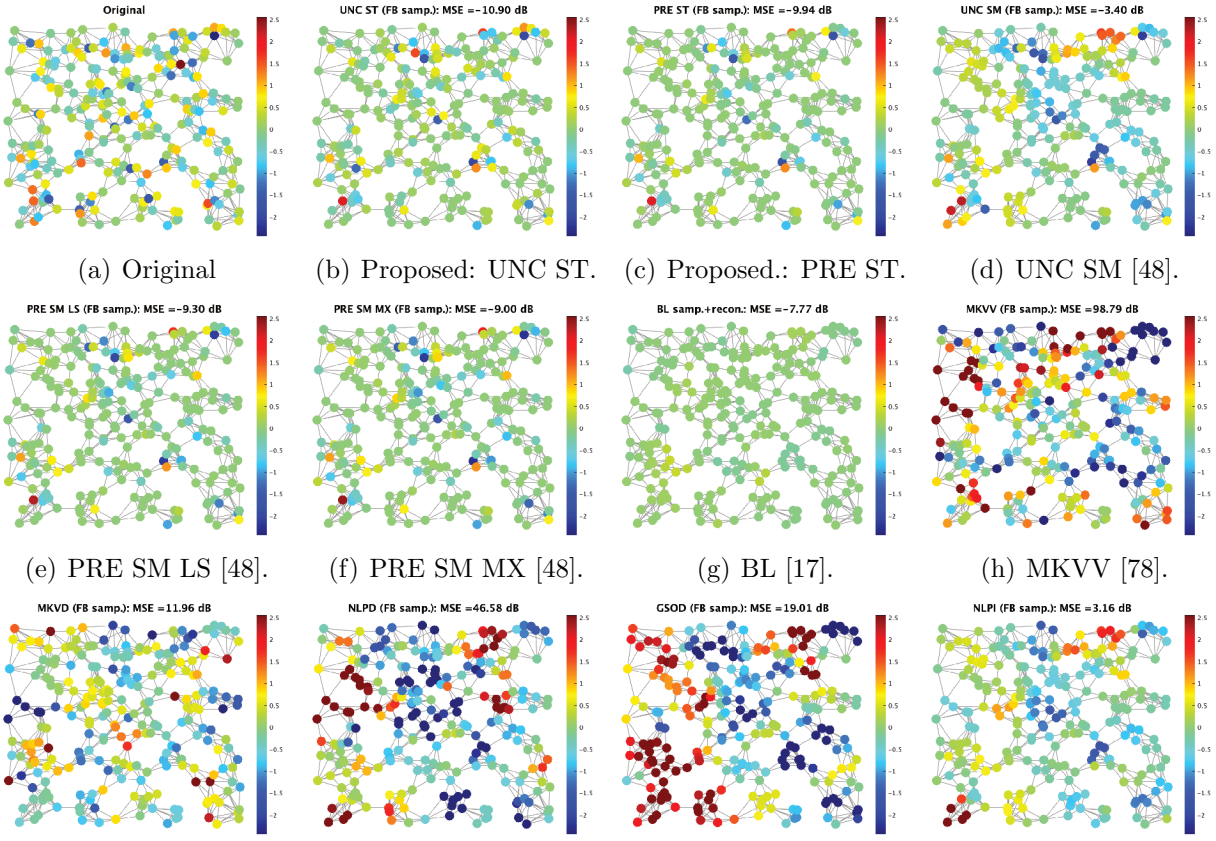
We use the global sea surface temperature dataset [82]. It records snapshots of sea surface temperatures for every month from 2004 to 2021. They are spatially sampled at intersections of 1-degree latitude-longitude grids. For the experiment, we used data for 24 months from 2018 to 2020. For this dataset, we randomly sample $N = 291$ intersections and they are regarded as nodes. We then construct a 5-nearest neighbor graph based on Euclidean distance. Edge weights are set to be unweighted. We then remove the edges crossing land areas. The number of samples is set to $K = 73$.

For each time instance, we estimate the covariance from the data in the previous year. We use the method proposed in [26]. The estimated covariance is used for the proposed graph Wiener filter. We calculate the average MSEs of 24 months for samplings in node and graph frequency domains, and compare with the existing methods in the previous subsection. For node domain sampling, we perform the experiments on 20 random sampling patterns of nodes for calculating the average MSE.

Results

Table 3.3 summarizes the average MSEs with the standard deviations (in decibels). Recovered graph signals are visualized in Fig. 3.5. From Table 3.3, we observe that the proposed method outperforms existing methods for almost all cases. This is also validated in Fig. 3.5.

Table 3.3 occasionally reveals a significant disparity in recovery performance between the proposed method and other methods. This discrepancy stems from the extensive dynamic range of global sea surface temperatures (approximately 0 to 30 degrees Celsius), which is considerably larger than that observed in typical signal communication applications. In Table 3.3, we would like to highlight that our method demonstrates relatively superior performance compared to the alternatives.



(a) Original (b) Proposed: UNC ST. (c) Proposed.: PRE ST. (d) UNC SM [48].

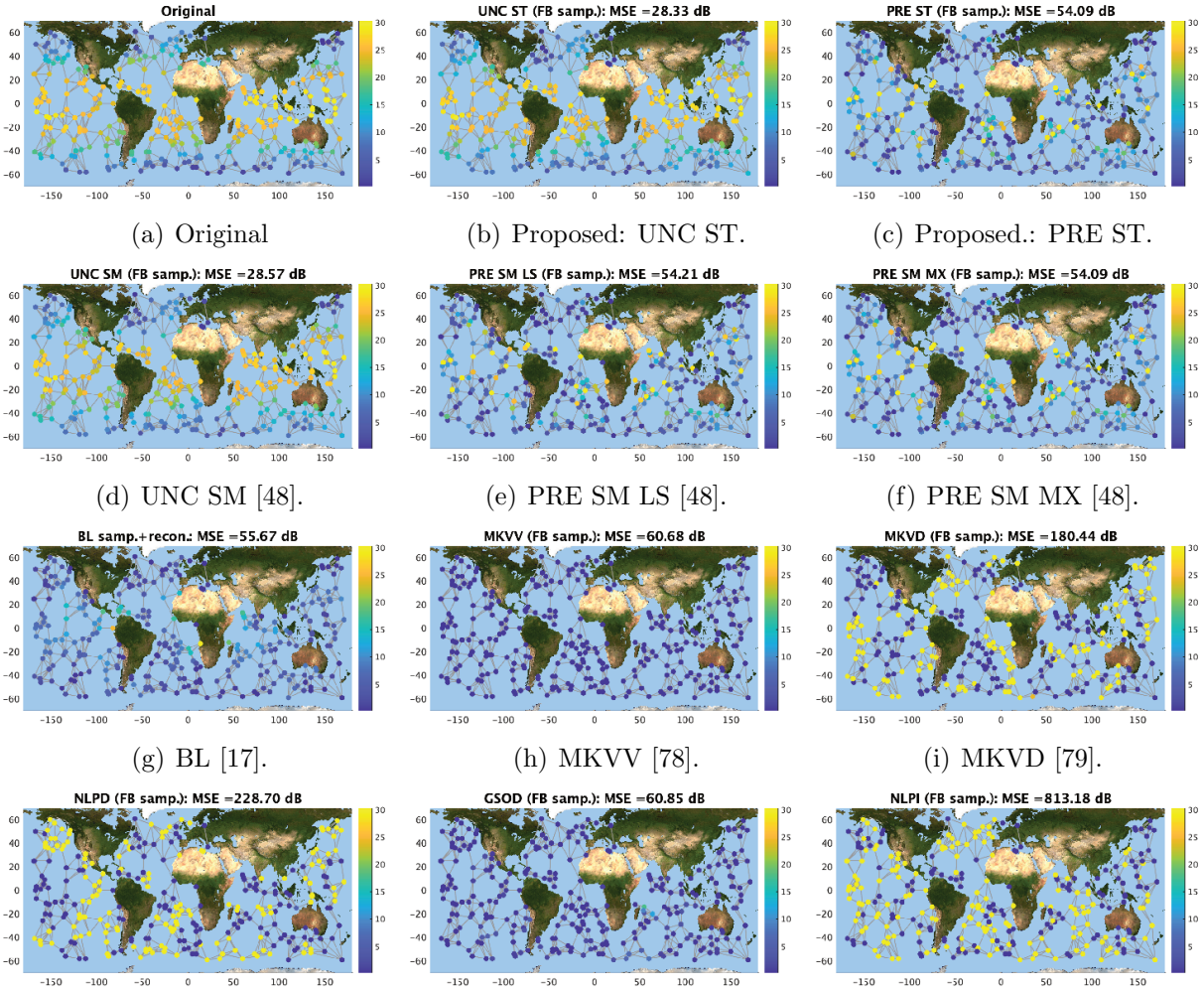
(e) PRE SM LS [48]. (f) PRE SM MX [48]. (g) BL [17]. (h) MKVV [78].

(i) MKVD [79]. (j) NLPD [10]. (k) GSOD [80]. (l) NLPI [81].

Figure 3.4: Signal recovery experiments for noisy graph signals on a random sensor graph with $N = 256$. Sampling is performed by (3.45) in the node domain. UNC and PRE denote the unconstrained and predefined solutions. SM and ST denote recovery with smoothness and stochastic priors. LS and MX denote least-squares and minimax criteria.

Table 3.3: Average MSEs of the reconstructed signals for the sea surface temperature data (in decibels). Notations are the same as Table 3.2.

Prior		Stochastic		Smoothness [48]			BL	MKVV	MKVD	NLPD	GSOD	NLPI
Reconst./	Criterion	UNC	PRE	UNC	PRE LS	PRE MX	[17]	[78]	[79]	[10]	[80]	[81]
BL	VS	**20.6**±4.05	54.4±3.17	232±15.5	154±8.71	232±15.5	56.0±3.14	60.9±3.13	209±15.8	206±12.3	61.0±3.13	827±7.05
	SS	**14.1**±1.16	55.1±0.15	**14.1**±1.16	63.3±0.14	551±0.15	14.2±1.15	-	-	-	-	-
FB	VS	**28.3**±4.34	54.4±3.17	32.0±3.93	54.5±3.17	54.4±3.17	-	60.9±3.13	210±15.8	206±12.0	61.0±3.13	828±6.88
	SS	**19.5**±1.11	55.1±0.14	22.0±0.94	55.1±0.14	55.1±0.14	-	-	-	-	-	-



(a) Original (b) Proposed: UNC ST. (c) Proposed.: PRE ST.

(d) UNC SM [48]. (e) PRE SM LS [48]. (f) PRE SM MX [48].

(g) BL [17]. (h) MKVV [78]. (i) MKVD [79].

(j) NLPD [10]. (k) GSOD [80]. (l) NLPI [81].

Figure 3.5: Signal recovery experiments for the sea surface temperature data on a 5-nearest neighbor graph. Sampling is performed by (3.45) in the node domain. Abbreviations are the same as those in Fig. 3.4.

3.8 Conclusion

Generalized graph signal sampling with stochastic priors is proposed in this chapter. We define the stationarity of random graph signals based on classical WSS. We then derive the graph Wiener filter for unconstrained and predefined reconstruction, based on the minimization of the expectation of the squared norm error between the original and reconstructed graph signals. We show that, when sampling is performed in the graph frequency domain, spectral responses of the graph Wiener filter parallel those in generalized sampling for WSS signals. We also reveal a theoretical relationship between the proposed graph Wiener filter and existing signal recovery under different priors. In the recovery experiments, we validate the MSE improvement of the proposed methods for various graphs and two sampling domains.

In this chapter, we focus on a GWSS process, where the covariance of graph signals remains invariant with sinusoidal modulation over time. Generally, however, graph signals could be non-stationary and they require more insights and provide more challenges. Furthermore, while we have not specified the specific designs of probabilistic density functions, the covariance matrix should have some properties depending on them. We leave these open questions as future work.

Appendix 3.A Wide sense stationarity

Stationarity of time-domain signals is reviewed here since this is useful to understand the connection between WSS and GWSS.

We consider a continuous-time signal $x(t)$. Its stationarity is defined as follows:

Definition 11 (Wide sense stationary for time-domain signals)**.** *Let $x(t)$ and $\gamma_x(t)$ be a stochastic signal in the time domain and its autocovariance function, respectively. The signal $x(t)$ is a wide sense stationary process if and only if the following two conditions are satisfied:*

1. $\mathbb{E}\left[x(t)\right] = \mu_x = \text{const},$ (3.49)
2. $\mathbb{E}\left[(x(t) - \mu_x)(x(\tau) - \mu_x)^*\right] = \gamma_x(t - \tau).$ (3.50)

A WSS process can be characterized in the Fourier domain by the Wiener-Khinchin theorem [37]. If $x(t)$ is a WSS process, then its power spectral density (PSD) function coincides with the CTFT of the autocovariance function of $x(t)$, $\gamma_x(t) \in L_1$, i.e.,

$$\Gamma_x(\omega) = \int_{-\infty}^{\infty} \gamma_x(t) e^{-j\omega t} d\omega. \tag{3.51}$$

Table 3.4: Comparison among WSS and GWSSs. OPE and COV denote operator and covariance, respectively.

	WSS	
	Shift	Modulation
OPE	$T_{t_0}\{x(t)\} = x(t-t_0)$	$x(t_0) * \delta_t = x(t)$
	GWSS	
	GWSS_T	GWSS_M
	$\mathbf{T}_\mathcal{G}\boldsymbol{x} = \mathbf{U}\exp(j\boldsymbol{\Pi})\mathbf{U}^*\boldsymbol{x}$	$\boldsymbol{x} \star \boldsymbol{\delta}_n = \mathbf{M}\text{diag}(\mathbf{U}^*\boldsymbol{\delta}_n)\mathbf{U}^*\boldsymbol{x}$

	WSS	
	Shift	Modulation
Mean	$\mathbb{E}\left[T_{t_0}\{x(t)\}\right] = \mu_x$	$\mathbb{E}\left[x(t_0) * \delta_\tau\right] = \mu_x$
	GWSS	
	GWSS_T	GWSS_M
	$\mathbb{E}[(\mathbf{T}_\mathcal{G}\boldsymbol{x})_n] = \mu_x$	$\mathbb{E}[(\boldsymbol{x} \star \boldsymbol{\delta}_n)_m] = \mu_x$

	WSS	
	Shift	Modulation
COV	$\mathbb{E}\left[T_{t_0}\{\bar{x}(t)\}T_{t_0}\{\bar{x}(\tau)\}^*\right] = \gamma_x(t-\tau)$	$\mathbb{E}\left[(\bar{x}(t_0) * \delta_t)(\bar{x}(t_0) * \delta_\tau)^*\right] = \gamma_x(t-\tau)$
	GWSS	
	GWSS_T	GWSS_M
	$\mathbb{E}[(\mathbf{T}_\mathcal{G}\bar{\boldsymbol{x}})_n(\mathbf{T}_\mathcal{G}\bar{\boldsymbol{x}})_k^*] = [\boldsymbol{\Gamma}_x]_{n,k}$	$\mathbb{E}[(\bar{\boldsymbol{x}} \star \boldsymbol{\delta}_n)_m(\bar{\boldsymbol{x}} \star \boldsymbol{\delta}_k)_m^*] = [\boldsymbol{\Gamma}_x]_{n,k}$

	WSS	
	Shift	Modulation
PSD	$\Gamma_x(\omega) = \int_{-\infty}^{\infty} \gamma_x(t)e^{-j\omega t}d\omega$ always	
	GWSS	
	GWSS_T	GWSS_M
	$\widehat{\Gamma}_x(\boldsymbol{\Lambda}) = \mathbf{U}^*\boldsymbol{\Gamma}_x\mathbf{U}$ if $\boldsymbol{\Lambda}$ is distinct	$\widehat{\Gamma}_x(\boldsymbol{\Lambda}) = \mathbf{U}^*\boldsymbol{\Gamma}_x\mathbf{U}$ always

We denote by $\delta_t = \delta(t_0 - t)$, $\bar{x}(t) = x(t) - \mu_x$, $\bar{\boldsymbol{x}} = \boldsymbol{x} - \eta_x\mathbf{1}$, and $\boldsymbol{\Pi} = \pi\sqrt{\boldsymbol{\Lambda}/\rho_\mathcal{G}}$.

Not surprisingly, WSS has several equivalent expressions because of the correspondence between time-shift and frequency-modulation, i.e., $x(t-\tau) \leftrightarrow e^{-j\tau\omega}X(\omega)$. Here, we present another two expressions of WSS to show the connection with GWSSs.

Corollary 3 (WSS by shift)**.** *Let $T_{t_0}\{\cdot\}$ be the shift operator that delays the signal by t_0, i.e., $T_{t_0}\{x(t)\} = x(t-t_0)$. Definition 11 can be written as follows:*

1. $\mathbb{E}\left[T_{t_0}\{x(t)\}\right] = \mu_x = \text{const},$ (3.52)
2. $\mathbb{E}\left[(T_{t_0}\{x(t)\} - \mu_x)(T_{t_0}\{x(\tau)\} - \mu_x)^*\right] = \gamma_x(t-\tau).$ (3.53)

The equivalence of conditions (3.52) and (3.49) is easy to verify. Since the autocovariance only depends on the time difference, (3.53) is identical to (3.50).

Noting that

$$x(\tau) = x(t) * \delta(t-\tau), \tag{3.54}$$

leads to the following corollary.

Corollary 4 (WSS by modulation)**.** *Definition 11 can be expressed equivalently by using $\delta(t)$ as follows:*

1. $\mathbb{E}\left[x(t_0) * \delta(t_0 - \tau)\right] = \mu_x = \text{const},$ (3.55)
2. $\mathbb{E}\left[(x(t_0) * \delta(t_0 - t) - \mu_x)(x(t_0) * \delta(t_0 - \tau) - \mu_x)^*\right] = \gamma_x(t-\tau).$ (3.56)

In the chapter, we utilize these expressions of WSS for formally defining GWSS.

Appendix 3.B Graph wide sense stationarity

In this appendix, we describe and compare some definitions of GWSS, including ours. These definitions differ in whether Corollary 3 or 4 in Appendix 3.A is used for the baseline. They coincide for time-domain signals, but this is not the case for graph signals, leading to slightly different definitions of GWSS. The definitions are mainly divided according to whether the covariance is diagonalizable by the GFT basis $\mathbf{U}$. These differences are summarized in Table 3.4.

GWSS followed by Corollary 3

First, we show the definition of GWSS [29] as a counterpart of Corollary 3. In the literature of GSP, the graph Laplacian $\mathbf{L}$ is often referred to as a counterpart of the time-shift operator (translation operator) [56,72]. However, $\mathbf{L}$ changes the signal energy, i.e., $\|\mathbf{L}\boldsymbol{x}\| \neq \|\boldsymbol{x}\|$, while time shift does not, i.e., $\|T_\tau\{x[n]\}\| = \|x[n]\|$.

The definition of GWSS introduced here is based on a graph-translation operator which preserves the signal energy.

Definition 12 (Graph wide sense stationary by translation ($\mathrm{GWSS_T}$) [83])**.** *Let $\boldsymbol{x}$ be a graph signal on a graph $\mathcal{G}$. Suppose that a graph-translation operator $\mathbf{T}_\mathcal{G}$ is defined by*

$$\mathbf{T}_\mathcal{G} \coloneqq \exp\left(j\pi\sqrt{\frac{\mathbf{L}}{\rho_\mathcal{G}}}\right), \tag{3.57}$$

where $\rho_\mathcal{G} = \max_{m\in\mathcal{V}} \sqrt{2d_m(d_m + \overline{d}_m)}$, $\overline{d}_m = \frac{\sum_{n=0}^{N-1} a_{mn}d_n}{d_m}$ and $\rho_\mathcal{G} \geq \lambda_{\max}$ [84]. Then, $\boldsymbol{x}$ is a graph wide sense stationary process by translation, if and only if it the following two conditions are satisfied,

1. $\mathbb{E}\left[(\mathbf{T}_\mathcal{G}\boldsymbol{x})_n\right] = \mu_x = \text{const}$, (3.58)
2. $\mathbb{E}\left[(\mathbf{T}_\mathcal{G}\boldsymbol{x} - \mu_x\mathbf{1})_n(\mathbf{T}_\mathcal{G}\boldsymbol{x} - \mu_x\mathbf{1})_k^*\right] = [\boldsymbol{\Gamma}_x]_{n,k}$. (3.59)

The condition (3.58) implies that $\mathbb{E}[\boldsymbol{x}]$ is also constant. To see this, note that (3.58) can be expressed as

$$\mathbf{T}_\mathcal{G}\mathbb{E}[\boldsymbol{x}] = \mathbf{U}\exp(j\pi\sqrt{\boldsymbol{\Lambda}/\rho_\mathcal{G}})\mathbf{U}^*\mathbb{E}[\boldsymbol{x}]. \tag{3.60}$$

Since $\lambda_0 = 0$ and $\boldsymbol{u}_0 = \mathbf{1}$ are always satisfied, i.e., $\mathbf{L}\mathbf{1} = 0\cdot\mathbf{1}$, (3.60) holds if and only if $\mathbb{E}[\boldsymbol{x}] = \mu_x\boldsymbol{u}_0 = \mu_x\mathbf{1}$. It also implies that $\mathbf{T}_\mathcal{G}\mathbf{1} = \mathbf{1}$.

Table 3.5: Diagonalizability of the covariance for GWSS definitions. If all eigenvalues in $\mathbf{\Lambda}$ are distinct, they coincide with each other.

$\mathbf{\Gamma}_x$ is not necessarily diagonalizable by $\mathbf{U}$	$\mathrm{GWSS_T}$ [29]
$\mathbf{\Gamma}_x$ is diagonalizable by $\mathbf{U}$	$\mathrm{GWSS_M}$
$\mathbf{\Gamma}_x$ is a polynomial in $\mathbf{L}$	[26, 72]

When $\{\lambda_i\}_{i=0,\dots,N-1}$ are distinct, the condition (3.59) implies that $\mathbf{\Gamma}_x$ is diagonalizable by $\mathbf{U}$ because (3.59) can be expressed as

$$\begin{aligned}\mathbf{\Gamma}_x &= \mathbf{T}_{\mathcal{G}}\mathbb{E}[(\boldsymbol{x}-\mu_x\mathbf{1})(\boldsymbol{x}-\mu_x\mathbf{1})^*]\mathbf{T}_{\mathcal{G}}^* \\ &= \mathbf{U}\exp(j\pi\sqrt{\mathbf{\Lambda}/\rho_{\mathcal{G}}})\mathbf{U}^*\mathbf{\Gamma}_x\mathbf{U}\exp(-j\pi\sqrt{\mathbf{\Lambda}/\rho_{\mathcal{G}}})\mathbf{U}^* \\ &= \mathbf{U}(\widehat{\mathbf{\Gamma}}_x \circ \mathbf{\Theta})\mathbf{U}^*, \end{aligned} \tag{3.61}$$

where $\widehat{\mathbf{\Gamma}}_x = \mathbf{U}^*\mathbf{\Gamma}_x\mathbf{U}$ and $[\mathbf{\Theta}]_{i,l} = \exp\{j\pi(\sqrt{\lambda_i/\rho_{\mathcal{G}}} - \sqrt{\lambda_l/\rho_{\mathcal{G}}})\}$. The third equality follows by the relationship $\mathrm{diag}(\boldsymbol{a})\mathbf{X}\mathrm{diag}(\boldsymbol{b}^*) = \mathbf{X}\circ\boldsymbol{a}\boldsymbol{b}^*$. Since $[\mathbf{\Theta}]_{i,l} = 1$ for $\lambda_i = \lambda_l$ and $[\mathbf{\Theta}]_{i,l} \neq 1$ otherwise, the equality holds if and only if $[\widehat{\mathbf{\Gamma}}_x]_{i,l} = 0$ for $\lambda_i \neq \lambda_l$. If the eigenvalues are distinct, then this condition is equivalent that $\mathbf{\Gamma}_x$ is diagonalizable by $\mathbf{U}$.

Relationship among GWSS definitions

We now discuss the relationship among some representative definitions of GWSS from the viewpoint of the PSD. These results are summarized in Table 3.5.

Existing definitions of GWSS are defined as counterparts of the classical WSS definitions in Corollaries 3 and 4. In terms of diagonalizability, $\mathrm{GWSS_M}$ (the counterpart of Corollary 3) is stricter than $\mathrm{GWSS_T}$ (the counterpart of Corollary 4). This is because the covariance is always diagonalizable by $\mathbf{U}$ in $\mathrm{GWSS_M}$, while that is not the case with $\mathrm{GWSS_T}$ in general.

Next, we compare $\mathbf{\Theta}$ in (3.61) with $\mathbf{\Xi}$ in (3.18). In (3.18), off-diagonal entries in $\mathbf{\Xi}$ are not equal to 1. In contrast, those in $\mathbf{\Theta}$ can take the value 1 in the case $\lambda_i = \lambda_l$ for $i \neq l$, i.e., eigenvalues with multiplicity greater than 1. Therefore, $\mathrm{GWSS_T}$ allows the existence of non-zero off-diagonal elements of $\widehat{\mathbf{\Gamma}}_x$ for some graphs having repeated eigenvalues, while $\mathrm{GWSS_M}$ always yields the diagonal $\widehat{\mathbf{\Gamma}}_x$ regardless of graphs. In fact, $\mathrm{GWSS_M}$ and $\mathrm{GWSS_T}$ coincide with each other if all eigenvalues of $\mathbf{L}$ are distinct (cf. Lemma 1). Therefore, $\mathrm{GWSS_T}$ effectively assumes distinct eigenvalues of $\mathbf{L}$ in [29].

It is often assumed that $\mathbf{\Gamma}_x$ is a polynomial in $\mathbf{L}$ [26, 72][4]. This is sufficient for the diagonalizability of $\mathbf{\Gamma}_x$. It is noteworthy that the polynomial assumption is equivalent to

[4]In this dissertation, we use $\mathbf{L}$ as a counterpart of time-shift for simplicity. Nevertheless, we can easily extend the proposed GWSS for other graph variation operators, including $\mathbf{A}$ and $\boldsymbol{\mathcal{L}} \coloneqq \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$.

$\widehat{\Gamma}_x(\lambda_i) = \widehat{\Gamma}_x(\lambda_l)$ for all $\lambda_i = \lambda_l$ [26]. This is a special case of GWSS_M, since GWSS_M allows for $\widehat{\Gamma}_x(\lambda_i) \neq \widehat{\Gamma}_x(\lambda_l)$ for any $\lambda_i = \lambda_l$. As a result, GWSS_M is a good compromise between applicability and a rigorous relationship to the WSS definition. Therefore, we use GWSS_M as our GWSS definition.

Chapter 4

Sampling set selection for graph signals under arbitrary signal priors

4.1 Introduction

Sampling of graph signals is one of the key topics in graph signal processing (GSP) [10–12, 14, 15, 20, 56]. Many studies of graph signal sampling only focus on bandlimited graph signal models [10–12,14,20,46,56,56]. However, the bandlimited setting is *one of the possible subspaces of deterministic graph signals*. Many other full-band graph signal models are also useful for applications. For example, we often encounter piecewise constant/smooth graph signals [49] that are full-band in graph Fourier domain. Other full-band graph signal subspaces are also under consideration like periodic graph spectrum signals [48]. Unfortunately, sampling for arbitrary signal subspaces has not been studied without a few exceptions [48,67].

In this chapter, we propose a sampling set selection (SSS) algorithm for graph signals under arbitrary graph signal models. It is known that graph signal sampling (and recovery) with arbitrary signal models is well represented by a *generalized sampling* framework [15, 16, 63]. In generalized sampling, a correction operator is placed between sampling and reconstruction operators to compensate signal subspaces. An important condition to design the optimal correction operator is called the *direct sum* (DS) condition. This can be represented with two subspaces: The sampled signal subspace and the reconstruction subspace. Roughly speaking, we seek the optimal sampling set such that the two subspaces are maximally close to each other in the proposed SSS method.

The proposed SSS algorithm is deterministic and selects nodes with a greedy algorithm. For fast implementation, we also develop the sampling set search along with the Neumann series approximation for the matrix inversion in the objective function. In the recovery experiments, we demonstrate that the proposed SSS exhibits the lower MSE than existing SSS algorithms for various signal models.

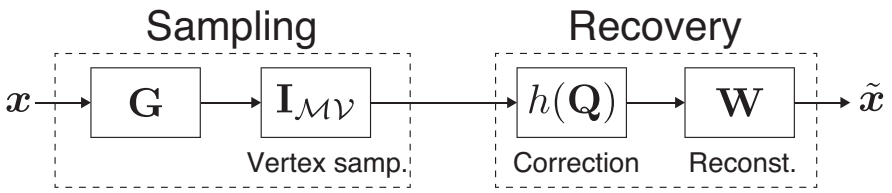


Figure 4.1: Generalized sampling framework. The dotted left and right boxes are sampling and recovery phases, respectively.

4.2 Generalized sampling of graph signals

In this section, we briefly review generalized graph signal sampling [48, 70] because the proposed SSS algorithm is designed based on this framework.

Framework of graph signal sampling

First of all, we define a nodal domain sampling operator as follows:

Definition 13 (Nodal domain sampling)**.** *Let* $\mathbf{I}_{\mathcal{MV}} \in \{0,1\}^{K\times N}$ *be the submatrix of the identity matrix indexed by* $\mathcal{M} \subset \mathcal{V}$ $(|\mathcal{M}| = K)$ *and* $\mathcal{V}$*. The sampling operator is defined by*

$$\mathbf{S}^{\mathsf{T}} \coloneqq \mathbf{I}_{\mathcal{MV}}\mathbf{G}, \tag{4.1}$$

where $\mathbf{G} \in \mathbb{R}^{N\times N}$ *is an arbitrary graph filter. A sampled graph signal is thus given by* $\boldsymbol{y} = \mathbf{S}^{\mathsf{T}}\boldsymbol{x}$*.*

Note that (4.1) is a natural extension of the sampling operator for the standard time domain sampling to the graph setting. In the graph setting, the reconstruction operator depends on the chosen sampling operator $\mathbf{S}^{\mathsf{T}}$ because $\mathbf{I}_{\mathcal{MV}}$ is not unique in general.

The sampling and recovery framework based on generalized sampling is illustrated in Fig. 4.1. The input signal $\boldsymbol{x}$ is sampled by $\mathbf{S}^{\mathsf{T}}$, and the sample is corrected by a *correction operator* $h(\mathbf{Q}) : \mathbb{R}^{N\times K} \mapsto \mathbb{R}^{K\times K}$. Note that $\mathbf{Q}$ plays the key role of this chapter, which is derived from signal priors. Hereafter, we refer to $\mathbf{Q}$ as a *prior operator*. After correction, the corrected samples are reconstructed by a reconstruction operator $\mathbf{W} \in \mathbb{R}^{N\times K}$.[1] This framework explains many existing graph signal sampling, including sampling for bandlimited signals [15].

As a result, the reconstructed graph signal is represented as follows:

$$\tilde{\boldsymbol{x}} = \mathbf{W}h(\mathbf{Q})\boldsymbol{y} = \mathbf{W}h(\mathbf{Q})\mathbf{S}^{\mathsf{T}}\boldsymbol{x}. \tag{4.2}$$

[1]For simplicity, we suppose that $\mathbf{W}^{\mathsf{T}}\mathbf{W}$ is invertible.

In this scenario, the recovery problem turns out to be seeking the best possible $h(\mathbf{Q})$ based on signal priors.

In the following, we introduce representative signal models and recovery methods corresponding to the models (i.e., designs of $h(\mathbf{Q})$ and $\mathbf{W}$).

Graph signal models and signal recovery

In this chapter, we consider three representative signal priors:

Subspace prior [48]: The generation subspace $\mathcal{A}$ spanned by $\mathbf{A}$ is known. It is formally defined as $\mathcal{A} = \left\{\boldsymbol{x} \in \mathbb{R}^N : \ \boldsymbol{x} = \mathbf{A}\boldsymbol{d}\right\}$, where $\boldsymbol{d}$ is an expansion coefficient. The well-known bandlimited setting is categorized into this prior.

Smoothnes prior [48]: The signal energy that is measured by the smoothness measuring function (i.e., high-pass filter) $\mathbf{V}$ is bounded by $\sigma \in \mathbb{R}_+$, i.e., $\mathcal{V} = \left\{\boldsymbol{x} \in \mathbb{R}^N : \ \|\mathbf{V}\boldsymbol{x}\| \leq \sigma\right\}$.

Stochastic prior [70]: The covariances of graph signals and noise, $\mathbf{\Gamma}_x$ and $\mathbf{\Gamma}_\eta$, are known.

We will describe the recovery problem for each prior.

Suppose that $\mathbf{S}^\mathsf{T}$ is given. The correction operator $h(\mathbf{Q})$ is often designed based on three well-known criteria as follows.

Least squares (LS) criterion: The LS criterion seeks the signal that minimizes the error-in-sample, i.e.,

$$\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathcal{T}} \|\mathbf{S}^\mathsf{T}\boldsymbol{x} - \boldsymbol{y}\|^2. \tag{4.3}$$

where $\mathcal{T}$ is the set of signals under consideration.

Minimax (MX) criterion: The MX criterion corresponds to the reconstruction that minimizes the worst-case error, i.e.,

$$\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \max_{\tilde{\boldsymbol{x}} \in \mathcal{T}, \mathbf{S}^\mathsf{T}\tilde{\boldsymbol{x}} = \boldsymbol{y}} \|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2. \tag{4.4}$$

The LS and MX criteria are used for signals under subspace and smoothness priors.

Minimum MSE (MMSE) criterion: When $\boldsymbol{x}$ is a random process under the stochastic prior, LS and MX may be ill-posed. Instead, the MMSE criterion seeks the signal that minimizes the MSE, i.e.,

$$\tilde{\boldsymbol{x}} = \arg\min_{\tilde{\boldsymbol{x}}} \mathbb{E}[\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2]. \tag{4.5}$$

Table 4.1: Filter designs for generalized graph signal sampling. The following definitions are used: $\mathbf{P} = (\mathbf{W}^\mathsf{T}\mathbf{W})^{-1}\mathbf{W}^\mathsf{T}\mathbf{Q}$, $\widehat{\mathbf{W}} = \mathbf{W}(\mathbf{W}^\mathsf{T}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{W})^{-1}\mathbf{W}^\mathsf{T}\mathbf{S}$, $\widetilde{\mathbf{W}} = (\mathbf{V}^\mathsf{T}\mathbf{V})^{-1}\mathbf{S}$ and $\overline{\mathbf{W}} = \mathbf{\Gamma}_x\mathbf{S}$. LS and MX solutions coincide with each other for the unconstrained case.

Prior	Criteria	Predefined		Unconstrained		
		$\mathbf{Q}$	$h(\mathbf{Q})$	$\mathbf{Q}$	$h(\mathbf{Q})$	$\mathbf{W}$
Subspace	LS	$\mathbf{W}$	$(\mathbf{S}^\mathsf{T}\mathbf{Q})^{-1}$	$\mathbf{A}$	$(\mathbf{S}^\mathsf{T}\mathbf{Q})^{-1}$	$\mathbf{A}$
	MX	$\mathbf{A}$	$\mathbf{P}(\mathbf{S}^\mathsf{T}\mathbf{Q})^{-1}$			
Smoothness	LS	$\widehat{\mathbf{W}}$	$(\mathbf{S}^\mathsf{T}\mathbf{Q})^{-1}$	$\widetilde{\mathbf{W}}$	$(\mathbf{S}^\mathsf{T}\mathbf{Q})^{-1}$	$\widetilde{\mathbf{W}}$
	MX	$\widetilde{\mathbf{W}}$	$\mathbf{P}(\mathbf{S}^\mathsf{T}\mathbf{Q})^{-1}$			
Stochastic	MMSE	$\overline{\mathbf{W}}$	$\mathbf{P}(\mathbf{S}^\mathsf{T}\mathbf{Q}+\mathbf{\Gamma}_\eta)^{-1}$	$\overline{\mathbf{W}}$	$(\mathbf{S}^\mathsf{T}\mathbf{Q}+\mathbf{\Gamma}_\eta)^{-1}$	$\overline{\mathbf{W}}$

Along with the design of $h(\mathbf{Q})$, we may consider the following two cases for $\mathbf{W}$: 1) $\mathbf{W}$ is determined before sampling and recovery (predefined case), or 2) $\mathbf{W}$ is optimally chosen (unconstrained case). Table 4.1 summarizes $\mathbf{Q}$, $h(\mathbf{Q})$, and $\mathbf{W}$ (for the unconstrained case) for the three priors. Their detailed derivations are omitted and can be found in [15, 16, 85].

In the following, we describe the important condition for the best possible recovery, i.e., requirements for $h(\mathbf{Q})$. In fact, all the recovery conditions in Table 4.1 are explained by the characteristics of $\mathbf{S}$ and $\mathbf{Q}$.

Direct sum condition

In any priors and recovery criteria mentioned above, $h(\mathbf{Q})$ contains matrix inversion (see Table 4.1). Hence, for the best possible recovery, we need to consider whether the corresponding matrix including $\mathbf{Q}$ is invertible. This condition is represented as the following DS condition[2]:

Definition 14 (Direct sum condition [16])**.** *Let $\mathcal{S}$ and $\mathcal{Q}$ be the range spaces of $\mathbf{S}$ and $\mathbf{Q}$, respectively. When $\mathcal{S}^\perp$ (the orthogonal complement of $\mathcal{S}$) and $\mathcal{Q}$ span the entire Euclidean spaces $\mathbb{R}^N$ and intersect only at the origin, the direct sum condition is satisfied, i.e., $\mathbb{R}^N = \mathcal{Q}\oplus\mathcal{S}^\perp$. If so, $\mathbf{S}^\mathsf{T}\mathbf{Q}$ is invertible. This condition is equivalent that any singular value of $\mathbf{S}^\mathsf{T}\mathbf{Q}$ is nonzero, i.e., $\sigma_i(\mathbf{S}^\mathsf{T}\mathbf{Q}) \neq 0$ for all i.*

In the next section, we formulate an optimization problem for SSS based on the DS condition in Definition 14. Note that our problem does not depend on the choice of $\mathbf{Q}$. Therefore, the proposed SSS is applicable for arbitrary signal priors as long as $\mathbf{Q}$ is appropriately chosen.

[2]When the DS condition is not satisfied, the inverse in $h(\mathbf{Q})$ can be replaced by the Moore–Penrose pseudo inverse, which corresponds to least squares or minimal norm solutions.

4.3 Proposed sampling set selection

In this section, we introduce the proposed SSS algorithm for arbitrary graph signal priors. As mentioned in the previous section, the recovery performance essentially depends on the correction filter $h(\mathbf{Q})$, and it is based on the DS condition. Therefore, we consider selecting nodes so that the node subset maximally satisfies the DS condition. Interestingly, our approach provides an intuitive geometric interpretation of the node selection, which is relevant to the DS condition. We also investigate the theoretical convergence of the proposed algorithm.

Problem formulation

First, we define the "goodness" of sampling sets through the DS condition. Suppose that the prior operator $\mathbf{Q}$ (Table 4.1) is given. Since the DS condition is satisfied if all singular values of $\mathbf{S}^\mathsf{T}\mathbf{Q}$ are nonzero, it is natural to consider maximizing $|\det(\mathbf{S}^\mathsf{T}\mathbf{Q})|$ as a criterion of the sampling set because $|\det(\mathbf{S}^\mathsf{T}\mathbf{Q})| = \prod_{i=1}^{N} \sigma_i(\mathbf{S}^\mathsf{T}\mathbf{Q})$ [86].

Recalling that $\mathbf{S}^\top = \mathbf{I}_{\mathcal{M}\mathcal{V}}\mathbf{G}$ in (4.1), let us define $\mathbf{Z}$ as follows.

$$\mathbf{Z} \coloneqq \mathbf{G}^\mathsf{T}\mathbf{Q}\mathbf{Q}^\mathsf{T}\mathbf{G}. \tag{4.6}$$

With $\mathbf{Z}$, we consider the following problem as a sampling set selection:

$$\mathcal{M}^* = \arg\max_{\mathcal{M}\subset\mathcal{V}} \det(\mathbf{Z}_\mathcal{M}). \tag{4.7}$$

Since $\det(\mathbf{Z}_\mathcal{M}) = \det(\mathbf{S}^\mathsf{T}\mathbf{Q}\mathbf{Q}^\mathsf{T}\mathbf{S}) = |\det(\mathbf{S}^\mathsf{T}\mathbf{Q})|^2$, (4.7) considers the DS condition directly. It is different from the other approaches based on the bandlimited assumption [10, 11]. Note that this cost function is applicable for any signal model (including the bandlimited one).

The direct maximization of (4.7) is combinatorial and is practically intractable. Therefore, we use a greedy algorithm like those used in the previous SSS methods. The greedy selection is formulated as follows.

$$y^* = \arg\max_{y\in\mathcal{M}^c} \det(\mathbf{Z}_{\mathcal{M}\cup\{y\}}). \tag{4.8}$$

That is, we select a node y^* one by one.

Figure 4.2: Geometric interpretation of the proposed SSS. We denote the subspaces of sampled graph signals at the n-th and $(n+1)$-th selections by $\mathcal{S}n$ and $\mathcal{S}n+1$, respectively.

Geometric interpretation

Here, we consider the geometric relationship between our objective function and the DS condition. The principal angle between $\mathcal{S}$ and $\mathcal{Q}$ is defined by [87][3]

$$\cos^2(\mathcal{S}, \mathcal{Q}) \coloneqq \frac{\det(\mathbf{S}^\mathsf{T}\mathbf{Q}\mathbf{Q}^\mathsf{T}\mathbf{S})}{\det(\mathbf{S}^\mathsf{T}\mathbf{S})\det(\mathbf{Q}^\mathsf{T}\mathbf{Q})}. \tag{4.9}$$

The DS condition in Definition 14 can be expressed by $\cos^2(\mathcal{S}^\perp, \mathcal{Q}) \neq 1$. Since $\cos^2(\mathcal{S}^\perp, \mathcal{Q}) = 1-\cos^2(\mathcal{S}, \mathcal{Q})$, the DS condition is satisfied if $\mathcal{S}$ is close to $\mathcal{Q}$. In other words, the maximization of $\det(\mathbf{Z}_\mathcal{M})$ implies the maximum separation between $\mathcal{S}^\perp$ and $\mathcal{Q}$.

The geometric interpretation of two subspace angles is depicted in Fig. 4.2. We can observe that selecting one node is equivalent to an increment in $\cos^2_{\text{pri}}(\mathcal{S}, \mathcal{Q})$. Generally speaking, $\cos^2_{\text{pri}}(\mathcal{S}, \mathcal{Q})$ measures the discrepancy between $\mathcal{Q}$ and $\mathcal{S}^\perp$. If $\cos^2_{\text{pri}}(\mathcal{S}, \mathcal{Q})$ is large, $\mathcal{Q}$ and $\mathcal{S}^\perp$ are far apart from each other. Therefore, maximizing $\cos^2_{\text{pri}}(\mathcal{S}^\perp, \mathcal{Q})$ promotes satisfaction of the DS condition as defined in Definition 14.

Proposed algorithm

The determinant term in (4.8) requires $O(|\mathcal{M}|!)$ computational complexity in the worst case if we perform to compute it straightforwardly [86]. Therefore, we rewrite (4.8) into a more efficient form.

Suppose that $\text{rank}(\mathbf{Z}) \geq K$ and $\det(\mathbf{Z}_\mathcal{M}) > 0$. By applying the Schur determinant

[3]For simplicity, we suppose that $\mathbf{S}^\mathsf{T}\mathbf{S}$ and $\mathbf{Q}^\mathsf{T}\mathbf{Q}$ are invertible.

formula [88] to (4.8), it can be rewritten as

$$\begin{aligned} y^* &= \arg\max_{y \in \mathcal{M}^c} \det(\mathbf{Z}_{\mathcal{M}\cup\{y\}}) \\ &= \arg\max_{y \in \mathcal{M}^c} \det(\mathbf{Z}_{\mathcal{M}}) \cdot (\mathbf{Z}_{y,y} - \mathbf{Z}_{y,\mathcal{M}}(\mathbf{Z}_{\mathcal{M}})^{-1}\mathbf{Z}_{\mathcal{M},y}) \\ &= \arg\max_{y \in \mathcal{M}^c} \mathbf{Z}_{y,y} - \mathbf{Z}_{y,\mathcal{M}}(\mathbf{Z}_{\mathcal{M}})^{-1}\mathbf{Z}_{\mathcal{M},y}, \end{aligned} \tag{4.10}$$

where we omit the multiplication with $\det(\mathbf{Z}_{\mathcal{M}})$ in the third equivalence because it does not depend on y^*. For the first selection, i.e., $\mathcal{M} = \emptyset$, the node y maximizing $\mathbf{Z}_{y,y}$ is chosen because it immediately follows (4.8).

Still, (4.10) is computationally expensive due to the matrix inversion, which typically requires $O(|\mathcal{M}|^3)$ computational complexity. To alleviate this, we utilize the Neumann series for $(\mathbf{Z}_{\mathcal{M}})^{-1}$, i.e.,

$$(\mathbf{Z}_{\mathcal{M}})^{-1} = \alpha \sum_{k=0}^{\infty} (\mathbf{I} - \alpha \mathbf{Z}_{\mathcal{M}})^k, \tag{4.11}$$

where α is chosen such that $\|\mathbf{I} - \alpha \mathbf{Z}_{\mathcal{M}}\| \leq 1$. Let us define $\boldsymbol{\varepsilon}_y := (\mathbf{Z}_{\mathcal{M}})^{-1}\mathbf{Z}_{\mathcal{M},y}$ in (4.10). Then, substitution of (4.11) into $\boldsymbol{\varepsilon}_y$ leads to

$$\boldsymbol{\varepsilon}_y = \alpha \sum_{k=0}^{\infty} (\mathbf{I} - \alpha \mathbf{Z}_{\mathcal{M}})^k \mathbf{Z}_{\mathcal{M},y} = \alpha \sum_{k=0}^{\infty} (\mathbf{I} - \alpha \mathbf{Z}_{\mathcal{M}})^k \mathbf{Z}_{\mathcal{M}} \boldsymbol{\delta}_y, \tag{4.12}$$

where $\boldsymbol{\delta}_y$ is the Kronecker delta centered at the node y.

In this chapter, we truncate (4.12) until the mth term. We can update (4.12) by the following rule [16]:

$$\begin{aligned} \boldsymbol{\varepsilon}_y^{m+1} &= \alpha \sum_{k=0}^{m+1} (\mathbf{I} - \alpha \mathbf{Z}_{\mathcal{M}})^k \mathbf{Z}_{\mathcal{M}} \boldsymbol{\delta}_y \\ &= \boldsymbol{\varepsilon}_y^0 + (\mathbf{I} - \alpha \mathbf{Z}_{\mathcal{M}}) \boldsymbol{\varepsilon}_y^m, \end{aligned} \tag{4.13}$$

where $\boldsymbol{\varepsilon}_y^0 = \alpha \mathbf{Z}_{\mathcal{M}} \boldsymbol{\delta}_{\boldsymbol{y}}$. Since $\boldsymbol{\varepsilon}_y^\infty = \boldsymbol{\varepsilon}_y = (\mathbf{Z}_{\mathcal{M}})^{-1}\mathbf{Z}_{\mathcal{M},y}$, the error $\boldsymbol{\varepsilon}_y^0 - \alpha \mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m$ converges to zero.

For fast convergence, we need to select an optimal α in (4.13) in each update. In particular, we seek the best step size at the $(m+1)$th iteration as follows:

$$\begin{aligned} \alpha_{m+1}^* &= \arg\min_{\alpha} \|\boldsymbol{\varepsilon}_y^{m+1} - \boldsymbol{\varepsilon}_y^m\|^2 \\ &= \arg\min_{\alpha} \|(\mathbf{I} - \alpha \mathbf{Z}_{\mathcal{M}}) \boldsymbol{\varepsilon}_y^m\|^2. \end{aligned} \tag{4.14}$$

The solution of (4.14) results in

$$\alpha_{m+1}^{*} = \frac{(\boldsymbol{\varepsilon}_y^m)^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m}{\|\mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m\|^2}. \tag{4.15}$$

In the initial step, α_0^* is arbitrarily set such that $\|\mathbf{I} - \alpha_0^* \mathbf{Z}_{\mathcal{M}}\| \leq 1$. While α is tuned in the adaptive manner, (4.13) converges to $(\mathbf{Z}_{\mathcal{M}})^{-1}\mathbf{Z}_{\mathcal{M},y}$ [16].

We present the proposed sampling set selection method in Algorithm 1. When $\mathbf{Z}$ is diagonalizable by $\mathbf{U}$, we can approximate $\mathbf{Z}$ by the Chebyshev polynomial approximation (CPA) without directly calculating $\mathbf{U}$ [89]. This could result in a further speed-up of the algorithm.

Algorithm 1: Greedy SSS for arbitrary signal prior

Input: $\mathbf{Z}, \mathcal{M} = \emptyset, K, m = 0$
while $|\mathcal{M}| < K$ **do**
  Compute $\boldsymbol{\varepsilon}_y^0 = \alpha_0^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\delta}_y$
  while $\|\boldsymbol{\varepsilon}_y^0 - \alpha_m^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m\| \geq \beta$ *for some* $\beta > 0$ **do**
    $\alpha_{m+1}^* \leftarrow \frac{(\boldsymbol{\varepsilon}_y^m)^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m}{\|\mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m\|^2}$
    $\boldsymbol{\varepsilon}_y^{m+1} \leftarrow \boldsymbol{\varepsilon}_y^0 + (\mathbf{I} - \alpha_{m+1}^* \mathbf{Z}_{\mathcal{M}}) \boldsymbol{\varepsilon}_y^m$
    $m \leftarrow m + 1$
  $y^* \rightarrow \arg\max_{y \in \mathcal{M}^c} \mathbf{Z}_{yy} - \mathbf{Z}_{y,\mathcal{M}} \boldsymbol{\varepsilon}_y^m$
  $\mathcal{M} \leftarrow \mathcal{M} \cup \{y\}$
Output: $\mathcal{M}$

Convergence analysis

In this section, we analyze the error convergence of Algorithm 1. We consider the error between $\boldsymbol{\varepsilon}_y^m$ and $\boldsymbol{\varepsilon}_y^{m+1}$, i.e., $\|\boldsymbol{\varepsilon}_y^0 - \alpha_{m+1}^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^{m+1}\|$. First of all, we have the following transformation.

$$\begin{aligned}
&\boldsymbol{\epsilon}_y^0 - \kappa \alpha_m^* \mathbf{Z}_{\mathcal{M}} (\boldsymbol{\varepsilon}_y^0 + (\mathbf{I} - \kappa \mathbf{Z}_{\mathcal{M}}) \boldsymbol{\varepsilon}_y^m) \\
=&(\mathbf{I} - \kappa \alpha_m^* \mathbf{Z}_{\mathcal{M}}) \boldsymbol{\epsilon}_y^0 - \kappa \mathbf{Z}_{\mathcal{M}} (\mathbf{I} - \kappa \mathbf{Z}_{\mathcal{M}}) \boldsymbol{\varepsilon}_y^m \\
=&(\mathbf{I} - \kappa \alpha_m^* \mathbf{Z}_{\mathcal{M}}) (\boldsymbol{\varepsilon}_y^0 - \kappa \alpha_m^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\epsilon}_y^m) \\
=&(\mathbf{I} - \kappa \alpha_m^* \mathbf{Z}_{\mathcal{M}}) (\kappa (\boldsymbol{\varepsilon}_y^0 - \alpha_m^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\epsilon}_y^m) + (1 - \kappa) \boldsymbol{\varepsilon}_y^0),
\end{aligned} \tag{4.16}$$

where $\kappa_{m+1} = \frac{\alpha^*_{m+1}}{\alpha^*_m}$ and $\kappa_0 = 1$. Then, we can bound above (4.16) as

$$\begin{aligned}\|\boldsymbol{\varepsilon}_y^0 - \alpha^*_{m+1}\mathbf{Z}_{\mathcal{M}}\boldsymbol{\varepsilon}_y^{m+1}\| \leq& \sigma_{\max}(\mathbf{I} - \kappa_m\alpha^*_m\mathbf{Z}_{\mathcal{M}})(\kappa\|\boldsymbol{\varepsilon}_y^0 - \alpha^*_m\mathbf{Z}_{\mathcal{M}}\boldsymbol{\epsilon}_y^m\| + (1-\kappa_m)\|\boldsymbol{\varepsilon}_y^0\|)\\ =&\|\boldsymbol{\varepsilon}_y^0\|\sum_{l=0}^{m}(1-\kappa_{m-l})\left(\prod_{k=1}^{l}\kappa_{m-k+1}\sigma_{\max}(\mathbf{I} - \kappa_{m-k+1}\alpha^*_{m-k+1}\mathbf{Z}_{\mathcal{M}})\right)\\ \leq&\|\boldsymbol{\varepsilon}_y^0\|\sum_{l=0}^{m}(1-\kappa_{m-l})\left(\prod_{k=1}^{l}\kappa_{m-k+1}\left(1 + \kappa_{m-k+1}\alpha^*_{m-k+1}\sigma_{\max}(\mathbf{Z}_{\mathcal{M}})\right)\right).\end{aligned} \tag{4.17}$$

In the first inequality in (4.17), we use the fact that $\|\boldsymbol{a} + \boldsymbol{b}\| \geq \|\boldsymbol{a}\| + \|\boldsymbol{b}\|$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are arbitrary vectors. In the second inequality in (4.17), we use the fact that $\sigma_{\max}(\mathbf{I} - \mathbf{Z}) \leq 1 + \sigma_{\max}(\mathbf{Z})$, where $\mathbf{Z}$ is an arbitrary matrix. Since (4.17) is a polynomial in $\sigma_{\max}(\mathbf{Z}_{\mathcal{M}})$, we can express it as

$$\begin{aligned}\|\boldsymbol{\varepsilon}_y^0 - \alpha^*_{m+1}\mathbf{Z}_{\mathcal{M}}\boldsymbol{\varepsilon}_y^{m+1}\| =&\|\boldsymbol{\varepsilon}_y^0\|\sum_{l=0}^{m}\tilde{\kappa}_l\sigma_{\max}(\mathbf{Z}_{\mathcal{M}})^l\\ \leq& C\|\boldsymbol{\varepsilon}_y^0\|\sum_{l=0}^{m}(\rho\sigma_{\max}(\mathbf{Z}_{\mathcal{M}}))^l,\end{aligned} \tag{4.18}$$

where $\tilde{\kappa}_l = (1-\kappa_{m-l})\prod_{k=1}^{l}\binom{m}{l}\kappa_{m-k+1}\alpha_{m-k+1}$ is coefficients of the polynomial. By definition of $\tilde{\kappa}_l$, there exists some constants C and $\rho < 1$ such that $\tilde{\kappa}_l < C\rho^l$. As a result, (4.18) is convergent as $m \to \infty$ at a rate proportional to $\sigma_{\mathcal{M}}(\mathbf{Z}_{\mathcal{M}})^m$.

One of resonable choices of the stopping criterion β in Algorithm 1 is utilizing the upper bound in (4.18).

$$\beta = C_{\text{user}}\|\boldsymbol{\varepsilon}_y^0\|\sum_{l=0}^{m}(\rho\sigma_{\max}(\mathbf{Z}_{\mathcal{M}}))^l, \tag{4.19}$$

where C_{user} is the user-defined convergent ratio. With (4.19), we can select an appropriate β tailored to the maximum iteration number m and the spectral norm of $\mathbf{Z}_{\mathcal{M}}$.

4.4 Recovery experiments

In this section, we demonstrate the effectiveness of the proposed method by sampling graph signals with several graph signal models. We perform recovery experiments on a random sensor graph with $N = 256$ for six graph signal models:

Subspace prior

- Bandlimited (BL) graph signals with the bandwidth $\mathcal{B} = \{1, \ldots, K/4\}$ [12].

Table 4.2: Average reconstruction MSEs (in decibels) for 100 independent runs.

Signal Model	Subspace			Smoothness		Stochastic	
	BL	PGS	PWC	GMRF	PWL	GMRF w/ noise	BP w/ noise
Proposed	**-733.9**	**-632.3**	**-684.5**	**-22.88**	**-18.12**	**-18.74**	**-25.43**
MPV [11]	-718.0	-12.57	-7.29	-18.13	-16.68	-17.64	-24.01
AVM [43]	-724.3	-12.88	-9.33	-18.66	-17.70	-18.49	-23.10
SP [12]	-718.9	-11.65	-6.88	-17.19	-14.81	-15.51	-21.45
FSSS [20]	-703.3	-14.26	-19.46	-21.98	-14.81	-13.89	-14.25

- Periodic graph spectrum (PGS) signals with the generator response $A(\lambda) = \exp(-1.5\lambda/\lambda_{\max})$ [48].
- Piecewise constant (PWC) signals with five pieces [49].

Smoothness prior

- Gaussian Markov random field (GMRF) signals with the power spectrum $\Gamma_x(\lambda) \propto 1/(\lambda+\epsilon)$ [50].
- Piecewise linear (PWL) signals with the density $p = K/N$ [18].

Stochastic prior

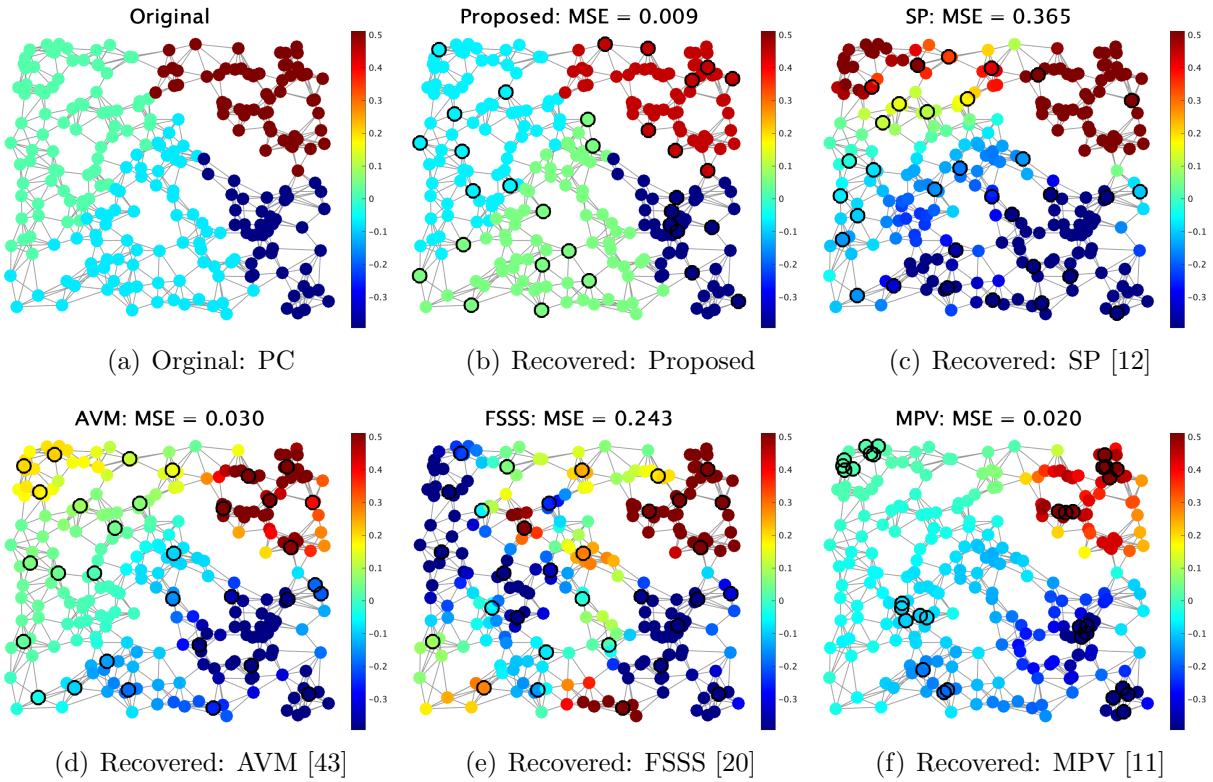
- GMRF with $\Gamma_x(\lambda) \propto 1/(\lambda + \epsilon)$.
- Bandpass (BP) graph signals with power spectrum $\Gamma_x(\lambda) \propto \exp(\{(2\lambda - \lambda_{\max})/(\sqrt{\lambda+\epsilon})\}^2)$.

For the subspace priors, we set to $\boldsymbol{d} \sim \mathcal{N}(1, 1)$. The sampling ratio is set to $K = N/8$. Further, the spectral response of the sampling filter $\mathbf{G}$ is defined by

$$\hat{g}(\lambda) := \begin{cases} 1 & \lambda \leq 1 \\ 2 - 2\lambda/(\lambda_{\max} + \epsilon) & \lambda > 1. \end{cases} \tag{4.20}$$

Under the stochastic prior, noise conforms to $\boldsymbol{\eta} \sim \mathcal{N}(0, 0.2)$. We set $\epsilon = 0.1$.

The proposed method is compared with four existing methods [11, 12, 20, 43]. Table 4.2 summarizes averaged MSEs in decibels for 100 runs. Although we only show the results in the unconstrained case due to the limitation of space, the predefined case also presents the same tendency. It is observed that the proposed method outperforms the alternative methods for all signal models. In particular, the reconstructed signals with the subspace priors are perfect recovery in machine precision. While the existing approaches based on the bandlimited



(a) Orginal: PC (b) Recovered: Proposed (c) Recovered: SP [12]

(d) Recovered: AVM [43] (e) Recovered: FSSS [20] (f) Recovered: MPV [11]

Figure 4.3: Examples of recovery for PWC graph signals on sensor graphs. Black circles represent selected nodes. Colors on nodes represent magnitudes of signal values.

model show perfect recovery for bandlimited signals, they do not perfectly recover full-band signals like PGS and PWC signals. For the other priors, the proposed method also presents consistently lower reconstruction errors than the other methods. The numerical experiments are conducted in MATLAB environments using 64-bit machine precision ($\pm 2.23 \times 10^{-308}$ to $\pm 1.80 \times 10^{308}$).

While most existing methods rely on bandlimited models, this chapter first presents perfect recovery with node domain sampling beyond the bandlimited signal model. Examples of recovered signals are visualized in Fig. 4.3 along with the selected sampling sets.

4.5 Conclusion

In this chapter, an SSS algorithm for arbitrary graph signal models is proposed. In contrast to existing approaches, we directly focus on the DS condition for signal recovery based on the generalized sampling framework. We seek the optimal sampling set such that the DS

condition is maximally satisfied. The proposed greedy SSS algorithm utilizes the Neumann series approximation for fast implementation. The recovery experiments demonstrate that the proposed method is effective for various graph signal models.

Although we assume that the signal subspace or alternative priors are exactly given, they are usually obtained by some estimation algorithms, e.g., dictionary learning and bandwidth estimation. Therefore, they generally involve an estimation error and it may cause a large error in recovered graph signals. This requires a more robust criterion of SSS. It remains as future work.

Chapter 5

Sensor placement problem for sensors with multiple specifications

5.1 Introduction

Environmental monitoring is crucial for urban infrastructures to maintain the safety and quality of our lives. Sensors may be placed at junctions/intersections of sensor networks [3, 90]. Many infrastructures can be considered as sensor networks, such as water supply [91], gas pipeline [92], and IP networks [93]. Meanwhile, monitoring all nodes at the same time is impractical due to the cost of sensors and the size of networks. Therefore, we need to select an optimal set of sensor locations from the available candidates to effectively monitor the entire network with a minimal number of sensors. This is called *sensor placement problem on graphs* (SPPG) [94, 95].[1]

SPPG seeks the optimal subset of nodes based on some criteria. One of the important criteria in the literature is the least-squares estimation error which plays a key role in existing approaches [96]. However, the least-squares criterion can only be applied to limited signal models. In particular, it needs to assume the subspace of original signals is very close to that of measured signals [16, 48].

We can achieve the best possible recovery beyond the least-squares with the minimax criterion. It has been studied in the context of *generalized sampling theory* [16, 48, 97]. Its framework is formulated so that the *worst case* estimation error is minimized, in contrast to the *average case* estimation of the least-squares. The SPPG with the minimax criterion has been studied in [15, 98].

We face another challenge in applications of SPPG: The existence of sensors having differ-

[1] This can also viewed as an SSS, as mentioned in preceding chapters. In this dissertation, we use the term SPPG, differentiating it from SSS when addressing specific constraints related to sensors.

ent specifications. In practice, we may have a set of sensors with two or more specifications while most SPPGs only assume the single specification. SPPG becomes challenging if multiple sensor types must be considered simultaneously since the deployment costs and sensing area coverage can vary depending on the sensors.

In this chapter, we propose the SPPG using a set of sensors having two or more specifications, ensuring the best possible recovery for arbitrary signal models. We select sensor placements such that they simultaneously maximize the overall coverage area and minimize the total sensing budget. We formulate the SPPG as a difference-of-convex (DC) optimization on the basis of generalized sampling theory. It is efficiently solved by the primal-dual splitting (PDS) algorithm [99]. We demonstrate the effectiveness of our approach through recovery experiments.

Notation: We let $\mathbf{A} \in \mathbb{R}^{N \times M}$ be a generator matrix of $\boldsymbol{x}$. Throughout this chapter, we define a signal subspace as follows [48]:

$$\mathcal{A} \coloneqq \left\{ \boldsymbol{x} \mid \boldsymbol{x} = \mathbf{A}\boldsymbol{d} \text{ for some } \boldsymbol{d} \in \mathbb{R}^{M} \right\}, \tag{5.1}$$

where $\boldsymbol{d}$ is some expansion coefficient.

5.2 Related work

In this section, we introduce the classical formulation for SPPG and its relaxation as a convex optimization problem.

First of all, we define a sampling operator in the nodal domain as follows [15]:

Definition 15 (Sampling in the nodal domain)**.** *Let $\mathbf{I}_{\mathcal{M}\mathcal{V}} \in \{0,1\}^{K \times N}$ be the submatrix of the identity matrix indexed by $\mathcal{M} \subset \mathcal{V}$ ($|\mathcal{M}| = K$) and $\mathcal{V}$. The sampling operator is defined by*

$$\mathbf{S}^{\top} \coloneqq \mathbf{I}_{\mathcal{M}\mathcal{V}}\mathbf{G}, \tag{5.2}$$

where $\mathbf{G} \in \mathbb{R}^{N \times N}$ is an arbitrary graph filter. A sampled graph signal is thus given by $\boldsymbol{y} = \mathbf{S}^{\top}\boldsymbol{x}$.

Note that (5.2) is a natural extension of the sampling operator for the standard time domain sampling to the graph setting. We refer to the subspace $\mathcal{S}$ spanned by $\mathbf{S}$ as sampling subspace.

Let us consider the following measurements:

$$\boldsymbol{y} = \mathbf{S}^{\top}\boldsymbol{x} + \boldsymbol{\eta}, \tag{5.3}$$

where $\boldsymbol{\eta}$ is additive white Gaussian noise. The least-squares reconstruction is obtained by solving the following problem [48]:

$$\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\mathbf{S}^{\top}\boldsymbol{x} - \boldsymbol{y}\|^2 = \mathbf{S}(\mathbf{S}^{\top}\mathbf{S})^{\dagger}\boldsymbol{y}, \tag{5.4}$$

where † denotes the Moore-Penrose pseudo inverse.

Sensor placement problem is generally formulated as

$$\mathcal{M}^* = \arg\max f(\mathcal{M}), \tag{5.5}$$

where $f : \mathcal{M} \to \mathbb{R}$ is a properly-designed cost function. There are several designs of f in the literature. Optimal experimental designs are one of the major choices, including A-, E-, and D-optimal designs (see [100]). In this chapter, we introduce the A-optimal one, which is naturally derived from the error analysis.

Suppose that the signal subspace in (5.1) is identical with the sampling subspace, i.e., $\mathcal{A} = \mathcal{S}$. Then, the expected value of the least-squares error can be expressed by

$$\begin{aligned}\mathbb{E}[\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2] =& \mathbb{E}[\|\mathbf{S}(\mathbf{S}^\top\mathbf{S})^\dagger\boldsymbol{\eta}\|^2] \\ =& \operatorname{tr}(\mathbf{S}(\mathbf{S}^\top\mathbf{S})^\dagger\boldsymbol{\Gamma}_\eta(\mathbf{S}^\top\mathbf{S})^\dagger\mathbf{S}^\top) \\ =& \operatorname{tr}(\boldsymbol{\Gamma}_\eta(\mathbf{S}^\top\mathbf{S})^\dagger) \\ \leq& \operatorname{tr}(\boldsymbol{\Gamma}_\eta)\operatorname{tr}((\mathbf{S}^\top\mathbf{S})^\dagger),\end{aligned} \tag{5.6}$$

where $\boldsymbol{\Gamma}_\eta$ is the covariance matrix of $\boldsymbol{\eta}$, i.e., $\boldsymbol{\Gamma}_\eta = \mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^\top]$. Therefore, the least-squares error is bounded by $\operatorname{tr}((\mathbf{S}^\top\mathbf{S})^\dagger)$.

Accordingly, we can design the SPPG such that the error is minimized, i.e.,

$$\begin{aligned}\mathcal{M}^* =& \arg\min_{\mathcal{M}\subset\mathcal{V}} \operatorname{tr}((\mathbf{S}^\top\mathbf{S})^\dagger) \\ =& \arg\max_{\mathcal{M}\subset\mathcal{V}} \operatorname{tr}(\mathbf{S}^\top\mathbf{S}) \\ =& \arg\max_{\mathcal{M}\subset\mathcal{V}} \operatorname{tr}(\mathbf{G}^\top \operatorname{blkdiag}(\mathbf{I}_\mathcal{M}, \mathbf{O})\mathbf{G}),\end{aligned} \tag{5.7}$$

where $\operatorname{blkdiag}(\mathbf{B}_1, \mathbf{B}_2, \ldots)$ is the block diagonal matrix composed of $\{\mathbf{B}_\ell\}_{\ell=1,\ldots}$, and $\mathbf{O}$ is an all-zero matrix. By introducing $\boldsymbol{m} : \mathcal{V} \to \{0, 1\}$ as an indicator vector corresponding to $\mathcal{M}$, we can rewrite (5.7) as

$$\boldsymbol{m}^* = \arg\max_{\boldsymbol{m}\in\{0,1\}^N} \operatorname{tr}(\mathbf{G}^\top \operatorname{diag}(\boldsymbol{m})\mathbf{G}) \quad \text{s.t. } \mathbf{1}^\top\boldsymbol{m} = K, \tag{5.8}$$

where $\mathbf{1}$ is an all-one vector. Since (5.8) is combinatorial, which is NP hard, it is impractical to solve it in practice. Therefore, SPPG usually resorts to alternative relaxation/approximation approaches. Many alternatives have been studied in the literature, including convex relaxation [96, 101], greedy algorithm [10], and quadratic programming [102], to name a few.

This chapter utilizes a convex relaxation of (5.8), which is advantageous to formulate the optimization with multiple constraints. In the relaxation, we use $\boldsymbol{s} \in [0, 1]^N$ instead of

$\boldsymbol{m} \in \{0,1\}^N$, i.e., the relaxation of binary labeling to continuous labeling of sensor locations. With this modification, (5.8) is cast into the convex counterpart as

$$\boldsymbol{s}^* = \underset{\boldsymbol{s}\in[0,1]^N}{\arg\min} - \mathrm{tr}(\mathbf{G}^\top \mathrm{diag}(\boldsymbol{s})\mathbf{G}) \quad \text{s.t. } \mathbf{1}^\top \boldsymbol{s} = K. \tag{5.9}$$

In fact, sensor placement problem in (5.9) is widely accepted in the literature, which has been first considered in the seminal work in [96].

Note that all of the above-mentioned methods implicitly assume all available sensors have one single specification: In the next section, we formulate an SPPG for a set of sensors having multiple specifications.

5.3 Proposed SPPG

In this section, we propose SPPG for a set of sensors with multiple specifications. First, we formulate SPPG based on the minimax criterion, which is naturally derived from the generalized sampling framework of graph signals. Next, we incorporate the constraints on sensor specifications into the problem.

SPPG based on minimax recovery

In general, the signal subspace differs from the sampling subspace. In this case, SPPG based on the least-squares mentioned in Sec. 5.2 may not provide the best possible recovery [16]. Herein, we suppose that $\mathcal{A} \neq \mathcal{S}$.

To achieve the best possible recovery, we consider the following minimax recovery problem [48]:

$$\tilde{\boldsymbol{x}} = \underset{\tilde{\boldsymbol{x}}\in\mathcal{A}}{\arg\min} \max_{\mathbf{S}^\top \boldsymbol{x} = \boldsymbol{y}} \|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2 = \mathbf{A}(\mathbf{S}^\top \mathbf{A})^\dagger \boldsymbol{y}, \tag{5.10}$$

where $\boldsymbol{y}$ is defined in (5.3). Here, we suppose that $\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S}$ is invertible for simplicity[2]. Recall that the original signal is defined as $\boldsymbol{x} = \mathbf{A}\boldsymbol{d}$ in (5.1). Then, the error of the minimax recovery is given by

$$\begin{aligned}
&\mathbb{E}[\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2] \\
&= \mathrm{tr}(\boldsymbol{x}\boldsymbol{x}^\top - \mathbf{E}\boldsymbol{x}\boldsymbol{x}^\top \mathbf{E}^\top) + \mathrm{tr}(\mathbf{A}(\mathbf{S}^\top \mathbf{A})^\dagger \boldsymbol{\Gamma}_\eta (\mathbf{A}^\top \mathbf{S})^\dagger \mathbf{A}^\top) \\
&\leq \mathrm{tr}(\boldsymbol{\Sigma}_d)\,\mathrm{tr}(\mathbf{A}^\top \mathbf{A})\,\mathrm{tr}(\mathbf{I} - \mathbf{A}^\top \mathbf{S}(\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S})^{-1}\mathbf{S}^\top \mathbf{A}) \\
&\quad + \mathrm{tr}(\boldsymbol{\Gamma}_\eta)\,\mathrm{tr}(\mathbf{A}^\top \mathbf{A})\,\mathrm{tr}((\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S})^{-1}) \\
&= \mathrm{tr}(\boldsymbol{\Gamma}_\eta)\,\mathrm{tr}(\mathbf{A}^\top \mathbf{A})\,\mathrm{tr}((\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S})^{-1}),
\end{aligned} \tag{5.11}$$

[2]The same formulation may also be derived in the non-invertible case.

where $\mathbf{E} = \mathbf{A}(\mathbf{S}^\top \mathbf{A})^\dagger \mathbf{S}^\top$ and $\mathbf{\Sigma}_d = \boldsymbol{d}\boldsymbol{d}^\top$. In the third equality in (5.11), we use the fact that the first term becomes zero, i.e.,

$$\begin{aligned} &\operatorname{tr}(\mathbf{I} - \mathbf{A}^\top \mathbf{S}(\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S})^{-1}\mathbf{S}^\top \mathbf{A})) \\ &= \operatorname{tr}(\mathbf{I} - (\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S})^{-1}\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S})) = 0. \end{aligned} \tag{5.12}$$

If we assume the noise-free case, i.e., $\mathbf{\Gamma}_\eta = \mathbf{0}$ in (5.11) and (5.12), $\boldsymbol{x}$ is perfectly recovered, i.e., $\mathbb{E}[\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2] = 0$.

Accordingly, we can formulate the SPPG that minimizes (5.11) as follows:

$$\begin{aligned} \mathcal{M}^* &= \arg\min_{\mathcal{M}\subset\mathcal{V}} \operatorname{tr}((\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S})^{-1}) \\ &= \arg\max_{\mathcal{M}\subset\mathcal{V}} \operatorname{tr}(\mathbf{S}^\top \mathbf{A}\mathbf{A}^\top \mathbf{S}) \\ &= \arg\max_{\mathcal{M}\subset\mathcal{V}} \operatorname{tr}(\mathbf{A}^\top \mathbf{G}^\top \operatorname{blkdiag}(\mathbf{I}_\mathcal{M}, \mathbf{O})\mathbf{G}\mathbf{A}). \end{aligned} \tag{5.13}$$

We can rewrite (5.13) as follows with the same fashion as (5.8):

$$\boldsymbol{m}^* = \arg\max_{\boldsymbol{m}\in\{0,1\}^N} \operatorname{tr}(\mathbf{A}^\top \mathbf{G}^\top \operatorname{diag}(\boldsymbol{m})\mathbf{G}\mathbf{A}) \quad \text{s.t. } \mathbf{1}^\top \boldsymbol{m} = K. \tag{5.14}$$

As a result, the convex relaxation of (5.14) results in

$$\boldsymbol{s}^* = \arg\max_{\boldsymbol{s}\in[0,1]^N} \operatorname{tr}(\mathbf{A}^\top \mathbf{G}^\top \operatorname{diag}(\boldsymbol{s})\mathbf{G}\mathbf{A}) \quad \text{s.t. } \mathbf{1}^\top \boldsymbol{s} = K. \tag{5.15}$$

(5.15) is the generalization of (5.9) to an arbitrary signal model since we do not assume any specific $\mathbf{A}$. In the following, we further extend (5.15) into SPPG for a set of sensors with multiple specifications.

SPPG for sensors having multiple specifications

Suppose that sensors have different deployment costs and area coverage. We formulate SPPG with the assumption based on a DC function and then derive the PDS algorithm to solve it.

For simplicity, we focus on SPPG for a set of sensors with two different specifications. We can easily extend it to any number of specifications straightforwardly. Let $\mathbf{S}_1^\top = \mathbf{I}_{\mathcal{M}_1\mathcal{V}}\mathbf{G}_1$ and $\mathbf{S}_2^\top = \mathbf{I}_{\mathcal{M}_2\mathcal{V}}\mathbf{G}_2$ be the sampling operators corresponding to the two sensor placements with Specifications 1 and 2 (see Definition 15). Similar to (5.10), the minimax recovery results in

$$\tilde{\boldsymbol{x}} = \arg\min_{\tilde{\boldsymbol{x}}\in\mathcal{A}} \max_{\substack{\mathbf{S}_1^\top \boldsymbol{x} = \boldsymbol{y}_1 \\ \mathbf{S}_2^\top \boldsymbol{x} = \boldsymbol{y}_2}} \|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2 = \begin{bmatrix}\mathbf{A} & \mathbf{A}\end{bmatrix} \begin{bmatrix}\mathbf{S}_1^\top \mathbf{A} \\ \mathbf{S}_2^\top \mathbf{A}\end{bmatrix}^\dagger \begin{bmatrix}\boldsymbol{y}_1 \\ \boldsymbol{y}_2\end{bmatrix}, \tag{5.16}$$

where $\boldsymbol{y}_1 = \mathbf{S}_1^\top \boldsymbol{x} + \boldsymbol{\eta}$ and $\boldsymbol{y}_2 = \mathbf{S}_2^\top \boldsymbol{x} + \boldsymbol{\eta}$ are measured graph signals for each sensor.

Here, we also consider the convex formulation of the SPPG as presented in (5.15). Let $\boldsymbol{s}_1 \in [0,1]^N$ and $\boldsymbol{s}_2 \in [0,1]^N$ be the sensor positions of sensors with Specifications 1 and 2, respectively (see (5.15)). In the same manner as (5.15), we easily obtain the SPPG for two sensors as follows:

$$\begin{bmatrix} \boldsymbol{s}_1^* \\ \boldsymbol{s}_2^* \end{bmatrix} = \underset{\boldsymbol{s}_1, \boldsymbol{s}_2 \in [0,1]^N}{\arg\min} \; -\mathrm{tr}\left(\mathbf{A}^\top \begin{bmatrix} \mathbf{G}_1^\top & \mathbf{G}_2^\top \end{bmatrix} \mathrm{diag}(\boldsymbol{s}_1, \boldsymbol{s}_2) \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} \mathbf{A} \right), \tag{5.17}$$

By comparing (5.15) and (5.17), $\boldsymbol{s}$ and $\mathbf{G}$ in (5.15) can viewed as $\boldsymbol{s} = \begin{bmatrix} \boldsymbol{s}_1^\top & \boldsymbol{s}_2^\top \end{bmatrix}^\top$ and $\mathbf{G} = \begin{bmatrix} \mathbf{G}_1^\top & \mathbf{G}_2^\top \end{bmatrix}^\top$, respectively.

Note that (5.17) is still an incomplete formulation since it does not take into account the difference in the sensors' specifications. In addition, sensors having different specifications are possible to be placed at the same position as this formulation. Below, we design the constraints for their deployment costs, coverage areas, and sensor positions.

Deployment costs

While the deployment cost of sensors has been assumed to be uniform so far, we now assume that there are two different deployment costs of sensors. Let $c_1 > 0$ and $c_2 > 0$ be the costs of sensors with Specifications 1 and 2, and suppose $c_1 < c_2$ without loss of generality. The budget constraint is characterized by

$$\begin{bmatrix} c_1 \mathbf{1}_N^\top & c_2 \mathbf{1}_N^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \end{bmatrix} \leq K. \tag{5.18}$$

Note that K can be viewed as the total budget of sensors herein, while K is used as the number of sensors in (5.9).

Coverage area

Recall that $\mathbf{I}_{\mathcal{M}\mathcal{V}}$ in (5.2) specifies the sensor positions on $\mathcal{G}$. A graph filter $\mathbf{G}$ can be viewed as a coverage area of the sensors: For example, we consider

$$[\mathbf{G}]_{l,k} = \sum_{i=0}^{N-1} \sum_{p=0}^{P} \alpha_p \lambda^p u_i[l] u_i[k] = \left[\mathbf{U} \left(\sum_{p=0}^{P} \alpha_p \mathbf{\Lambda}^p \right) \mathbf{U}^\top \right]_{l,k}, \tag{5.19}$$

where $\{\alpha_p\}$ are arbitrary polynomial filter coefficients. In graph signal processing, (5.19) is referred to as a P-hop localized filter [42], whose nonzero response is limited within P-hop neighbors of the target node. This implies that the coverage area of all sensors is limited to P-hop.

In a practical scenario, $\mathbf{G}$ would be designed to aggregate the data from P-hop neighbors. For example, in TCP/IP networks, master (hub) nodes usually aggregate TCP/IP packets around neighbor slave nodes [93]. An aggregating filter can be implemented by [56]

$$\mathbf{G} = \mathbf{\Psi}\mathrm{diag}(u_i[0], u_i[1], \ldots)\mathbf{U}^\top, \tag{5.20}$$

where $[\mathbf{\Psi}]_{k,l} = \lambda_l^k$.

Here, we let $\mathbf{G}_k$ be graph filters with the P_k-hop localization for the sensor with Specification k. In practice, there should be a trade-off between the cost and coverage: A more costly sensor will have a wider coverage area. Since we assume $c_1 < c_2$ in (5.18), we set to $P_1 < P_2$.

Orthogonality of selection

Since $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ should not be overlapped in practice, we can assume they are distinct, i.e., $\boldsymbol{s}_1 \perp \boldsymbol{s}_2$. However, it is difficult to directly force the orthogonality condition in a convex optimization we consider. To this aim, we exploit the alternative constraint based on a DC function in the following theorem:

Theorem 7. *Let $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}_+^N$ be arbitrary vectors. Then, the following inequality holds*

$$\mathcal{L}(\boldsymbol{s}_1, \boldsymbol{s}_2) \coloneqq \|\boldsymbol{s}_1 + \boldsymbol{s}_2\|_1 - \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|_1 \geq 0, \tag{5.21}$$

where the equality holds if and only if $\boldsymbol{s}_1 \perp \boldsymbol{s}_2$.

We omit the proof due to the limitation of space. Therefore, the minimization of $\mathcal{L}(\boldsymbol{s}_1, \boldsymbol{s}_2)$ facilitates $\boldsymbol{s}_1 \perp \boldsymbol{s}_2$.

SPPG with PDS

We now incorporate the structure of the different specification sensors, presented in Sec. 5.3, to (5.17). We can rewrite (5.17) as

$$\begin{aligned}\begin{bmatrix}\boldsymbol{s}_1^* \\ \boldsymbol{s}_2^*\end{bmatrix} = \underset{\boldsymbol{s}_1, \boldsymbol{s}_2 \in [0,1]^N}{\arg\min} \; &- \mathrm{tr}\Big(\mathbf{A}^\top \widetilde{\mathbf{G}}^\top \mathrm{diag}(\boldsymbol{s}_1, \boldsymbol{s}_2)\widetilde{\mathbf{G}}\mathbf{A}\Big) + \mu\mathcal{L}(\boldsymbol{s}_1, \boldsymbol{s}_2), \\ &\text{s.t. } \boldsymbol{c}^\top \begin{bmatrix}\boldsymbol{s}_1 \\ \boldsymbol{s}_2\end{bmatrix} \leq K,\end{aligned} \tag{5.22}$$

where $\widetilde{\mathbf{G}} = \begin{bmatrix}\mathbf{G}_1^\top & \mathbf{G}_2^\top\end{bmatrix}^\top$, $\boldsymbol{c} = \begin{bmatrix}c_1\mathbf{1}^\top & c_2\mathbf{1}^\top\end{bmatrix}^\top$, and μ is used for encouraging the orthogonal constraint (see Thereom 7). Note that (5.22) is no longer convex but DC, which stems from

$\mathcal{L}$ (see Theorem 7). While DC optimizations are usually challenging to solve, we derive an efficient algorithm based on PDS.

We transform (5.22) into the applicable form to PDS [99]. We omit the preliminary of the PDS algorithm for brevity (see [99] and references therein). We let $\tilde{\boldsymbol{s}}$ be $\tilde{\boldsymbol{s}} = \begin{bmatrix} \boldsymbol{s}_1^\top & \boldsymbol{s}_1^\top \end{bmatrix}^\top$, and let $\mathbf{T}$ be

$$\mathbf{T} \coloneqq \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \boldsymbol{e}_i^{N\top} \mathbf{A}^\top \widetilde{\mathbf{G}}^\top \boldsymbol{e}_j^{2N} \boldsymbol{e}_j^{2N\top} \otimes \boldsymbol{e}_i^N \boldsymbol{e}_i^{N\top} \mathbf{A}^\top \widetilde{\mathbf{G}}^\top \boldsymbol{e}_j^{2N}, \tag{5.23}$$

where $\{\boldsymbol{e}_i^D\}$ is the canonical bases of $\mathbb{R}^D$. We use $\mathbf{T}$ to express the first term in (5.22) as the equivalent linear transform: $\mathbf{1}^\top \mathbf{T} \tilde{\boldsymbol{s}} = \mathrm{tr}(\mathbf{A}^\top \widetilde{\mathbf{G}}^\top \mathrm{diag}(\tilde{\boldsymbol{s}}) \mathbf{A}^\top \widetilde{\mathbf{G}}^\top)$. In addition, let $\iota_{\mathcal{C}}$ be the indicator function over a convex set $\mathcal{C}$, i.e.,

$$\iota_{\mathcal{C}}(\tilde{\boldsymbol{s}}) = \begin{cases} \tilde{\boldsymbol{s}} & \text{if } \tilde{\boldsymbol{s}} \in \mathcal{C} \\ +\infty & \text{otherwise.} \end{cases} \tag{5.24}$$

By introducing the dual variables $\boldsymbol{r}_1 = \boldsymbol{s}_1 + \boldsymbol{s}_2$, $\boldsymbol{r}_2 = \boldsymbol{s}_1 - \boldsymbol{s}_2$, and $\boldsymbol{r}_3 = \tilde{\boldsymbol{s}}$, we can rewrite (5.22) as follows:

$$\begin{aligned} \tilde{\boldsymbol{s}}^* = \arg\min_{\tilde{\boldsymbol{s}}} & -\mathbf{1}^\top \mathbf{T} \tilde{\boldsymbol{s}} + \mu(\|\boldsymbol{r}_1\|_1 - \|\boldsymbol{r}_2\|_1) + \iota_{\mathcal{C}_1}(\tilde{\boldsymbol{s}}) + \iota_{\mathcal{C}_2}(\boldsymbol{r}_3) \\ \text{s.t.} & \begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \boldsymbol{r}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \\ \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \end{bmatrix} = \mathbf{R} \tilde{\boldsymbol{s}}, \end{aligned} \tag{5.25}$$

where $\mathcal{C}_1 = [0, 1]^{2N}$ and $\mathcal{C}_2 = \{\tilde{\boldsymbol{s}} \mid \boldsymbol{c}^\top \tilde{\boldsymbol{s}} \leq K\}$.

In the PDS algorithm, we use the proximal operators of from the second to five terms in (5.25). Formally, the proximal operator of a convex function g is defined as [23]

$$\mathrm{prox}_{\mu g}(\boldsymbol{v}) \coloneqq \arg\min_{\boldsymbol{\xi}} h(\boldsymbol{\xi}) + \frac{1}{2\mu} \|\boldsymbol{\xi} - \boldsymbol{v}\|_2^2. \tag{5.26}$$

We can easily compute the proximal operators in (5.25) and summarize their analytical solutions in Table 5.1. Consequently, we describe the proposed method in Algorithm 2.

PDS is usually used for convex optimization and there exist several DC solvers, such as [103]. These DC algorithms are guaranteed to converge to a critical point, however, they would require an inner loop in every iteration to solve subproblems. As a result, the standard DC algorithms may be relatively computationally costly. To alleviate this, we approximately but efficiently solve (5.22) based on PDS.

Table 5.1: Proximal operators used in the proposed method.

$\mathrm{prox}_{\iota_{\mathcal{C}_1}}([\tilde{\boldsymbol{s}}]_\ell) = \max(\min([\tilde{\boldsymbol{s}}]_\ell, 1), 0)$
$\mathrm{prox}_{\pm\mu\|\cdot\|_1}([\boldsymbol{r}]_\ell) = \mathrm{sgn}([\boldsymbol{r}]_\ell)\max(\lvert[\boldsymbol{r}]_\ell\rvert \mp \mu, 0)$
$\mathrm{prox}_{\iota_{\mathcal{C}_2}}(\boldsymbol{r}_3) = \boldsymbol{r}_3 - \frac{\max(\boldsymbol{c}^\top \boldsymbol{r}_3 - K, 0)}{4N^2}\mathbf{1}$

Algorithm 2: SPPG algorithm for solving (5.25)

Input: $\tilde{\boldsymbol{s}}_0, \boldsymbol{r}_{3,0} \in \mathbb{R}^{2N}$, $\boldsymbol{r}_{1,0}, \boldsymbol{r}_{2,0} \in \mathbb{R}^N$, and $\tau \in]0, +\infty[$

for $n = 0, 1, \ldots$ **do**

$\boldsymbol{u}_n = \tilde{\boldsymbol{s}}_n - \tau(-\mathbf{T}^\top \mathbf{1} + \mathbf{R}^\top [\boldsymbol{r}_{1,n}^\top \boldsymbol{r}_{2,n}^\top \boldsymbol{r}_{3,n}^\top]^\top)$

$\boldsymbol{u}_{1,n} = \boldsymbol{r}_{1,n} + \tau[\mathbf{I}\,\mathbf{I}]\tilde{\boldsymbol{s}}_{1,n}$

$\boldsymbol{u}_{2,n} = \boldsymbol{r}_{2,n} + \tau[\mathbf{I}\ -\mathbf{I}]\tilde{\boldsymbol{s}}_{2,n}$

$\boldsymbol{u}_{3,n} = \boldsymbol{r}_{3,n} + \tau\,\mathrm{blkdiag}(\mathbf{I}, \mathbf{I})\tilde{\boldsymbol{s}}_{2,n}$

$\boldsymbol{p}_n = \mathrm{prox}_{\tau\iota_{[0,1]^{2N}}}(\boldsymbol{u}_n)$

$\boldsymbol{p}_{1,n} = \boldsymbol{u}_{1,n} - \tau(\mathrm{prox}_{\mu/\tau\|\cdot\|_1}(\boldsymbol{u}_{1,n}/\tau))$

$\boldsymbol{p}_{2,n} = \boldsymbol{u}_{2,n} - \tau(\mathrm{prox}_{-\mu/\tau\|\cdot\|_1}(\boldsymbol{u}_{2,n}/\tau))$

$\boldsymbol{p}_{3,n} = \boldsymbol{u}_{3,n} - \tau(\mathrm{prox}_{\iota_{\boldsymbol{c}^\top \cdot \leq K}}(\boldsymbol{u}_{3,n}/\tau))$

$\boldsymbol{q}_n = \boldsymbol{p}_n - \tau(-\mathbf{T}^\top \mathbf{1} + \mathbf{R}^\top [\boldsymbol{p}_{1,n}^\top \boldsymbol{p}_{2,n}^\top \boldsymbol{p}_{3,n}^\top]^\top)$

$\boldsymbol{q}_{1,n} = \boldsymbol{p}_{1,n} + \tau[\mathbf{I}\,\mathbf{I}]\boldsymbol{p}_n$

$\boldsymbol{q}_{2,n} = \boldsymbol{p}_{2,n} + \tau[\mathbf{I}\ -\mathbf{I}]\boldsymbol{p}_n$

$\boldsymbol{q}_{3,n} = \boldsymbol{p}_{3,n} + \tau\,\mathrm{blkdiag}(\mathbf{I}, \mathbf{I})\boldsymbol{p}_n$

$(\tilde{\boldsymbol{s}}_{n+1}, \boldsymbol{r}_{i,n+1}) = (\tilde{\boldsymbol{s}}_n, \boldsymbol{r}_{i,n}) - (\boldsymbol{u}_n, \boldsymbol{u}_{i,n}) + (\boldsymbol{q}_n, \boldsymbol{q}_{i,n}) \quad (i = 1, \ldots, 3)$

Output: $\tilde{\boldsymbol{s}}_n$

5.4 Discussion toward practical implementation

In this section, we discuss the practical designs of parameters indicating sensor specifications in our method: Sensing budget and coverage area (hop counts).

Wireless sensor networks are a compelling application of SPPG, where their edges indicate the communication among sensors [1, 4, 90]. In this context, selected and non-selected nodes are referred to as *sink* and *source* sensors, respectively.

Sensor prices

Usually, the deployment cost of sensors is designed depending on the market price of sensors. Many factors of sensors vary the price, including communication ability/performance and types of measurements [104].

Suppose that the objective measurement types are uniform in the sensor network. The communication ability is typically designed by the communicating distance on the graph [104]. For example, in sensor networks, all source nodes must be capable of communicating with at least one nearest sink node while some source nodes may be located near multiple sink nodes (i.e., intermediate sensors) and able to communicate with them. Since the number of adjacent sink nodes of the source sensors depends on the number of interfaces, the intermediate sensors are more costly than normal ones.

Hop counts

Sink sensors are desired to collect the measurements from all source sensors efficiently. If the shortest distance (i.e., hop count) that sink sensors can communicate is large, its coverage region becomes large while its communication cost should be expensive. Therefore, the hop count is usually determined depending on the trade-off.

The energy consumption model is widely used as a metric of communication cost in the literature [105]. Let us consider one-bit data communication from the source and sink nodes, which are R meters away from each other. Denoting a hop count by $P \in \mathbb{N}$, it is defined as follows [105]:

$$E_{\text{total}}(P) = \begin{cases} k\left[E_{\text{elec}} + E_{\text{fs}}\left(\frac{R}{P}\right)^2\right] + (k-1)E_{\text{elec}} & \text{if } \frac{R}{P} < D_0 \\ k\left[E_{\text{elec}} + E_{\text{mp}}\left(\frac{R}{P}\right)^4\right] + (k-1)E_{\text{elec}} & \text{if } \frac{R}{P} \geq D_0, \end{cases} \tag{5.27}$$

where E_{elec}, E_{fs}, E_{mp}, and D_0 are the energy cost to transmit or receive to one-bit signal, amplification coefficient for the free space model, that for the multipath fading model, and is a threshold value. Following from (5.27), the energy consumption is a convex function. By minimizing (5.27), we have the optimal hop count with

$$P^\star = \left\lceil \frac{R}{D_0} \right\rceil, \tag{5.28}$$

where $\lceil \cdot \rceil$ is the ceiling function.

In summary, we would practically implement SPPG with a sensor having multiple specifications in three steps:

1. We list up J available sensors and calculate their optimal hop count $\{P_j^\star\}_{j=1,\ldots,J}$ in (5.28).

2. We design analysis filters $\{\mathbf{G}_j\}_{j=1,\ldots,J}$ so that their polynomial orders accord with the number of interfaces of the J sensors.

3. We perform SPPG by Algorithm 2.

5.5 Experiments

In this section, we perform a signal recovery experiment for synthetic graph signals. We construct a K-nearest neighbor graph from randomly distributed points in 2-D space $[0,1] \times [0,1]$. The number of nodes is set to $N = 256$. We consider the bandlimited graph signal model, i.e., $\boldsymbol{x} = \mathbf{U}_{\mathcal{VB}}\boldsymbol{d}$, where $\boldsymbol{d} \sim \mathcal{N}(1,1)$, the bandwidth is $B = 32$, and $\mathbf{U}_{\mathcal{VB}}$ is the first B eigen-vectors of $\mathbf{L}$.

In the same fashion as the previous section, we assume two sensors with different specifications. We suppose that the total sensing budget is $K = 32$ for the proposed method, and the deployment costs of sensors are $c_1 = 1$ and $c_2 = 2$, respectively. Both of the graph spectral responses of $\mathbf{G}_1$ and $\mathbf{G}_2$ are given by $\widehat{\mathbf{G}}_\ell(\mathbf{\Lambda}) = \exp(-\rho\mathbf{\Lambda})$, where ρ is empirically set to $\rho = 8.52$. Then, we derive its two different P_k-hop-localized graph filters with the Chebyshev polynomial approximation [89], whose orders are $P_1 = 5$ and $P_2 = 7$, respectively (see (5.19)). For brevity, we call $\mathbf{G}_1$ the *cheap* sensor and $\mathbf{G}_2$ the *costly* sensor. Additive noise $\boldsymbol{\eta}$ conforms to $\boldsymbol{\eta} \sim \mathcal{N}(0, 10^{-3})$. Since the resulting sensor position of the proposed method, $\tilde{\boldsymbol{s}}^*$ in (5.25), is not binary, we transform $\tilde{\boldsymbol{s}}^*$ into the binary with thresholding: We select the greatest values in $\tilde{\boldsymbol{s}}^*$ such that it stays within K.

We calculate the average MSE of reconstruction errors for 10 independent runs. Since there are no prior works on SPPG for multi-type sensors, we use the SPPG for a single-type sensor introduced in (5.9) as a baseline, and compare the average MSE with each case of $\mathbf{G}_1$ and $\mathbf{G}_2$. Furthermore, we consider two standards of comparison below:

Budget consensus: The budgets of all methods are commonly set to 32.

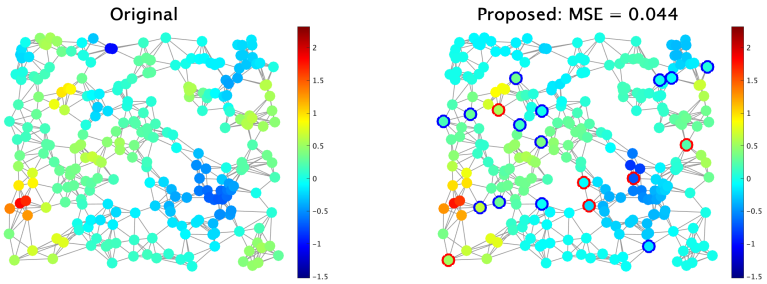
Cardinarity consensus: The numbers of sensors of all the methods are set to 19.

The parameters of Algorithm 2 are tuned by grid search.

We summarize the results in decibels in Table 5.2. We also visualize a recovered graph signal in Fig. 5.1. For the noise-free case, the proposed method exhibits lower MSE than other methods in both consensuses. It also shows the same tendency for the noisy case. This indicates that the proposed approach can adapt the tradeoff between the deployment cost and coverage of sensors.

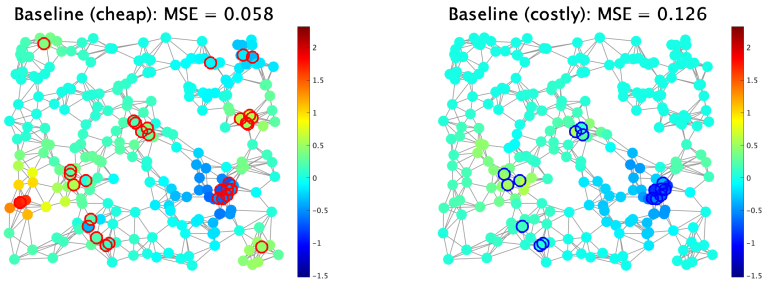
Table 5.2: Average MSEs (in decibels) for 10 independent runs.

Method	Proposed	Budget consensus		Cardinarity consensus	
		Baseline		Baseline	
		Cheap	Costly	Cheap	Costly
Clean	-24.54	-21.70	-16.63	-19.52	-17.54
Noisy	-22.14	-18.72	-13.48	-17.80	-13.96



(a) Orginal

(b) Recovered: Proposed



(c) Recovered: Baseline (cheap)

(d) Recovered: Baseline (costly)

Figure 5.1: Examples of recovered graph signals. Red- and blue-circled nodes denote the cheap and costly sensors, respectively.

5.6 Conclusion

This chapter presents an SPPG method under the constraints of sensors having multiple specifications. We generalize the classical SPPG so that it guarantees the best possible recovery under an arbitrary signal model. We devise the constraints for incorporating the structure of multiple sensor specifications into the SPPG. We derive the corresponding PDS algorithm for solving the problem. In the experiments, we demonstrate that the proposed

method can outperform existing methods. Future work includes experiments with real sensor networks.

Chapter 6

Multi-channel sampling on graphs

6.1 Introduction

Sampling of graph signals is one of the central research topics in GSP [15]. Most studies on sampling theory for graph signals, hereafter, we call it *graph sampling theory*, focus on the bandlimited graph signal model as an analog of the classical sampling theory for time-domain signals [10–12,56]. However, we often encounter full-band graph signals in many applications. For example, piecewise smooth graph signals and multi-band graph signals are classified into full-band signals. While some works study graph signal sampling beyond the bandlimited model [48, 97, 106], they consider signals under one signal model: Signals with the mixture of two or more signal models cannot be recovered properly.

Generally, full-band signals can be represented by a mixture of multiple signals conforming to different generation models. For recovering such signals, we need to consider multiple sampling systems, i.e., *multi-channel sampling* (MCS), where its single-channel sampling and recovery correspond to one signal model.

For standard time-domain signals, MCS has been studied as the Papoulis' sampling theorem [107]: The ideally-bandlimited signal can be recovered from samples obtained by MCS with arbitrary M sampling methods, e.g., non-uniform sampling and bandpass sampling. Later, it was extended into the full-band case [108–110]. This can be viewed as a special case of generalized sampling [16]. It is composed of sampling, correction, and reconstruction transforms. Sampling and reconstruction transforms can be arbitrarily chosen while the correction transform compensates for their non-ideal behaviors, ensuring that the reconstructed signal is in some sense close to the original signal. From a generalized sampling perspective, MCS can be viewed as one of the possible sampling transforms. While numerous works have been presented for time-domain MCS [108–112], there has been no approach to MCS in the graph setting in spite of having various full-band graph signals in many applications.

In this chapter, we consider MCS for GSP to recover full-band graph signals. The pro-

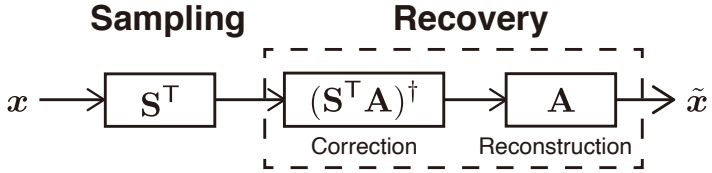


Figure 6.1: The framework of single-channel sampling.

posed MCS is derived from the above-mentioned generalized sampling by extending the sampling transform for the graph setting [16]. We also design the sampling transform for MCS on graphs. It requires the selection of a subset of nodes, i.e., sampling set selection (SSS). We select the sampling set such that graph signals are best recovered.

One can notice that MCS is related to filter banks. In fact, sampling of full-band graph signals has been studied in a different line of research: Graph filter bank (GFB) designs [54, 113–122]. GFBs are composed of multiple (typically low- and high-pass) graph filters and down- and up-sampling operators, which are also components in MCS. Typically, perfect reconstruction (PR) GFBs are designed based on the properties of the given graph operator (e.g., adjacency matrix or graph Laplacian).

Bipartite graph filter banks (BGFB) are one of the PR GFBs and they are designed so that graph signals on bipartite graphs are perfectly recovered [54, 114, 117–120]. While BGFBs can satisfy several desirable properties of GFB, they have two major limitations: 1) their PR property is limited to signals on the bipartite graph, and 2) BGFBs as well as many GFBs require the eigendecomposition of the graph operator to implement analysis and synthesis filters [54, 117–120][1], which is computationally expensive for large graphs.

The proposed method overcomes the limitations of BGFBs by viewing GFBs as a special case of MCS for graph signals: MCS can guarantee PR for arbitrary graph signals independent of the graph operator. Therefore, it does not require the graph simplification (typically bipartition). Furthermore, MCS allows for the use of arbitrary graph filters and down- and up- sampling operators. Our MCS can be implemented without eigendecomposition by utilizing polynomial filters for both analysis and synthesis, while many existing methods require eigendecomposition for achieving PR. Recovery experiments demonstrate that the proposed method outperforms exiting GFBs.

[1]There are few exceptions like methods in [114, 121], but they typically need careful filter designs.

6.2 Sampling framework for graph signals

In this section, we briefly review preliminary works on single-channel graph signal sampling [106]. First, we introduce a generalized sampling framework on graphs [15]. Second, we introduce the SSS under subspace priors.

Sampling under subspace priors

Suppose that graph signals are characterized by the following linear model:

$$\boldsymbol{x} \coloneqq \mathbf{A}\boldsymbol{d}, \tag{6.1}$$

where $\mathbf{A} \in \mathbb{R}^{N\times K}$ $(K \leq N)$ is a known generator matrix and $\boldsymbol{d} \in \mathbb{R}^K$ are expansion coefficients. The generator matrix $\mathbf{A}$ specifies the signal subspace $\mathcal{A}$.

Let $\mathbf{S}^\mathsf{T} \in \mathbb{R}^{K\times N}$ be a sampling operator. Then, sampled graph signals are given by

$$\boldsymbol{c} = \mathbf{S}^\top \boldsymbol{x}. \tag{6.2}$$

Regardless of the choice of $\mathbf{S}^\mathsf{T}$ (and $\mathbf{A}$), the best possible recovery is always given by [16,48,97]

$$\tilde{\boldsymbol{x}} = \mathbf{A}\mathbf{H}\boldsymbol{c} = \mathbf{A}(\mathbf{S}^\mathsf{T}\mathbf{A})^\dagger \mathbf{S}^\mathsf{T}\boldsymbol{x}, \tag{6.3}$$

where $\dagger$ is the Moore-Penrose inverse. The framework is illustrated in Fig. 6.1. The sampling matrix $\mathbf{S}^\mathsf{T}$ specifies the sampling subspace $\mathcal{S}$. The correction matrix is therefore represented as $\mathbf{H} = (\mathbf{S}^\mathsf{T}\mathbf{A})^\dagger$. If $\mathcal{A}$ and $\mathcal{S}$ together span $\mathbb{R}^N$ and only intersect at the origin, perfect recovery, i.e., $\tilde{\boldsymbol{x}} = \boldsymbol{x}$, is obtained with $(\mathbf{S}^\mathsf{T}\mathbf{A})^{-1}$. We refer to this condition as the *direct sum* (DS) condition [15].

While this chapter focuses on the case that the signal subspace is known, we can recover signals without the exact knowledge of $\mathcal{A}$ with appropriate priors such as smoothness and stochastic priors [16, 48, 67, 97].

Sampling set selection for full-band graph signals

We introduce nodal domain sampling in the single-channel setting. Since there is no regular sampling in the graph setting, it requires the selection of a subset of nodes, which is hereafter referred to as sampling set selection (SSS). nodal domain sampling operator is defined as follows:

Definition 16 (Nodal domain sampling)**.** *Let $\mathbf{I}_{\mathcal{M}\mathcal{V}} \in \{0,1\}^{K\times N}$ be the submatrix of the identity matrix indexed by $\mathcal{M} \subset \mathcal{V}$ $(|\mathcal{M}| = K)$ and $\mathcal{V}$. The sampling operator is defined by*

$$\mathbf{S}^\mathsf{T} \coloneqq \mathbf{I}_{\mathcal{M}\mathcal{V}}\mathbf{G}, \tag{6.4}$$

where $\mathbf{G} \in \mathbb{R}^{N\times N}$ *is an arbitrary graph filter. A sampled graph signal is thus given by* $\boldsymbol{y} = \mathbf{S}^\mathsf{T}\boldsymbol{x}$.

There exist several approaches of SSS, i.e., the design of $\mathbf{I}_{\mathcal{M}\mathcal{V}}$, however, most methods assume the bandlimited graph signals. In this chapter, we consider sampling of full-band graph signals since the analysis of graph signals beyond the bandlimited assumption is generally necessary in a multi-channel setting. To this aim, we introduce the quality of sampling based on the DS condition: We consider the following problem as a sampling set selection:

$$\mathcal{M}^* = \arg\max_{\mathcal{M}\subset\mathcal{V}} \det(\mathbf{Z}_\mathcal{M}), \tag{6.5}$$

where $\mathbf{Z} = \mathbf{G}\mathbf{A}\mathbf{A}^\mathsf{T}\mathbf{G}^\mathsf{T}$. Since $\det(\mathbf{Z}_\mathcal{M}) = \det(\mathbf{S}^\mathsf{T}\mathbf{A}\mathbf{A}^\mathsf{T}\mathbf{S}) = |\det(\mathbf{S}^\mathsf{T}\mathbf{A})|^2$, (6.5) encourages that the DS condition is satisfied. The cost function in (6.5) is designed based on the D-optimal design, ensuring that the direct sum condition is satisfied. Unlike other SSS methods, (6.5) is applicable to a wide range of signal models beyond the bandlimited assumption. Further details can be found in [106].

The direct maximization of (6.5) is combinatorial and is practically intractable. Therefore, we apply a greedy method to (6.5). Suppose that $\mathrm{rank}(\mathbf{Z}) \geq K$. By applying the Schur determinant formula [88] to (6.5), it results in

$$\begin{aligned} y^* &= \arg\max_{y\in\mathcal{M}^c} \det(\mathbf{Z}_{\mathcal{M}\cup\{y\}}) \\ &= \arg\max_{y\in\mathcal{M}^c} \det(\mathbf{Z}_\mathcal{M}) \cdot (\mathbf{Z}_{y,y} - \mathbf{Z}_{y,\mathcal{M}}(\mathbf{Z}_\mathcal{M})^{-1}\mathbf{Z}_{\mathcal{M},y}) \\ &= \arg\max_{y\in\mathcal{M}^c} \mathbf{Z}_{y,y} - \mathbf{Z}_{y,\mathcal{M}}(\mathbf{Z}_\mathcal{M})^{-1}\mathbf{Z}_{\mathcal{M},y}, \end{aligned} \tag{6.6}$$

where we omit the multiplication with $\det(\mathbf{Z}_\mathcal{M})$ in the third equivalence because it does not depend on y^*.

Still, (6.6) is computationally expensive due to the matrix inversion, which typically requires $O(|\mathcal{M}|^3)$ computational complexity. To alleviate this, we utilize the Neumann series for $(\mathbf{Z}_\mathcal{M})^{-1}$. We omit the detail due to the limitation of the space. Please refer to [106]. As a result, the SSS algorithm is described by Algorithm 3.

In the following, we extend the single-channel sampling to the multi-channel sampling. The MCS parallels most of the formulation of the single-channel one.

6.3 Proposed multi-channel sampling on graphs

In this section, we build a MCS framework by extending the single channel sampling introduced in the previous section. The framework is illustrated in Fig. 6.2. In addition, we

Algorithm 3: Single-channel SSS

Input: $\mathbf{Z}, \mathcal{M} = \emptyset, K, m = 0$
Set α_0^* such that $\|\mathbf{I} - \alpha_0^* \mathbf{Z}_{\mathcal{M}}\|_2 \leq 1$
while $|\mathcal{M}| < K$ **do**
  Compute $\boldsymbol{\varepsilon}_y^0 = \alpha_0^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\delta}_y$
  while $\|\boldsymbol{\varepsilon}_y^0 - \alpha_m^* \mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m\| \geq \beta$ *for some* $\beta > 0$ **do**
    $\alpha_{m+1}^* \leftarrow \frac{(\boldsymbol{\varepsilon}_y^m)^\mathsf{T} \mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m}{\|\mathbf{Z}_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m\|^2}$
    $\boldsymbol{\varepsilon}_y^{m+1} \leftarrow \boldsymbol{\varepsilon}_y^0 + (\mathbf{I} - \alpha_{m+1}^* \mathbf{Z}_{\mathcal{M}}) \boldsymbol{\varepsilon}_y^m$
    $m \leftarrow m + 1$
  $y^* \rightarrow \arg\max_{y \in \mathcal{M}^c} \mathbf{Z}_{yy} - \mathbf{Z}_{y,\mathcal{M}} \boldsymbol{\varepsilon}_y^m$
  $\mathcal{M} \leftarrow \mathcal{M} \cup \{y\}$
Output: $\mathcal{M}$

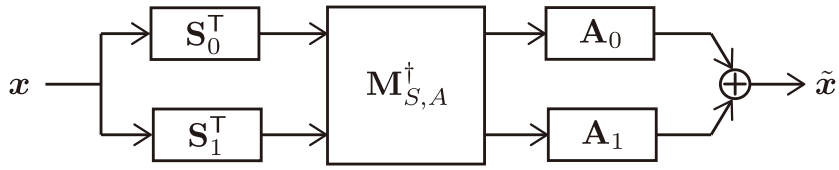


Figure 6.2: The framework of MCS in the case of $J = 2$.

convert the MCS into an equivalent subband-wise expression. This allows for an efficient computation of the recovery transform. Based on the framework, we develop the SSS for our MCS such that full-band graph signals are best recovered.

Framework of multi-channel sampling

Suppose that the number of the sampling set in the ℓth channel is K_ℓ. We now assume that graph signals are generated with J generators, i.e.,

$$\boldsymbol{x} = \sum_{\ell=0}^{J-1} \mathbf{A}_\ell \boldsymbol{d}_\ell, \tag{6.7}$$

where $\mathbf{A}_\ell \in \mathbb{R}^{N \times K_\ell}$ and $\boldsymbol{d}_\ell \in \mathbb{R}^{K_\ell}$ are the ℓth generator and expansion coefficients, respectively.

We often encounter the graph signal model in (6.7) for many applications. For example, in multiscale analysis of graph signals [55], piecewise-smooth graph signals are often considered.

These signals are composed of a combination of piecewise constant and globally smooth signal models. This situation includes sensor network data [27].

The sampling operator $\mathbf{S}^{\mathsf{T}} \in \mathbb{R}^{(K_0+\cdots+K_{J-1})\times N}$ is given by

$$\mathbf{S}^{\mathsf{T}} = \begin{bmatrix} \mathbf{S}_0^{\mathsf{T}} \\ \vdots \\ \mathbf{S}_{J-1}^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{\mathcal{M}_0 \mathcal{V}} \mathbf{G}_0 \\ \vdots \\ \mathbf{I}_{\mathcal{M}_{J-1} \mathcal{V}} \mathbf{G}_{J-1} \end{bmatrix}, \tag{6.8}$$

where $\mathcal{M}_\ell$ is the ℓth sampling set and $\mathbf{G}_\ell \in \mathbb{R}^{N\times N}$ is an arbitrary graph filter for the ℓth subband.

According to (6.3), the recovered graph signals are given by [16, pp. 226–235]

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} \mathbf{A}_0 & \cdots & \mathbf{A}_{J-1} \end{bmatrix} \mathbf{M}_{S,A}^{\dagger} \begin{bmatrix} \mathbf{S}_0^{\mathsf{T}} \\ \vdots \\ \mathbf{S}_{J-1}^{\mathsf{T}} \end{bmatrix} \boldsymbol{x}, \tag{6.9}$$

where

$$\mathbf{M}_{S,A} = \begin{bmatrix} \mathbf{S}_0^{\mathsf{T}}\mathbf{A}_0 & \cdots & \mathbf{S}_0^{\mathsf{T}}\mathbf{A}_{J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{S}_{J-1}^{\mathsf{T}}\mathbf{A}_0 & \cdots & \mathbf{S}_{J-1}^{\mathsf{T}}\mathbf{A}_{J-1} \end{bmatrix}. \tag{6.10}$$

This implies recovery performance essentially depends on the invertibility of $\mathbf{M}_{S,A}$.

Hereafter, we focus on the two-channel sampling, i.e., $J = 2$, for simplicity. To extend the following MCS for $J > 2$, we may cascade the two-channel MCS [116], and perform the proposed sampling scheme recursively.

Subband-wise representation of MCS

Since the MCS requires the matrix inversion of (6.10), it could be computationally consuming, especially for large graphs. To reduce its computational cost, we rewrite (6.10) as a computationally-efficient form.

For simplicity, this chapter focuses on the critically-sampled case, i.e., $\mathcal{M}_1 = \mathcal{M}_0^c$ (please refer to the notation in Sec.6.2), but any sampling ratio can be applied to our MCS including over- and under-sampled cases. Suppose that $\mathbf{S}_0^{\mathsf{T}}\mathbf{A}_0$ and $\mathbf{S}_1^{\mathsf{T}}\mathbf{A}_1$ are invertible. Under this condition, $\mathbf{M}_{S,A}^{\dagger}\mathbf{S}^{\mathsf{T}}$ can be rewritten [88] as

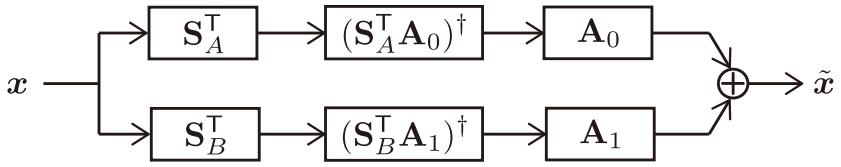


Figure 6.3: The framework of the subband-wise MCS.

$$
\begin{aligned}
&\mathbf{M}_{S,A}^{\dagger}\mathbf{S}^{\mathsf{T}} \\
&= \begin{bmatrix} \mathbf{S}_A^{\mathsf{T}}\mathbf{A}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B^{\mathsf{T}}\mathbf{A}_1 \end{bmatrix}^{\dagger} \\
&\quad \times \begin{bmatrix} \mathbf{I} & -\mathbf{S}_0^{\mathsf{T}}\mathbf{A}_1(\mathbf{S}_1^{\mathsf{T}}\mathbf{A}_1)^{-1} \\ -\mathbf{S}_1^{\mathsf{T}}\mathbf{A}_0(\mathbf{S}_0^{\mathsf{T}}\mathbf{A}_0)^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{S}_0^{\mathsf{T}} \\ \mathbf{S}_1^{\mathsf{T}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{S}_A^{\mathsf{T}}\mathbf{A}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B^{\mathsf{T}}\mathbf{A}_1 \end{bmatrix}^{\dagger} \begin{bmatrix} \mathbf{S}_A^{\mathsf{T}} \\ \mathbf{S}_B^{\mathsf{T}} \end{bmatrix},
\end{aligned} \tag{6.11}
$$

where

$$
\begin{aligned}
\mathbf{S}_A^{\mathsf{T}} &\coloneqq \mathbf{S}_0^{\mathsf{T}} - \mathbf{S}_0^{\mathsf{T}}\mathbf{A}_1(\mathbf{S}_1^{\mathsf{T}}\mathbf{A}_1)^{-1}\mathbf{S}_1^{\mathsf{T}} \\
\mathbf{S}_B^{\mathsf{T}} &\coloneqq \mathbf{S}_1^{\mathsf{T}} - \mathbf{S}_1^{\mathsf{T}}\mathbf{A}_0(\mathbf{S}_0^{\mathsf{T}}\mathbf{A}_0)^{-1}\mathbf{S}_0^{\mathsf{T}}.
\end{aligned} \tag{6.12}
$$

By viewing $\mathbf{S}_A^{\mathsf{T}}$ and $\mathbf{S}_B^{\mathsf{T}}$ as new sampling operators, (6.9) can be expressed by

$$
\tilde{\boldsymbol{x}} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_1 \end{bmatrix} \widetilde{\mathbf{M}}_{S,A}^{\dagger} \begin{bmatrix} \mathbf{S}_A^{\mathsf{T}} \\ \mathbf{S}_B^{\mathsf{T}} \end{bmatrix} \boldsymbol{x}, \tag{6.13}
$$

where

$$
\widetilde{\mathbf{M}}_{S,A} = \begin{bmatrix} \mathbf{S}_A^{\mathsf{T}}\mathbf{A}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B^{\mathsf{T}}\mathbf{A}_1 \end{bmatrix}. \tag{6.14}
$$

In comparison with (6.9), we notice that (6.13) can be viewed as the subband-wise MCS. The modified framework is illustrated in Fig. 6.3. Obviously, the inverse of (6.14) requires the lower computational complexity than that of (6.10).

In contrast to the single channel setting, we need to simultaneously consider the best $\mathcal{M}_0$ and $\mathcal{M}_1$ in the multi-channel case. In the following, we extend the SSS to that for MCS.

SSS for MCS

Based on (6.13), we design the SSS for the graph MCS. Recall that we assume that $\mathbf{S}_0^\mathsf{T}\mathbf{A}_0$ and $\mathbf{S}_1^\mathsf{T}\mathbf{A}_1$ in (6.12) are invertible. To satisfy this condition, we maximize the product of $\det(\mathbf{S}_0^\mathsf{T}\mathbf{A}_0\mathbf{A}_0^\mathsf{T}\mathbf{S}_0)$ and $\det(\mathbf{S}_1^\mathsf{T}\mathbf{A}_1\mathbf{A}_1^\mathsf{T}\mathbf{S}_1)$ (see (6.5)). We consider the following problem:

$$\mathcal{M}^* = \arg\max_{\mathcal{M}\subset\mathcal{V}} \det([\mathbf{Z}_0]_{\mathcal{M}}) \det([\mathbf{Z}_1]_{\mathcal{M}^c}), \tag{6.15}$$

where $\mathbf{Z}_0 = \mathbf{G}_0\mathbf{A}_0\mathbf{A}_0^\mathsf{T}\mathbf{G}_0^\mathsf{T}$ and $\mathbf{Z}_1 = \mathbf{G}_1\mathbf{A}_1\mathbf{A}_1^\mathsf{T}\mathbf{G}_1^\mathsf{T}$. By maximizing $\det([\mathbf{Z}_0]_{\mathcal{M}}) \det([\mathbf{Z}_1]_{\mathcal{M}^c})$, the inverses of $\mathbf{S}_0^\mathsf{T}\mathbf{A}_0$ and $\mathbf{S}_1^\mathsf{T}\mathbf{A}_1$ become stable. The cost function in (6.15) is designed such that the graph signal is maximally separated into two subbands based on the expression in (6.13). While the PR condition is not structurally guaranteed, it is known that PR can be practically realized in many cases [10].

By applying the Schur determinant formula [88] to (6.15), the greedy SSS algorithm of the proposed graph MCS selects a node y^* one-by-one that maximizes the following equation:

$$\begin{aligned} y^* &= \arg\max_{y\in\mathcal{M}^c} \det([\mathbf{Z}_0]_{\mathcal{M}\cup\{y\}}) \det([\mathbf{Z}_1]_{\mathcal{M}\setminus\{y\}}) \\ &= \arg\max_{y\in\mathcal{M}^c} \det([\mathbf{Z}_0]_{\mathcal{M}}) \det([\mathbf{Z}_1]_{\mathcal{M}}) \\ &\quad \cdot \left(\frac{[\mathbf{Z}_0]_{y,y} - [\mathbf{Z}_0]_{y,\mathcal{M}}([\mathbf{Z}_0]_{\mathcal{M},\mathcal{M}})^{-1}[\mathbf{Z}_0]_{\mathcal{M},y}}{[\mathbf{Z}_1]_{y,y} - [\mathbf{Z}_1]_{y,\overline{\mathcal{M}}}([\mathbf{Z}_1]_{\overline{\mathcal{M}},\overline{\mathcal{M}}})^{-1}[\mathbf{Z}_1]_{\overline{\mathcal{M}},y}} \right) \\ &= \arg\max_{y\in\mathcal{M}^c} \frac{[\mathbf{Z}_0]_{y,y} - [\mathbf{Z}_0]_{y,\mathcal{M}}([\mathbf{Z}_0]_{\mathcal{M},\mathcal{M}})^{-1}[\mathbf{Z}_0]_{\mathcal{M},y}}{[\mathbf{Z}_1]_{y,y} - [\mathbf{Z}_1]_{y,\overline{\mathcal{M}}}([\mathbf{Z}_1]_{\overline{\mathcal{M}},\overline{\mathcal{M}}})^{-1}[\mathbf{Z}_1]_{\overline{\mathcal{M}},y}}, \end{aligned} \tag{6.16}$$

where $\overline{\mathcal{M}} = \mathcal{V}\setminus(\mathcal{M}\cup\{y\})$. We omit the multiplication with $\det([\mathbf{Z}_0]_{\mathcal{M}}) \det([\mathbf{Z}_1]_{\mathcal{M}})$ in the third equivalence because it does not depend on y^*.

We apply the Neumann series approximation to the invereses of $[\mathbf{Z}_0]_{\mathcal{M},\mathcal{M}}$ and $[\mathbf{Z}_1]_{\overline{\mathcal{M}},\overline{\mathcal{M}}}$ in (6.16) (see Algorithm 3). As a result, the proposed algorithm is described by Algorithm 4.

In the following, we clarify relationship between the above MCS and existing BFBs.

6.4 Relationship between MCS and BGFB

In this section, we reveal that existing BGFBs are a special case of our MCS in (6.9). We depict the framework of PR BGFBs in Fig. 6.4.

Let $\mathcal{G}_{\mathrm{bpt}} = (\mathcal{V}_L, \mathcal{V}_H, \mathcal{E})$ be a bipartite graph, where $\mathcal{V}_L$ and $\mathcal{V}_H$ are two disjoint sets of nodes such that every edge is connected between a node in $\mathcal{V}_L$ and that in $\mathcal{V}_H$. In other words, no edges exist within $\mathcal{V}_L$ and $\mathcal{V}_H$. We denote the normalized Laplacian matrix for

Algorithm 4: Two-channel SSS

Input: $\mathbf{Z}_0, \mathbf{Z}_1, \mathcal{M} = \emptyset, K, m = 0$
Set α_0^* such that $\|\mathbf{I} - \alpha_0^* \mathbf{Z}_{\mathcal{M}}\|_2 \leq 1$
Set β_0^* such that $\|\mathbf{I} - \beta_0^* \mathbf{Z}_{\overline{\mathcal{M}}}\|_2 \leq 1$
while $|\mathcal{M}| < K$ **do**
  Compute $\boldsymbol{\varepsilon}_y^0 = \alpha_0^* [\mathbf{Z}_0]_{\mathcal{M}} \boldsymbol{\delta}_y$
  Compute $\boldsymbol{\vartheta}_y^0 = \beta_0^* [\mathbf{Z}_1]_{\overline{\mathcal{M}}} \boldsymbol{\delta}_y$
  while $\|\boldsymbol{\varepsilon}_y^0 - \alpha_m^* [\mathbf{Z}_0]_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m\| \geq \gamma \wedge \|\boldsymbol{\vartheta}_y^0 - \beta_m^* [\mathbf{Z}_1]_{\overline{\mathcal{M}}} \boldsymbol{\vartheta}_y^m\| \geq \gamma$ *for some* $\gamma > 0$ **do**
    $\alpha_{m+1}^* \leftarrow \frac{(\boldsymbol{\varepsilon}_y^m)^{\mathsf{T}} [\mathbf{Z}_0]_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m}{\|[\mathbf{Z}_0]_{\mathcal{M}} \boldsymbol{\varepsilon}_y^m\|^2}$
    $\beta_{m+1}^* \leftarrow \frac{(\boldsymbol{\vartheta}_y^m)^{\mathsf{T}} [\mathbf{Z}_1]_{\overline{\mathcal{M}}} \boldsymbol{\vartheta}_y^m}{\|[\mathbf{Z}_1]_{\overline{\mathcal{M}}} \boldsymbol{\vartheta}_y^m\|^2}$
    $\boldsymbol{\varepsilon}_y^{m+1} \leftarrow \boldsymbol{\varepsilon}_y^0 + (\mathbf{I} - \alpha_{m+1}^* [\mathbf{Z}_0]_{\mathcal{M}}) \boldsymbol{\varepsilon}_y^m$
    $\boldsymbol{\vartheta}_y^{m+1} \leftarrow \boldsymbol{\vartheta}_y^0 + (\mathbf{I} - \beta_{m+1}^* [\mathbf{Z}_1]_{\overline{\mathcal{M}}}) \boldsymbol{\vartheta}_y^m$
    $m \leftarrow m + 1$
  $y^* \rightarrow \arg\max_{y \in \mathcal{M}^c} \frac{[\mathbf{Z}_0]_{yy} - [\mathbf{Z}_0]_{y,\mathcal{M}} \boldsymbol{\varepsilon}_y^m}{[\mathbf{Z}_1]_{yy} - [\mathbf{Z}_1]_{y,\overline{\mathcal{M}}} \boldsymbol{\vartheta}_y^m}$
  $\mathcal{M} \leftarrow \mathcal{M} \cup \{y\}$
Output: $\mathcal{M}$

$\mathcal{G}_{\text{bpt}}$ by $\boldsymbol{\mathcal{L}}_{\text{bpt}}$. Its eigendecomposition is given by $\boldsymbol{\mathcal{L}}_{\text{bpt}} = \mathbf{V} \boldsymbol{\Lambda}_{\text{bpt}} \mathbf{V}^{\mathsf{T}}$, where

$$\mathbf{V} \coloneqq \begin{bmatrix} \mathbf{U}_{LL} & \mathbf{U}_{LL} \\ \mathbf{U}_{HL} & -\mathbf{U}_{HL} \end{bmatrix}. \tag{6.17}$$

Let $\mathbf{S}_{\text{ana},\ell}^{\mathsf{T}}$ and $\mathbf{S}_{\text{syn},\ell}$ be the ℓth analysis and synthesis transforms, respectively. Following Theorem 2 in [116], analysis and synthesis transforms in a BGFB can be expressed by

$$\begin{aligned} \mathbf{S}_{\text{ana},\ell}^{\mathsf{T}} &\coloneqq \mathbf{I}_{\mathcal{M}\mathcal{V}} \mathbf{H}_\ell = \mathbf{U}_{LL} \begin{bmatrix} \mathbf{I}_{N/2} & \mathbf{J}_{N/2} \end{bmatrix} \widehat{\mathbf{H}}_\ell \mathbf{V}^{\mathsf{T}} \\ \mathbf{S}_{\text{syn},\ell} &\coloneqq \mathbf{G}_\ell \mathbf{I}_{\mathcal{M}^c\mathcal{V}}^{\mathsf{T}} = \mathbf{V} \widehat{\mathbf{G}}_\ell \begin{bmatrix} \mathbf{I}_{N/2} \\ -\mathbf{J}_{N/2} \end{bmatrix} \mathbf{U}_{LL}^{\mathsf{T}} \end{aligned} \qquad \text{for } \ell = 1, 2, \tag{6.18}$$

where $\widehat{\mathbf{H}}_\ell = H_\ell(\boldsymbol{\Lambda}_{\text{bpt}})$, $\widehat{\mathbf{G}}_\ell = G_\ell(\boldsymbol{\Lambda}_{\text{bpt}})$, and $\mathbf{J}$ is the counter-identity matrix. Note that the PR condition in GFBs can be expressed by [54, 114]

$$\mathbf{S}_{\text{syn},0} \mathbf{S}_{\text{ana},0}^{\mathsf{T}} + \mathbf{S}_{\text{syn},1} \mathbf{S}_{\text{ana},1}^{\mathsf{T}} = \mathbf{I}. \tag{6.19}$$

In graph filter bank designs, $\widehat{\mathbf{H}}_\ell$ and $\widehat{\mathbf{G}}_\ell$ in (6.18) are designed so that (6.19) is satisfied.

From a perspective of generalized sampling [16] (cf. Fig. 6.3), we can view that $\mathbf{S}_{\text{ana},0}^{\mathsf{T}} = \mathbf{S}_A^{\mathsf{T}}$, $\mathbf{S}_{\text{ana},1}^{\mathsf{T}} = \mathbf{S}_B^{\mathsf{T}}$, $\mathbf{S}_{\text{syn},0} = \mathbf{A}_0 (\mathbf{S}_A^{\mathsf{T}} \mathbf{A}_0)^{-1}$, and $\mathbf{S}_{\text{syn},1} = \mathbf{A}_1 (\mathbf{S}_B^{\mathsf{T}} \mathbf{A}_1)^{-1}$. Consequently, the PR

condition of GFBs in (6.19) is rewritten as a MCS as follows:

$$\widetilde{\mathbf{A}}_0\mathbf{S}_A^\mathsf{T} + \widetilde{\mathbf{A}}_1\mathbf{S}_B^\mathsf{T} = \mathbf{I}, \tag{6.20}$$

where $\widetilde{\mathbf{A}}_0 = \mathbf{A}_0(\mathbf{S}_A^\mathsf{T}\mathbf{A}_0)^{-1}$ and $\widetilde{\mathbf{A}}_1 = \mathbf{A}_1(\mathbf{S}_B^\mathsf{T}\mathbf{A}_1)^{-1}$. While the GFB form (6.19) and the MCS form (6.20) are not identical in general, they coincide each other with bipartite graphs. This can be stated in the following theorem.

Theorem 8. *Let $\mathbf{S}_0^\mathsf{T} \in \mathbb{R}^{K\times N}$ and $\mathbf{S}_1^\mathsf{T} \in \mathbb{R}^{(N-K)\times N}$ be two sampling operators, and let $\mathbf{A}_0 \in \mathbb{R}^{N\times K}$ and $\mathbf{A}_1 \in \mathbb{R}^{K\times(N-K)}$ be generation operators such that $\mathbf{S}_0^\mathsf{T}\mathbf{A}_0$ and $\mathbf{S}_1^\mathsf{T}\mathbf{A}_1$ are invertible. If sampling and reconstruction are performed on the bipartite graph defined by* (6.18)*, it follows that*

$$\widetilde{\mathbf{A}}_0\mathbf{S}_A^\mathsf{T} + \widetilde{\mathbf{A}}_1\mathbf{S}_B^\mathsf{T} = \mathbf{S}_{\text{syn},0}\mathbf{S}_{\text{ana},0}^\mathsf{T} + \mathbf{S}_{\text{syn},1}\mathbf{S}_{\text{ana},1}^\mathsf{T} = \mathbf{I}. \tag{6.21}$$

That is, the PR condition for the graph MCS is reduced to that of the conventional BGFBs as long as the graph is bipartite and the normalized Laplacian is used as the graph operator.

Proof. By definition in (6.12), we need to show $\widetilde{\mathbf{A}}_0\mathbf{S}_0^\mathsf{T}\widetilde{\mathbf{A}}_1\mathbf{S}_1^\mathsf{T} + \widetilde{\mathbf{A}}_1\mathbf{S}_1^\mathsf{T}\widetilde{\mathbf{A}}_0\mathbf{S}_0^\mathsf{T} = \mathbf{0}$. This can be verified by

$$\begin{aligned}
&\mathbf{V}^\mathsf{T}(\widetilde{\mathbf{A}}_0\mathbf{S}_0^\mathsf{T}\widetilde{\mathbf{A}}_1\mathbf{S}_1^\mathsf{T} + \widetilde{\mathbf{A}}_1\mathbf{S}_1^\mathsf{T}\widetilde{\mathbf{A}}_0\mathbf{S}_0^\mathsf{T})\mathbf{V} \\
&= 2(\widehat{\mathbf{G}}_0\begin{bmatrix}\mathbf{I} & \mathbf{J}\end{bmatrix}\widehat{\mathbf{H}}_0\widehat{\mathbf{G}}_1\begin{bmatrix}\mathbf{I} & -\mathbf{J}\end{bmatrix}\widehat{\mathbf{H}}_1) \\
&= 2(\widehat{\mathbf{G}}_0\widehat{\mathbf{H}}_0\widehat{\mathbf{G}}_1\widehat{\mathbf{H}}_1 + \widehat{\mathbf{G}}_0\widehat{\mathbf{H}}_0'\widehat{\mathbf{G}}_1\widehat{\mathbf{H}}_1 \\
&\quad - \widehat{\mathbf{G}}_0\widehat{\mathbf{H}}_0\widehat{\mathbf{G}}_1\widehat{\mathbf{H}}_1' - \widehat{\mathbf{G}}_0\widehat{\mathbf{H}}_0\widehat{\mathbf{G}}_1\widehat{\mathbf{H}}_1) \\
&= 2(\widehat{\mathbf{G}}_0\widehat{\mathbf{H}}_0'\widehat{\mathbf{G}}_1\widehat{\mathbf{H}}_1 - \widehat{\mathbf{G}}_0\widehat{\mathbf{H}}_0\widehat{\mathbf{G}}_1\widehat{\mathbf{H}}_1') \\
&= 2\widehat{\mathbf{G}}_0\widehat{\mathbf{G}}_1(\widehat{\mathbf{H}}_0'\widehat{\mathbf{H}}_1 - \widehat{\mathbf{H}}_0\widehat{\mathbf{H}}_1') \\
&= 2\widehat{\mathbf{G}}_0\widehat{\mathbf{G}}_1(\widehat{\mathbf{H}}_0\mathbf{J}\widehat{\mathbf{H}}_1 - \widehat{\mathbf{H}}_0\mathbf{J}\widehat{\mathbf{H}}_1) = \mathbf{0},
\end{aligned} \tag{6.22}$$

where $\widehat{\mathbf{H}}' = \mathbf{J}\widehat{\mathbf{H}} = \widehat{\mathbf{H}}\mathbf{J}$. This completes the proof. □

Note that (6.21) can be expressed as

$$\begin{aligned}
&\begin{bmatrix}\mathbf{A}_0 & \mathbf{A}_1\end{bmatrix}\begin{bmatrix}\mathbf{S}_A^\mathsf{T}\mathbf{A}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B^\mathsf{T}\mathbf{A}_1\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{S}_A^\mathsf{T} \\ \mathbf{S}_B^\mathsf{T}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{A}_0 & \mathbf{A}_1\end{bmatrix}\begin{bmatrix}\mathbf{S}_0^\mathsf{T}\mathbf{A}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_1^\mathsf{T}\mathbf{A}_1\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{S}_0^\mathsf{T} \\ \mathbf{S}_1^\mathsf{T}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{S}_{\text{syn},0} & \mathbf{S}_{\text{syn},1}\end{bmatrix}\begin{bmatrix}\mathbf{S}_{\text{ana},0}^\mathsf{T} \\ \mathbf{S}_{\text{ana},1}^\mathsf{T}\end{bmatrix} \\
&= \mathbf{I}.
\end{aligned} \tag{6.23}$$

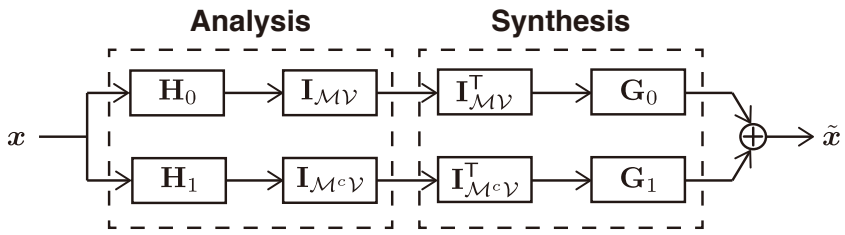
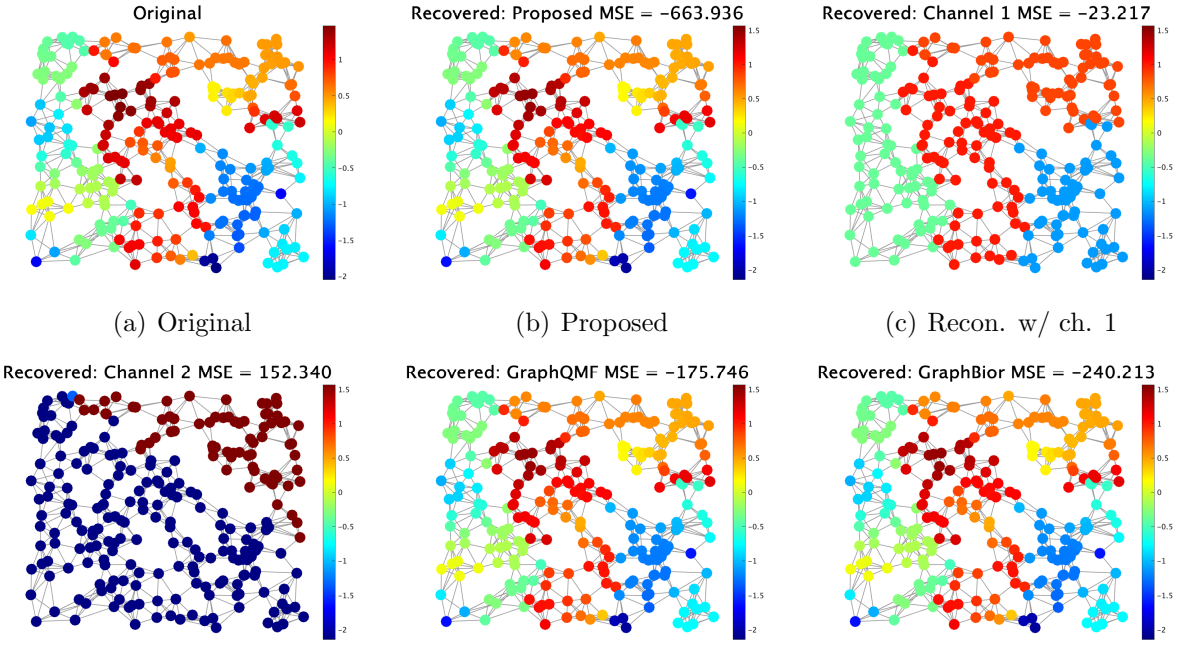


Figure 6.4: Framework of two-channel PR GFBs.

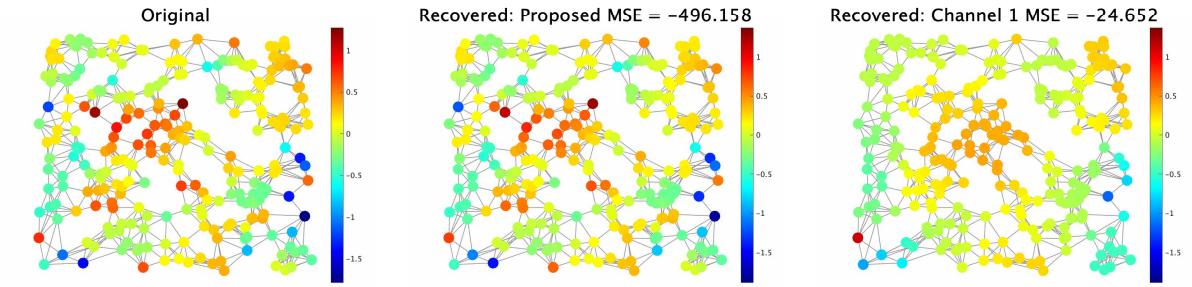


(a) Original (b) Proposed (c) Recon. w/ ch. 1

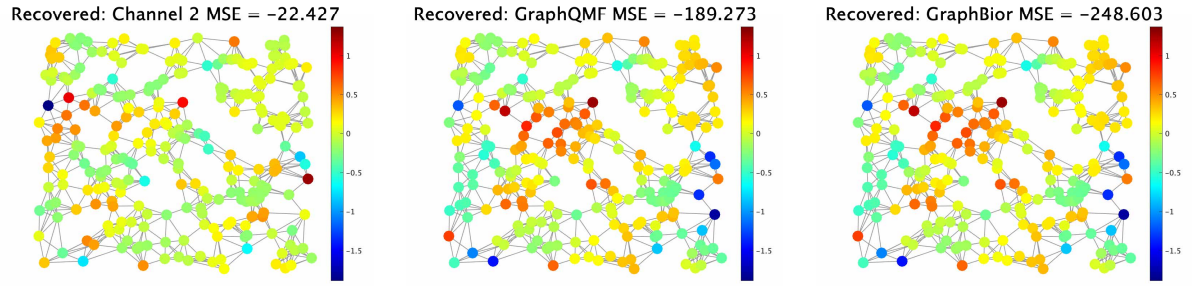
(d) Recon. w/ ch. 2 (e) GraphQMF (f) GraphBior

Figure 6.5: Examples of recovery for PWS graph signals on sensor graphs.

For the second equality, we replace the notations with $\mathbf{S}^{\mathsf{T}}_{\text{ana},0} = \mathbf{S}^{\mathsf{T}}_0$, $\mathbf{S}^{\mathsf{T}}_{\text{ana},1} = \mathbf{S}^{\mathsf{T}}_1$, $\mathbf{S}_{\text{syn},0} = \mathbf{A}_0(\mathbf{S}^{\mathsf{T}}_0\mathbf{A}_0)^{-1}$, and $\mathbf{S}_{\text{ana},1} = \mathbf{A}_1(\mathbf{S}^{\mathsf{T}}_1\mathbf{A}_1)^{-1}$ [66]. According to Theorem 1 and (6.23), we directly obtain $\mathbf{S}^{\mathsf{T}}_A = \mathbf{S}^{\mathsf{T}}_0$ and $\mathbf{S}^{\mathsf{T}}_B = \mathbf{S}^{\mathsf{T}}_1$ if sampling and reconstruction are performed on the bipartite graph. This statement also follows for graph spectral sampling [116].



(a) Original (b) Proposed (c) Recon. w/ ch. 1



(d) Recon. w/ ch. 2 (e) GraphQMF (f) GraphBior

Figure 6.6: Examples of recovery for UBP graph signals on sensor graphs.

Table 6.1: Average MSEs of 30 independent runs. RS is the random sensor graph and SR is the Swiss roll graph.

Methods		Proposed	Recon. w/ ch. 1	Recon. w/ ch. 2	GraphQMF	GraphBior
RS	PWS	**-619.04**	-18.81	194.68	-168.25	-230.60
	UBP	**-465.98**	-18.89	24.04	-186.39	-249.12
SR	PWS	**-654.56**	-19.93	174.18	-172.95	-231.03
	UBP	**-382.10**	-17.00	-9.40	-189.39	-248.31

6.5 Recovery experiments

In this section, we validate performances of the proposed SSS for the graph MCS. We perform graph signal recovery experiments with synthetic and real-world graphs.

Synthetic graphs

Setup

The experiments are performed on random sensor and Swiss roll graphs with $N = 256$. The sampling ratio is set to $K = |\mathcal{M}_0| = N/2$. For both graphs, we generate two synthetic graph signals:

Piecewise smooth (PWS) graph signals [31]: It is composed of piecewise constant components and smooth components:

$$\boldsymbol{x} = [\mathbf{1}_{\mathcal{T}_1} \cdots \mathbf{1}_{\mathcal{T}_P}]\boldsymbol{d}_1 + \mathbf{U}_{\mathcal{B}}\boldsymbol{d}_2, \tag{6.24}$$

where the number of clusters is set to $P = 4$ and the bandwidth is set to $|\mathcal{B}| = K/4$.

Union of band-pass (UBP) graph signals: It is composed of several band-pass components:

$$\boldsymbol{x} = \mathbf{U}\sum_{\ell=1}^{2} G_\ell(\boldsymbol{\Lambda})\boldsymbol{d}_\ell, \tag{6.25}$$

where generation (synthesis) filters $G_\ell(\boldsymbol{\Lambda})$ are implemented by Meyer wavelet kernel [77].

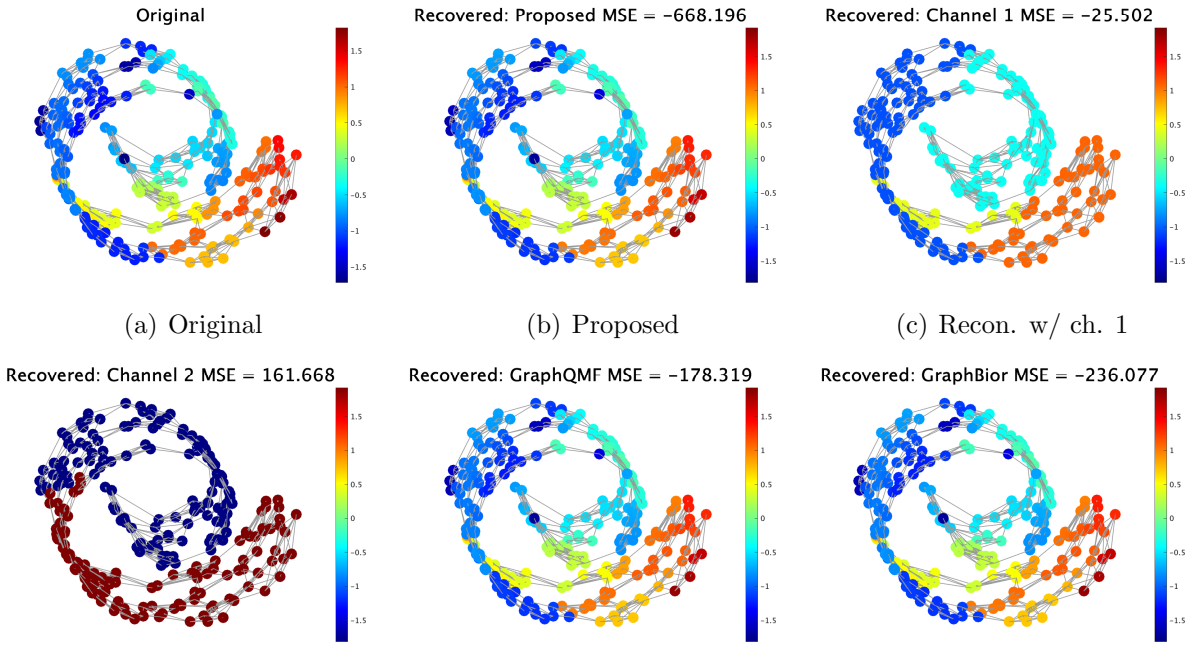
In both cases, analysis filters are given by Mexican hat wavelet kernel [77].

We calculate the average MSE of reconstructed graph signals for 30 independent runs. We compare the result to the well-known BGFBs, graphQMF [54] and graphBior [114]. GraphQMF is an orthogonal graph filter bank, which requires eigendecomposition to achieve exact PR. This is because the polynomial approximation of filters results in reconstruction errors. In contrast, GraphBior is a PR graph filter bank that utilizes polynomial graph filters and can be implemented without requiring eigen-decomposition of the graph operator.

We apply the Harary's decomposition algorithm [123] to those BGFBs for the graph bipartition. We also perform reconstruction with the single channel sampling as a benchmark. For all methods, we implement analysis and synthesis graph filters with the 50th order polynomial approximation [89]. The numerical experiments are conducted in MATLAB environments using 64-bit machine precision ($\pm 2.23 \times 10^{-308}$ to $\pm 1.80 \times 10^{308}$).

Results

The reconstruction MSEs in decibels are summarized in Table 6.1 and examples of the reconstructed graph signals are visualized in Figs. 6.5-6.8. As observed in Table 6.1, the proposed MCS best recovers graph signals for both signal models and both graphs. The single-channel sampling failed to recover. That is, the conventional single-channel sampling does not work



(a) Original (b) Proposed (c) Recon. w/ ch. 1

(d) Recon. w/ ch. 2 (e) GraphQMF (f) GraphBior

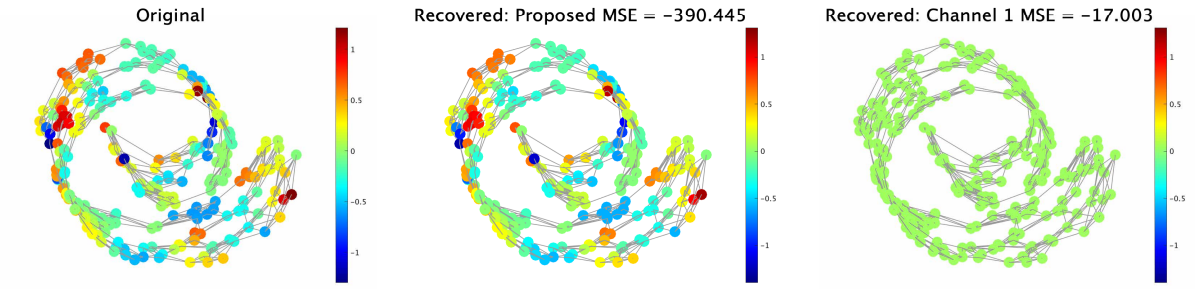
Figure 6.7: Examples of recovery for PWS graph signals on swiss roll graphs.

well for a mixture of multiple graph signal models. GraphQMF presents the low MSE but it involves slight errors caused by polynomial approximations of graph filters. GraphBior and the proposed method can be regarded as PR in machine precision but our method presents the lowest MSE for all methods regardless of the graphs. Although GraphQMF and GraphBior exhibit sufficient performance, with MSE values that can also be considered nearly zero in practice, the significant performance gap in favor of our method is theoretically important. This gap suggests that selecting the optimal sampling set based on our method contributes to improvements in rigorous recovery performance compared to other methods.

Note that our MCS can achieve PR with a user-specified graph operator, while the other methods require to the use of the normalized graph Laplacian or graph simplification prior to sampling.

Real-world graph

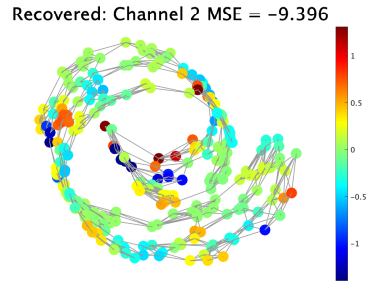
We then perform a sampling experiment for a real-world graph.



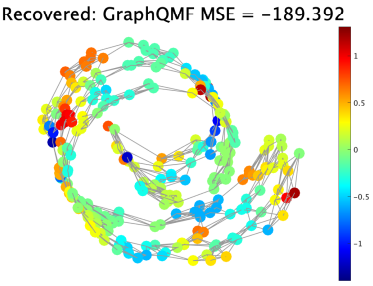
(a) Original

(b) Proposed

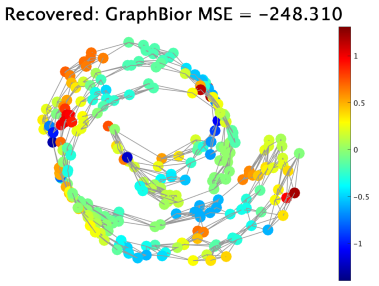
(c) Recon. w/ ch. 1



(d) Recon. w/ ch. 2



(e) GraphQMF



(f) GraphBior

Figure 6.8: Examples of recovery for UBP graph signals on swiss roll graphs.

Setup

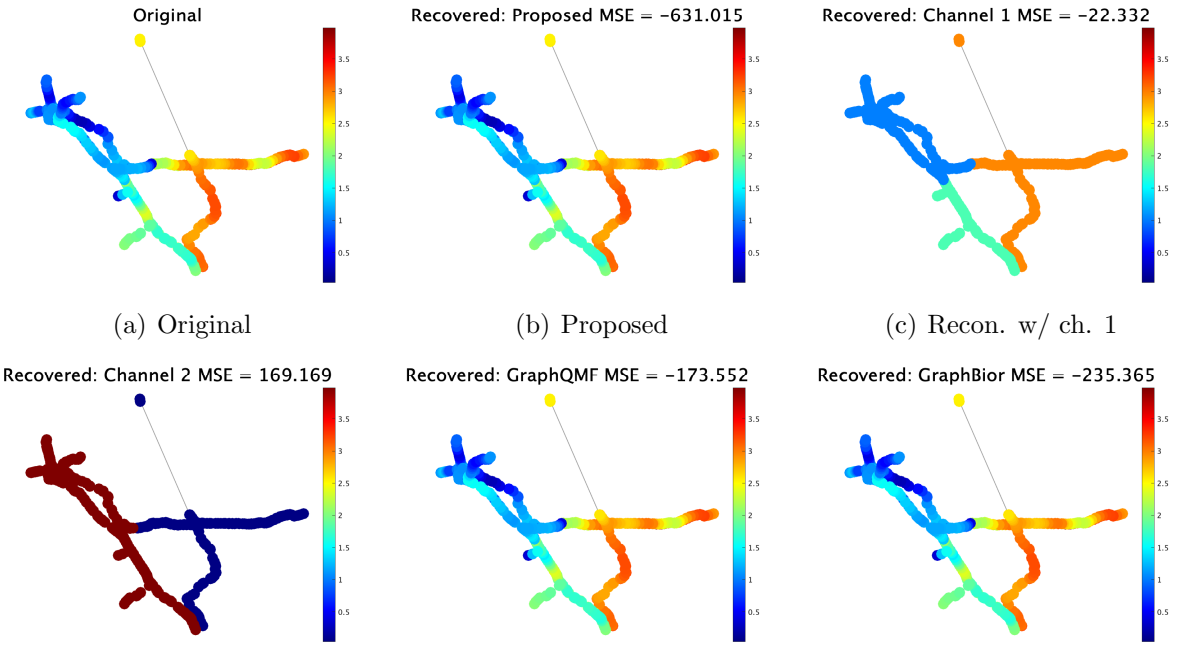
We utilize a traffic network dataset from the Caltrans Performance Measurement System[2]. Nodes represent stations of 17 highways in Alameda county, CA, where $N = 593$. Edges connect the nodes if the stations are adjacent on the same highway or if there is a junction close to the stations on different highways. Hereafter, we refer to this graph as Alameda graph.

We consider a synthetic signal on Alameda graph to objectively measure the reconstruction quality since there is no ground-truth available[3]. In this experiment, we generate PWS graph signals according to (6.24) where we divide Alameda graph into three clusters with spectral clustering [125]. The sampling ratio is set to $K = |\mathcal{M}_0| = 297$. Analysis filters are the same as the previous experiment.

We calculate the average MSE of reconstructed graph signals for 30 independent runs, and compare it with the existing methods from the previous experiment.

[2]This dataset is publicly available http://pems.dot.ca.gov.

[3]Note that estimating generator functions (both for single- and multi-channel cases) is an open problem while some recent studies are undergoing [97, 124].



(a) Original (b) Proposed (c) Recon. w/ ch. 1

(d) Recon. w/ ch. 2 (e) GraphQMF (f) GraphBior

Figure 6.9: Examples of recovery for PWS graph signals on Alameda graph.

Table 6.2: Average MSEs of 30 independent runs for recovery on Alameda graph.

Proposed	Recon. w/ ch. 1	Recon. w/ ch. 2	GraphQMF	GraphBior
-633.29	-20.73	196.12	-172.59	-234.86

Results

The MSEs in decibels are summarized in Table 6.2 and reconstructed graph signals are visualized in Fig. 6.9. Similar to the previous experiment, we observe that the proposed MCS exhibits the lowest MSEs among all the methods. This suggests that our MCS can be applied to real-world datasets, regardless of the topology of graphs.

6.6 Conclusion

In this chapter, we develop the first multi-channle sampling (MCS) framework for graph signals, by extending the single-channel graph signal sampling to the multi-channel setting.

We present an SSS method for our MCS such that graph signals are best recovered. We reveal that existing BGFBs are a special case of the proposed MCS. We demonstrate the effectiveness of the proposed method by showing that our MCS outperforms the existing BGFBs and the single-channel sampling.

In this chapter, we assume that the number of generators of graph signals accords with the number of channels used in the analysis step. In general, they may differ, since the exact number of generators may be unknown. This requires error analysis and a more flexible framework in such challenging scenarios. We leave them as future work.

Chapter 7

Conclusion

This dissertation has systematically addressed the challenges associated with sampling for graph signals. We establish a unified graph signal sampling paradigm from the perspective of the generalized sampling theory framework. Through the three inter-connected projects, we have not only advanced theoretical foundations but also demonstrated practical applications in diverse settings.

Chapter 2 lays the groundwork by developing the generalized graph sampling theory (GGST), which offers a flexible framework for sampling under arbitrary signal priors. This framework enhances the adaptability of graph signal processing to handle various graph structures and signal characteristics, ensuring the best possible recovery and robust performance in real-world scenarios.

Chapter 3 extends classical sampling theory to the graph setting, focusing on developing a GGST for random graph signals characterized by graph wide sense stationarity (GWSS). This framework, which consists of sampling, correction, and reconstruction transforms, is designed to minimize mean-squared error (MSE). The proposed approach is applicable to any linear sampling method and complements existing deterministic GGSTs by introducing stochastic priors. Experiments with synthetic signals validate the effectiveness and robustness.

Chapter 4 proposes an SSS method that operates beyond the traditional bandlimited framework. Leveraging the GGST, the proposed method is designed to accommodate arbitrary linear graph signal models and utilize the Neumann series approximation for computational efficiency. Experiments validates the effectiveness of our approach, highlighting its robustness across different graph signal models.

Chapter 5 expands the scope of SSS by tackling a sensor placement problem on graphs (SPPG) with diverse sensor specifications, crucial for applications like environmental monitoring. Our approach, based on difference-of-convex programming, optimizes sensor placements to maximize the overall coverage and minimize the sensing budget. Results from various network scenarios underscored the versatility and efficiency of our proposed methodology.

Chapter 6 develops a multi-channel sampling (MCS) framework for full-band graph signals, which generally consist of mixtures of multiple signal generation models. Building on GGST, we design an SSS method tailored for MCS, where SSS is performed independently for each channel. Additionally, we reveal the theoretical relationship to alternative multi-channel systems: Existing graph filter banks are categorized as a special case of our MCS framework. Experimental validations demonstrate that our framework can effectively recover full-band graph signals beyond single-channel sampling.

Ubiquitous sensors have significantly transformed our lives, services, and businesses. This has led to a paradigm shift in signal processing, from using sensor data for convenience to enhancing the quality of social and living conditions through extensive sensor data. Despite the exploding popularity of signal processing, the scale of analysis techniques has not sufficiently adapted to many applications. As a result, many researchers employ distributed processing, approximation, and heuristics to alleviate the enormous computational burden without considering the reduction of input data. Sampling can directly address this "obstacle" of modern signal processing with theoretical support. Therefore, our work has the potential to be a solid foundation for signal processing, machine learning, and data science, contributing to a highly sophisticated society.

This dissertation primarily focused on expanding applicable signal models towards sampling paradigms for graph signals, starting from deterministic, random, and their composite models. We began by laying the fundamentals of sampling in general vector spaces. We gradually shifted to addressing several challenges caused by the irregular distribution of nodes: the unambiguity of GWSS (Chapter 3), the non-uniqueness of SSS (Chapter 4), and the rich variety of graph signals (Chapter 6). In the context of SPPG, in Chapter 5, we also considered SSS in a more practical scenario where sensors have heterogeneous specifications. Numerical results demonstrated the theoretical consistency of the proposed methods.

Since GSP has rapidly developed in recent years, its analysis techniques, including sampling, are said to be approaching their maturity. The advancement of GSP has resulted from numerous studies adhering to the conventions of modern signal processing techniques. However, we have demonstrated indications of overcoming these limitations by recasting sampling within general vector spaces. I believe that mathematics is powerful and versatile across disciplines. Demonstrating the impact of non-conventional mathematical techniques is one of the themes of this dissertation.

This dissertation unified existing sampling techniques and developed them into a generalized framework that accommodates several signal models within the context of GSP. While this dissertation laid a substantial theoretical foundation, enhancing the accuracy of GSP applications, we also encountered the following three challenges from different perspectives of sampling.

Accuracy of prior information: Our approaches are valid if appropriate signal priors are available. Two open questions remain: how to obtain exact prior information in practical

applications and how to address cases where signal priors involve errors.

Graph construction: Our approaches are always optimal for a given graph and its associated signal models. However, we do not consider which graph is the best for sampling.

Constraints on practice and applications: This dissertation mainly focused on applications to sensor networks. Nevertheless, there are many applications, each with unique challenges.

In future work, I aim to tackle these challenges to achieve my ultimate goal: developing an interdisciplinary seminal work that people, including non-researchers, youth, and elders, *unconsciously* use in their lives, similar to Shannon's sampling theorem. Therefore, my research interests extend to several disciplines, including the analysis and measurement of sensors, machine learning, fundamental mathematics, and beyond signal processing.

Bibliography

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Comput. Netw.*, vol. 51, no. 4, pp. 921–960, Mar. 2007.

[2] S. Chen, B. Liu, C. Feng, C. Vallespi-Gonzalez, and C. Wellington, "3D point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception," *IEEE Signal Process. Mag.*, vol. 38, no. 1, pp. 68–86, Jan. 2021.

[3] L. M. L. Oliveira and J. J. P. C. Rodrigues, "Wireless sensor networks: A survey on environmental monitoring," *J. Commun.*, vol. 6, no. 2, pp. 143–151, 2011.

[4] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare," *Proc. of the IEEE*, vol. 98, no. 11, pp. 1947–1960, Jan. 2010.

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

[6] E. Ceci and S. Barbarossa, "Graph signal processing in the presence of topology uncertainties," *IEEE Trans. Signal Process.*, vol. 68, pp. 1558–1573, 2020.

[7] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 4, May 2001, pp. 2033–2036.

[8] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges and applications," *Proc. IEEE*, Mar. 2018.

[9] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, May 2019.

[10] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.

[11] M. Tsitsvero, S. Barbarossa, and P. D. Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.

[12] 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, Jul. 2016.

[13] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, "Fast graph sampling set selection using Gershgorin disc alignment," *IEEE Trans. Signal Process.*, vol. 68, pp. 2419–2434, 2020.

[14] 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, Mar. 2018.

[15] Y. Tanaka, Y. C. Eldar, A. Ortega, and G. Cheung, "Sampling signals on graphs: From theory to applications," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 14–30, Nov. 2020.

[16] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems.* Cambridge Univ. Press, Apr. 2015.

[17] Y. Tanaka, "Spectral domain sampling of graph signals," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3752–3767, Jul. 2018.

[18] A. Gadde, A. Anis, and A. Ortega, "Active semi-supervised learning using sampling theory for graph signals," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining.* New York, New York, USA: ACM Press, 2014, pp. 492–501.

[19] C. Dinesh, G. Cheung, and I. Bajic, "Point cloud sampling via graph balancing and Gershgorin disc alignment," *arXiv:2103.06153 [eess]*, Mar. 2021.

[20] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, "Eigendecomposition-free sampling set selection for graph signals," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2679–2692, May 2019.

[21] S. Nomura, J. Hara, H. Higashi, and Y. Tanaka, "Dynamic sensor placement based on graph sampling theory," *arXiv:2211.04019 [eess]*, Nov. 2022.

[22] A. Sadeghi, A. B. Asghar, and S. L. Smith, "Approximation algorithms for distributed multi-robot coverage in non-convex environments," *arXiv:2005.02471 [cs]*, May 2020.

[23] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, Mar. 2004.

[24] Y. C. Eldar and T. G. Dvorkind, "A minimum squared-error framework for generalized sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2155–2167, Jun. 2006.

[25] A. Anis, A. E. Gamal, A. S. Avestimehr, and A. Ortega, "A sampling theory perspective of graph-based semi-supervised learning," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2322–2342, Apr. 2019.

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

[27] S. Chen, Y. Yang, C. Faloutsos, and J. Kovacevic, "Monitoring Manhattan's traffic at 5 intersections?" in *Proc. 2016 IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*. IEEE, Dec. 2016, pp. 1270–1274.

[28] D. U. Leonzio, P. Bestagini, M. Marcon, G. P. Quarta, and S. Tubaro, "Water leak detection and localization using convolutional autoencoders," in *Proc. 2023 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[29] B. Girault, "Stationary graph signals using an isometric graph translation," in *Proc. Europ. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 1516–1520.

[30] Y. Sun and S. Halgamuge, "Minimum-cost heterogeneous node placement in wireless sensor networks," *IEEE Access*, vol. 7, pp. 14 847–14 858, 2019.

[31] S. Chen, A. Singh, and J. Kovačević, "Multiresolution representations for piecewise-smooth signals on graphs," *arXiv:1803.02944 [eess]*, Mar. 2018.

[32] M. Unser, "Sampling—50 Years After Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[33] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, May 2012.

[34] X. Yan, S. Z. Gilani, H. Qin, M. Feng, L. Zhang, and A. Mian, "Deep keyframe detection in human action videos," *arXiv:1804.10021 [cs]*, Apr. 2018.

[35] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *In Proc. IEEE/CVF Conf. Comput. Vis. Patt. Recognit. (CVPR)*. Salt Lake City, UT: IEEE, Jun. 2018, pp. 6479–6488.

[36] A. Stuhlsatz, J. Lippel, and T. Zielke, "Feature extraction with deep neural networks by a generalized discriminant analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 596–608, Apr. 2012.

[37] N. Wiener, "Generalized harmonic analysis," *Acta Mathematica*, vol. 55, pp. 117–258, 1930.

[38] M. Unser and J. Zerubia, "Generalized sampling: Stability and performance analysis," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 2941–2950, Dec. 1997.

[39] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Prentice Hall, 1993.

[40] Y. C. Eldar and T. Werther, *Genral Framework for Consistent Sampling in Hilbert Spaces*, 3rd ed. Int. J. Wavelets, Multiresol. Inf. Process., Sep. 2005, vol. 3.

[41] A. G. Garcıa, "Sampling theory and RKHSs," Ph.D. dissertation, Universidad. Carlos III de Madrid, 2015.

[42] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, Mar. 2016.

[43] A. Jayawant and A. Ortega, "Practical graph signal sampling with log-linear size scaling," *arXiv:2102.10506 [eess]*, Feb. 2021.

[44] E. Clark, T. Askham, S. L. Brunton, and J. Nathan Kutz, "Greedy sensor placement with cost constraints," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2642–2656, Apr. 2019.

[45] D. Sharma, A. Deshpande, and A. Kapoor, "On greedy maximization of entropy," in *Proc. the 32nd Int. Conf. Mach. Learn. (ICML)*, ser. ICML'15. JMLR.org, 2015, pp. 1330–1338.

[46] N. Tremblay, P.-O. Amblard, and S. Barthelmé, "Graph sampling with determinantal processes," *arXiv:1703.01594 [cs, stat]*, Mar. 2017.

[47] N. Perraudin, B. Ricaud, D. I. Shuman, and P. Vandergheynst, "Global and local uncertainty principles for signals on graphs," *APSIPA Trans. Signal Inf. Process.*, vol. 7, no. 1, p. e3, 2018.

[48] Y. Tanaka and Y. C. Eldar, "Generalized sampling on graphs with subspace and smoothness priors," *IEEE Trans. Signal Process.*, vol. 68, pp. 2272–2286, Mar. 2020.

[49] S. Chen, R. Varma, A. Singh, and J. Kovacevic, "Representations of piecewise smooth signals on graphs," in *Proc. 2016 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. IEEE, Mar. 2016, pp. 6370–6374.

[50] A. Gadde and A. Ortega, "A probabilistic interpretation of sampling theory of graph signals," *arXiv:1503.06629 [cs]*, Mar. 2015.

[51] 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, May 2013.

[52] A. Agaskar and Y. M. Lu, "A spectral graph uncertainty principle," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4338–4356, Jul. 2013.

[53] A. Sandryhaila and J. M. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sep. 2014.

[54] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.

[55] D. I. Shuman, M. J. Faraji, and P. Vandergheynst, "A multiscale pyramid transform for graph signals," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2119–2134, Apr. 2016.

[56] 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, Apr. 2016.

[57] L. F. O. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 34–47, Jan. 2018.

[58] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[59] B. S. Dees, L. Stankovic, M. Dakovic, A. G. Constantinides, and D. P. Mandic, "Unitary shift operators on a graph," Sep. 2019.

[60] A. Gavili and X.-P. Zhang, "On the shift operator, graph frequency, and optimal filtering in graph signal processing," *IEEE Trans. Signal Process.*, vol. 65, no. 23, pp. 6303–6318, Feb. 2017.

[61] J. Shi and J. M. F. Moura, "Graph signal processing: Modulation, convolution, and sampling," Dec. 2019.

[62] A. J. Jerri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.

[63] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.

[64] S. Li, Y. Jin, and D. I. Shuman, "Scalable M-channel critically sampled filter banks for graph signals," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3954–3969, Aug. 2019.

[65] R. Varma and J. Kovacevic, "Passive and active sampling for piecewise-smooth graph signals," in *Proc. Int. Conf. Samp. Theory and Appl. (SampTA)*. IEEE, Jul. 2019, pp. 1–5.

[66] O. Christensen and Y. C. Eldar, "Oblique dual frames and shift-invariant spaces," *Appl. Comput. Harmon. Anal.*, vol. 17, no. 1, pp. 48–68, Jul. 2004.

[67] S. P. Chepuri, Y. C. Eldar, and G. Leus, "Graph sampling with and without input priors," in *Proc. 2018 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 4564–4568.

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

[69] Y. C. Eldar and M. Unser, "Nonideal sampling and interpolation from noisy observations in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2636–2651, Jul. 2006.

[70] J. Hara, Y. Tanaka, and Y. C. Eldar, "Generalized graph spectral sampling with stochastic priors," in *Proc. 2020 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2020, pp. 5680–5684.

[71] A. Sakiyama and Y. Tanaka, "Graph reduction method using localization operator and its application to pyramid transform," in *Proc. 2017 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*. IEEE, Dec. 2017, pp. 050–055.

[72] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Stationary graph processes: Parametric power spectral estimation," in *Proc. 2017 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. IEEE, Mar. 2017, pp. 4099–4103.

[73] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes, Fourth Edition*. McGraw-Hill High. Educ., 2002.

[74] X. Jian and W. P. Tay, "Wide-sense stationarity in generalized graph signal processing," *IEEE Trans. Signal Process.*, vol. 70, pp. 3414–3428, 2022.

[75] A. Hasanzadeh, X. Liu, N. Duffield, and K. R. Narayanan, "Piecewise stationary modeling of random processes over graphs with an application to traffic prediction," in *Proc. 2019 IEEE Int. Conf. Big Data (Big Data)*, Feb. 2019, pp. 3779–3788.

[76] E. M. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, ser. Princeton Lectures in Analysis. Princeton University Press, 2005, no. v. 3.

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

[78] A. Heimowitz and Y. C. Eldar, "Smooth graph signal interpolation for big data," *IEEE Trans. Signal Process.*, Jun. 2018.

[79] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *Proc. 2013 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2013, pp. 5445–5449.

[80] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Proces.*, vol. 64, no. 16, pp. 4363–4378, Aug. 2016.

[81] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Proc. 2013 IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2013, pp. 491–494.

[82] N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, "Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century," *J. Geophys. Res. Atmos.*, vol. 108, no. D14, 2003.

[83] B. Girault, "Signal processing on graphs - contributions to an emerging field," Ph.D. dissertation, École normale supérieure de lyon, Lausanne, Swizerland, Dec. 2015.

[84] K. C. Das, "Extremal graph characterization from the bounds of the spectral radius of weighted graphs," *Appl. Math. Comput.*, vol. 217, no. 18, pp. 7420–7426, May 2011.

[85] J. Hara, K. Yamada, S. Ono, and Y. Tanaka, "Design of graph signal sampling matrices for arbitrary signal subspaces," in *2021 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 5275–5279.

[86] F. Zhang, *Matrix Theory: Basic Results and Techniques*. Springer Science & Business Media, Aug. 2011.

[87] H. Gunawan and O. Neswan, "On angles between subspaces of inner product spaces," *J. Indo. Math. Soc.*, vol. 11, pp. 129–135, 2005.

[88] F. Zhang, *The Schur Complement and Its Applications*. Springer Science & Business Media, Mar. 2006.

[89] D. I. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev polynomial approximation for distributed signal processing," in *2011 Int. Conf. Distrib. Comput. Sensor Syst. Workshops (DCOSS)*, Jun. 2011, pp. 1–8.

[90] T. Alhmiedat, "A survey on environmental monitoring systems using wireless sensor networks," *Journal of Networks*, vol. 10, no. 11, pp. 606–615, Jan. 2016.

[91] N. Nasser, A. Ali, L. Karim, and S. Belhaouari, "An efficient wireless sensor network-based water quality monitoring system," in *2013 ACS Int. Conf. Comput. Syst. Appl. (AICCSA)*, May 2013, pp. 1–4.

[92] T. T. T. Zan, P. Gupta, M. Wang, J. Dauwels, and A. Ukil, "Multi-objective optimal sensor placement for low-pressure gas distribution networks," *IEEE Sens. J.*, vol. 18, no. 16, pp. 6660–6668, Aug. 2018.

[93] J. J. P. C. Rodrigues and P. A. C. S. Neves, "A survey on IP-based wireless sensor network solutions," *Int. J. Commun. Syst.*, vol. 23, no. 8, pp. 963–981, 2010.

[94] S. Rathi and R. Gupta, "Sensor placement methods for contamination detection in water distribution networks: A review," *Procedia Eng.*, vol. 89, pp. 181–188, Jan. 2014.

[95] M. M. Aral, J. Guan, and M. L. Maslia, "Optimal design of sensor placement in water distribution networks," *J. Water Res. Plan. Manage.*, vol. 136, no. 1, pp. 5–18, Jan. 2010.

[96] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

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

[98] J. Hara and Y. Tanaka, "Multi-channel sampling on graphs and its relationship to graph filter banks," *IEEE Open J. Signal Process.*, vol. 4, pp. 148–156, 2023.

[99] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, Jan. 2015.

[100] F. Pukelsheim, *Optimal Design of Experiments*. Philadelphia: Soc. Indust. Appl. Math., Apr. 2006.

[101] T. Nagata, T. Nonomura, K. Nakai, K. Yamada, Y. Saito, and S. Ono, "Data-driven sparse sensor selection based on a-optimal design of experiment with ADMM," *IEEE Sensors J.*, vol. 21, no. 13, pp. 15 248–15 257, Jul. 2021.

[102] Y. H. Kim and A. Ortega, "Toward optimal rate allocation to sampling sets for bandlimited graph signals," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1364–1368, Sep. 2019.

[103] J. C. O. Souza, P. R. Oliveira, and A. Soubeyran, "Global convergence of a proximal linearized algorithm for difference of convex functions," *Optimization Letters*, vol. 10, no. 7, pp. 1529–1539, Oct. 2016.

[104] M. Ma, Z. Zhong, Z. Zhai, and R. Sun, "A novel optimal sensor placement method for optimizing the diagnosability of liquid rocket engine," *Aerospace*, vol. 11, no. 3, p. 239, Mar. 2024.

[105] J. Wang, Y. Gao, X. Yin, F. Li, and H.-J. Kim, "An enhanced PEGASIS algorithm with mobile sink support for wireless sensor networks," *Wirel. Commun. Mob. Comput.*, vol. 2018, pp. 1–9, Dec. 2018.

[106] J. Hara and Y. Tanaka, "Sampling set selection for graph signals under arbitrary signal priors," in *ICASSP 2022 - 2022 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2022, pp. 5732–5736.

[107] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. 24, no. 11, pp. 652–654, Nov. 1977.

[108] I. Djbkovic and P. Vaidyanathan, "Generalized sampling theorems in multiresolution subspaces," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 583–599, Mar. 1997.

[109] O. Christensen and Y. C. Eldar, "Generalized shift-invariant systems and frames for subspaces," *J. Fourier Anal. Appl.*, vol. 11, no. 3, pp. 299–313, Jun. 2005.

[110] A. Aldroubi, "Oblique projections in atomic spaces," *Proceedings of the American Mathematical Society*, vol. 124, no. 7, pp. 2051–2060, 1996.

[111] J. Brown, "Multi-channel sampling of low-pass signals," *IEEE Trans. Circuits Syst.*, vol. 28, no. 2, pp. 101–106, Feb. 1981.

[112] Y. Eldar and A. Oppenheim, "Filterbank reconstruction of bandlimited signals from nonuniform and generalized samples," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2864–2875, Oct. 2000.

[113] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.

[114] S. K. Narang and A. Ortega, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4673–4685, Oct. 2013.

[115] Y. Tanaka and A. Sakiyama, "M-channel oversampled graph filter banks," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3578–3590, Jul. 2014.

[116] A. Sakiyama, K. Watanabe, Y. Tanaka, and A. Ortega, "Two-channel critically sampled graph filter banks with spectral domain sampling," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1447–1460, Mar. 2019.

[117] A. Sakiyama and Y. Tanaka, "Oversampled graph Laplacian matrix for graph filter banks," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6425–6437, Dec. 2014.

[118] D. B. H. Tay and J. Zhang, "Techniques for constructing biorthogonal bipartite graph filter banks," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5772–5783, Nov. 2015.

[119] D. B. H. Tay, Y. Tanaka, and A. Sakiyama, "Critically sampled graph filter banks with polynomial filters from regular domain filter banks," *Signal Processing*, vol. 131, pp. 66–72, Feb. 2017.

[120] D. B. H. Tay and A. Ortega, "Bipartite graph filter banks: Polyphase analysis and generalization," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4833–4846, Sep. 2017.

[121] N. Tremblay and P. Borgnat, "Subgraph-based filterbanks for graph signals," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3827–3840, Aug. 2016.

[122] E. Pavez, B. Girault, A. Ortega, and P. A. Chou, "Two channel filter banks on arbitrary graphs with positive semi definite variation operators," *arXiv:2203.02858*, Mar. 2022.

[123] F. Harary, D. Hsu, and Z. Miller, "The biparticity of a graph," *J. Graph Theory*, vol. 1, no. 2, pp. 131–133, 1977.

[124] H. Behjat, U. Richter, D. V. D. Ville, and L. Sörnmo, "Signal-adapted tight frames on graphs," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 6017–6029, Nov. 2016.

[125] U. von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

List of Publications

This dissertation is based on the following publications:

Jounal Paper

[A1] Junya Hara, Yuichi Tanaka, Yonina C. Eldar, Graph signal sampling under stochastic priors, *IEEE Transactions on Signal Processing*, volume 71, pages 1421-1434, April 2023.

[A2] Junya Hara, Yuichi Tanaka, Multi-channel sampling on graphs and its relationship to graph filter banks, *IEEE Open Journal of Signal Processing*, volume 4, pages 148-156, February 2023.

International Conference

[B1] Junya Hara, Yuichi Tanaka, Sampling set selection for graph signals under arbitrary signal priors, *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pages 5732–5736.

[B2] Junya Hara, Shunsuke Ono, Hiroshi Higashi, Yuichi Tanaka, Sensor placement problem on networks for sensors with multiple specifications, *European Signal Processing Conference (EUSIPCO)*, August 2024, [*accepted for the presentation*].

This dissertation is also related to the following publications:

International Conference

[C1] Junya Hara, Yuichi Tanaka, Yonina C. Eldar, Generalized graph spectral sampling with stochastic priors, *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pages 5680-5684.

[C2] Junya Hara, Koki Yamada, Shunsuke Ono, Yuichi Tanaka, Design of graph signal sampling matrices for arbitrary signal subspaces, *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, June 2021, pages 5275-5279.

[C3] Junya Hara, Yuichi Tanaka, Graph signal sampling for multiple generator functions, *in Proceedings of Asia Pacific Signal and Information Processing Association 14th Annual Summit and Conference (APSIPA ASC 2022)*, November 2022.

Domestic Conference

[D1] 原惇也, 田中雄一, Stochastic prior を用いた一般化グラフサンプリング, 第 34 回信号処理シンポジウム, 2019 年 11 月, [口頭発表].

[D2] 原惇也, 山田宏樹，小野峻佑，田中雄一, 任意の信号部分空間に対応したグラフ信号のサンプリング, 信号処理研究会, 2021 年 3 月, [口頭発表].

[D3] 原惇也, 田中雄一, 任意のグラフ信号モデルに対応したサンプリング集合選択, 信号処理研究会, 2022 年 8 月, [口頭発表].

[D4] 原惇也, 田中雄一, 多様なセンサ種類を考慮したネットワーク上のセンサ配置決定法, 信号処理研究会, 2023 年 7 月, [口頭発表].

Acknowledgments

I would like to express my deep gratitude to Profs. Yuichi Tanaka and Hiroshi Higashi, my thesis advisors, for their patient guidance, enthusiastic encouragement, and useful critiques of this research work. I am also immensely thankful to my current lab mates at Osaka University and Tokyo University of Agriculture and Technology, as well as to our alumni, for their insightful comments and suggestions.

My sincere thanks also go to the Graduate School of Engineering at Osaka University for providing the necessary facilities and support needed to complete this project. Additionally, I would like to acknowledge the financial support received from the Japan Society for the Promotion of Science, which was invaluable for the completion of this dissertation.

A special acknowledgment goes to my co-authors, Profs. Shunsuke Ono and Yonina Eldar, without whom this research would not have been possible. Their expertise, insight, and dedication were instrumental in bringing this project to fruition. Collaborating with them has been an immensely enriching experience.

I would like to express my gratitude to Ms. Yamashita and Ms. Mase for their help in accounting and general clerical support.

On a personal note, I wish to thank my family for their understanding and endless love throughout the duration of my studies.

Lastly, I would like to extend my deepest thanks to my friends for their unwavering support, understanding, and companionship throughout my academic journey.