On the Impact of Sample Size in Reconstructing Noisy Graph Signals: A Theoretical Characterisation

Baskaran Sripathmanathan, Xiaowen Dong, *Member, IEEE*, and Michael Bronstein

***Abstract*—Reconstructing a signal on a graph from noisy observations of a subset of the vertices is a fundamentals problem in the field of graph signal processing. This paper investigates how sample size affects reconstruction error in the presence of noise via an in-depth theoretical analysis of the two most common reconstruction methods in the literature, least-squares reconstruction (LS) and graph-Laplacian regularised reconstruction (GLR). Our theorems show that at sufficiently low signal-to-noise ratios (SNRs), under these reconstruction methods we may simultaneously decrease sample size and decrease average reconstruction error. We further show that at sufficiently low SNRs, for LS reconstruction we have a Λ-shaped error curve and for GLR reconstruction, a sample size of $\mathcal{O}(\sqrt{N})$, where N is the total number of vertices, results in lower reconstruction error than near full observation. We present thresholds on the SNRs, τ and τ_{GLR}, below which the error is non-monotonic, and illustrate these theoretical results with experiments across multiple random graph models, sampling schemes and SNRs. These results demonstrate that any decision in sample-size choice has to be made in light of the noise levels in the data.**

***Index Terms*—Graph signal processing, sampling, reconstruction, least squares, graph-laplacian regularisation.**

I. Introduction

REAL-WORLD signals, such as brain fMRIs [1], urban air pollution [2], and political preferences [3], are often noisy and incomplete, making analysis of the signals harder. The reconstruction of these signals from limited observation is of practical importance, and can benefit from the fact that they can be treated as graph signals, signals defined on a network domain. Graph signal processing (GSP) generalises the highly successful tools of sampling and reconstruction in classical signal processing by extending the classical shift operator to a graph shift operator [4] such as the adjacency matrix [5] or the graph Laplacian, enabling us to extrapolate the full data across the graph from observations on a subset of vertices [6].

In the literature, the vast majority of studies on graph-based sampling focus on designing efficient sampling schemes that are approximately optimal under certain criteria [7, Chapter 6], because optimal vertex choice under noise is in general NP-hard [8], [9]. While these studies provide useful understanding in terms of sampling and reconstruction for graph signals, they focus on the performance of sampling schemes at fixed sample sizes [10], [11], [12], [13], [14], [15], [16], [17], while much less attention has been paid to understanding the impact of varying sample size on the mean squared error (MSE) of reconstruction. Sample size is an important parameter in both understanding and using sampling schemes, especially in the common setting of a fixed sample budget.

The literature studies the impact of sample size both empirically and theoretically (summarised in Table I), and can be further divided by the setting considered: whether the observations are noisy or noiseless, and by which reconstruction method. Empirical results, linked to sampling schemes, are usually obtained in the noisy setting under least squares (LS) reconstruction or its variants [9], [13], [14], [15] or graph-Laplacian regularised (GLR) reconstruction [13], [16]. These results show that MSE decreases as sample size increases in a restricted range of noise level and sample size. An exception is [17, Fig. 1] which empirically shows non-monotonicity of MSE with sample size under LS as it considers the full range of possible sample sizes, including sample sizes less than bandwidth, and a relatively high noise level. Of the three main theoretical results on the impact of sample size in the literature, two focus on slightly different settings: [12] presents sample size bounds for perfect signal reconstruction in the noiseless setting, while [9, Lemma 1] proves that MSE decreases as sample size increases in the noisy setting and provides bounds on the impact of sample size on MSE, but assumes a specific form of regularised LS which is Bayesian optimal. While these theoretical results provide valuable insight, the settings they are based on do not always agree with those in the empirical studies above, hence a generic understanding is still lacking. The third theoretical result [11, Theorem 6] assumes unregularised LS and noisy reconstruction, and proves an MSE upper bound that is tight and decreasing in sample size; however, this result is practically constrained to the case where sample size is greater than or equal to bandwidth [11, Eq. 3].

Received 12 July 2024; revised 29 March 2025 and 24 February 2026; accepted 28 February 2026. Date of publication 18 March 2026; date of current version 12 May 2026. The associate editor coordinating the review of this manuscript and approving it for publication was David I. Shuman. (*Corresponding author: Baskaran Sripathmanathan.*)

Baskaran Sripathmanathan and Xiaowen Dong are with the Department of Engineering Science, University of Oxford, OX2 6ED Oxford, U.K. (e-mail: baskaran.mobile@gmail.com; xdong@robots.ox.ac.uk).

Michael Bronstein is with the Department of Computer Science, University of Oxford, OX1 3QD Oxford, U.K. (e-mail: michael.bronstein@cs.ox.ac.uk).

This article has supplementary downloadable material available at https://doi.org/10.1109/TSP.2026.3674580, provided by the authors.

Digital Object Identifier 10.1109/TSP.2026.3674580

1053-587X © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

TABLE I
STUDIES ON THE IMPACT OF SAMPLE SIZE ON MSE

Reference	Considers Noise?	Reconstruction Method	Range of Possible Sample Sizes
Tremblay, 2017 [13]	✓	LS, GLR	Limited
Wang, 2018 [14]	✓	LS	Limited
Wang, 2019 [15]	✓	LS	Limited
Bai, 2020 [16]	✓	GLR	Limited
Anis, 2016 [17]	✓	LS	Full

(a) Empirical Results

Reference	Considers Noise?	Reconstruction Method	Range of Possible Sample Sizes
Puy, 2018 [11]	✓	LS	Limited
Chamon, 2017 [9]	✓	LS (Regularised)	Full
Shomorony, 2014 [12]	×	LS	Full
This Paper	✓	**LS, GLR**	**Full**

(b) Theoretical Results

In this paper, we fill the gap in the literature by providing a theoretical characterisation of the impact of sample size on the average MSE (or MMSE – see Section II-D) across the full range of possible sample sizes in the most common settings, i.e. noisy observations and LS or GLR reconstruction. More specifically, we focus on whether under sufficiently low SNRs, decreasing sample size may actually decrease MSE. Furthermore, we investigate both the full range of sample sizes and all possible levels of noise, which is important for the application of graph sampling to domains where the noise may be greater than the signal (e.g. finance, with typically low signal-to-noise ratio [18], or environmental sciences, with inaccurate measurements from low-cost sensors [19]) or when observing more samples might not be feasible (e.g. resource-constrained settings). This breadth is only possible through our rigorous theoretical characterisation which allows us to understand behaviour at high noise levels, to characterise behaviour on arbitrarily large graphs, and to show when certain behaviours of MSE happen and why.

Our results begin by using a Bias-Variance decomposition to explain why decreasing sample size may decrease MSE for any linear reconstruction method (Section III-B), and we then specialise to LS and GLR (Section III-C). We prove that under LS reconstruction of a k-bandlimited signal, if the samples were chosen to be optimal in the noiseless case then we can always reduce MSE under high noise by reducing sample size from k to $k-1$ (Section III-D, Theorem 2). We prove that under GLR, if certain graph invariants hold on a graph with N vertices, reducing sample size from almost N vertices to $\mathcal{O}(\sqrt{N})$ vertices will reduce MSE at high noise levels (Section III-F, Proposition 3), and that these invariants hold for large Erdős–Rényi graphs with high probability (Section III-F, Proposition 4). Our experiments (Section IV) validate this for large Stochastic Blockmodel and Barabasi-Albert graphs as well, demonstrating the applicability of our results to graphs with community structure and large scale-free networks. We also investigate how sensitive our results are to different kinds of noise by presenting variants of our theoretical and empirical results under both bandlimited (Sections III-E) & III-G and full-band noise (Sections III-D & III-F). Although our results are mainly of theoretical nature, they may provide useful guidance on choosing the appropriate sample size in light of noise levels in the data.

Our paper presents four primary contributions:

1) A theoretical characterisation of how decreasing sample size may decrease MSE, not only under LS but also under GLR, a regularised method.
2) Analysis of both LS and GLR under bandlimited noise to show non-monotonicity of the MSE is not caused by just the high frequency component of the noise.
3) Asymptotic analysis, showing how the non-monotonicity of the MSE with sample size persists as $N \to \infty$.
4) Extensive experimental simulations illustrating the theoretical results under LS and GLR, and bandlimited noise.

This work significantly extends a previous conference paper [20], which presented preliminary versions of Corollary 1.1, Proposition 2, and a weaker version of Theorem 2, corresponding to the LS part of contribution 1) above. Lemmas 4–7 closely follow those in the conference version [20].

II. BACKGROUND AND PROBLEM FORMULATION

In this section, we first introduce basic definitions of graphs, and the signal and noise models we focus on. We then provide notations for sampling, discuss reconstruction methods, and evaluation criteria. Finally, we present our problem setting.

A. Graphs and Graph Signals

A graph $\mathcal{G}$ consists of a set of N vertices, a set of edges between these vertices, and associated edge weights. We assume $\mathcal{G}$ is connected and undirected, and that the combinatorial graph Laplacian $\boldsymbol{L}$ is real positive semidefinite with N distinct eigenvalues $0 = \lambda_1 < \lambda_2 < \ldots < \lambda_N$ which are also called *graph frequencies* [4][1]. Write the eigendecomposition of $\boldsymbol{L}$ as $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ where $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_N)$ and the columns of $\boldsymbol{U}$ are the eigenbasis of $\boldsymbol{L}$ and form an orthonormal basis of $\mathbb{R}^N$.

The most common signal model used in the graph signal processing literature is the bandlimited signal model, where a k-*bandlimited signal* is a linear combination of the first k columns of $\boldsymbol{U}$ [5]. This is a common smoothness model for graph signals. To obtain this, we define $\boldsymbol{\Pi}_{bl(K)}$ as the projection operator that k-bandlimits a signal (we provide a symbolic definition in Section II-B).

It is rare for observed signals to be perfectly bandlimited. While this can be modelled by assuming the underlying signal to be reconstructed is not bandlimited but rather 'approximately bandlimited signals' [21], [22], or from other more general priors [23], [24], we take the more common approach [11], [13], [14], [15], [16] of assuming the underlying signal is k-bandlimited with additive observation noise. We assume we observe a corrupted signal $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ where

[1]Although we focus on the combinatorial graph Laplacian, our results on LS also hold for the normalised graph Laplacian or any graph shift operator that is positive semidefinite.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

- $\boldsymbol{x}$ is a random k-bandlimited signal with $\mathbb{E}[\boldsymbol{x}] = 0$ and $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Pi}_{bl(\mathcal{K})}$,
- $\boldsymbol{n} = \sigma \cdot \boldsymbol{\epsilon}$ is noise where $\sigma > 0$, $\mathbb{E}[\boldsymbol{\epsilon}] = 0$ and either
  1) $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{I}_N$ ('full-band noise')
  2) $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Pi}_{bl(\mathcal{K})}$ ('k-bandlimited noise')

Our results apply to any noise with these properties (e.g. Gaussian or Rademacher noise). We refer to the $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{I}_N$ case as 'full-band noise' as the associated corrupted signal $\boldsymbol{y}$ has high frequency components, whereas the other case does not. In the literature, noise levels are often described using the SNR $= \frac{\mathbb{E}[||\boldsymbol{x}||_2^2]}{\mathbb{E}[||\boldsymbol{n}||_2^2]}$, a ratio of norms which must be positive[2]. We note that under full-band noise we have $\sigma^2 = \frac{k}{N \cdot \text{SNR}}$ and under k-bandlimited noise we have $\sigma^2 = \frac{1}{\text{SNR}}$.

B. Notation for Sampling

We use the same submatrix notation as [25]. For any matrix $\boldsymbol{X}$ and sets $\mathcal{A}, \mathcal{B}$, we write $[\boldsymbol{X}]_{\mathcal{A},\mathcal{B}}$ to be the submatrix of $\boldsymbol{X}$ with row indices in $\mathcal{A}$ and column indices in $\mathcal{B}$. We define the subvector $[\boldsymbol{x}]_{\mathcal{A}}$ of a vector $\boldsymbol{x}$ similarly. We define a specific shorthand for taking a principal submatrix:

$$[\boldsymbol{X}]_{\mathcal{A}} = [\boldsymbol{X}]_{\mathcal{A},\mathcal{A}}. \tag{1}$$

We let $\mathcal{N} = \{1, \dots, N\}$ and $\mathcal{K} = \{1, \dots, k\}$. We also define two pieces of notation for projections. Let

$$\boldsymbol{\Pi}_{\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{N},\mathcal{B}}[\boldsymbol{I}]_{\mathcal{B},\mathcal{N}}, \quad \boldsymbol{\Pi}_{bl(\mathcal{K})} = [\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}[\boldsymbol{U}]^T_{\mathcal{N},\mathcal{K}}. \tag{2}$$

Finally, $\mathcal{A} \backslash \mathcal{B} = \{i \in A \mid i \notin B\}$ and $\mathcal{A}^c = \mathcal{N} \backslash \mathcal{A}$. In general, we adhere to standard set notation.

We will mainly use $\mathcal{N}$ to index the vertices of $\mathcal{G}$ and use $\mathcal{K}$ to index the first k columns of $\boldsymbol{U}$. $\boldsymbol{\Pi}_{bl(\mathcal{K})}$ can be understood as an ideal low-pass filter.

C. Reconstruction Methods

We define a *reconstruction method* (or 'interpolation operator' [9]) to take potentially noisy observations on a vertex sample set $\mathcal{S}$ and reconstruct the signal across all vertices. In this paper we focus on LS and GLR, which have objectives:

$$\text{LS:} \quad \hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \text{span}([\boldsymbol{U}]_{\mathcal{N},\mathcal{K}})}{\arg\min} ||[\boldsymbol{x}]_{\mathcal{S}} - \boldsymbol{y}||_2 \tag{3}$$

$$\text{GLR:} \quad \hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathbb{R}^N}{\arg\min} ||[\boldsymbol{x}]_{\mathcal{S}} - \boldsymbol{y}||_2^2 + \mu \boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} \tag{4}$$

We summarise their differences in Table II, labelling the input parameters into the reconstruction, whether they are biased and whether they require computation of the eigenbasis $[\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}$.

Our analysis of LS also applies to the commonly used iterative reconstruction method, Projection onto Convex Sets [26], as POCS converges to LS.

We call a reconstruction method *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 $\boldsymbol{R}_{\mathcal{S}} \in \mathbb{R}^{N \times |\mathcal{S}|}$.

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

TABLE II
THE LS AND GLR RECONSTRUCTION METHODS

Method	Parameters	Biased	Requires Computing $[\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}$
LS	k	×	✓
GLR	μ	✓	×

Remark 1: LS and GLR are both linear:

$$\text{LS:} \quad \boldsymbol{R}_{\mathcal{S}} = [\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}[\boldsymbol{U}]^{\dagger}_{\mathcal{S}} \tag{5}$$

$$\text{GLR:} \quad \boldsymbol{R}_{\mathcal{S}} = [(\boldsymbol{\Pi}_{\mathcal{S}} + \mu \boldsymbol{L})^{-1}]_{\mathcal{N},\mathcal{S}} \tag{6}$$

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

Across all linear models under the noisy setting, LS leads to the minimum-variance unbiased estimator of $\boldsymbol{x}$ [27] for k-bandlimited signals and full-band noise, which theoretically justifies us focusing our analysis on LS. On the other hand, GLR implicitly assumes a multivariate Gaussian signal with covariance $\boldsymbol{L}^{\dagger}$ [28], which is slightly mismatched with the k-bandlimited signal model we focus on in this paper. Nevertheless, in practice GLR is still often used for large graphs as computing $[\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}$ for LS is slow [11], [16]. It is also a typical regression model that promotes signal smoothness and has been widely adopted in the literature [29]. Because of these considerations we believe GLR is still a meaningful case to study.

Finally, we clarify what we mean when we consider LS with sample size less than bandwidth. In this case if LS is defined as the minimisation of the objective (3) there are multiple solutions. Thus, we follow [17] and define the LS reconstruction as the unique minimum-norm solution [30, Sect. 5.5.1], hence (5) applies regardless of sample size.

Remark 2: We note that if $\boldsymbol{x}, \boldsymbol{\epsilon}$ and σ are Gaussian, we can compute the best predictor of $\boldsymbol{x}$ in MSE terms. This is the expectation of $\boldsymbol{x}$ conditional on the observation, and is also known as the 'Optimal Bayesian Reconstruction'. It has been proven in this case that decreasing sample size must increase MSE [9]. We do not attempt to characterise this or other noise-adaptive methods; we assume that $\boldsymbol{R}_{\mathcal{S}}$ is not dependent on σ.

D. Optimality Criteria for Sampling

To meaningfully contrast choices of vertex sample set size and selection, we need to evaluate reconstruction performance, and we do so by certain optimality criteria. In the noiseless case, the main optimality criterion for a vertex sample set $\mathcal{S}$ is whether it is a *uniqueness set* [31], that is, if we can perfectly reconstruct any k-bandlimited signal observed on $\mathcal{S}$. Such a set always exists and $\mathcal{S}$ is a uniqueness set for a bandwidth k if and only if $\text{rank}([\boldsymbol{U}]_{\mathcal{S},\mathcal{K}}) = k$ [17].

In the case of additive observation noise, there are multiple common optimality criteria [7, Chapter 6]:

- *MMSE criterion*: Minimise average MSE. [14], [15], [32]
- *Confidence Ellipsoid criterion*: Minimise the confidence ellipsoid around eigenbasis co-efficients. [13], [32], [33]
- *WMSE criterion*: Minimise worst-case MSE. [5], [16]

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

Under LS, these criteria have the following names and forms:

$$(MMSE)\ A\text{-Optimality: minimise } \mathrm{tr}(\boldsymbol{P}^{-1}) \tag{7}$$

$$(Conf.Ellips.)\ D\text{-Optimality: maximise } \det(\boldsymbol{P}) \tag{8}$$

$$(WMSE)\ E\text{-Optimality: maximise } \lambda_{min}(\boldsymbol{P}) \tag{9}$$

$$\text{where} \quad \boldsymbol{P} = \begin{cases} [\boldsymbol{\Pi}_{bl(\mathcal{K})}]S & \text{if } |\mathcal{S}| < k \\ [\boldsymbol{U}]^T_{\mathcal{S},\mathcal{K}}[\boldsymbol{U}]_{\mathcal{S},\mathcal{K}} & \text{if } |\mathcal{S}| \geq k \end{cases} \tag{10}$$

and we define $\mathrm{tr}\left(\boldsymbol{P}^{-1}\right) = +\infty$ in (7) if $\boldsymbol{P}$ is not invertible.

E. Problem Setting

In this paper, we are interested in a theoretical characterisation of the impact of sample size on MSE under all possible SNRs. For our theoretical results and experiments, we assume:

- A known graph $\mathcal{G}$ which is connected and undirected.
- A known bandwidth k.
- A clean underlying k-bandlimited signal $\boldsymbol{x}$ drawn from a known distribution.
- Observations of $\boldsymbol{x}$ are corrupted by noise which is either:
  - full-band, so we observe a non-bandlimited signal.
  - k-bandlimited, so we observe a k-bandlimited signal.
- Linear reconstruction, in particular LS and GLR.

In this paper, we study the behaviour of the MMSE criterion, that is, the MSE averaged over a known signal model and known noise model, which we write as

$$\mathrm{MSE}_{\mathcal{S}} = \mathbb{E}[\boldsymbol{x}, [\boldsymbol{\epsilon}]||\hat{\boldsymbol{x}} - \boldsymbol{x}||_2^2 \mid \mathcal{S} \text{ observed}]. \tag{11}$$

III. Main Theoretical Results

A. Overview and Proof Approaches

In this section, we prove theorems showing how the relationship between sample size and MSE changes with different levels of observation noise, with a focus on showing when reducing sample size reduces MSE. We first present a high level sketch of our approach. To study the effect of noise, we perform a Bias-Variance decomposition on the MSE:

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

where the bias term $\xi_1(\mathcal{S}) = \mathrm{MSE}_{\mathcal{S}}$ when $\sigma^2 = 0$ is the MSE attributable to reconstruction of the clean signal, and $\xi_2(\mathcal{S})$ can be understood as a noise-sensitivity term (see Section III-B for derivations). With this decomposition, the relationship between sample size and MSE under different levels of noise reduces to how the bias and noise-sensitivity vary with respect to sample size.

The focus of the paper is not to characterise all the cases where decreasing sample size decreases MSE, but rather to clearly show that it does happen in a wide variety of cases. In service of this, we focus on certain broad cases that are more tractable, which we call 'simplifications'. For example, we only compare a sample set $\mathcal{S}$ to a subset of it, i.e., $\mathcal{T} \subset \mathcal{S}$.

Our general approach per reconstruction method is:

- Choose a simplification ('simplify');
- Under this simplification, characterise all conditions when decreasing sample size can decrease MSE ('characterise');
- Show that these conditions actually happen ('existence');
- Study these conditions as $N \to \infty$, to prove the conditions may happen on large graphs ('asymptotics').

For LS, we simplify the problem by only considering decreasing the sample size by one at a time, which we call the 'single vertex' simplification. We pay particular attention to subsets sampled by sampling schemes that are optimal in the noiseless setting (Subsection III-D). For GLR, we compare observing the full graph to observing a subset of the vertices. We call this the 'full observation' simplification. We focus on graphs which satisfy certain graph invariants (Subsection III-F). We justify these simplifications in the relevant subsections below.

We then consider reconstruction under bandlimited noise (Subsections III-E and III-G) to show that the reduction in MSE from reducing sample size is not sensitive to our noise model, nor due to the high frequency component of the noise.

An overview of the four steps in our general approach together with the main results are summarised in Table III.

TABLE III
Structure of Theoretical Results

	General	LS		GLR	
Noise	Any	Full-band	Bandlimited	Full-band	Bandlimited
Simplification	$\mathcal{S} \supset \mathcal{T}$	$\mathcal{S}$ vs $\mathcal{S}\backslash\{v\}$		$\mathcal{N}$ vs $\mathcal{S}$	
Characterisation	Thm 1	Corr 1.1	Corr 3.2	Corr 2.1	
Existence		Thm 2	Corr 3.3	Thm 3	Thm 4
Asymptotics		Remark 8	Remark 10	Propn 4	Propn 5

B. General Results

We start with general results. All results in this Section III-B apply to a very general signal model; we only assume that $\boldsymbol{x}$ and $\boldsymbol{\epsilon}$ are drawn from independent distributions, that $\mathbb{E}[\boldsymbol{\epsilon}] = 0$, $\mathbb{E}[||\boldsymbol{x}||_2^2] < \infty$ and $\mathbb{E}[||\boldsymbol{\epsilon}||_2^2] < \infty$. We then specialise to the signal and noise models introduced in Section II-A, based on which we present the analysis in Section III-C and results in Sections III-D, III-F, III-E and III-G.

To understand the effect of changing the sample size on MSE at different levels of noise, we use a variant of the Bias-Variance decomposition [34] on the MSE to separate out the effect of noise. Let $\hat{\boldsymbol{x}}$ be a reconstruction of the signal $\boldsymbol{x}$, then

$$\mathrm{MSE}_{\mathcal{S}} = \mathbb{E}_{\boldsymbol{x}}\left[\underbrace{||\boldsymbol{x} - \mathbb{E}_{\boldsymbol{\epsilon}}[\hat{\boldsymbol{x}}]||_2^2}_{\mathrm{Bias}(\hat{\boldsymbol{x}},\boldsymbol{x})^2} + \underbrace{\mathbb{E}_{\boldsymbol{\epsilon}}\left[||\mathbb{E}_{\boldsymbol{\epsilon}}[\hat{\boldsymbol{x}}] - \hat{\boldsymbol{x}}||_2^2\right]}_{\mathrm{Var}(\hat{\boldsymbol{x}})}\right] \tag{13}$$

This decomposition applies to *any* reconstruction $\hat{\boldsymbol{x}}$ of $\boldsymbol{x}$.

Proof: See Appendix D or [35, Chapter 5.4.4]. □

We consider reconstructing a signal with a linear reconstruction method. We first consider a generic $\boldsymbol{R}_{\mathcal{S}}$. For linear reconstruction, we have $\hat{\boldsymbol{x}} = \boldsymbol{R}_{\mathcal{S}}[\boldsymbol{x} + \sigma \cdot \boldsymbol{\epsilon}]_{\mathcal{S}}$. By assumption, $\mathbb{E}[\boldsymbol{\epsilon}] = 0$, so $\mathbb{E}\boldsymbol{\epsilon}[\hat{\boldsymbol{x}}] = \boldsymbol{R}_{\mathcal{S}}[\boldsymbol{x}]_{\mathcal{S}}$. Therefore

$$\mathrm{Bias}(\hat{\boldsymbol{x}}, \boldsymbol{x})^2 = \mathbb{E}_{\boldsymbol{x}}\left[\left|\left|\left(\boldsymbol{I} - \boldsymbol{R}_{\mathcal{S}}[\boldsymbol{I}]_{\mathcal{S},\mathcal{N}}\right)\boldsymbol{x}\right|\right|_2^2\right] = \xi_1(\mathcal{S}) \tag{14}$$

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

$$\mathrm{Var}(\hat{\boldsymbol{x}}) = \sigma^2 \cdot \mathbb{E}_{\boldsymbol{\epsilon}}\left[||\boldsymbol{R}_{\mathcal{S}}[\boldsymbol{\epsilon}]_{\mathcal{S}}||_2^2\right] = \sigma^2 \cdot \xi_2(\mathcal{S}) \tag{15}$$

where $\xi_1(\mathcal{S})$ and $\xi_2(\mathcal{S})$ are averaged over $\boldsymbol{x}$ and $\boldsymbol{\epsilon}$, so are not random. Therefore

$$\mathrm{MSE}_{\mathcal{S}} = \underbrace{\xi_1(\mathcal{S})}_{\mathbb{E}[\mathrm{Bias}(\hat{\boldsymbol{x}},\boldsymbol{x})^2]} + \underbrace{\sigma^2 \cdot \xi_2(\mathcal{S})}_{\mathbb{E}[\mathrm{Var}(\hat{\boldsymbol{x}})]}. \tag{16}$$

We will refer to $\xi_1(\mathcal{S})$ as the 'bias' of $\boldsymbol{R}_{\mathcal{S}}$ and $\xi_2(\mathcal{S})$ as the 'noise-sensitivity' of $\boldsymbol{R}_{\mathcal{S}}$.

In Statistical Learning Theory, the standard Bias-Variance decomposition is used to show how increasing the complexity of a model often increases its ability to fit the data (reducing 'bias') while increasing its noise-sensitivity (increasing 'variance'), and that the optimum model complexity minimising MSE balances these two components. In the rest of the paper, we will use our Bias-Variance decomposition to show that while decreasing the sample size for a reconstruction method might increase bias it can also decrease noise-sensitivity hence reducing the variance, and that the optimal sample size minimising MSE balances these two components. In one sense, this is analogous to avoiding 'overfitting to noise' in machine learning, where increasing the number of parameters can increase variance more than it decreases bias.

We provide the definitions and a theoretical result to quantify this. Both the 'single vertex' and 'full observation' simplifications considered in the paper compare an observed set $\mathcal{S}$ to its subset $\mathcal{T} \subset \mathcal{S}$ which is reflected in our definitions.

Definition 1: Let $\mathcal{T} \subset \mathcal{S}$. We say that

$$\begin{array}{ll} \boldsymbol{R}_{\mathcal{T}} \text{ is} \, less \, biased \, than \boldsymbol{R}_{\mathcal{S}} & \text{if } \xi_1(\mathcal{T}) < \xi_1(\mathcal{S}) \\ \boldsymbol{R}_{\mathcal{T}} \text{is} less - noise - sensitive \, than \boldsymbol{R}_{\mathcal{S}} & \text{if } \xi_2(\mathcal{T}) < \xi_2(\mathcal{S}) \end{array} \tag{17}$$

Furthermore, we say that

$$\begin{array}{ll} \mathcal{T} \text{is} \, better \, than \, \mathcal{S} & \text{if } \mathrm{MSE}_{\mathcal{T}} < \mathrm{MSE}_{\mathcal{S}} \\ \mathcal{T} \text{is} \, as \, good \, or \, better \, than \, \mathcal{S} & \text{if } \mathrm{MSE}_{\mathcal{T}} \leq \mathrm{MSE}_{\mathcal{S}} \\ \mathcal{T} \text{is} \, worse \, than \, \mathcal{S} & \text{if} \mathrm{MSE}_{\mathcal{T}} > \mathrm{MSE}_{\mathcal{S}}. \end{array} \tag{18}$$

For $i \in \{1, 2\}$ and $\mathcal{T} \subset \mathcal{S}$, let

$$\Delta_i(\mathcal{S}, \mathcal{T}) = \xi_i(\mathcal{S}) - \xi_i(\mathcal{S} \backslash \mathcal{T}). \tag{19}$$

$\Delta_1(\mathcal{S}, \mathcal{T}) > 0$ means $\boldsymbol{R}_{\mathcal{S}\backslash\mathcal{T}}$ is less biased than $\boldsymbol{R}_{\mathcal{S}}$ and $\Delta_2(\mathcal{S}, \mathcal{T}) > 0$ means $\boldsymbol{R}_{\mathcal{S}\backslash\mathcal{T}}$ is less sensitive to noise than $\boldsymbol{R}_{\mathcal{S}}$. Then, by (16), the change in MSE from reducing sample size is

$$\mathrm{MSE}_{\mathcal{S}} - \mathrm{MSE}_{\mathcal{S}\backslash\mathcal{T}} = \Delta_1(\mathcal{S}, \mathcal{T}) + \sigma^2 \cdot \Delta_2(\mathcal{S}, \mathcal{T}) \tag{20}$$

so $\mathcal{S}\backslash\mathcal{T}$ is better than $\mathcal{S}$ if and only if

$$\Delta_1(\mathcal{S}, \mathcal{T}) > -\sigma^2 \cdot \Delta_2(\mathcal{S}, \mathcal{T}). \tag{21}$$

Remark 3: If either $\Delta_1(\mathcal{S}, \mathcal{T})$ or $\Delta_2(\mathcal{S}, \mathcal{T})$ are positive, we can always pick σ^2 so $\mathcal{S}\backslash\mathcal{T}$ is better than $\mathcal{S}$. If both $\Delta_1(\mathcal{S}, \mathcal{T})$ and $\Delta_2(\mathcal{S}, \mathcal{T})$ are negative then $\mathcal{S}\backslash\mathcal{T}$ is never better than $\mathcal{S}$.

The following Theorem characterises our bias/variance trade-off by computing the noise level at which an increase in bias is outweighed by an decrease in noise-sensitivity (or vice-versa) on average.

Theorem 1: Assume a linear reconstruction method and consider $\mathcal{S} \supset \mathcal{T}$. Let

$$\tau(\mathcal{S}, \mathcal{T}) = \frac{\mathbb{E}[||\boldsymbol{x}||_2^2]}{\mathbb{E}[||\boldsymbol{\epsilon}||_2^2]} \cdot \frac{\Delta_2(\mathcal{S}, \mathcal{T})}{-\Delta_1(\mathcal{S}, \mathcal{T})} \tag{22}$$

then $\mathcal{S}\backslash\mathcal{T}$ is better than $\mathcal{S}$ if and only if one of the following conditions is met:

$$\begin{cases} \mathrm{SNR} < \tau(\mathcal{S}, \mathcal{T}) & \text{and } \Delta_1(\mathcal{S}, \mathcal{T}) < 0 \quad \text{(23a)} \\ \mathrm{SNR} > \tau(\mathcal{S}, \mathcal{T}) & \text{and } \Delta_1(\mathcal{S}, \mathcal{T}) > 0 \quad \text{(23b)} \\ 0 < \Delta_2(\mathcal{S}, \mathcal{T}) & \text{and } \Delta_1(\mathcal{S}, \mathcal{T}) = 0. \quad \text{(23c)} \end{cases}$$

Proof Sketch: We first get $\frac{\mathbb{E}[||\boldsymbol{x}||_2^2]}{\mathbb{E}[||\boldsymbol{\epsilon}||_2^2]} \cdot \Delta_2(\mathcal{S}, \mathcal{T}) > -\Delta_1(\mathcal{S}, \mathcal{T}) \cdot \mathrm{SNR}$ from (21) and then divide by $-\Delta_1(\mathcal{S}, \mathcal{T})$, case-splitting on its different possible signs. See Appendix G for a full proof. □

We interpret each of these conditions as follows:

(23a) If $\boldsymbol{R}_{\mathcal{S}\backslash\mathcal{T}}$ is more biased than $\boldsymbol{R}_{\mathcal{S}}$, then $\mathcal{S}\backslash\mathcal{T}$ is better than $\mathcal{S}$ if SNR is low enough (below a threshold τ).

(23b) If $\boldsymbol{R}_{\mathcal{S}\backslash\mathcal{T}}$ is less biased than $\boldsymbol{R}_{\mathcal{S}}$, then $\mathcal{S}\backslash\mathcal{T}$ is better than $\mathcal{S}$ if SNR is *high* enough (above a threshold τ).

(23c) If $\boldsymbol{R}_{\mathcal{S}\backslash\mathcal{T}}$ and $\boldsymbol{R}_{\mathcal{S}}$ are equally biased and $\boldsymbol{R}_{\mathcal{S}\backslash\mathcal{T}}$ is less noise-sensitive than $\boldsymbol{R}_{\mathcal{S}}$, then $\mathcal{S}\backslash\mathcal{T}$ is better than $\mathcal{S}$ at every non-zero noise level.

Theorem 1 does not guarantee that any of the conditions for $\mathcal{S}\backslash\mathcal{T}$ being better than $\mathcal{S}$ would happen; for example, $\boldsymbol{R}_{\mathcal{S}}$ could be both less biased and less sensitive to noise than $\boldsymbol{R}_{\mathcal{S}\backslash\mathcal{T}}$. To show reducing sample size can reduce MSE, i.e. Δ_1 and Δ_2 are not always both non-positive, we look at specific reconstruction methods below.

To show that the conditions in Theorem 1 happen, we focus on showing that Δ_2 can be positive, i.e., decreasing sample size can decrease noise-sensitivity of the reconstruction method. This focus on the noise model means results about Δ_2 under one signal model imply results for *every* finite variance signal model, even if the signal model is not bandlimited:

Proposition 1: Fix a mean zero noise model. Assume that $\boldsymbol{R}_{\mathcal{S}}$ is not a funtion of σ and suppose that we prove that $\Delta_2(\mathcal{S}, \mathcal{T}) > 0$ under a specific signal model. If instead $\boldsymbol{x}$ is drawn from *any* signal model where the signal is independent to the noise and $\mathbb{E}[||\boldsymbol{x}||_2^2] < \infty$, then there exists a threshold $\tau'(\mathcal{S}, \mathcal{T}) > 0$ (which is dependent on the choice of signal model) such that if

$$\mathrm{SNR} < \tau'(\mathcal{S}, \mathcal{T}) \tag{24}$$

then under that new signal model,

$$\mathrm{MSE}_{\mathcal{S}\backslash\mathcal{T}} < \mathrm{MSE}_{\mathcal{S}}. \tag{25}$$

Proof: See Appendix E. □

This includes the signal model where $\boldsymbol{x}$ is some fixed signal. While Proposition 1 means that our approach applies to any signal model, any result computing τ is specific to the chosen signal model. This is the focus of the rest of the section.

C. Specialisation to Signal and Noise Models

We now move to our specific signal and noise models introduced in Section II-A. Firstly, we note that in Theorem 1, $\frac{\mathbb{E}[||\boldsymbol{x}||_2^2]}{\mathbb{E}[||\boldsymbol{\epsilon}||_2^2]}$ is $\frac{k}{N}$ for full-band noise and 1 for bandlimited noise.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
POSSIBLE SIGNS OF $\Delta_1(\mathcal{S}, v)$ AND $\Delta_2(\mathcal{S}, v)$

	$\Delta_2 > 0$	$\Delta_2 \leq 0$
$\Delta_1 > 0$	×	×
$\Delta_1 < 0$	✓	×
$\Delta_1 = 0$	×	✓

(a) LS Reconstruction

	$\Delta_2 > 0$	$\Delta_2 \leq 0$
$\Delta_1 > 0$	✓	✓
$\Delta_1 < 0$	✓	✓
$\Delta_1 = 0$	∼	∼

(b) GLR Reconstruction

For any zero-mean random vector $\boldsymbol{g}$ and compatible matrix $\boldsymbol{A}$, write $\text{Cov}(\boldsymbol{g}) = \boldsymbol{X}\boldsymbol{X}^T$. Using

$$\mathbb{E}_{\boldsymbol{g}}\left[\,||\boldsymbol{A}\boldsymbol{g}||_2^2\right] = \mathbb{E}_{\boldsymbol{g}}\left[\text{tr}\left(\boldsymbol{A}\boldsymbol{g}\boldsymbol{g}^T\boldsymbol{A}^T\right)\right] \tag{26}$$

$$= \text{tr}\left(\boldsymbol{A}\text{Cov}(\boldsymbol{g})\boldsymbol{A}^T\right) = ||\boldsymbol{A}\boldsymbol{X}||_F^2 \tag{27}$$

gives

$$\xi_1(\mathcal{S}) = ||[\boldsymbol{U}]_{\mathcal{N},\mathcal{K}} - \boldsymbol{R}_{\mathcal{S}}[\boldsymbol{U}]_{\mathcal{S},\mathcal{K}}||_F^2 \tag{28}$$

$$\xi_2(\mathcal{S}) = \begin{cases} ||\boldsymbol{R}_{\mathcal{S}}||_F^2 & \text{if full-band noise} \\ ||\boldsymbol{R}_{\mathcal{S}}[\boldsymbol{U}]_{\mathcal{S},\mathcal{K}}||_F^2 & \text{if } k\text{-bandlimited noise} \end{cases} \tag{29}$$

We note that given Theorem 1 and these terms, for any given reconstruction method and two sets $\mathcal{S} \supset \mathcal{T}$ we can now analytically compute if a given $\mathcal{T}$ is better than a set $\mathcal{S}$ under our signal model, and at what noise levels. The rest of the paper focuses on how to find such sets.

To show when $\Delta_2 > 0$, we first consider a 'single vertex' simplification where $\mathcal{T} = \{v\}$. This is the approach we take for LS, and also motivates the 'full observation' simplification for GLR. For a vertex v, as shorthand we write:

$$\tau(\mathcal{S}, v) = \tau(\mathcal{S}, \{v\}), \quad \Delta_i(\mathcal{S}, v) = \Delta_i(\mathcal{S}, \{v\}). \tag{30}$$

We start by looking at possible combinations of signs of $\Delta_1(\mathcal{S}, v)$ and $\Delta_2(\mathcal{S}, v)$, which we present in Table IV.

In Table IV, × will not happen, ✓ may happen and ∼ may theoretically happen but are unlikely. A more detailed explanation of the two tables, and proofs of the '×' subcases, are presented in Appendix F. We now discuss LS and GLR separately in the following sections.

D. LS With Full-Band Noise

In this subsection we show how decreasing sample size by one can decrease MSE under LS. From Table IV(a), we see that reducing sample size never reduces bias under LS, so we focus on when reducing sample size reduces noise-sensitivity.

1) Overview and simplification: Our approach in this subsection is as follows. We consider the 'single vertex' simplification. For a sample set $\mathcal{S}$ and $v \in \mathcal{S}$, we first characterise under exactly what conditions $\mathcal{S}\backslash v$ is better than $\mathcal{S}$ (Corollary 1.1). We then show that the conditions must occur under sampling schemes which are optimal in the noiseless case (Theorem 2). Finally, we comment on how the conditions persist as $N \to \infty$.

2) Characterisation: By Table IV(a) we can eliminate conditions (23b) and (23c) in Theorem 1 for the single-vertex simplification under LS.

We simplify Theorem 1 to the following:

Corollary 1.1: Assume LS reconstruction. Then

$$\tau(\mathcal{S}, v) = \frac{k}{N} \cdot \Delta_2(\mathcal{S}, v) \tag{31}$$

and $\mathcal{S}\backslash\{v\}$ is better than $\mathcal{S}$ if and only if

$$\text{SNR} < \tau(\mathcal{S}, v). \tag{32}$$

Proof Sketch: Under LS, $\Delta_1(\mathcal{S}, v)$ can only be -1 or 0. This simplifies τ in Theorem 1 to (31):

If $\Delta_1(\mathcal{S}, v) = 0$*:* We have $\Delta_2(\mathcal{S}, v) \leq 0$ by Table IV(a). By Theorem 1, $\mathcal{S}\backslash\{v\}$ is never better than $\mathcal{S}$.

If $\Delta_1(\mathcal{S}, v) = -1$*:* We have $\Delta_2(\mathcal{S}, v) > 0$ by Table IV(a) so $\tau(\mathcal{S}, v) > 0$. Theorem 1's conditions reduce to this case.

See Appendix H for a full proof. □

We note that the definition of τ in Corollary 1.1 is not completely equivalent to the definition of τ in Theorem 1 – specifically, if $\Delta_1 = 0$ then τ in Theorem 1 is undefined, and τ in Corollary 1.1 is negative. We make this change for ease of presentation – with this new definition $\tau(\mathcal{S}, v) > 0$ if and only if there is a noise level where $\mathcal{S}\backslash\{v\}$ is better than $\mathcal{S}$.

Corollary 1.1 says that if SNR is too low (below a threshold τ that depends on the bandwidth and the chosen samples), then using a smaller sample set improves the average reconstruction error. Note that we have not yet proven that condition (32) is ever satisfied, i.e., reducing sample size can reduce MSE. We outline why it is not immediately obvious that (32) can hold, and then show situations where it does hold and thus reducing sample size will reduce MSE.

Remark 4: If $\Delta_2 \leq 0$, we have $\tau(\mathcal{S}, v) \leq 0 <$ SNR, so (32) cannot hold and so $\mathcal{S}\backslash\{v\}$ is never better than $\mathcal{S}$ for any SNR.

3) Existence: We first note that Corollary 1.1 as stated leaves room for a clever sampling scheme which picks $\mathcal{S}_i$ where $\tau(\mathcal{S}_i, v)$ is always non-positive and so condition (32) never holds, and hence $\mathcal{S}\backslash\{v\}$ would never be better than $\mathcal{S}$ for any $v \in \mathcal{S}$.

Corollary 1.1 uses that $\Delta_2(\mathcal{S}, v) > 0$ if and only if $\text{rank}\left([\boldsymbol{U}]_{\mathcal{S},\mathcal{K}}\right) = 1 + \text{rank}\left([\boldsymbol{U}]_{\mathcal{S}\backslash\{v\},\mathcal{K}}\right)$; this condition can be understood as observing v being informative in noise-free reconstruction. Note that the addition of a row to a matrix can only change the rank by 0 or 1. This equivalence also means that the 'clever sampling scheme' must never increase said rank as it picks nodes, which is impossible – else we could pick $\mathcal{N}$ sequentially s.t. $\text{rank}\left([\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}\right) = 0$. We now spell out these consequences formally and contextualise them.

Most sampling schemes in the literature construct sample sets by adding vertices one-by-one (e.g. greedy schemes, which are 'near-optimal' [36]). We call such schemes 'sequential'. We now show that for any graph, sequential sampling schemes always eventually add a vertex where (32) can be satisfied.

Proposition 2: Consider a sequential sampling scheme that constructs sample sets $\mathcal{S}_1, \ldots, \mathcal{S}_N$ where $\mathcal{S}_i = \mathcal{S}_{i-1} \cup \{v_i\}$. Then there are exactly k indices $1 \leq I_1, \ldots, I_k \leq N$ where

$$\forall 1 \leq j \leq k : \tau(\mathcal{S}_{I_j}, v_{I_j}) > 0, \tag{33}$$

and so $\mathcal{S}_{I_j}\backslash\{v_{I_j}\}$ is better than $\mathcal{S}_{I_j}$ at some SNR.

Proof Sketch: By Table IV(a), $\tau \propto \Delta_2 > 0 \iff \Delta_1 < 0$. As $\Delta_1(\mathcal{S}, v) = \xi_1(\mathcal{S}) - \xi_1(\mathcal{S}\backslash\{v\}) \in \{0, -1\}$ under LS and

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

$\xi_1(\emptyset) = k$ and $\xi_1(\mathcal{N}) = 0$, we have that $\Delta_2 > 0$ exactly k times. See Appendix I for a full proof. □

Proposition 2 still leaves room for a hypothetical sequential sampling scheme where $\tau(\mathcal{S}, v) > 0$ only for the last k chosen vertices, and therefore if such a scheme selects $|\mathcal{S}| \leq N - k$ then $\mathcal{T} \subset \mathcal{S}$ is never better than $\mathcal{S}$. We now show that there is a trade-off between such a property and performance in the noiseless case, namely that any scheme which is optimal in the noiseless case, like most deterministic schemes in the literature, could not have this property. We first define optimality in the noiseless case.

Definition 2: A sampling scheme is *noiseless-optimal* for LS reconstruction of k-bandlimited signals if the first k vertices it samples form a uniqueness set. That is, it finds the smallest possible uniqueness set [12].

Remark 5: A, D and E-optimal sampling are noiseless-optimal (see Appendix L for a proof).

We now show such schemes find $\tau(\mathcal{S}, v) > 0$ for the first k vertices they pick.

Theorem 2: Suppose we use a sequential noiseless-optimal scheme to select a vertex sample set $\mathcal{S}_m$ of size m. For $m \leq k$:

$$\forall v \in \mathcal{S}_m : \quad \tau(\mathcal{S}_m, v) \geq \frac{k}{N}, \tag{34}$$

i.e., for *any* vertex $v \in \mathcal{S}_m$, if SNR $< \tau(\mathcal{S}_m, v)$ (which always holds if SNR $< \frac{k}{N}$) then $\mathcal{S}\backslash\{v\}$ is better than $\mathcal{S}$. For $m > k$:

$$\forall v_+ \in \mathcal{S}_m \backslash \mathcal{S}_k : \quad \tau(\mathcal{S}_m, v_+) \leq 0. \tag{35}$$

That is, $\mathcal{S}_m\backslash\{v_+\}$ is never better than $\mathcal{S}_m$ for any of the vertices v_+ the sampling scheme adds beyond the first k.

Proof Sketch: For (34), $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}_m}$ and $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}_{m-1}}$ are full-rank so $\Delta_1 = -1$; we then show $\Delta_2 \geq 1$. For (35), as $m > k$, $\xi_1(\mathcal{S}_k) = \xi_1(\mathcal{S}_m)$. See Appendix K for a full proof. □

Theorem 2 explicitly demonstrates a case where reducing sample size reduces MSE: namely if SNR $\leq \frac{k}{N}$, when the sample is picked by an A-, D- or E-optimal scheme, and $|\mathcal{S}| \leq k$, then reducing sample size will always reduce MSE. On the other hand, for $|\mathcal{S}| > k$ we transition into a regime where reducing sample size *cannot* reduce MSE as long as the smaller set also has size $\geq k$.

Theorem 2 also gives us the shape of the MSE curve as sample size varies at high noise levels. At high noise levels, $\text{MSE}_{\mathcal{S}} \approx \sigma^2 \cdot \xi_2(\mathcal{S})$ so $\text{MSE}_{\mathcal{S}} - \text{MSE}_{\mathcal{S}\backslash\{v\}} \approx \sigma^2 \cdot \Delta_2 = \sigma^2 \cdot \frac{k}{N}\tau$ by Corollary 1.1. Theorem 2 gives the sign of τ, which therefore shows whether the MSE increases or decreases with sample size.

Remark 6: Using a sequential noiseless-optimal sampling scheme, such as greedy A-, D- or E- optimal sampling, under sufficiently high noise leads to the MSE being Λ-shaped with regards to sample size with a peak at $|\mathcal{S}| = k$.

The intuition behind a Λ-shaped MSE under noiseless-optimal sampling is as follows. When the sample size is below k and we increase it, we infer more about the signal—which can be seen in the noiseless case, as each of these samples improves our prediction—and inevitably incorporate some of the noise into our reconstruction. Beyond the first k, samples provide no new information about the signal in the noiseless case. These 'additional samples' (corresponding to (35)) are much like getting repetitive observations of already-seen vertices, which we can average to reduce the effect of noise. This is what causes the transition at $|\mathcal{S}| = k$.

Remark 7: If the MSE is Λ-shaped, then even if removing one vertex does not improve $\mathcal{S}$, removing multiple vertices might decrease MSE. This happens if we reduce sample size enough to transition from the right side of the peak to the left side of the peak of Λ.

4) Asymptotics: Finally, we discuss the large graphs case.

Remark 8: If $\frac{k}{N}$ is fixed, the lower bound on τ in Theorem 2 is constant. Therefore, at a fixed SNR (e.g., $\frac{k}{2N}$), decreasing sample size decreases MSE on arbitrarily large graphs.

E. LS With k-Bandlimited Noise

We have observed that MSE can decrease when sample size decreases under LS and full band noise. This raises the question of whether the decrease is caused by some sort of interference effect between the high-frequency components of the noise and the bandlimited (low-frequency) signal. In Appendix A we show that MSE can decrease when sample size decreases under LS with k-bandlimited noise, demonstrating that this is not the case.

F. GLR With Full-Band Noise

In this subsection, we show how decreasing sample size can decrease MSE under GLR reconstruction and full-band noise.

1) Overview and simplification: We start by trying to simplify Theorem 1. Table IV(b) contains no × scenarios and so the 'single vertex' simplification cannot eliminate any conditions in Theorem 1: surprisingly, $\boldsymbol{R}_{\mathcal{S}\backslash\{v\}}$ can be *less* biased than $\boldsymbol{R}_{\mathcal{S}}$ for GLR, which can be observed experimentally. Instead, as we focus on tractability and showing that it is possible to reduce sample size to reduce MSE, rather than fully characterizing all such cases, we pick a situation where $\Delta_1 \geq 0$ so we can simplify Theorem 1. Specifically, we compare the full observation set $\mathcal{S} = \mathcal{N}$ to a subset of it, which we call the 'full observation' simplification. As it is hard to interpret what reconstruction means under full observation [37], our results should be understood as approximately showing that reducing the sample size from nearly full observation to some smaller size may reduce MSE.

Our approach is then as follows. We first characterise under exactly which conditions a sample set $\mathcal{S} \subset \mathcal{N}$ is better than $\mathcal{N}$ (Corollary 2.1). We then show that these conditions must occur if certain graph invariants hold (Theorem 3). Finally, we analyse the parameters in Theorem 3 to obtain a 'suggested sample size' (Remark 9) and show the conditions still occur as $N \to \infty$ (Proposition 4).

2) Characterisation: We first present the following Corollary of Theorem 1.

Corollary 2.1: Assume GLR and that $k > 1$. Consider a non-empty sample set $\mathcal{S} \subset \mathcal{N}$. Then

$$\tau(\mathcal{N}, \mathcal{S}^c) = \frac{k}{N} \cdot \frac{\Delta_2(\mathcal{N}, \mathcal{S}^c)}{-\Delta_1(\mathcal{N}, \mathcal{S}^c)} \tag{36}$$

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

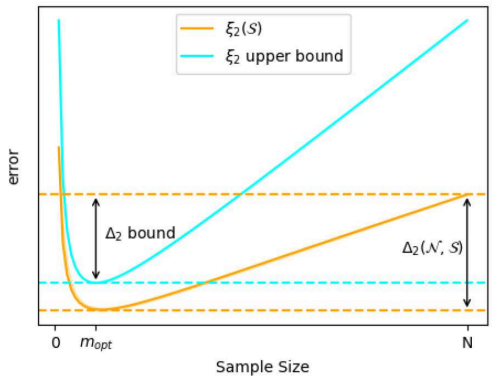


Fig. 1. Proof direction to show $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$.

and $\mathcal{S}$ is better than $\mathcal{N}$ if and only if one of the following conditions is met:

$$\begin{cases} \text{SNR} < \tau(\mathcal{N}, \mathcal{S}^c) & \text{and } [\boldsymbol{U}]_{\mathcal{S}^c,\{2,\dots,k\}} \neq \mathbf{0} \quad (37a) \\ 0 < \Delta_2(\mathcal{N}, \mathcal{S}^c) & \text{and } [\boldsymbol{U}]_{\mathcal{S}^c,\{2,\dots,k\}} = \mathbf{0}. \quad (37b) \end{cases}$$

where $[\boldsymbol{U}]_{\mathcal{S}^c,\{2,\dots,k\}} = \mathbf{0}$ corresponds to any k-bandlimited signal always being constant on all of $\mathcal{S}^c$.

Proof Sketch: We use Cauchy-Schwartz and Lemma 8 to lower bound $\xi_1(\mathcal{S})$ and show that either all k columns of $[\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}$ are eigenvectors of $\boldsymbol{\Pi}_{\mathcal{S}} + \mu\boldsymbol{L}$ (which is exactly when $[\boldsymbol{U}]_{\mathcal{S}^c,\{2,\dots,k\}} = \mathbf{0}$) and $\Delta_1(\mathcal{N}, \mathcal{S}^c) = 0$, corresponding to (37b), or $\Delta_1(\mathcal{N}, \mathcal{S}^c) < 0$, corresponding to (37a). We then apply Theorem 1. See Appendix M for a full proof. □

We now explain the conditions in Corollary 2.1. Condition (37a) corresponds to the single case we see in Corollary 1.1. Condition (37b) is more of an edge case, e.g., if $\lambda_k < N$ and every vertex in $\mathcal{S}^c$ has degree $N-1$ [38, Corollary 2.3]. In (37b), $[\boldsymbol{U}]_{\mathcal{S}^c,\{2,\dots,k\}} = \mathbf{0}$ means any k-bandlimited signal will be constant on all of $\mathcal{S}^c$. Our proof shows that in the noiseless case this implies that observing $\mathcal{S}^c$ will not improve the MSE, i.e., $\text{MSE}_{\mathcal{S}\cup\mathcal{T}_c} = \text{MSE}_{\mathcal{S}}$ for any $\mathcal{T}_c \subseteq \mathcal{S}^c$. The other condition in (37b), i.e., $0 < \Delta_2(\mathcal{N}, \mathcal{S}^c)$, corresponds to an increase in noise-sensitivity from reconstructing from those additional vertices. Therefore, condition (37b) says that if $\mathcal{S}^c$ reveals nothing new about the underlying signal and also makes the reconstruction more sensitive to noise, one should not observe $\mathcal{S}^c$ and only observe $\mathcal{S}$.

3) Existence: Like Corollary 1.1, Corollary 2.1 does not show that any set $\mathcal{S}$ is ever better than $\mathcal{N}$, i.e., that $\tau(\mathcal{N}, \mathcal{S}^c) > 0$ ever happens. As described in Section III-C, we will find $\mathcal{S}$ s.t. $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$. We illustrate our proof method in Fig. 1. Fundamentally, our approach rests on the fact that ξ_2, which is approximately proportional to $\text{MSE}_{\mathcal{S}}$ at high noise levels, is approximately U-shaped, and we can use an upper bound to approximate the minimum and thus bound $\Delta_2(\mathcal{N}, \mathcal{S}^c)$ below. Explicitly, our method to show $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ is:

1) Find an upper bound $B(\cdot)$ for $\xi_2(\mathcal{S})$ that only depends on $\boldsymbol{L}$ and $|\mathcal{S}|$, not on μ or the choice of $\mathcal{S}$.
2) Find the sample size m_{opt} that minimises the bound and show $m_{opt} < N$.
3) Identify when $\xi_2(\mathcal{N}) - B(m_{opt}) > 0$ (this is the 'Δ_2 bound' in Fig. 1). As this quantity is a lower bound for $\Delta_2(\mathcal{N}, \mathcal{S}^c)$ this shows that $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ for any $\mathcal{S}$ of size m_{opt}.
4) Bound Δ_1 using the same tools and use these to lower bound $\tau(\mathcal{N}, \mathcal{S}^c)$

We see that m_{opt} suggests an 'optimal sample size', in that the ξ_2 upper bound approximates ξ_2, and we can optimise it as a surrogate. Following step 1, we now bound $\xi_2(\mathcal{S})$.

Lemma 1: Let λ_i be the eigenvalues of $\boldsymbol{L}$, the combinatorial Laplacian. Let $|\mathcal{S}| = m$. Then

$$\xi_2(\mathcal{S}) \leq r\frac{N}{m} + \sum_{i=2}^{m} \omega\left(\max\left[1, \frac{\lambda_{N+2-i}}{\lambda_i}\right]\right) \quad (38)$$

$$\leq r\frac{N}{m} + r(m-1). \quad (39)$$

where we define $B(m)$ to be the RHS of (38) and where

$$\omega(x) = \frac{1}{4}\left(\sqrt{x} + \frac{1}{\sqrt{x}}\right)^2, \quad r = \omega\left(\frac{\lambda_N}{\lambda_2}\right) \quad (40)$$

and the summation in (38) is 0 if $m = 0$ or 1.

Proof: Decompose $\boldsymbol{R}_{\mathcal{S}}$ and apply Kantorovich's inequality. See Appendix O. □

The function ω arises from the Kantorovich inequality, commonly used in variance bounds [39], [40]. ω is an increasing function and for $x > 1, \omega(x) \leq x$; this means r increases with $\frac{\lambda_N}{\lambda_2}$. We discuss r in more detail in Appendix S.

As mentioned, in step 4 we need a bound on $\xi_1(\mathcal{S})$ to lower bound $\tau(\mathcal{N}, \mathcal{S}^c)$. The shape and accuracy of this bound do not affect the core of our argument, only how tight our approximation of τ is. Therefore, we leave it in Appendix P; the term $B_k(m)$ it introduces in the proof is a tightening of $B(m)$ ($B_k(m) < B(m)$) and is explained in Appendix B.

By steps 2–4, using the bound $B(m)$ from Lemma 1 yields:

Theorem 3: Let $B(m)$ be defined as in Lemma 1 and let m_{opt} be the sample size minimising B.

$$\text{If} \qquad B(m_{opt}) < N \quad (41)$$

then $m_{opt} < \lceil \frac{N+1}{2} \rceil$. Furthermore, $\exists \mu_{ub} > 0$ and $\tau_{GLR}(\mu) > 0$ s.t. under GLR with parameter $\mu \in (0, \mu_{ub})$,

$$\text{if} \qquad \text{SNR} < \tau_{GLR}(\mu) \quad (42)$$

then, *any* sample set $\mathcal{S}$ of size m_{opt} is better than $\mathcal{N}$.

Proof Sketch: We prove that for any μ, $\xi_1(\mathcal{S}) \leq k + B_k(m)$ and $\xi_2(\mathcal{S}) \leq B(m)$ via Lemmas 1 and 3. We also show $\xi_1(\mathcal{N}) = \sum_{i=1}^{k}\left(1 - \frac{1}{1+\mu\lambda_i}\right)^2$ and $\xi_2(\mathcal{N}) = \sum_{i=1}^{N}(1+\mu\lambda_i)^{-2}$. We use these to bound Δ_i, and finally apply Corollary 2.1 as our conditions mean $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$. See Appendix Q for a full proof. □

We contrast Theorems 2 and 3. Theorem 2 provides necessary and sufficient conditions. Theorem 3, while still useful, only provides sufficient conditions. Theorem 2 applies to any graph, but only to sample sets chosen under noiseless-optimal sampling schemes, while Theorem 3 has no requirement on sampling schemes, but only applies to graphs which fulfill certain graph invariants.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

Because of the difficulty in writing down an analytic form for m_{opt}, the parameters in Theorem 3 – m_{opt}, μ_{ub} and τ_{GLR} – are difficult to analyse. To better understand we present a version of Theorem 3 instead based on the weaker bound (39). This bound is a linear transformation of $\frac{N}{m} + m$ and so is clearly U-shaped and has an minimum at $m = \sqrt{N}$, and thus following steps 2, 3 and 4 yields the following proposition:

Proposition 3: Let $\bar{\lambda} = \frac{1}{N}\sum_{i=1}^{N} \lambda_i$ be the mean eigenvalue of $\boldsymbol{L}$ and let r be defined as in Lemma 1.

$$\text{If} \quad 2r\sqrt{N} < N, \tag{43}$$

$$\text{then let} \quad \mu_{ub_weak} = \bar{\lambda}^{-1}\left(\sqrt[4]{N}\cdot(2r)^{-\frac{1}{2}} - 1\right) \tag{44}$$

$$\text{and} \quad \tau_{GLR_weak} = \frac{\sqrt{N}(1+\mu\bar{\lambda})^{-2} - 2r}{\sqrt{N} + 2r\cdot\frac{N}{k}} \tag{45}$$

then under GLR with parameter $\mu \in (0, \mu_{ub_weak})$, if

$$\text{SNR} < \tau_{GLR_weak}(\mu) \tag{46}$$

then *any* sample set $\mathcal{S}$ of size $\left\lceil\sqrt{N}\right\rceil$ is better than $\mathcal{N}$.

Proof Sketch: We bound $\xi_1(\mathcal{S})$ and $\xi_2(\mathcal{S})$ in terms of $r\left(\frac{N}{m} + m - 1\right)$, which is minimised at $m = \sqrt{N}$. We explicitly compute $\xi_i(\mathcal{N})$ and apply Corollary 2.1. See Appendix S for a full proof. □

Proposition 3 is sufficient to show that decreasing sample size can decrease MSE on a restricted class of graph models (e.g. Erdős–Rényi graphs). Proposition 3 involves several parameters: r, $2r\sqrt{N}$, μ_{ub_weak} and τ_{GLR_weak} which can be interpreted as graph properties (see Appendix R for a detailed explanation and sensitivity analysis).

Intuitively, we expect the equivalent parameters in Proposition 3 and Theorem 3 to behave similarly; for m_{opt} to behave like $\left\lceil\sqrt{N}\right\rceil$, for $B(m_{opt})$ to behave like $2r\sqrt{N}$ and for μ_{ub} and τ_{GLR} to behave like their weak counterparts. We can explicitly show that m_{opt} is $\mathcal{O}(\sqrt{N})$:

Remark 9: If condition (41) in Theorem 3 holds then $m_{opt} \in \left[\left\lfloor\sqrt{N}\right\rfloor, \left\lceil\sqrt{rN}\right\rceil\right]$ and $m_{opt} \leq \left\lceil\frac{N+1}{2}\right\rceil$.

Proof: See Appendix R. □

The proof of Theorem 3 leads to an upper bound of $\text{MSE}_{\mathcal{S}}$. We present this upper bound to link the experiments in Section IV with Fig. 1, which can be considered to be the case where $\sigma^2 \to \infty$.

Corollary 3.1: Let $B(m)$ be defined as in Lemma 1 and $B_k(m)$ be defined as it will be in Lemma 3.

For a sample set $\mathcal{S}$ of size m,

$$\text{MSE}_{\mathcal{S}} \leq (k + B_k(m)) + \sigma^2 \cdot B(m). \tag{47}$$

Proof: By Lemma 1 $\xi_2(\mathcal{S}) \leq B(m)$. By Lemma 9 in Appendix P, $\xi_1(\mathcal{S}) \leq k + B_k(m)$. Combining these using (16) gives the bound. □

4) Asymptotics: Finally, we consider the large graph case.

While our sensitivity analysis suggests that $\tau_{GLR} \to 0$ as N gets large, this is an oversimplification from assuming indepence of parameters in our analysis. We present a Proposition to show that that $\tau_{GLR} \not\to 0$ as $N \to \infty$ for ER graphs.

Proposition 4: For Erdős–Rényi graphs, as $N \to \infty$, condition (41) in Theorem 3 holds w.h.p. and τ_{GLR} tends to a positive constant even under optimal choice of μ [37].

Proof: Full statement and proof in Appendix T. □

This shows that reducing sample size can reduce MSE on arbitrarily large graphs under GLR even with optimal μ.

G. GLR With k-Bandlimited Noise

Once more, one might ask whether the MSE increasing with sample size under GLR is caused by some sort of interference effect between the high-frequency components of the noise and the bandlimited (low-frequency) signal. We present a variant of Theorem 3 in Appendix B to disprove this.

IV. Experiments

Our theoretical results show how the relationship between sample size and MSE changes with the level of noise, focusing on how reducing sample size will reduce MSE if the SNR is below a threshold. In this section, we demonstrate the applicability and validity of these results via empirical experiments.

We first demonstrate the applicability of our results with plots of the thresholds $\tau(\mathcal{S}, v)$, τ_{GLR} and τ_{GLR_bl} against sample size (Figs. 2 and 3). These plots show concrete SNR values for the thresholds, giving a practical understanding of how high noise levels need to be for reducing sample size to reduce MSE for different random graph models and parameters. Additionally, we empirically tabulate the probabilities that graphs from each model satisfy the conditions of our theorems (Table VI), helping the reader evaluate the impact of our theorems across different applications. We then demonstrate the validity of our results by plotting $\text{MSE}_{\mathcal{S}}$ against sample size (Figs. 4 and 5) at SNRs below, near and above the derived thresholds, showing that the behaviour of $\text{MSE}_{\mathcal{S}}$ follows our theoretical results. We finally present similar results on real-world datasets, validating the applicability of our results to real-world datasets.

A. Experimental Setup

We now present the setup of the experiments. All results are presented with 90% confidence intervals and all experiments use the combinatorial Laplacian $\boldsymbol{L}$ and its eigenbasis.

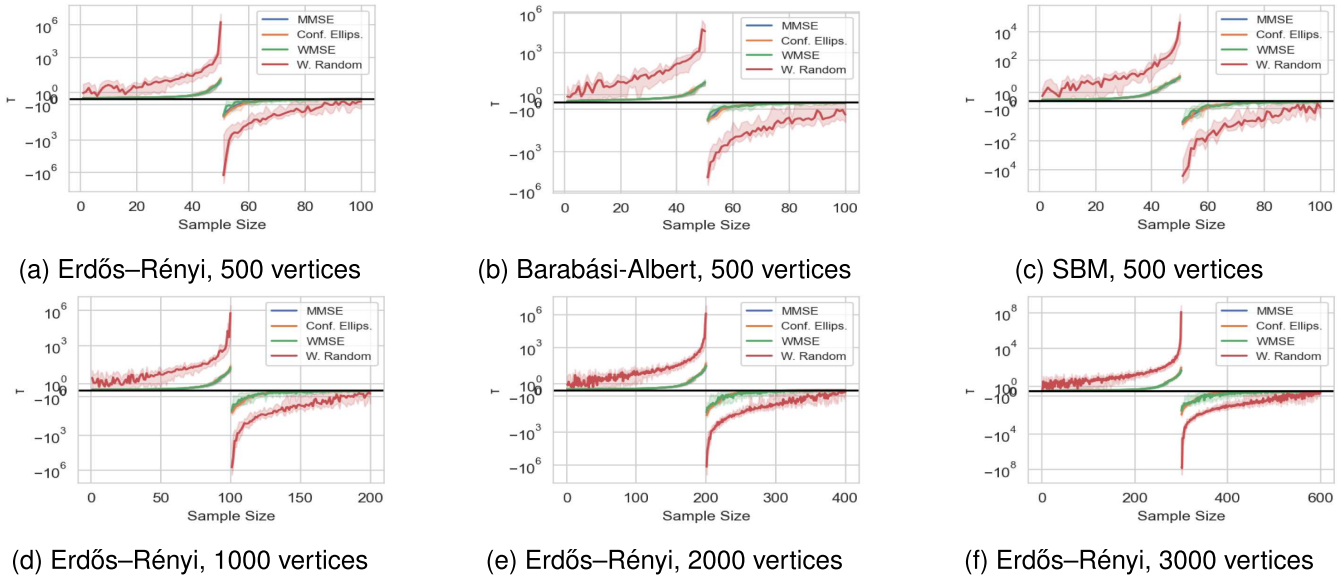
1) Synthetic graph generation: We consider each of the following unweighted random graph models:

- Erdős–Rényi (ER) with edge probability $p = 0.8$ (experiments with other values of p show similar results)
- 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

We consider 10 instantiations of each model for plots, and 1000 instantiations of each model to assess the probability the graph invariant conditions in our Theorems are met.

We present threshold plots for graphs with 500, 1000, 2000 and 3000 vertices. We only present MSE plots for graphs with 500 vertices (like [16, Fig. 8]) as they are intended as an accompaniment to our threshold plots and theorems to demonstrate their validity, and a single graph size suffices.

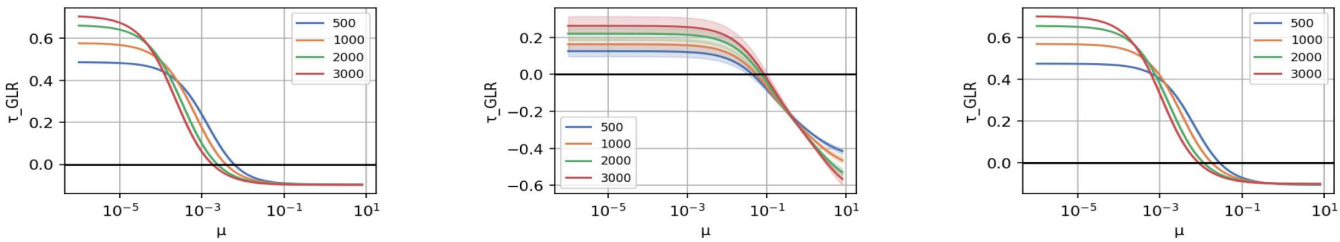
Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.



(a) Erdős–Rényi, 500 vertices (b) Barabási-Albert, 500 vertices (c) SBM, 500 vertices

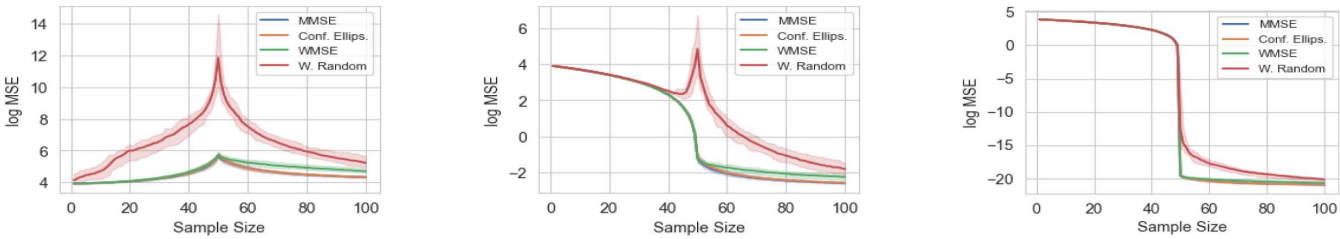
(d) Erdős–Rényi, 1000 vertices (e) Erdős–Rényi, 2000 vertices (f) Erdős–Rényi, 3000 vertices

Fig. 2. $\tau(\mathcal{S}, v)$ for different random graph models and different N under LS and full-band noise (bandwidth $= \frac{\#\text{ vertices}}{10}$).



(a) Erdős–Rényi (τ_{GLR}) (b) Barabási-Albert (τ_{GLR}) (c) SBM (τ_{GLR})

Fig. 3. τ_{GLR} for different random graph models (#vertices $=$ colour, bandwidth $= \frac{\#\text{ vertices}}{10}$).



(a) Full-band noise, SNR = 10^{-1} (b) Full-band noise, SNR = 10^{2} (c) Full-band noise, SNR = 10^{10}

Fig. 4. Average MSE under LS on ER graphs (#vertices $= 500$, bandwidth $= 50$).

2) Synthetic signal generation: We set the bandwidth $k = \lfloor \frac{N}{10} \rfloor$, per [16]. We consider these SNRs for full-band noise:

- LS: $10^{-1}, 10^{2}, 10^{10}$ (i.e. $-10dB, 20dB, 100dB$)
- GLR: $10^{-1}, 0.5, 10^{10}$ (i.e. $-10dB, -3dB, 100dB$)

and the following SNRs for bandlimited noise:

- LS: $10^{-1}, 1, 10^{10}$ (i.e. $-10dB, 0dB, 100dB$)
- GLR: $10^{-2}, 0.5, 10^{10}$ (i.e. $-20dB, -3dB, 100dB$)

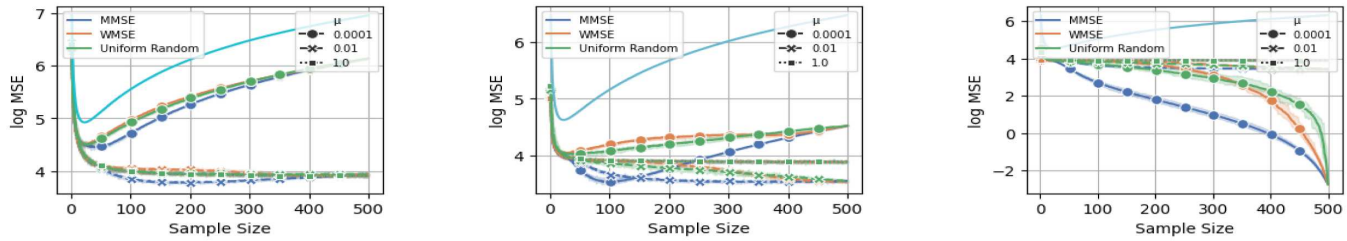
These SNRs are chosen to demonstrate that there are three regimes for MSE with distinctive properties—the high noise regime, the transitionary regime and the approximately noiseless regime—and that τ captures when the regimes change. Suitable values of SNRs to demonstrate this vary between reconstruction methods and noise types, hence our choices.

To test the MSE in reconstructing signals from samples on random graph models, we generate 200 signals by sampling $\boldsymbol{y} = \boldsymbol{x}_{raw} + \sigma \cdot \epsilon_{raw}$ where:

1) $\boldsymbol{x}_{raw} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Pi}_{bl(\mathcal{K})})$

2a) If full-band noise, $\boldsymbol{\epsilon}_{raw} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$, $\sigma = \sqrt{\frac{k}{N \cdot \text{SNR}}}$

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.



(a) Full-band noise, SNR = 10^{-1} (b) Full-band noise, SNR = $\frac{1}{2}$ (c) Full-band noise, SNR = 10^{10}

Fig. 5. Average MSE under GLR on ER graphs (#vertices = 500, bandwidth = 50). Line without markers is an upper bound.

2b) If bandlimited noise, $\boldsymbol{\epsilon}_{raw} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Pi}_{bl(\mathcal{K})})$, $\sigma = \frac{1}{\sqrt{\text{SNR}}}$

3) Real-world datasets: We also consider two real-world datasets as in [41]. The first is an FMRI dataset, where the original graph consists of 4465 nodes corresponding to voxels of the cerebellum, with 292 Blood-Oxygen-Level-Dependent (BOLD) signals derived from FMRI. We sample a connected subgraph of 367 nodes via Neighbourhood Sampling [42]. The second is a Weather dataset, where a 10-nearest neighbours graph of 45 cities in Sweden is constructed, with 95 signals derived from the temperature. See [41], [43], [44] for details on graph construction and signal generation in both cases. We generate $\boldsymbol{x}$ by k-bandlimiting the original signals, where For FMRI we set $k = 36$ and for Weather dataset we set $k = 8$. We otherwise follow the above methodology in Synthetic Signal Generation for generating signals with full-band noise.

4) Sample-set selection: We generate sample sets greedily using exact analytic forms and by exactly computing $\mathbf{\Pi}_{bl(\mathcal{K})}$.

- LS: We use (7)-(9) to exactly compute the MMSE [14], [15], [32], Confidence Ellipsoid [13], [32], [33] and WMSE criteria [16], which are deterministic and guaranteed to be noiseless-optimal. We also look at Weighted Random sampling [11], which is neither deterministic nor guaranteed to be noiseless-optimal.
- GLR: We exactly compute the MMSE criterion, which is a function of SNR and noise type, and the WMSE criterion [5], which is not. We also consider uniform random sampling.

Note that sampling schemes in the literature tend to differ from ours mainly in that they approximate our setup for computational efficiency reasons; e.g. approximating the projection matrix $\mathbf{\Pi}_{bl(K)}$ with a polynomial in $\boldsymbol{L}$ [14], and approximating optimality criteria [16]. As these differences are for efficiency reasons, we do not expect them to matter in our experiments.

5) Parameters of reconstruction methods: We consider LS with $k = \lfloor \frac{N}{10} \rfloor$ (as above) and GLR with $\mu \in \{10^{-4}, 10^{-2}, 1\}$.

B. Experimental Results on Synthetic Graphs

For full-band noise, we present threshold plots for all graphs and MSE plots for ER graphs in the main body of the paper. MSE plots for BA and SBM graphs are presented in Appendix X. For bandlimited noise, we present threshold plots, MSE plots and discussion of how the experimental and theoretical results correspond in Appendix C. We also present Table V showing how our plots correspond to our theoretical results, which also corresponds to the summary of theoretical results in Table III.

TABLE V
CORRESPONDENCE BETWEEN THEORETICAL AND EMPIRICAL RESULTS

	LS		GLR	
	Full-band	Bandlimited	Full-band	Bandlimited
Characterisation	Corr 1.1 (Figs. 2a-2c & 4a-4c)	Corr 3.2 (Figs. 11a-11c)	Corr 2.1 (Figs. 5a-5c)	
Existence	Thm 2 (Figs. 2a-2c & 4a-4c)	Corr 3.3 (Figs. 11a-11c)	Thm 3 (Figs. 3a-3c & 5a-5c & Table VI)	Thm 4 (Figs. 10a-10c & 12a-12c & Table VII
Asymptotics	Rmk 8 (Figs. 2d-2f)	Rmk 10	Propn 4 (Figs. 3a-3c)	Propn 5 (Figs. 10a-10c)

1) τ plots (LS): Fig. 2 shows $\tau(\mathcal{S}, v)$ as sample size varies for sequential sampling methods under LS, where v is the latest node added to $\mathcal{S}$. For ER graphs, for sample size smaller than the bandwidth, $\tau(\mathcal{S}, v) > 0$ and beyond that $\tau(\mathcal{S}, v) \leq 0$. The maximum of $\tau(\mathcal{S}, v)$ observed is approximately 10^6 $(60dB)$ for weighted random sampling, and approximately 10 $(10dB)$ for the deterministic sampling methods. The confidence intervals for weighted random sampling is much wider than for the deterministic sampling methods. Next, we observe the same phenomenon as with ER for BA and SBM graphs, with maxima of approximately 10^5 $(50dB)$ for weighted random sampling and maxima of approximately 10 $(10dB)$ for the deterministic sampling methods. Finally, Fig. 2d–2f show the same phenomenon as Fig. 2a happens for ER graphs at sizes of 1000, 2000 and 3000 vertices.

We now correlate our experiments and our theoretical results. As Corollary 1.1 is necessary and sufficient, the sign of $\tau(\mathcal{S}, v)$ tells us exactly when removing a vertex improves $\mathcal{S}$. Therefore τ being negative is an informative statement, telling us that $\mathcal{S}\backslash\{v\}$ is *never* better than $\mathcal{S}$. Concretely, if SNR is below the maximum $\tau(\mathcal{S}, v)$ observed, then when sample size equals bandwidth we can reduce sample size to reduce MSE.

Theorem 2 proves that noiseless-optimal methods (MMSE, Confidence Ellipsoid and WMSE in our experiments) will have $\tau(\mathcal{S}, v) > 0$ for sample sizes smaller than the bandwidth, and then $\tau(\mathcal{S}, v) \leq 0$ afterwards, and Fig. 2 validates this. Note that even though this pattern holds for Weighted Random Sampling in our experiments, Theorem 2 does not guarantee it always holds for Weighted Random Sampling.

While we prove that $\tau(\mathcal{S}, v) \not\to 0$ as $N \to \infty$, we conjecture the stronger claim that at a sample size equal to bandwidth,

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

$\tau(\mathcal{S}, v)$ might actually increase with graph size, which is supported (but not proven) by Fig. 2d–2f.

2) τ_{GLR} plots (GLR): In Fig. 3a–3c, we plot τ_{GLR}, where if SNR $< \tau_{GLR}$, then there is a sample size $m_{opt} < N$ where *any* sample set of size m_{opt} is better than $\mathcal{N}$. Unlike with LS, our theorems are only sufficient so SNR $> \tau_{GLR}$ is uninformative. For ER graphs, we see that τ_{GLR} is decreasing in μ, and that $\tau_{GLR} > 0$ for sufficiently small μ for all graph sizes. The maximum value in this case for τ_{GLR} ranges between 0.4 $(-4dB)$ and 0.7 $(-1.5dB)$. We see a similar pattern to ER graphs for BA graphs, the main differences being that $\tau_{GLR} > 0$ for larger values of μ and that the maxmimum of τ_{GLR} is approximately 0.2 $(-7dB)$. τ_{GLR} for SBM graphs behaves very similarly to ER graphs in our experiments. Note that the confidence intervals for ER and SBM graphs are so tight as to not be clearly seen in Fig. 3, while being much wider for BA graphs. As graph properties, when we sample from a random graph model, r is a random variables. We observe wider confidence intervals when $\mathcal{G}$ is sampled from the BA graph model as $\text{Var}_{\mathcal{G}}(r)$ is much higher than when we sample from the ER or SBM graph models.

As $\tau_{GLR} > 0$, Fig. 3a–3c show we can reduce sample size to reduce MSE for all examined graph models. We see τ_{GLR} is only positive for small enough μ, motivating μ_{ub} in Theorem 3. The value of μ below which $\tau_{GLR} > 0$ is at least μ_{ub}.

Finally, Proposition 4 proves that $\tau_{GLR} \not\to 0$ as $N \to \infty$ for $\mu = \frac{c}{\lambda_2}$, or $\frac{c}{\lambda_N}$, or $\frac{c}{\sqrt{\lambda_2 \lambda_N}}$ on ER graphs, i.e. for decreasing μ as N increases. We note that Fig. 3a–3c provide empirical evidence that this might hold for all graph models tested.

3) MSE plots (LS): The MSE plots demonstrate the validity of our theoretical results linking MSE and sample size.

Fig. 4a–4c show log MSE against sample size for LS under full-band noise. For high noise (a), we see MSE increases with sample size no larger than bandwidth, and decreases afterwards – that is, it is Λ-shaped, as described in Remark 6. In (b), for our three deterministic noiseless-optimal sampling schemes (orange/green/blue), MSE is decreasing in sample size. For weighted random sampling, we see for sample sizes no larger than bandwidth, MSE first decreases and then increases, attaining a maximum with sample size at bandwidth, and then decreases with sample size. In (c), the almost noiseless case, we see MSE is decreasing in sample size for all sampling schemes, with a large drop when sample size is at bandwidth.

Comparing Figs. 4a–4c to 2a, Fig. 4a corresponds to when SNR $< \tau(\mathcal{S}, v)$, Fig. 4c corresponds to SNR $> \tau(\mathcal{S}, v)$ and Fig. 4b corresponds to when SNR lies between some values of $\tau(\mathcal{S}, v)$ for weighted random sampling. We see that MSE increases with sample size when SNR $< \tau(\mathcal{S}, v)$ and decreases otherwise, proving the validity of Corollary 1.1. Fig. 4a (green, orange, blue curves) shows that for low SNRs, optimal sampling schemes lead MSE to monotonically increase with each additional sample until the sample size reaches the bandwidth, illustrating Theorem 2 and Remark 6. We also validate Remark 7: if we are slightly above the bandwidth (50 for Fig. 4a), i.e., to the right of the peak, then reducing sample size by one does not reduce MSE; however, if we significantly reduce sample size, i.e., transitioning from just right of the peak to left of the peak, we can reduce MSE.

Interestingly, Fig. 4a shows that at a low SNR of 10^{-1}, the optimal sample size under LS is zero. Although this might appear counter-intuitive at a first glance, it makes concrete the idea that reconstruction does not work if there is too much noise: at this noise level letting $\hat{\boldsymbol{x}} = \mathbf{0}$ rather than fitting with LS with any number of observed vertices will result in a lower MSE on average. We can also formalise this in terms of our Bias-Variance decomposition; a $\mathbf{0}$ reconstruction has high bias but zero variance, and reconstructing from a non-zero number of samples has lower bias but positive variance. At a high enough noise the variance term in the MSE will dominate, and the MSE from taking $\hat{\boldsymbol{x}} = \mathbf{0}$ will be lower than reconstructing from any non-zero number of samples. The same reasoning applies to why the MSE at a smaller number of samples (e.g. 5 samples) is better than a larger number (e.g. 100 samples).

On the other hand, for high SNRs (Fig. 4c), MSE decreases monotonically as sample size increases for all sampling schemes, showing Corollary 1.1 is necessary and sufficient. Finally, Fig. 4b illustrates the situation between the two cases.

4) MSE plots (GLR): As with LS, the MSE plots demonstrate the validity of our theoretical results linking MSE and sample size. Fig. 5a–5c show how MSE changes with sample size for different values of μ under full-band noise, along with an upper bound (light blue) which is not dependent on μ. This bound is approximately U-shaped in all cases. For SNR $= 10^{-1}$, we see for $\mu = 10^{-4}$, under all sampling schemes, MSE is minimised at a sample size around 16 to 22. For the MMSE sampling scheme with $\mu = 0.01$, MSE is minimised at a sample size of approximately 200. In all other cases where SNR $= 10^{-1}$, MSE is minimised at full observation $(|\mathcal{S}| = 500)$. For SNR $= \frac{1}{2}$ and $\mu = 10^{-4}$, MSE is minimised at a sample size a bit less than 100. For larger μ, we see MSE is minimised at full observation. In the nearly noiseless case, MSE decreases with sample size under all parameter choices. In all cases where the MSE is minimised at a sample size less than N, the MSE is approximately U-shaped like our bound and our diagram Fig. 1.

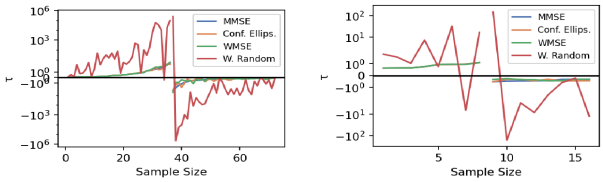
Fig. 5a–5c illustrate Corollary 2.1, Theorem 3 and Corollary 3.1 in the following ways. First, the MSE upper bound corresponds to Corollary 3.1, and we can see it is greater than any observed MSE at each sample size. The sample size which minimises our upper bound is m_{opt} (Remark 9) and as $r \in (1, 1.01]$ in our Erdős–Rényi experiments, $m_{opt} \in [22, 23]$, which empirically well approximates the sample size that minimises MSE in our low μ and low SNR experiments. Second, Fig. 5a–5c show that MSE can decrease with increasing sample size, and Fig. 5c shows that at high SNR full observation is best, illustrating the necessary and sufficient nature of Corollary 2.1. Finally, Fig. 5a–5c illustrate the dependence on SNR and μ in Theorem 3, i.e. at low μ and SNR the optimal sample size is less than N, but at high enough μ or SNR this no longer holds.

Fig. 5a–5c also demonstrate some limitations of the characterisation in Corollary 3.1 and Theorem 3. We see from Fig. 5b that even though the MSE at m_{opt} is lower than at N, our bound never goes below the maximum observed MSE, so is too loose to show this. This is partly because the Theorem and the Corollary bound $\xi_2(\mathcal{S})$ only as a function of $|\mathcal{S}|$, ignoring μ and the composition of $\mathcal{S}$. This limitation corresponds to a gap

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
PROBABILITY THEOREM CONDITIONS ARE MET

	ER	SBM	BA
Theorem 3 conditions met	100%	100%	99.9%



(a) FMRI (b) Weather

Fig. 6. $\tau(\mathcal{S}, v)$ for real-world data under LS & full-band noise.

between τ_{GLR} and $\tau(\mathcal{N}, \mathcal{S}^c)$, demonstrating τ_{GLR} is a lower bound for $\tau(\mathcal{N}, \mathcal{S}^c)$ rather than an exact characterisation.

5) Checking conditions: While our theorems for LS apply to all graphs, Theorem 3 for GLR relies on conditions around graph invariants. We sample graphs from each random graph model to empirically show the probability the conditions of Theorem 3 are met at a sample size of m_{opt} for some $\mu > 0$:

Proposition 4 shows the conditions in Theorem 3 hold w.h.p. for ER graphs as $N \to \infty$. However, the proposition does not say whether the conditions hold for a graph of a given size, or other graph models. The results in Table VI show empirically that the conditions hold under full-band noise (Theorem 3) for ER, BA and SBM graphs with 500 vertices. This outlines the applicability of our theorems.

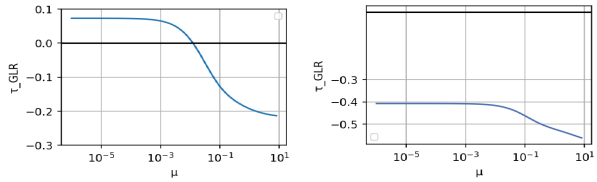
If the conditions on our Theorems are not met, they provide no information about the shape of the MSE. However, Figs. 18 and 22 in Appendix X show empirically that for BA and SBM graphs under bandlimited noise with SNR $\in \{10^{-2}, \frac{1}{2}\}$ under GLR with $\mu \in \{10^{-2}, 10^{-4}\}$, even if the conditions of Theorem 4 are not met, reducing sample size from N to below N reduces MSE under the presented sampling schemes. We leave further investigation of this as future work.

C. Experimental Results on Real-World Datasets

We first discuss the plots of τ and τ_{GLR} for real-world datasets. In practice, one cannot know for sure what the signal model of a real-world signal is; however, computation of τ is dependent on our choice of theoretical signal model. We have therefore computed τ under the signal model assumptions given in Section II-A.

1) τ plots (LS): We examine Fig. 6. By Proposition 1, the sign of $\tau(\mathcal{S}, v)$ we have plotted is provably correct for *any* signal model τ. We validate the actual value of τ with MSE experiments (Figs. 8 and 9).

2) τ plots (GLR): By Theorem 3, τ_{GLR} being positive means $\tau(\mathcal{N}, \mathcal{S}^C) > 0$ under the bandlimited signal model. As Theorem 3 functions by showing $\Delta_2 > 0$, by Proposition 1, τ_{GLR} being positive under one signal model means $\tau(\mathcal{N}, \mathcal{S}^C)$ is positive under any signal model. We use this to interpret Fig. 7. We can see from Fig. 7a that if $\mu < 0.01$ then at some noise level, any set of size m_{opt} is better than $\mathcal{N}$. Fig. 7b which shows



(a) FMRI (b) Weather

Fig. 7. τ_{GLR} for real-world data.

$\tau_{GLR} < 0$ for all μ is completely uninformative – τ_{GLR} is a lower bound for $\tau(\mathcal{N}, \mathcal{S}^C)$, and a negative lower bound cannot tell us whether the bounded quantity is or isn't positive. We validate this in the MSE section.

3) MSE plots (LS): At high noise levels, Fig. 8d–8a and show a Λ-shaped MSE, validating Proposition 1 applied to Remark 6.

In both plots in Fig. 6, all of the lines are above 10^{-1}. This corresponds to Fig. 8d and 8a being Λ-shaped. In Fig. 6, all of the lines are under 10^{10}. This corresponds to Fig. 8f and 8f being decreasing, which we observe.

In Fig. 6a, we see the red line (Weighted Random sampling) is broadly above 10^2 and the other lines (MMSE, WMSE and Confidence Ellipsoid Sampling) are below 10^2. This corresponds to the red MSE line being Λ-shaped and the other lines decreasing in Fig. 8b.

In Fig 6b all lines are below 10^2 and so we expect at SNR $= 10^2$ for all lines to be decreasing. For MMSE, WMSE and Confidence Ellipsoid sampling this is consistent with Fig. 8e; however, it is clear our covariance assumption has underestimated τ for Weighted Random Sampling as the red line is Λ-shaped in Fig. 9e.

We see that the real-world results are broadly in line with the synthetic experiments, validating our use of the signal model presented in Section II-A to calculate τ for LS.

4) MSE plots (GLR): All four figures in Fig. 9 corresponding to SNR $\leq \frac{1}{2}$ and $\mu < 1$ show the minimum MSE at sample size of significantly less than N, which validates Theorem 3 combined with Proposition 1. This is expected for the FMRI graph for $\mu < 0.01$, as $\tau_{GLR} > 0$ and we had proof that decreasing sample size from N would decrease MSE. However, it is the case for the Weather graph, even though our computation of τ_{GLR} was negative; this is not unexpected as Theorem 3 is sufficient but not necessary.

V. DISCUSSION

In this paper we studied the impact of sample size on linear reconstruction of noisy k-bandlimited graph signals. We showed theoretically and experimentally, in the same settings as much of the sample set selection literature, that reconstruction error is not always monotonic in sample size, i.e., at sufficiently low SNRs, reconstruction error can sometimes be improved by *reducing* sample size. Our finding reveals that existing results in the literature for the noiseless setting may not necessarily generalise to the noisy case, even when considering regularised reconstruction methods. It also demonstrates the need to consider both optimal sample size selection and reconstruction

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

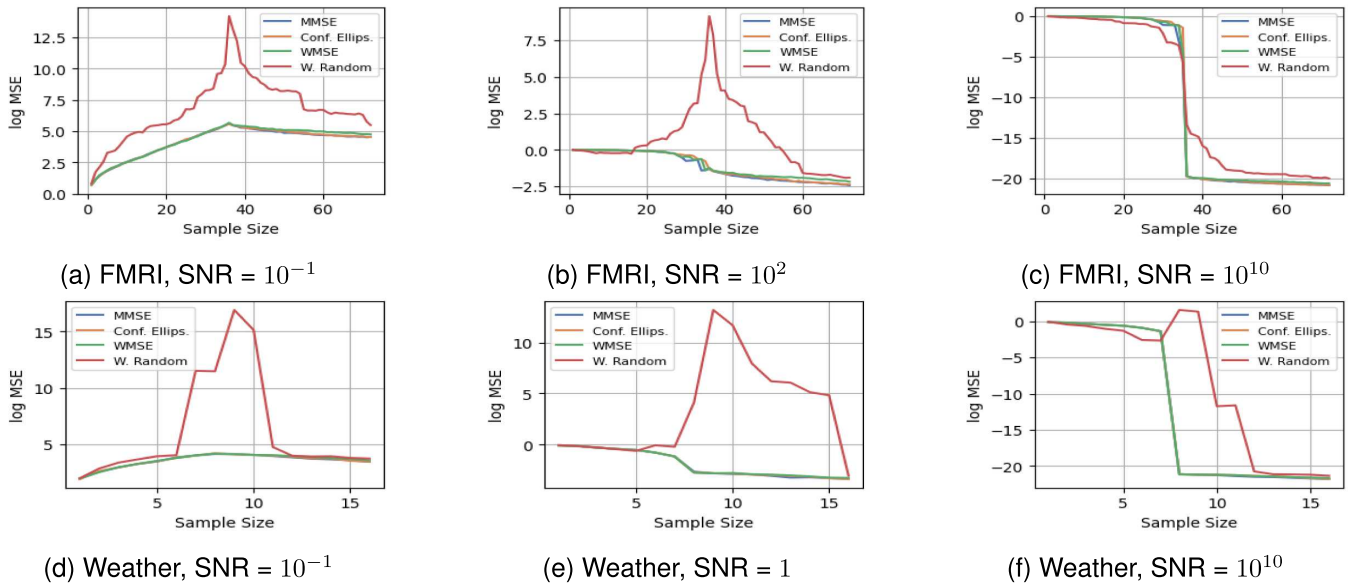


Fig. 8. Average MSE under LS on real-world data and full-band noise.

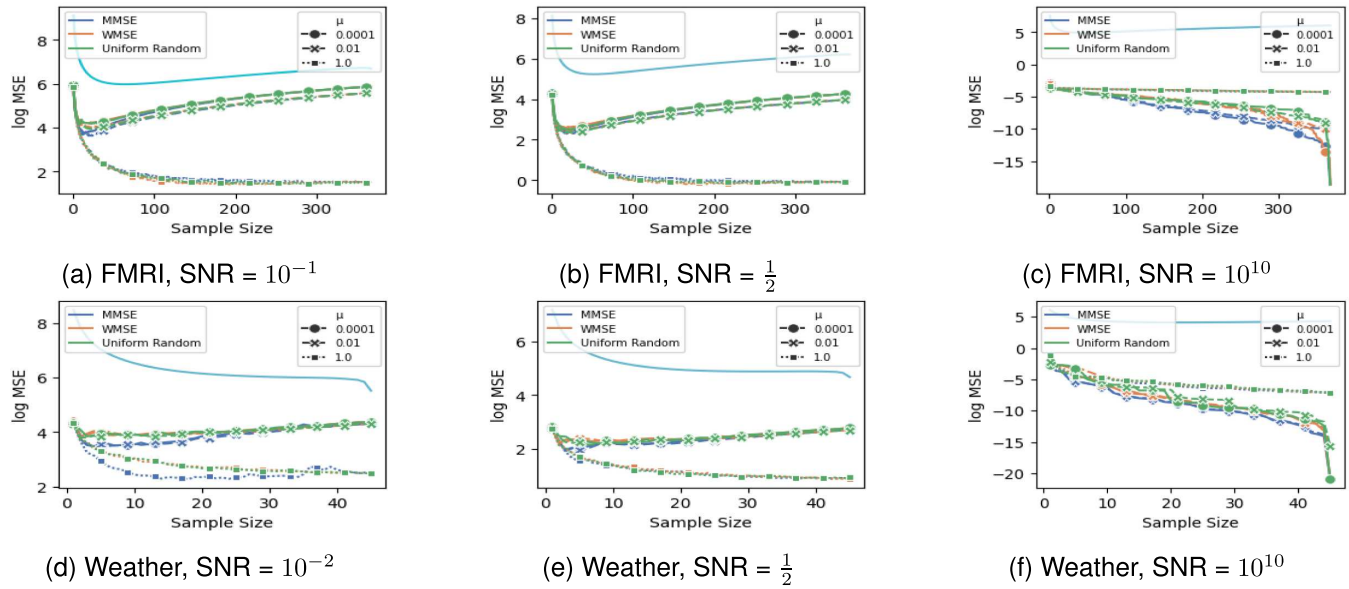


Fig. 9. Average MSE under GLR on real-world data and full-band noise. Line without markers is an upper bound.

methods at the same time, and motivates assessment of noise levels in datasets to do so.

One practical implication of our theoretical results is that it can provide useful guidance on how to make use of the sampling budget available. For example, if one is aware of potentially high noise present in the observed signals, to minimise MSE under LS reconstruction, it might be preferable to not use all of the provided sampling budget.

Another use case could be a practical sampling algorithm, for both the LS and GLR reconstructions. For LS reconstruction, if one has chosen k or fewer nodes by some noiseless-optimal scheme, one can calculate $\tau(\mathcal{S}, v)$ for each element of $\mathcal{S}$, remove v if SNR $< \tau(\mathcal{S}, v)$, and repeat until it is no longer possible to find a node v in our remaining set $\mathcal{S}_{rem}$ where SNR $< \tau(\mathcal{S}_{rem}, v)$.

For the GLR reconstruction, with a fixed μ and belief that the signal model in Section III-C applies (the latter could be assessed based on past observations and is required for computation of the SNR threshold τ), then our results suggest the following sampling algorithm can be considered. Given the graph structure, before observing any nodes:

1) Check whether condition 41 in Theorem 3 applies (this only depend on graph structure);
2) If they do, compute τ_{GLR};
3) If it is believed that SNR $< \tau_{GLR}$ holds, sample m_{opt} (approximately $\sqrt{N}$) nodes at random and reconstruct from those nodes.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.

We finally remark that future work includes extending the analysis on GLR to the normalised graph Laplacian, providing bounds on ξ_2 for LS, analysing other graph models such as Ring graphs or studying a more detailed "early-stopping" mechanism in sequential sampling schemes that do not use the full sample budget.

REFERENCES

[1] S. Itani and D. Thanou, "A graph signal processing framework for the classification of temporal brain data," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Piscataway, NJ, USA: IEEE Press, 2021, pp. 1180–1184.

[2] R. K. Jain, J. M. Moura, and C. E. Kontokosta, "Big data+ big cities: Graph signals of urban air pollution [Exploratory SP]," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 130–136, Sep. 2014.

[3] B. Renoust, G. Cheung, and S. Satoh, "Estimating political leanings from mass media via graph-signal restoration with negative edges," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Piscataway, NJ, USA: IEEE Press, 2017, pp. 1009–1014.

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

[5] 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.

[6] 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.

[7] F. Pukelsheim, *Optimal Design of Experiments.* Philadelphia, PA, USA: SIAM, 2006.

[8] A. Nikolov, M. Singh, and U. Tantipongpipat, "Proportional volume sampling and approximation algorithms for a-optimal design," *Math. Oper. Res.*, vol. 47, no. 2, pp. 847–877, 2022.

[9] L. F. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 34–47, Jan. 2018.

[10] X. Xie, J. Yu, H. Feng, and B. Hu, "Bayesian design of sampling set for bandlimited graph signals," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Piscataway, NJ, USA: IEEE Press, 2019, pp. 1–5.

[11] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Appl. Comput. Harmon. Anal.*, vol. 44, no. 2, pp. 446–475, 2018.

[12] H. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Piscataway, NJ, USA: IEEE Press, 2014, pp. 933–936.

[13] N. Tremblay, P.-O. Amblard, and S. Barthelmé, "Graph sampling with determinantal processes," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Piscataway, NJ, USA: IEEE Press, 2017, pp. 1674–1678.

[14] F. Wang, Y. Wang, and G. Cheung, "A-optimal sampling and robust reconstruction for graph signals via truncated Neumann series," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 680–684, May 2018.

[15] F. Wang, G. Cheung, and Y. Wang, "Low-complexity graph sampling with noise and signal reconstruction via Neumann series," *IEEE Trans. Signal Process.*, vol. 67, no. 21, pp. 5511–5526, Nov. 2019.

[16] 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.

[17] 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.

[18] O. Nabar and G. Shroff, "Conservative predictions on noisy financial data," in *Proc. ACM Int. Conf. AI Finance*, 2023, pp. 427–435.

[19] T. Bush et al., "The impact of COVID-19 public health restrictions on particulate matter pollution measured by a validated low-cost sensor network in Oxford, UK," *Build. Environ.*, vol. 237, Jun. 2023, Art. no. 110330.

[20] B. Sripathmanathan, X. Dong, and M. Bronstein, "On the impact of sample size in reconstructing graph signals," in *Proc. Int. Conf. Sampling Theory IEEE Int. Symp. Spread Spectr. Tech. Appl. (SampTA)*, Piscataway, NJ, USA: IEEE Press, 2023, pp. 1–6.

[21] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Fundamental limits of sampling strategies," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 539–554, Dec. 2016.

[22] S. Lin, X. Xie, H. Feng, and B. Hu, "Active sampling for approximately bandlimited graph signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP),* Piscataway, NJ, USA: IEEE Press, 2019, pp. 5441–5445.

[23] Y. Tanaka and Y. C. Eldar, "Generalized sampling on graphs with subspace and smoothness priors," *IEEE Trans. Signal Process.*, vol. 68, pp. 2272–2286, 2020.

[24] J. Hara and Y. Tanaka, "Sampling set selection for graph signals under arbitrary signal priors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP),* Piscataway, NJ, USA: IEEE Press, 2022, pp. 5732–5736.

[25] F. Zhang, "Schur complements and matrix inequalities in the Löwner ordering," *Linear Algebra Appl.*, vol. 321, nos. 1–3, pp. 399–410, 2000.

[26] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Piscataway, NJ, USA: IEEE Press, 2013, pp. 491–494.

[27] C.-F. Gauss, *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae.* Henricus Dieterich, 1823.

[28] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Dec. 2016.

[29] M. Belkin and P. Niyogi, "Semi-supervised learning on Riemannian manifolds," *Mach. Learn.*, vol. 56, pp. 209–239, 2004.

[30] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: John Hopkins Univ. Press, 2013.

[31] I. Pesenson, "Sampling in Paley-wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.

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

[33] A. Jayawant and A. Ortega, "Practical graph signal sampling with log-linear size scaling," *Signal Process.*, vol. 194, May 2022, Art. no. 108436.

[34] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Comput.*, vol. 4, no. 1, pp. 1–58, 1992.

[35] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* Cambridge, MA, USA: MIT Press, 2016.

[36] L. F. Chamon and A. Ribeiro, "Near-optimality of greedy set selection in the sampling of graph signals," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Piscataway, NJ, USA: IEEE Press, 2016, pp. 1265–1269.

[37] P.-Y. Chen and S. Liu, "Bias-variance tradeoff of graph Laplacian regularizer," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1118–1122, Aug. 2017.

[38] R. Merris, "Laplacian graph eigenvectors," *Linear Algebra Appl.*, vol. 278, nos. 1–3, pp. 221–236, 1998.

[39] C. Khatri and C. R. Rao, "Some generalizations of Kantorovich inequality," *Sankhyā, Indian J. Statist., A*, vol. 44, no. 1, pp. 91–102, 1982.

[40] A. Householder, "The Kantorovich and some related inequalities," *SIAM Rev.*, vol. 7, no. 4, pp. 463–473, 1965.

[41] Y.-C. Zhi, Y. C. Ng, and X. Dong, "Gaussian processes on graphs via spectral kernel learning," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 9, pp. 304–314, 2023.

[42] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1024–1034.

[43] A. Venkitaraman, S. Chatterjee, and P. Handel, "Gaussian processes over graphs," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP),* Piscataway, NJ, USA: IEEE Press, 2020, pp. 5640–5644.

[44] H. Behjat, U. Richter, D. Van De Ville, and L. Sörnmo, "Signal-adapted tight frames on graphs," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 6017–6029, Nov. 2016.

[45] R. Bhatia, *Matrix Analysis*, vol. 169. New York, NY, USA: Springer Science & Business Media, 2013.

[46] S. Liu and H. Neudecker, "Kantorovich and Cauchy-Schwarz inequalities involving positive semidefinite matrices, and efficiency comparisons for a singular linear model," *Linear Algebra Appl.*, vol. 259, no. 1, pp. 209–221, Jul. 1997.

[47] K. Nordström, "Convexity of the inverse and Moore–Penrose inverse," *Linear Algebra Appl.*, vol. 434, no. 6, pp. 1489–1512, 2011.

[48] M. Barahona and L. M. Pecora, "Synchronization in small-world systems," *Phys. Rev. Lett.*, vol. 89, no. 5, 2002, Art. no. 054101.

[49] T. Jiang, "Low eigenvalues of Laplacian matrices of large random graphs," *Probab. Theory Related Fields*, vol. 153, nos. 3–4, pp. 671–690, Aug. 2012.

[50] J. K. Baksalary and S. Puntanen, "Generalized matrix versions of the Cauchy-Schwarz and Kantorovich inequalities," *Aequationes Mathematicae*, vol. 41, no. 1, pp. 103–110, Mar. 1991.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on May 30,2026 at 10:51:34 UTC from IEEE Xplore. Restrictions apply.