On the Impact of Sample Size in Reconstructing Graph Signals

Baskaran Sripathmanathan, Xiaowen Dong, Michael Bronstein
University of Oxford

***Abstract*—Reconstructing a signal on a graph from observations on a subset of the vertices is a fundamental problem in the field of graph signal processing. It is often assumed that adding additional observations to an observation set will reduce the expected reconstruction error. We show that under the setting of noisy observation and least-squares reconstruction this is not always the case, characterising the behaviour both theoretically and experimentally.**

***Index Terms*—Graph signal processing, sampling, reconstruction, least squares, robustness.**

I. Introduction

Graph signal processing (GSP) has gained popularity owing to its ability to process and analyze signals on graphs, such as political preferences [1], brain fMRIs [2] and urban air pollution [3]. GSP generalises the highly successful tools of classical signal processing from regular domains such as grids to graphs. Similar to the classical case, the computational costs of processing and storing large volumes of graph signals can be prohibitive, and complete data may not be available owing to impractically high observation costs. Graph sampling provides a solution to these problems by efficiently extrapolating the full data across the graph from observations on a set of vertices or summaries of the data [4].

Sampling in the graph setting poses more challenges than classical sampling because of the irregularity of the graph domain. One such challenge is that periodic sampling, widely used in traditional signal processing, is not applicable. Instead, sample selection must adapt to the graph's topology. Optimal sample selection on graphs is in general NP-hard [5], [6]. Many works focus on providing efficient heuristics for selecting good sample sets under different optimality criteria. These studies also provide bounds to help practitioners manage the trade-off between observation cost and reconstruction loss while determining sample size [7]–[12]. One limitation of these bounds is the scope of their settings: some bounds are set in the noiseless setting [13], while most recent sample-set selection literature is set in the noisy observation setting. In the noisy observation setting, sample-size bounds can require optimal Bayesian reconstruction [5], as opposed to being generic over the various reconstruction methods presented in and benchmarked against in the sample set selection literature, e.g., least squares (LS) [10], [11], variants of LS [7], [8] or graph-Laplacian regularised (GLR) reconstruction [12]. Furthermore, driven by these bounds, many papers in the sample set literature only present experiments with sample sizes exceeding the bandwidth.

This paper presents two primary contributions. First, we demonstrate that the commonly held expectation that increasing sample size results in lower MSE (presented, for example, below equation (13) in [5]) does not hold under LS in many of the settings studied in the literature for signals with noise. Second, we show that it is possible to simultaneously reduce observation cost and reconstruction error compared to sampling and reconstruction schemes presented in the literature. We support our findings with theoretical evidence and experiments conducted under LS.

II. Related Work

Graph signal processing extends the fundamental problem of sampling and reconstruction from signals in the Euclidean domain to graph-structured data. It does so by generalising the graph shift operator [14] - the most common choices being the adjacency matrix of the graph, the graph Laplacian, or a normalised variant of those - and using it to define a signal model. While some work uses the adjacency matrix as the shift operator [15], and the theorems in [5] apply to all of these operators, most of the literature uses a normalised variant of the graph Laplacian. See [14] for a more complete consideration of the trade-offs involved in this choice.

The graph sampling literature is further divided by considerations around the signal model, the reconstruction method and the optimality objective, which we describe below.

A. Bandlimited Signals

The most common signal model used in the literature is the bandlimited signal model. For a graph $\mathcal{G}$ with a shift operator $\boldsymbol{L}$ with eigenvalues $\lambda_1 \leq .. \leq \lambda_N$, the space spanned by the first k eigenvectors of $\boldsymbol{L}$ is called a Paley-Wiener space $PW_\omega(\mathcal{G})$ (for any $\omega \in [\lambda_k, \lambda_{k+1})$) and its elements are called k-bandlimited signals. Pesenson [16] introduced the concept of a *uniqueness set* which is a vertex set capable of perfectly reconstructing any signal in $PW_\omega(\mathcal{G})$, and notes that it must include at least k vertices. This provides a unique optimality criterion for sample sets, for which multiple sampling schemes have been devised [17], [18].

B. Non-bandlimited signals

It is rare for observed signals to be perfectly bandlimited. While some work has focused on extending the class of underlying signals to 'approximately bandlimited signals' [19], [20], it is mostly assumed that there is a clean underlying bandlimited signal and our observations are corrupted by additive

noise. The extra error introduced by this noise is handled in two ways: noise-aware sampling criteria (and corresponding sampling schemes), and robust reconstruction algorithms.

While there is a unique optimality criterion in the noiseless case, there are multiple in the additive noise case:

- *MMSE criterion*: Minimise the average mean squared error (MSE), called *A-Optimality* under LS [7]–[9]
- *Confidence Ellipsoid criterion*: Minimise the confidence ellipsoid around the eigenbasis co-efficients, which is called *D-optimality* under LS [10], [11]
- *WMSE criterion*: Minimise the worst-case MSE, which is called *E-optimality* under LS [12], [15]

These criteria under LS reconstruction, and equivalences to other optimality criteria, are further studied in the literature on Optimal Design of Experiments (see [7] for more detail).

C. Reconstruction Methods

Reconstruction methods are also known as interpolation operators [5]. The most common methods of reconstructing noisy signals are LS reconstruction [10], [11], variants of LS reconstruction [7], [8] and GLR reconstruction [12]. The variants of LS reconstruction and GLR reconstruction are more robust than ordinary LS reconstruction. We provide more detail on these schemes in Section III-C.

III. Preliminaries

A. Graph Signals and bandlimitedness

We define a graph $\mathcal{G}$ to consist of a set of N vertices and a set of edges with associated edge weights. We assume the graph is connected and undirected. We consider a *bandlimited* signal $\boldsymbol{x}$ on $\mathcal{G}$, generalising the classical signal processing definition of bandlimitedness. We do so by considering a symmetric, positive semi-definite shift operator $\boldsymbol{L}$ on $\mathcal{G}$; commonly used examples of $\boldsymbol{L}$ include the combinatorial Laplacian and its normalised variants. We take its eigendecomposition

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$$

writing $\boldsymbol{\Lambda} = diag(\lambda_1, \dots, \lambda_N)$ where $0 = \lambda_1 \leq \dots \leq \lambda_N$ are the eigenvalues of $\boldsymbol{L}$, also known as the *graph frequencies* [14]. $\boldsymbol{U}$ forms an orthonormal basis of $\mathbb{R}^N$; let $\boldsymbol{U}_k$ be the k columns of $\boldsymbol{U}$ corresponding to the k lowest graph frequencies. We say that $\boldsymbol{x}$ is *k-bandlimited* if $\boldsymbol{x} \in span(\boldsymbol{U}_k)$.

B. Sampling

Inherent to the definition of bandlimitedness is that $\boldsymbol{x}$ comes from a low-dimensional subspace. This implies that we do not need to observe $\boldsymbol{x}$ on all N vertices. Indeed, there is some subset of vertices such that if we observe *any* k-bandlimited $\boldsymbol{x}$ on that subset we can reconstruct $\boldsymbol{x}$ fully and without error. Such a subset is called a *uniqueness set* [16].

Given a vertex sample set $\mathcal{S}$, let $\boldsymbol{M}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}| \times N}$ be the corresponding sampling matrix where

$$(\boldsymbol{M}_{\mathcal{S}})_{ij} = \begin{cases} 1 & \text{if } \mathcal{S}_i = j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

then $\mathcal{S}$ is a uniqueness set for a bandwidth k if and only if $\text{rank}(\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k) = k$ [17].

In practice, the signal we are given is often not perfectly bandlimited. We model this as observation noise; we observe a corrupted signal $\boldsymbol{y} = \boldsymbol{x} + \sigma \cdot \boldsymbol{\epsilon}$ where

- $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{U}_k\boldsymbol{U}_k^T)$ is a k-bandlimited Gaussian signal,
- $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ is i.i.d. Gaussian noise on each vertex,
- $\sigma > 0$ is some scaling of the noise

so the corrupted signal $\boldsymbol{y}$ has high-frequency components. Let the Signal-to-Noise ratio SNR = $\frac{\text{tr}(\text{cov}(\boldsymbol{x}))}{\text{tr}(\text{cov}(\sigma \cdot \boldsymbol{\epsilon}))}$ be the ratio of the variance of the signal to the variance of the noise. Then as a ratio of variances, SNR is positive[1] and $\sigma^2 = \frac{k}{N \cdot \text{SNR}}$.

There are multiple optimality criteria in the literature for the noisy setting; under LS they have the following forms:

$$\text{A-Optimality: minimise } \text{tr}(((\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)(\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)^T)^{-1}) \tag{2}$$

$$\text{D-Optimality: maximise } \det((\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)(\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)^T) \tag{3}$$

$$\text{E-Optimality: maximise } \lambda_{min}((\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)(\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)^T) \tag{4}$$

where $\lambda_{min}(\boldsymbol{A})$ is the minimum eigenvalue of $\boldsymbol{A}$.

In this paper, we use average MSE under our model as our loss, which corresponds to A-optimality for LS.

C. Reconstruction Methods

There exist two common reconstruction methods in the literature: LS reconstruction (a.k.a. the standard decoder [18]) and GLR reconstruction (as described in [12], [18]). We summarise the differences in Table I. Our analysis of LS also applies to the commonly used iterative method, Projection onto Convex Sets [21], as POCS converges to LS.

TABLE I: Reconstruction Methods

	Objective	Param	Bias	Needs $\boldsymbol{U}_k$
LS	$\min_{\boldsymbol{x} \in span(\boldsymbol{U}_k)} \|\|\boldsymbol{M}\boldsymbol{x} - \boldsymbol{y}\|\|_2$	bandwidth k	no	yes
GLR	$\min_{\boldsymbol{x} \in \mathbb{R}^N} \left(\|\|\boldsymbol{M}\boldsymbol{x} - \boldsymbol{y}\|\|_2 + \mu \boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}\right)$	μ	yes	no

It is well known that, for linear models with noise, LS reconstruction is the minimum-variance unbiased estimator of $\boldsymbol{x}$ [22]. This justifies us focusing our analysis of unbiased linear reconstruction methods on LS, at least theoretically. In practice, computing $\boldsymbol{U}_k$ is slow, so GLR reconstruction is used for large graphs instead [12], [18].

We define a *reconstruction method* to take observations on a vertex sample set $\mathcal{S}$ and reconstruct the signal across all vertices. We say that a reconstruction method is *linear* if it is linear in its observations. For a fixed vertex sample set $\mathcal{S}$ we can represent a linear reconstruction method by a matrix $\boldsymbol{R}_{\mathcal{S}} \in \mathbb{R}^{N \times |\mathcal{S}|}$.

[1] It is common in the literature to express the SNR in decibels, which may be negative, while its ratio form remains positive. We will only use the ratio form, so for example $-20dB$ would be written as $10^{-20/10} = 10^{-2} > 0$.

Remark 1. LS and GLR reconstruction are both linear:

$$\text{LS:} \quad \boldsymbol{R}_{\mathcal{S}} = \boldsymbol{U}_k(\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)^{\dagger} \tag{5}$$

$$\text{GLR:} \quad \boldsymbol{R}_{\mathcal{S}} = (\boldsymbol{M}_{\mathcal{S}}^T\boldsymbol{M}_{\mathcal{S}} + \mu\boldsymbol{L})^{-1}\boldsymbol{M}_{\mathcal{S}}^T \tag{6}$$

where for a matrix $\boldsymbol{A}$, $\boldsymbol{A}^{\dagger}$ is its Moore-Penrose pseudoinverse.

IV. Problem Setting

For our theoretical results and experiments, we assume:

- A clean underlying k-bandlimited signal $\boldsymbol{x}$.
- The bandwidth k is known.
- Observations of the signal are corrupted by flat-spectrum noise, meaning we observe a non-bandlimited signal.
- We focus on LS reconstruction.

Note that when the sample size is below the bandwidth, there are often multiple possible reconstructions. For example, when trying to minimise the LS criterion

$$\min_{z \in span(\boldsymbol{U}_k)} ||\boldsymbol{M}_{\mathcal{S}}\boldsymbol{z} - \boldsymbol{y}||$$

the following is a solution for any $\boldsymbol{\delta} \in \mathbb{R}^N$:

$$\boldsymbol{U}_k\left((\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)^{\dagger}\boldsymbol{z} + (\boldsymbol{I} - (\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)^{\dagger}(\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k)\boldsymbol{\delta}\right).$$

As we are mainly concerned in studying how sample size affects reconstruction error rather than recommending a specific reconstruction algorithm, for simplicity we pick the minimal 2-norm solution with $\boldsymbol{\delta} = 0$. This uniquely defines LS reconstruction even when $|\mathcal{S}| < k$ [23, Sect. 5.5.1].

V. Main Results

Consider reconstructing a signal with LS reconstruction. We observe the corrupted signal $\boldsymbol{y}$ at $\mathcal{S}$ and reconstruct $\boldsymbol{x}$ ($\hat{\boldsymbol{x}} = \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{y}$). We decompose the expected MSE from observing $\boldsymbol{y}$ at $\mathcal{S}$:

$$\begin{aligned}\mathbb{E}[\text{MSE}_{\mathcal{S}}] &= \mathbb{E}\left[||\boldsymbol{x} - \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{y}||_2^2\right] \\ &= \text{tr}(\text{cov}(\boldsymbol{x} - \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{y})) \\ &= \text{tr}(\text{cov}(\boldsymbol{x} - \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}(\boldsymbol{x} + \sigma\cdot\boldsymbol{\epsilon}))) \end{aligned} \tag{7}$$

As the underlying signal $\boldsymbol{x}$ and the noise $\boldsymbol{\epsilon}$ are independent, and as our sampling and reconstruction operators are linear:

$$\begin{aligned}\text{cov}(\boldsymbol{x} - \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}(\boldsymbol{x} + \sigma\cdot\boldsymbol{\epsilon})) = \text{cov}((\boldsymbol{x} - \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{x})) \\ + \text{cov}(\sigma\cdot(\boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{\epsilon})).\end{aligned} \tag{8}$$

Let $\boldsymbol{E} = \boldsymbol{I} - \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}$ and combine (7) and (8):

$$\begin{aligned}\mathbb{E}[\text{MSE}_{\mathcal{S}}] &= \text{tr}(\text{cov}(\boldsymbol{E}\boldsymbol{x})) + \sigma^2\cdot\text{tr}(\text{cov}(\boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{\epsilon})) \\ &= \text{tr}((\boldsymbol{E}\boldsymbol{U}_k)(\boldsymbol{E}\boldsymbol{U}_k)^T) + \sigma^2\cdot\text{tr}(\boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}^T\boldsymbol{R}_{\mathcal{S}}^T) \\ &= ||\boldsymbol{E}\boldsymbol{U}_k||_F^2 + \sigma^2\cdot\text{tr}(\boldsymbol{R}_{\mathcal{S}}\boldsymbol{R}_{\mathcal{S}}^T) \\ &= ||\boldsymbol{U}_k - \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k||_F^2 + \sigma^2\cdot||\boldsymbol{R}_{\mathcal{S}}||_F^2.\end{aligned} \tag{9}$$

We define

$$\xi_1(\mathcal{S}) = ||\boldsymbol{U}_k - \boldsymbol{R}_{\mathcal{S}}\boldsymbol{M}_{\mathcal{S}}\boldsymbol{U}_k||_F^2 \tag{10}$$

$$\xi_2(\mathcal{S}) = ||\boldsymbol{R}_{\mathcal{S}}||_F^2 \tag{11}$$

so that

$$\mathbb{E}[\text{MSE}_{\mathcal{S}}] = \xi_1(\mathcal{S}) + \sigma^2\cdot\xi_2(\mathcal{S}). \tag{12}$$

Remark 2. Setting $\sigma = 0$, we see that $\xi_1(\mathcal{S})$ can be interpreted as the reconstruction error in the absence of observation noise.

We use this decomposition to analyse changing the vertex sample set $\mathcal{S}$. We consider removing a vertex v from $\mathcal{S}$ to make $\mathcal{S}\backslash\{v\}$.

Definition V.1. Removing v *improves* $\mathcal{S}$ if

$$\mathbb{E}[\text{MSE}_{\mathcal{S}}] > \mathbb{E}[\text{MSE}_{\mathcal{S}\backslash\{v\}}].$$

For $i \in \{1, 2\}$, let

$$\Delta_i(\mathcal{S}, v) = \xi_i(\mathcal{S}) - \xi_i(\mathcal{S}\backslash\{v\}). \tag{13}$$

Then by (12), the change in MSE from removing v is

$$\mathbb{E}[\text{MSE}_{\mathcal{S}}] - \mathbb{E}[\text{MSE}_{\mathcal{S}\backslash\{v\}}] = \Delta_1(\mathcal{S}, v) + \sigma^2\cdot\Delta_2(\mathcal{S}, v). \tag{14}$$

If $\Delta_1(\mathcal{S}, v) + \sigma^2\cdot\Delta_2(\mathcal{S}, v) > 0$, removing v improves $\mathcal{S}$.

We note that Δ_1 is the change in MSE when there is no noise ($\sigma = 0$), so can be interpreted as learning more about $\boldsymbol{x}$. It is always non-positive under LS reconstruction (see Appendix A). On the other hand, Δ_2 is a noise-sensitivity term — its effect scales with σ^2 – and in many cases is positive. Under LS reconstruction, one can show that Δ_1 and Δ_2 are always of different signs (see Appendix B).

If the effect of increasing noise sensitivity exceeds the effect of learning more about the underlying signal, then we can decrease average MSE by *removing* a vertex from the observation set. This leads to our main result under LS reconstruction, which is summarised in the following theorem:

Theorem 1. Let

$$\tau(\mathcal{S}, v) = \frac{k}{N}\cdot\Delta_2(\mathcal{S}, v) \tag{15}$$

then removing v improves $\mathcal{S}$ if and only if

$$\text{SNR} < \tau(\mathcal{S}, v). \tag{16}$$

Proof. See Appendix C. □

This result says that if SNR is too low (below a threshold τ that depends on the bandwidth and the chosen samples), then we can remove a sample from our observation set to improve the average reconstruction error.

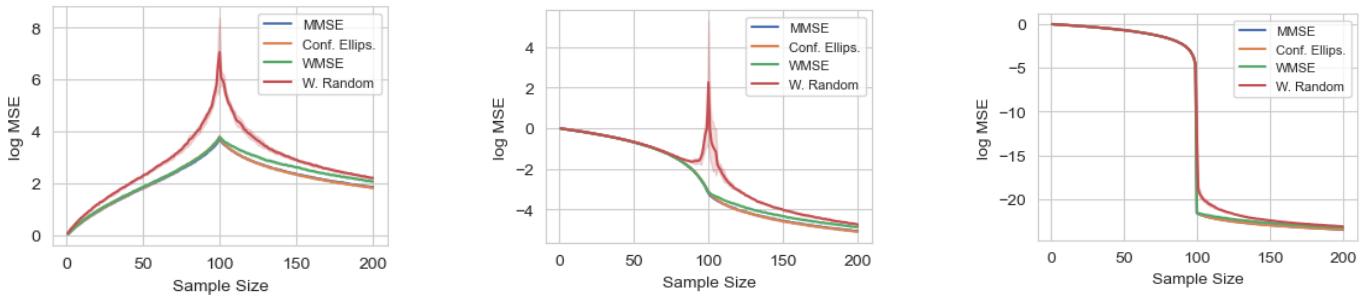
Remark 3. If Δ_2 is non-positive, we have $\tau(\mathcal{S}, v) \leq 0 <$ SNR. In this case (16) cannot hold and removing v will not improve $\mathcal{S}$ for any SNR.

Theorem 1 leaves room for a clever way to pick vertices such that the conditions on SNR in (16) would never be met, hence removing a vertex would never improve the sample set. We show that no such way exists.

Theorem 2. Consider a fixed vertex ordering $v_1, \ldots, v_N$ and let $\mathcal{S}_i$ be the set of the first i vertices. Then there are exactly k indices $1 \leq I_1, \ldots, I_k \leq N$ such that

$$\forall 1 \leq j \leq k : \tau(\mathcal{S}_{I_j}, v_{I_j}) > 0, \tag{17}$$

so removing v_{I_j} improves $\mathcal{S}_{I_j}$ at some SNR.



(a) SNR = 10^{-1} (b) SNR = 10^{2} (c) SNR = 10^{10}

Fig. 1: Average MSE for LS reconstruction on ER Graphs (#vertices=1000, bandwidth = 100) with different SNRs

Proof. See Appendix D. □

Theorem 2 suggests that any sampling scheme, interpreted as a sequential way of picking additional samples, must encounter exactly k instances where the additional vertex v picked on top of the current sample set $\mathcal{S}$ has $\tau(\mathcal{S} \cup v, v) > 0$, meaning at a high enough noise level it increases MSE on average. Schemes in the literature which are optimal in the noiseless case, such as A-, D- and E-optimal sampling schemes, see this happen for the first k vertices they pick.

Theorem 3. Suppose we have a greedy scheme which is optimal in the noiseless case: given the bandwidth k, the first k vertices it samples allow for perfect reconstruction of any clean k-bandlimited signal. Use this scheme to select a vertex sample set $\mathcal{S}_m$ with $|\mathcal{S}_m| = m \leq k$. Then

$$\forall m \leq k : \ \forall v \in \mathcal{S}_m : \ \ \tau(\mathcal{S}_m, v) > 0, \tag{18}$$

that is, for *any* vertex in $\mathcal{S}$, there exists some SNR such that removing that vertex would improve $\mathcal{S}$. Removal of any vertices which the scheme adds after this cannot improve the set:

$$\forall m' > k : \ \ \forall v \in \mathcal{S}_{m'} \backslash \mathcal{S}_k : \ \ \tau(\mathcal{S}_{m'}, v) \leq 0. \tag{19}$$

Proof. See Appendix E. □

Remark 4. Theorem 3 applies to A, D and E-optimal sampling as they are optimal in the noiseless case (see Appendix F).

Remark 5. Equation (19) says that removing one vertex from a sample set of size $m' > k$ chosen by a noiseless-optimal sampling scheme does not reduce error on average. Of course, if one removes *multiple* vertices to bring the sample size below k then the expected sample error may decrease.

VI. Experiments

A. Experimental Setup

We present two experiments to illustrate when removing vertices from the observation set can reduce MSE. For different types of graphs, we present plots of $\mathbb{E}[\text{MSE}_{\mathcal{S}_i}]$ (Fig. 1) and $\tau(\mathcal{S}_i, v_i)$ (Fig. 2) as the sample size i increases under different sampling schemes. Results are presented with 90% confidence intervals.

1) Sample Set Selection: The literature provides several approximations to make vertex sample set selection efficient. For example, approximating the projection matrix $\boldsymbol{U}_k\boldsymbol{U}_k^T$ [7] (subsets of which are used to compute optimality criteria) with a polynomial in $\boldsymbol{L}$, and approximating optimality criteria for easier computation [12].

For our experiments, we generate the vertex sample sets greedily using the exact analytical forms instead of approximations. We use the explicit forms of A/D/E optimality (see Eqns. (2), (3), (4)) and directly compute $\boldsymbol{U}_k\boldsymbol{U}_k^T$ throughout. We compare A/D/E optimal schemes (MMSE/Conf. Ellips./WMSE) to the Weighted Random sampling scheme in [18].

2) Graph Generation: We consider two graph sizes – small (100 vertices) and large (1000 vertices) – for 10 instantiations of each of the following unweighted random graph models:

- Erdős–Rényi (ER) with edge probability $p = 0.8$
- Barabási-Albert (BA) with a preferential attachment to 3 vertices at each step of its construction
- Stochastic Blockmodel (SBM) with intra- and inter-cluster edge probabilities of 0.7 and 0.1 respectively

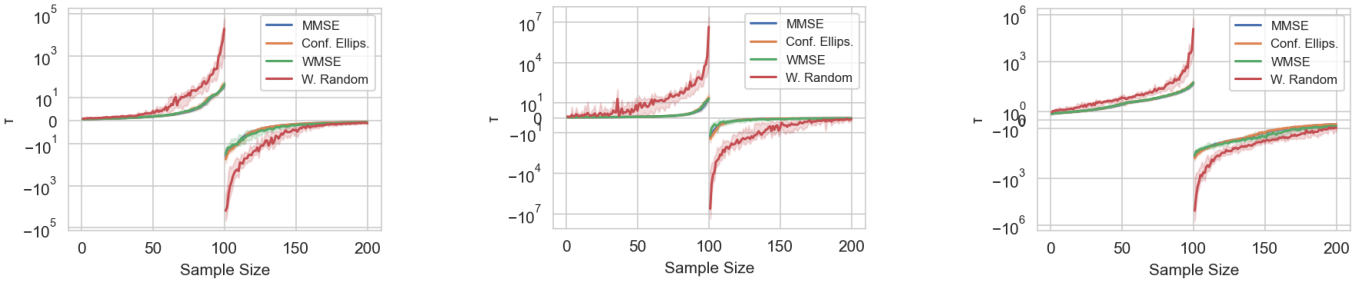
3) Signal Generation: To compute the MSE, we generate 200 signals as follows:

1) Generate $\boldsymbol{x}_{raw} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{U}_k\boldsymbol{U}_k^T)$, $\boldsymbol{\epsilon}_{raw} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$
2) Normalise: $\boldsymbol{x} = \frac{\boldsymbol{x}_{raw}}{||\boldsymbol{x}_{raw}||_2}$ and $\boldsymbol{\epsilon} = \frac{\boldsymbol{\epsilon}_{raw}}{||\boldsymbol{\epsilon}_{raw}||_2}$
3) Return $\boldsymbol{y} = \boldsymbol{x} + \frac{\boldsymbol{\epsilon}}{\sqrt{\text{SNR}}}$

4) Parameters: We set the bandwidth to $\lfloor N/10 \rfloor$, as per [12]. We pick various SNRs to demonstrate that the effect occurs below the threshold τ and disappears above it, which here means $10^{-1}, 10^{2}, 10^{10}$ (In dB: $-10dB$, $20dB$ and $100dB$).

B. Experimental Results

Fig. 1a shows that for low SNRs, optimal sampling schemes (the Green, Orange and Blue lines) lead MSE to increase with each additional sample until the sample size reaches the bandwidth, illustrating Theorem 3. On the other hand, for high SNRs (Fig. 1c), MSE decreases monotonically as sample size increases under all presented sampling schemes, illustrating Theorem 1. Fig 1b shows an intermediate case: for SNRs between these extremes some schemes (Red line) lead to increasing MSE with increasing sample size, while other schemes (Blue, Green, Orange lines) do not.



(a) Erdos-Renyi (b) Barabasi-Albert (c) SBM

Fig. 2: τ for different random graph models under LS reconstruction (#vertices = 1000, bandwidth = 100)

Interestingly, Fig. 1a shows that at a very low SNR of 10^{-1}, the optimal sample size under LS reconstruction is zero. One interpretation of this observation is that, under very high noise, if you throw away all of your samples and assume that your underlying signal is $\mathbf{0}$, you will on average have a lower MSE than if you reconstruct with LS from your observed samples. This follows from (12) and the positivity of ξ_1 and ξ_2 – if your error increases unboundedly with noise, at a sufficiently high noise level your MSE will be above the fixed MSE you would get by approximating your signal with $\mathbf{0}$.

Fig. 2a demonstrates experimentally what the SNR threshold τ looks like in practice. For the ER graphs ($N = 1000$), for signals with $k = 100$ and SNR < 10, under most 'optimal' schemes (Blue, Green, Orange) sampling 90 vertices rather than 100 vertices reduces both observation cost and reconstruction error. Similar results can be seen for Barabasi-Albert graphs (Fig. 2b) and different SNRs. This demonstrates that MSE increasing with sample size happens under conditions which might occur in practice, and is not simply a theoretical curiosity.

We present plots for the larger graph instances here; the smaller graphs (Fig. 3) follow the same pattern as the ER graphs and are presented in Appendix G.

VII. Discussion

In this paper we studied the impact of sample size on LS reconstruction of noisy k-bandlimited graph signals. We showed theoretically and experimentally that reconstruction error is not necessarily monotonic in sample size - that at sufficiently low SNRs, reconstruction error can sometimes be improved by *removing* a vertex from a sample set, even if the sample set was picked by a greedy optimal sampling scheme given a fixed sample size.

Our finding reveals that certain existing results in the literature for noiseless settings may not necessarily generalise to the noisy case. In addition, it further demonstrates the need to consider both optimal sample size selection and reconstruction methods at the same time. For example, the limitation of ordinary LS reconstruction may be mitigated by regularisation schemes such as that proposed in [5].

Future work includes extending the analysis in this paper to cover other reconstruction operators such as GLR reconstruction, providing bounds on ξ_1 and ξ_2 to create noise-aware sample size bounds, experimenting with other graph models such as Ring graphs or studying early-stopping schemes for LS reconstruction.

References

[1] B. Renoust, G. Cheung, and S. Satoh, "Estimating political leanings from mass media via graph-signal restoration with negative edges," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2017, pp. 1009–1014.

[2] S. Itani and D. Thanou, "A graph signal processing framework for the classification of temporal brain data," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1180–1184.

[3] R. K. Jain, J. M. Moura, and C. E. Kontokosta, "Big data+ big cities: Graph signals of urban air pollution [exploratory sp]," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 130–136, 2014.

[4] Y. Tanaka, Y. C. Eldar, A. Ortega, and G. Cheung, "Sampling signals on graphs: From theory to applications," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 14–30, 2020.

[5] L. F. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 34–47, 2017.

[6] A. Nikolov, M. Singh, and U. Tantipongpipat, "Proportional volume sampling and approximation algorithms for a-optimal design," *Mathematics of Operations Research*, vol. 47, no. 2, pp. 847–877, 2022.

[7] F. Wang, Y. Wang, and G. Cheung, "A-optimal sampling and robust reconstruction for graph signals via truncated neumann series," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 680–684, 2018.

[8] F. Wang, G. Cheung, and Y. Wang, "Low-complexity graph sampling with noise and signal reconstruction via neumann series," *IEEE Transactions on Signal Processing*, vol. 67, no. 21, pp. 5511–5526, 2019.

[9] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.

[10] A. Jayawant and A. Ortega, "Practical graph signal sampling with log-linear size scaling," *Signal Processing*, vol. 194, p. 108436, 2022.

[11] N. Tremblay, P.-O. Amblard, and S. Barthelmé, "Graph sampling with determinantal processes," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1674–1678.

[12] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, "Fast graph sampling set selection using gershgorin disc alignment," *IEEE Transactions on signal processing*, vol. 68, pp. 2419–2434, 2020.

[13] H. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 933–936.

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

[15] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE transactions on signal processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[16] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," *Transactions of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.

[17] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

[18] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446–475, 2018.

[19] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Fundamental limits of sampling strategies," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 539–554, 2016.

[20] S. Lin, X. Xie, H. Feng, and B. Hu, "Active sampling for approximately bandlimited graph signals," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5441–5445.

[21] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 491–494.

[22] C.-F. Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae*. Henricus Dieterich, 1823.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. John Hopkins University Press, 2013.

[24] R. Bhatia, *Matrix analysis*. Springer Science & Business Media, 2013, vol. 169.