

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

Baskaran Sripathmanathan, Xiaowen Dong, Michael Bronstein  
University of Oxford

**Abstract**—Reconstructing a signal on a graph from noisy observations of a subset of the vertices is a fundamental 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  $\Lambda$ -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,  $\tau$  and  $\tau_{GLR}$ , 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

**R**EAL-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]–[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.

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
<b>This Paper</b>	✓	<b>LS, GLR</b>	<b>Full</b>

(b) Theoretical Results

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]–[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].

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  $N \rightarrow \infty$ .
- 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  $\mathbf{L}$  is real positive semidefinite with  $N$  distinct eigenvalues  $0 = \lambda_1 < \lambda_2 < \dots < \lambda_N$  which are also called *graph frequencies* [4]<sup>1</sup>. Write the eigendecomposition of  $\mathbf{L}$  as  $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$  where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$  and the columns of  $\mathbf{U}$  are the eigenbasis of  $\mathbf{L}$  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  $\mathbf{U}$  [5]. This is a common smoothness model for graph signals. To obtain this, we define  $\mathbf{\Pi}_{bl(\mathcal{K})}$  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]–[16] of assuming the underlying signal is  $k$ -bandlimited with additive observation noise. We assume we observe a corrupted signal  $\mathbf{y} = \mathbf{x} + \mathbf{n}$  where

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

Our results apply to any noise with these properties (e.g. Gaussian or Rademacher noise). We refer to the  $\text{Cov}(\boldsymbol{\epsilon}) = \mathbf{I}_N$  case as ‘full-band noise’ as the associated corrupted signal  $\mathbf{y}$  has high frequency components, whereas the other case does not. In the literature, noise levels are often described using the  $\text{SNR} = \frac{\mathbb{E}[\|\mathbf{x}\|_2^2]}{\mathbb{E}[\|\mathbf{n}\|_2^2]}$ , a ratio of norms which must be positive<sup>2</sup>.

We note that under full-band noise we have  $\sigma^2 = \frac{k}{N \cdot \text{SNR}}$  and under  $k$ -bandlimited noise we have  $\sigma^2 = \frac{1}{\text{SNR}}$ .

<sup>1</sup>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.

<sup>2</sup>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  $-20\text{dB}$  will be written as  $10^{-20/10} > 0$ .

### B. Notation for sampling

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

$$[\mathbf{X}]_{\mathcal{A}} = [\mathbf{X}]_{\mathcal{A}, \mathcal{A}}. \quad (1)$$

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

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

Finally,  $\mathcal{A} \setminus \mathcal{B} = \{i \in \mathcal{A} \mid i \notin \mathcal{B}\}$  and  $\mathcal{A}^c = \mathcal{N} \setminus \mathcal{A}$ . 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  $\mathbf{U}$ .  $\mathbf{\Pi}_{bl(\mathcal{K})}$  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:

$$\text{LS:} \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \text{span}([\mathbf{U}]_{\mathcal{N}, \mathcal{K}})} \|\mathbf{x}\|_2 \quad (3)$$

$$\text{GLR:} \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_2^2 + \mu \mathbf{x}^T \mathbf{L} \mathbf{x} \quad (4)$$

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  $[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}$ .

TABLE II: The LS and GLR reconstruction methods.

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

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  $\mathbf{R}_{\mathcal{S}} \in \mathbb{R}^{N \times |\mathcal{S}|}$ .

**Remark 1.** LS and GLR are both linear:

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

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

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

Across all linear models under the noisy setting, LS leads to the minimum-variance unbiased estimator of  $\mathbf{x}$  [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  $\mathbf{L}^\dagger$  [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  $[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}$  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  $\mathbf{x}, \epsilon$  and  $\sigma$  are Gaussian, we can compute the best predictor of  $\mathbf{x}$  in MSE terms. This is the expectation of  $\mathbf{x}$  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  $\mathbf{R}_{\mathcal{S}}$  is not dependent on  $\sigma$ .

### 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  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = k$  [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]

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

$$\text{(MMSE) A-Optimality: minimise } \text{tr}(\mathbf{P}^{-1}) \quad (7)$$

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

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

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

and we define  $\text{tr}(\mathbf{P}^{-1}) = +\infty$  in (7) if  $\mathbf{P}$  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  $\mathbf{x}$  drawn from a known distribution.
- Observations of  $\mathbf{x}$  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

$$\text{MSE}_{\mathcal{S}} = \mathbb{E}_{\mathbf{x}, \epsilon} \left[ \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \mid \mathcal{S} \text{ observed} \right]. \quad (11)$$

### 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:

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

where the bias term  $\xi_1(\mathcal{S}) = \text{MSE}_{\mathcal{S}}$  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  $N \rightarrow \infty$ , 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 $\mathcal{S} \setminus \{v\}$		$\mathcal{N}$ 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  $\mathbf{x}$  and  $\epsilon$  are drawn from independent distributions, that  $\mathbb{E}[\epsilon] = 0$ ,  $\mathbb{E}[\|\mathbf{x}\|_2^2] < \infty$  and  $\mathbb{E}[\|\epsilon\|_2^2] < \infty$ . 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  $\hat{\mathbf{x}}$  be a reconstruction of the signal  $\mathbf{x}$ , then

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

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

*Proof.* See Appendix D or [35, Chapter 5.4.4].  $\square$

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

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

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

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

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

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

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

$$\begin{aligned} \mathbf{R}_{\mathcal{T}} \text{ is less biased than } \mathbf{R}_{\mathcal{S}} & \quad \text{if } \xi_1(\mathcal{T}) < \xi_1(\mathcal{S}) \\ \mathbf{R}_{\mathcal{T}} \text{ is less noise-sensitive than } \mathbf{R}_{\mathcal{S}} & \quad \text{if } \xi_2(\mathcal{T}) < \xi_2(\mathcal{S}) \end{aligned} \quad (17)$$

Furthermore, we say that

$$\begin{aligned} \mathcal{T} \text{ is better than } \mathcal{S} & \quad \text{if } \text{MSE}_{\mathcal{T}} < \text{MSE}_{\mathcal{S}} \\ \mathcal{T} \text{ is as good or better than } \mathcal{S} & \quad \text{if } \text{MSE}_{\mathcal{T}} \leq \text{MSE}_{\mathcal{S}} \\ \mathcal{T} \text{ is worse than } \mathcal{S} & \quad \text{if } \text{MSE}_{\mathcal{T}} > \text{MSE}_{\mathcal{S}}. \end{aligned} \quad (18)$$

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

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

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

$$\text{MSE}_{\mathcal{S}} - \text{MSE}_{\mathcal{S} \setminus \mathcal{T}} = \Delta_1(\mathcal{S}, \mathcal{T}) + \sigma^2 \cdot \Delta_2(\mathcal{S}, \mathcal{T}) \quad (20)$$

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

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

**Remark 3.** If either  $\Delta_1(\mathcal{S}, \mathcal{T})$  or  $\Delta_2(\mathcal{S}, \mathcal{T})$  are positive, we can always pick  $\sigma^2$  so  $\mathcal{S} \setminus \mathcal{T}$  is better than  $\mathcal{S}$ . If both  $\Delta_1(\mathcal{S}, \mathcal{T})$  and  $\Delta_2(\mathcal{S}, \mathcal{T})$  are negative then  $\mathcal{S} \setminus \mathcal{T}$  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

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

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

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

*Proof Sketch.* We first get  $\frac{\mathbb{E} \left[ \|\mathbf{x}\|_2^2 \right]}{\mathbb{E} \left[ \|\boldsymbol{\epsilon}\|_2^2 \right]} \cdot \Delta_2(\mathcal{S}, \mathcal{T}) > -\Delta_1(\mathcal{S}, \mathcal{T}) \cdot \text{SNR}$  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.  $\square$

We interpret each of these conditions as follows:

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

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

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

Theorem 1 does not guarantee that any of the conditions for  $\mathcal{S} \setminus \mathcal{T}$  being better than  $\mathcal{S}$  would happen; for example,  $\mathbf{R}_{\mathcal{S}}$  could be both less biased and less sensitive to noise than  $\mathbf{R}_{\mathcal{S} \setminus \mathcal{T}}$ . To show reducing sample size can reduce MSE, i.e.  $\Delta_1$  and  $\Delta_2$  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  $\Delta_2$  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  $\Delta_2$  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  $\mathbf{R}_{\mathcal{S}}$  is not a function of  $\sigma$  and suppose that we prove that  $\Delta_2(\mathcal{S}, \mathcal{T}) > 0$  under a specific signal model. If instead  $\mathbf{x}$  is drawn from *any* signal model where the signal is independent to the noise and  $\mathbb{E} \left[ \|\mathbf{x}\|_2^2 \right] < \infty$ , then there exists a threshold  $\tau'(\mathcal{S}, \mathcal{T}) > 0$  (which is dependent on the choice of signal model) such that if

$$\text{SNR} < \tau'(\mathcal{S}, \mathcal{T}) \quad (24)$$

then under that new signal model,

$$\text{MSE}_{\mathcal{S} \setminus \mathcal{T}} < \text{MSE}_{\mathcal{S}}. \quad (25)$$

*Proof.* See Appendix E.  $\square$

This includes the signal model where  $\mathbf{x}$  is some fixed signal. While Proposition 1 means that our approach applies to any signal model, any result computing  $\tau$  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,  $\frac{\mathbb{E} \left[ \|\mathbf{x}\|_2^2 \right]}{\mathbb{E} \left[ \|\boldsymbol{\epsilon}\|_2^2 \right]}$  is  $\frac{k}{N}$  for full-band noise and 1 for bandlimited noise.

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

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

$$= \text{tr}(\mathbf{A} \text{Cov}(\mathbf{g}) \mathbf{A}^T) = \|\mathbf{A} \mathbf{X}\|_F^2 \quad (27)$$

gives

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

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

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:

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

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.

TABLE IV: Possible signs of  $\Delta_1(\mathcal{S}, v)$  and  $\Delta_2(\mathcal{S}, v)$ .

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

(a) LS Reconstruction

(b) GLR Reconstruction

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 IVa, 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  $\mathcal{S} \setminus v$  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  $N \rightarrow \infty$ .

2) *Characterisation:* By Table IVa 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

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

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

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

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

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

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

See Appendix H for a full proof.  $\square$

We note that the definition of  $\tau$  in Corollary 1.1 is not completely equivalent to the definition of  $\tau$  in Theorem 1 – specifically, if  $\Delta_1 = 0$  then  $\tau$  in Theorem 1 is undefined, and  $\tau$  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  $\mathcal{S} \setminus \{v\}$  is better than  $\mathcal{S}$ .

Corollary 1.1 says that if SNR is too low (below a threshold  $\tau$  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  $\tau(\mathcal{S}, v) \leq 0 < \text{SNR}$ , so (32) cannot hold and so  $\mathcal{S} \setminus \{v\}$  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  $\mathcal{S} \setminus \{v\}$  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  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = 1 + \text{rank}([\mathbf{U}]_{\mathcal{S} \setminus \{v\}, \mathcal{K}})$ ; 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.  $\text{rank}([\mathbf{U}]_{\mathcal{N}, \mathcal{K}}) = 0$ . 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  $\mathcal{S}_1, \dots, \mathcal{S}_N$  where  $\mathcal{S}_i = \mathcal{S}_{i-1} \cup \{v_i\}$ . Then there are exactly  $k$  indices  $1 \leq I_1, \dots, I_k \leq N$  where

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

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

*Proof Sketch.* By Table IVa,  $\tau \propto \Delta_2 > 0 \iff \Delta_1 < 0$ . As  $\Delta_1(\mathcal{S}, v) = \xi_1(\mathcal{S}) - \xi_1(\mathcal{S} \setminus \{v\}) \in \{0, -1\}$  under LS and  $\xi_1(\emptyset) = k$  and  $\xi_1(\mathcal{N}) = 0$ , we have that  $\Delta_2 > 0$  exactly  $k$  times. See Appendix I for a full proof.  $\square$

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  $m \leq k$ :

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

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

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

That is,  $\mathcal{S}_m \setminus \{v_+\}$  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.  $\square$

Theorem 2 explicitly demonstrates a case where reducing sample size reduces MSE: namely if  $\text{SNR} \leq \frac{k}{N}$ , 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  $\text{MSE}_{\mathcal{S}} - \text{MSE}_{\mathcal{S} \setminus \{v\}} \approx \sigma^2 \cdot \Delta_2 = \sigma^2 \cdot \frac{k}{N} \tau$  by Corollary 1.1. Theorem 2 gives the sign of  $\tau$ , 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  $\Lambda$ -shaped with regards to sample size with a peak at  $|\mathcal{S}| = k$ .

The intuition behind a  $\Lambda$ -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  $\Lambda$ -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  $\Lambda$ .

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

**Remark 8.** If  $\frac{k}{N}$  is fixed, the lower bound on  $\tau$  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 IVb contains no  $\times$  scenarios and so the ‘single vertex’ simplification cannot eliminate any conditions in Theorem 1: surprisingly,  $\mathbf{R}_{\mathcal{S} \setminus \{v\}}$  can be *less* biased than  $\mathbf{R}_{\mathcal{S}}$  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  $N \rightarrow \infty$  (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

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

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

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

where  $[\mathbf{U}]_{\mathcal{S}^c, \{2, \dots, k\}} = \mathbf{0}$  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  $[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}$  are eigenvectors of  $\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L}$  (which is exactly when

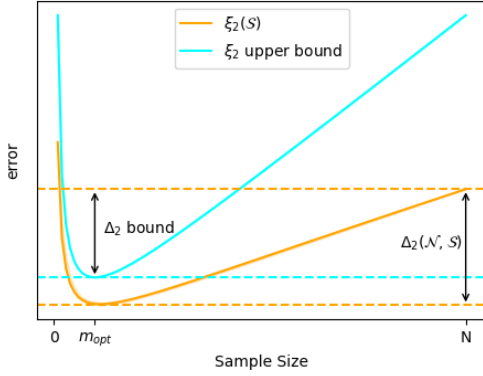


Fig. 1: Proof direction to show  $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ .

$[\mathbf{U}]_{\mathcal{S}^c, \{2, \dots, k\}} = \mathbf{0}$ ) 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.  $\square$

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),  $[\mathbf{U}]_{\mathcal{S}^c, \{2, \dots, k\}} = \mathbf{0}$  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  $\xi_2$ , 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  $\mathbf{L}$  and  $|\mathcal{S}|$ , not on  $\mu$  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 ‘ $\Delta_2$  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  $\Delta_1$  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 the sense that the  $\xi_2$  upper bound approximates  $\xi_2$ , and we can optimise it as a surrogate. Following step 1, we now bound  $\xi_2(\mathcal{S})$ .

**Lemma 1.** Let  $\lambda_i$  be the eigenvalues of  $\mathbf{L}$ , 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  $\mathbf{R}_{\mathcal{S}}$  and apply Kantorovich’s inequality. See Appendix O.  $\square$

The function  $\omega$  arises from the Kantorovich inequality, commonly used in variance bounds [39], [40].  $\omega$  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  $\tau$  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  $B$ .

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

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  $\mu \in (0, \mu_{ub})$ ,

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

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

*Proof Sketch.* We prove that for any  $\mu$ ,  $\xi_1(\mathcal{S}) \leq k + B_k(m)$  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  $\Delta_i$ , and finally apply Corollary 2.1 as our conditions mean  $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ . See Appendix N for a full proof.  $\square$

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.

Because of the difficulty in writing down an analytic form for  $m_{opt}$ , the parameters in Theorem 3 –  $m_{opt}$ ,  $\mu_{ub}$  and  $\tau_{GLR}$  – 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 a 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  $\mathbf{L}$  and let  $r$  be defined as in Lemma 1.

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

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

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

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

$$\text{SNR} < \tau_{GLR\_weak}(\mu) \quad (46)$$

then any sample set  $\mathcal{S}$  of size  $\lceil \sqrt{N} \rceil$  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 Q for a full proof.  $\square$

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:  $r$ ,  $2r\sqrt{N}$ ,  $\mu_{ub\_weak}$  and  $\tau_{GLR\_weak}$  which can be interpreted as graph properties (see Appendix S 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  $\lceil \sqrt{N} \rceil$ , for  $B(m_{opt})$  to behave like  $2r\sqrt{N}$  and for  $\mu_{ub}$  and  $\tau_{GLR}$  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  $m_{opt} \in \left[ \lceil \sqrt{N} \rceil, \lceil \sqrt{rN} \rceil \right]$  and  $m_{opt} \leq \lceil \frac{N+1}{2} \rceil$ .

*Proof.* See Appendix R.  $\square$

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  $\sigma^2 \rightarrow \infty$ .

**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  $m$ ,

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

*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.  $\square$

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

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

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

*Proof.* Full statement and proof in Appendix T.  $\square$

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

## 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  $\tau(\mathcal{S}, v)$ ,  $\tau_{GLR}$  and  $\tau_{GLR\_bl}$  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  $\mathbf{L}$  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.

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

and the following SNRs for bandlimited noise:

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

These SNRs are chosen to demonstrate that there are three regimes for MSE with distinctive properties—the high noise regime, the transitional regime and the approximately noiseless regime—and that  $\tau$  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  $\mathbf{y} = \mathbf{x}_{raw} + \sigma \cdot \epsilon_{raw}$  where:

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

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

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

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  $\mathbf{x}$  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  $\mathbf{\Pi}_{bl(\mathcal{K})}$  with a polynomial in  $\mathbf{L}$  [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
<b>Characterisation</b>	Corr 1.1 (Figs. 2a-2c & 4a-4c)	Corr 3.2 (Figs. 11a -11c)	Corr 2.1 (Figs. 5a-5c)	
<b>Existence</b>	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)
<b>Asymptotics</b>	Rmk 8 (Figs. 2d-2f)	Rmk 10	Propn 4 (Figs. 3a-3c)	Propn 5 (Figs. 10a-10c)

1)  $\tau$  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, Figs. 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  $\tau$  being negative is an informative statement, telling us that  $\mathcal{S} \setminus \{v\}$  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  $\tau(\mathcal{S}, v) \not\rightarrow 0$  as  $N \rightarrow \infty$ , we conjecture the stronger claim that at a sample size equal to bandwidth,  $\tau(\mathcal{S}, v)$  might actually increase with graph size, which is supported (but not proven) by Figs. 2d-2f.

2)  $\tau_{GLR}$  plots (GLR): In Figs. 3a-3c, we plot  $\tau_{GLR}$ , where if  $\text{SNR} < \tau_{GLR}$ , 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  $\text{SNR} > \tau_{GLR}$  is uninformative. For ER graphs, we see that  $\tau_{GLR}$  is decreasing in  $\mu$ , and that  $\tau_{GLR} > 0$  for sufficiently small  $\mu$  for all graph sizes. The maximum value in this case for  $\tau_{GLR}$  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  $\mu$  and that the maximum of  $\tau_{GLR}$  is approximately 0.2 (-7dB).  $\tau_{GLR}$  for SBM graphs

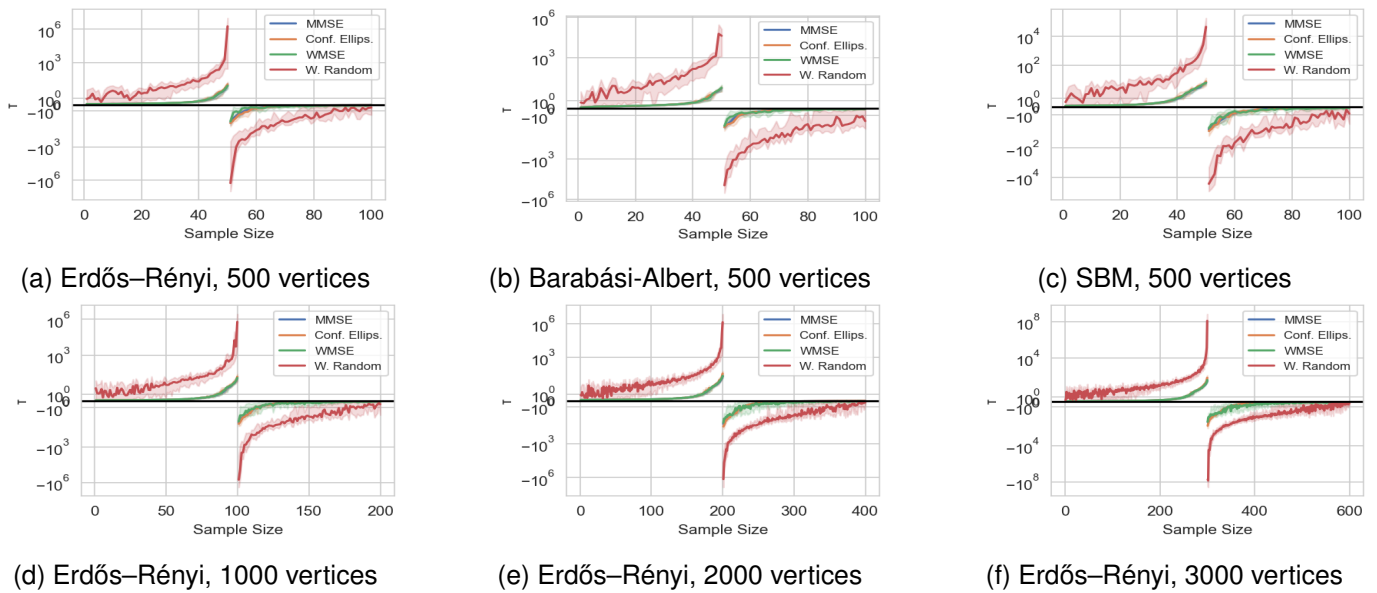


Fig. 2:  $\tau(\mathcal{S}, v)$  for different random graph models and different  $N$  under LS and full-band noise (bandwidth =  $\frac{\# \text{ vertices}}{10}$ ).

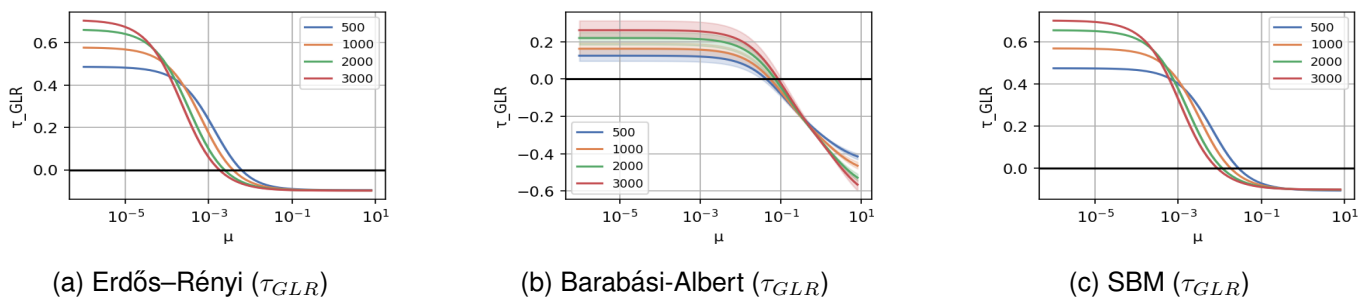


Fig. 3:  $\tau_{GLR}$  for different random graph models ( $\# \text{ vertices} = \text{colour}$ , bandwidth =  $\frac{\# \text{ vertices}}{10}$ ).

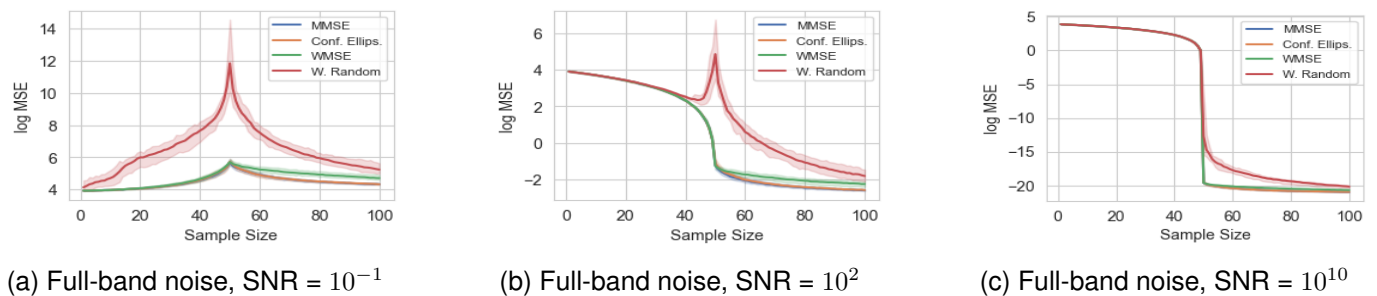


Fig. 4: Average MSE under LS on ER graphs ( $\# \text{ vertices} = 500$ , bandwidth = 50).

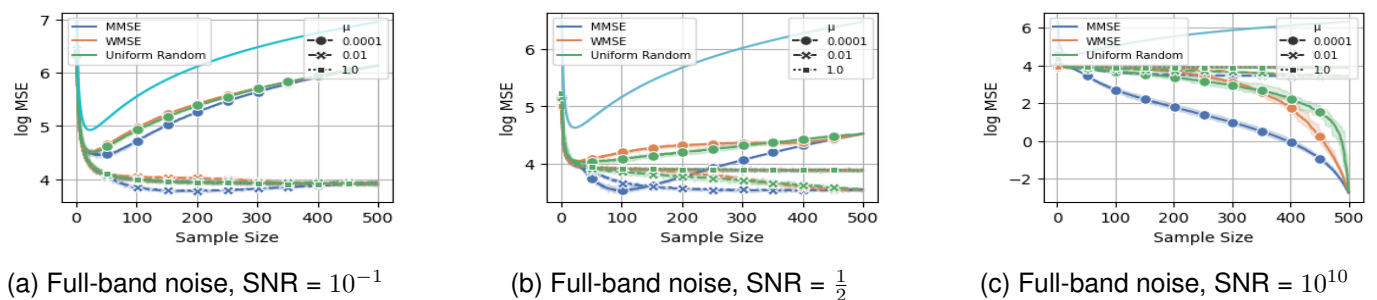


Fig. 5: Average MSE under GLR on ER graphs ( $\# \text{ vertices} = 500$ , bandwidth = 50). Line without markers is an upper bound.

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

Finally, Proposition 4 proves that  $\tau_{GLR} \not\rightarrow 0$  as  $N \rightarrow \infty$  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  $\mu$  as  $N$  increases. We note that Figs. 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.

Figs. 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  $\Lambda$ -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 Fig. 2a, Fig. 4a corresponds to when  $\text{SNR} < \tau(\mathcal{S}, v)$ , Fig. 4c corresponds to  $\text{SNR} > \tau(\mathcal{S}, v)$  and Fig. 4b corresponds to when  $\text{SNR}$  lies between some values of  $\tau(\mathcal{S}, v)$  for weighted random sampling. We see that MSE increases with sample size when  $\text{SNR} < \tau(\mathcal{S}, v)$  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  $\hat{x} = \mathbf{0}$  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  $\hat{x} = \mathbf{0}$  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. Figs. 5a-5c show how MSE changes with sample size for different values of  $\mu$  under full-band noise, along with an upper bound (light blue) which is not dependent on  $\mu$ . This bound is approximately U-shaped in all cases. For  $\text{SNR} = 10^{-1}$ , 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  $\text{SNR} = 10^{-1}$ , MSE is minimised at full observation ( $|\mathcal{S}| = 500$ ). For  $\text{SNR} = \frac{1}{2}$  and  $\mu = 10^{-4}$ , MSE is minimised at a sample size a bit less than 100. For larger  $\mu$ , 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.

Figs. 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  $\mu$  and low SNR experiments. Second, Figs. 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, Figs. 5a-5c illustrate the dependence on SNR and  $\mu$  in Theorem 3, i.e. at low  $\mu$  and SNR the optimal sample size is less than  $N$ , but at high enough  $\mu$  or SNR this no longer holds.

Figs. 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  $\mu$  and the composition of  $\mathcal{S}$ . This limitation corresponds to a gap between  $\tau_{GLR}$  and  $\tau(\mathcal{N}, \mathcal{S}^c)$ , demonstrating  $\tau_{GLR}$  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  $N \rightarrow \infty$ . However, the proposition does not say whether the conditions hold for a graph of a given size, or

TABLE VI: Probability theorem conditions are met.

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

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  $\text{SNR} \in \{10^{-2}, \frac{1}{2}\}$  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  $\tau$  and  $\tau_{GLR}$  for real-world datasets. In practice, one cannot know for sure what the signal model of a real-world signal is; however, computation of  $\tau$  is dependent on our choice of theoretical signal model. We have therefore computed  $\tau$  under the signal model assumptions given in Section II-A.

1)  $\tau$  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  $\tau$ . We validate the actual value of  $\tau$  with MSE experiments (Figs. 8 & 9).

2)  $\tau$  plots (GLR): By Theorem 3,  $\tau_{GLR}$  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,  $\tau_{GLR}$  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  $\tau_{GLR} < 0$  for all  $\mu$  is completely uninformative –  $\tau_{GLR}$  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, Figs. 8d and 8a show a  $\Lambda$ -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 Fig. 8a being  $\Lambda$ -shaped. In Fig. 6, all of the lines are under  $10^{10}$ . This corresponds to Fig. 8f and Fig. 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  $\Lambda$ -shaped and the other lines decreasing in Fig. 8b.

In Fig 6b all lines are below  $10^2$  and so we expect at  $\text{SNR} = 10^2$  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  $\tau$  for Weighted Random Sampling as the red line is  $\Lambda$ -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  $\tau$  for LS.

4) MSE plots (GLR): All four figures in Fig. 9 corresponding to  $\text{SNR} \leq \frac{1}{2}$  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  $\tau_{GLR}$  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 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  $\text{SNR} < \tau(\mathcal{S}, v)$ , and repeat until it is no longer possible to find a node  $v$  in our remaining set  $\mathcal{S}_{rem}$  where  $\text{SNR} < \tau(\mathcal{S}_{rem}, v)$ .

For the GLR reconstruction, with a fixed  $\mu$  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  $\tau$ ), 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  $\tau_{GLR}$ ;
- 3) If it is believed that  $\text{SNR} < \tau_{GLR}$  holds, sample  $m_{opt}$  (approximately  $\sqrt{N}$ ) nodes at random and reconstruct from those nodes.

We finally remark that future work includes extending the analysis on GLR to the normalised graph Laplacian, providing bounds on  $\xi_2$  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.

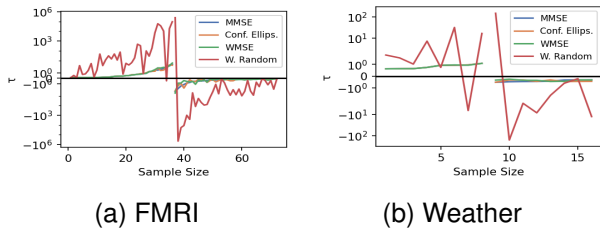


Fig. 6:  $\tau(\mathcal{S}, v)$  for real-world data under LS & full-band noise.

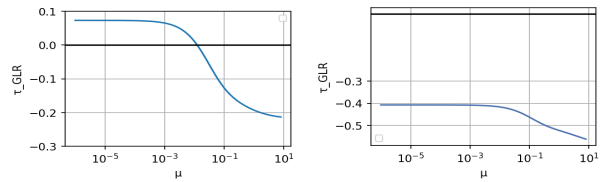


Fig. 7:  $\tau_{GLR}$  for real-world data.

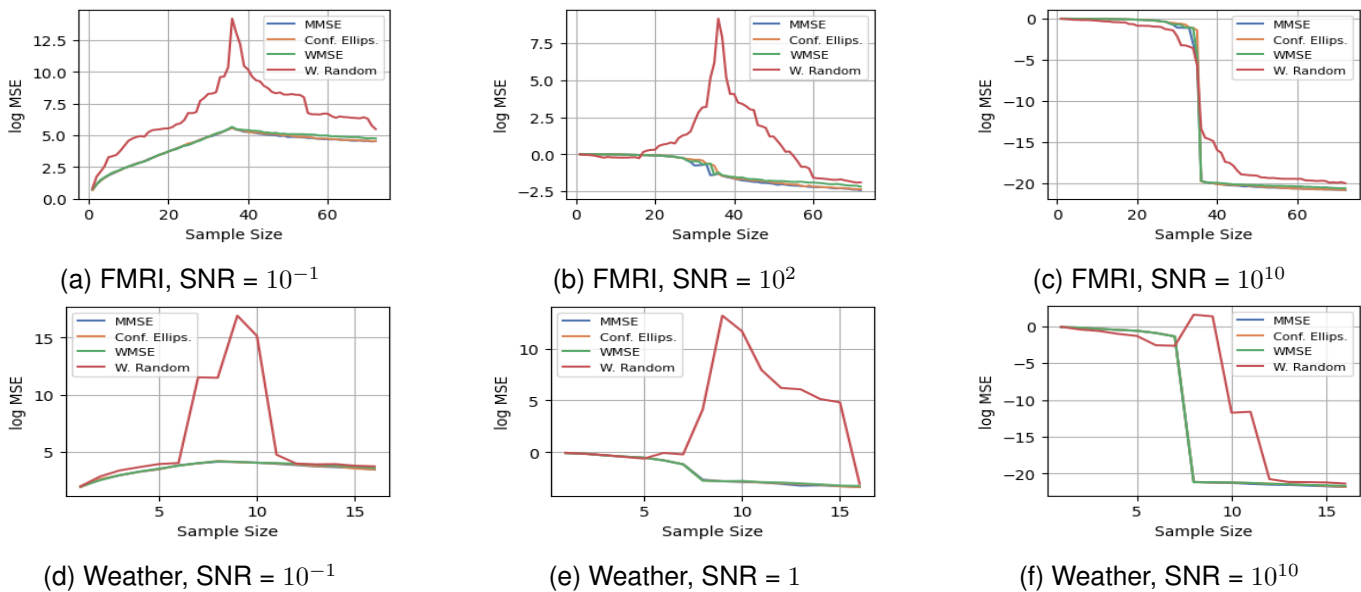


Fig. 8: Average MSE under LS on real-world data and full-band noise.

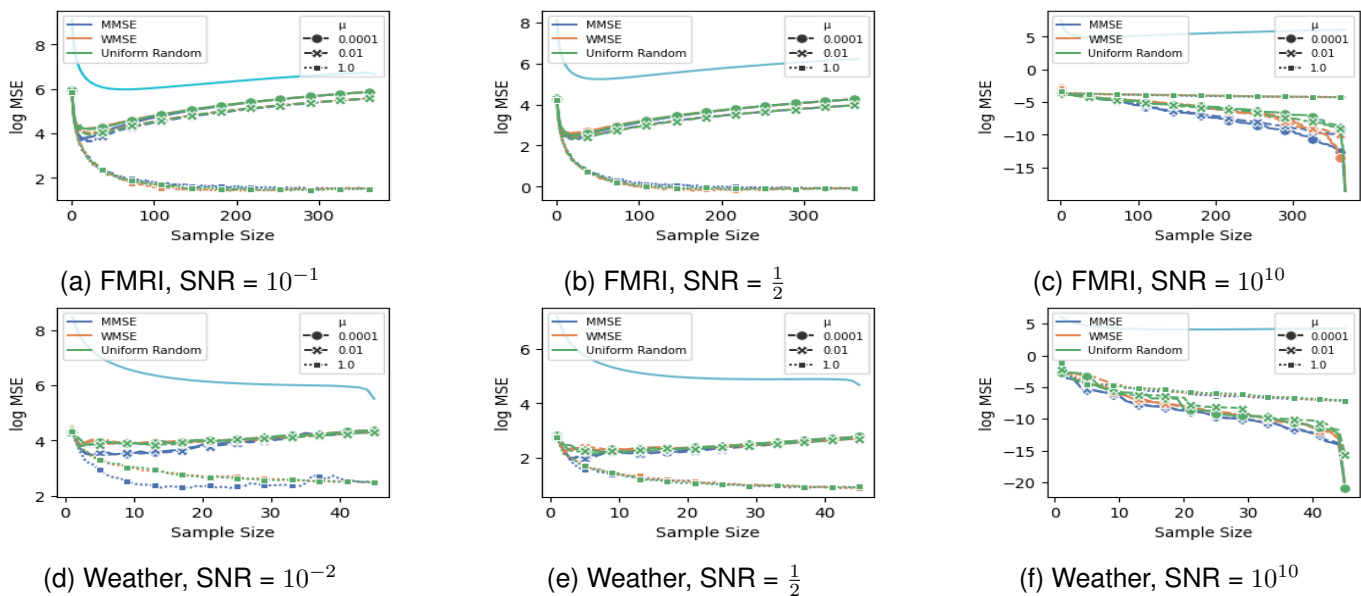


Fig. 9: Average MSE under GLR on real-world data and full-band noise. Line without markers is an upper bound.

## REFERENCES

- [1] S. Itani and D. Thanou, "A graph signal processing framework for the classification of temporal brain data," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1180–1184.
- [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 Processing Magazine*, vol. 31, no. 5, pp. 130–136, 2014.
- [3] B. Renoust, G. Cheung, and S. Satoh, "Estimating political leanings from mass media via graph-signal restoration with negative edges," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2017, pp. 1009–1014.
- [4] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.
- [5] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE transactions on signal processing*, vol. 63, no. 24, pp. 6510–6523, 2015.
- [6] Y. Tanaka, Y. C. Eldar, A. Ortega, and G. Cheung, "Sampling signals on graphs: From theory to applications," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 14–30, 2020.
- [7] F. Pukelsheim, *Optimal design of experiments*. SIAM, 2006.
- [8] A. Nikolov, M. Singh, and U. Tantipongpipat, "Proportional volume sampling and approximation algorithms for a-optimal design," *Mathematics of Operations Research*, vol. 47, no. 2, pp. 847–877, 2022.
- [9] L. F. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 34–47, 2017.
- [10] X. Xie, J. Yu, H. Feng, and B. Hu, "Bayesian design of sampling set for bandlimited graph signals," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2019, pp. 1–5.
- [11] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446–475, 2018.
- [12] H. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 933–936.
- [13] N. Tremblay, P.-O. Amblard, and S. Barthelmé, "Graph sampling with determinantal processes," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1674–1678.
- [14] F. Wang, Y. Wang, and G. Cheung, "A-optimal sampling and robust reconstruction for graph signals via truncated neumann series," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 680–684, 2018.
- [15] F. Wang, G. Cheung, and Y. Wang, "Low-complexity graph sampling with noise and signal reconstruction via neumann series," *IEEE Transactions on Signal Processing*, vol. 67, no. 21, pp. 5511–5526, 2019.
- [16] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, "Fast graph sampling set selection using gershgorin disc alignment," *IEEE Transactions on signal processing*, vol. 68, pp. 2419–2434, 2020.
- [17] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.
- [18] O. Nabar and G. Shroff, "Conservative predictions on noisy financial data," in *Proceedings of the Fourth ACM International Conference on AI in Finance*, 2023, pp. 427–435.
- [19] T. Bush, S. Bartington, F. D. Pope, A. Singh, G. N. Thomas, B. Stacey, G. Economides, R. Anderson, S. Cole, P. Abreu *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," *Building and Environment*, vol. 237, p. 110330, 2023.
- [20] B. Sripathmanathan, X. Dong, and M. Bronstein, "On the impact of sample size in reconstructing graph signals," in *2023 International Conference on Sampling Theory and Applications (SampTA)*. IEEE, 2023, pp. 1–6.
- [21] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Fundamental limits of sampling strategies," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 539–554, 2016.
- [22] S. Lin, X. Xie, H. Feng, and B. Hu, "Active sampling for approximately bandlimited graph signals," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5441–5445.
- [23] Y. Tanaka and Y. C. Eldar, "Generalized sampling on graphs with subspace and smoothness priors," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2272–2286, 2020.
- [24] J. Hara and Y. Tanaka, "Sampling set selection for graph signals under arbitrary signal priors," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 5732–5736.
- [25] F. Zhang, "Schur complements and matrix inequalities in the löwner ordering," *Linear Algebra and Its Applications*, vol. 321, no. 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 *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 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 Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.
- [29] M. Belkin and P. Niyogi, "Semi-supervised learning on riemannian manifolds," *Machine learning*, vol. 56, pp. 209–239, 2004.
- [30] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. John Hopkins University Press, 2013.
- [31] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," *Transactions of the American Mathematical Society*, 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 Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.
- [33] A. Jayawant and A. Ortega, "Practical graph signal sampling with log-linear size scaling," *Signal Processing*, vol. 194, p. 108436, 2022.
- [34] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural computation*, vol. 4, no. 1, pp. 1–58, 1992.
- [35] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, Massachusetts: MIT Press, 2016.
- [36] L. F. Chamon and A. Ribeiro, "Near-optimality of greedy set selection in the sampling of graph signals," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2016, pp. 1265–1269.
- [37] P.-Y. Chen and S. Liu, "Bias-variance tradeoff of graph laplacian regularization," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1118–1122, 2017.
- [38] R. Merris, "Laplacian graph eigenvectors," *Linear algebra and its applications*, vol. 278, no. 1-3, pp. 221–236, 1998.
- [39] C. Khatri and C. R. Rao, "Some generalizations of kantorovich inequality," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 91–102, 1982.
- [40] A. Householder, "The kantorovich and some related inequalities," *SIAM Review*, 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 Transactions on Signal and Information Processing over Networks*, vol. 9, pp. 304–314, 2023.
- [42] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.
- [43] A. Venkitaraman, S. Chatterjee, and P. Handel, "Gaussian processes over graphs," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5640–5644.
- [44] H. Behjat, U. Richter, D. Van De Ville, and L. Sörnmo, "Signal-adapted tight frames on graphs," *IEEE Transactions on Signal Processing*, vol. 64, no. 22, pp. 6017–6029, 2016.
- [45] R. Bhatia, *Matrix analysis*. Springer Science & Business Media, 2013, vol. 169.
- [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 and its applications*, vol. 259, pp. 209–221, 1997.
- [47] K. Nordström, "Convexity of the inverse and moore–penrose inverse," *Linear algebra and its applications*, vol. 434, no. 6, pp. 1489–1512, 2011.
- [48] M. Barahona and L. M. Pecora, "Synchronization in small-world systems," *Physical review letters*, vol. 89, no. 5, p. 054101, 2002.
- [49] T. Jiang, "Low eigenvalues of laplacian matrices of large random graphs," *Probability Theory and Related Fields*, vol. 153, pp. 671–690, 2012.
- [50] J. K. Baksalary and S. Puntanen, "Generalized matrix versions of the cauchy-schwarz and kantorovich inequalities," *Aequationes Mathematicae*, vol. 41, pp. 103–110, 1991.

APPENDIX A  
THEORETICAL RESULTS FOR LS UNDER  $k$ -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 this subsection, by showing that MSE can decrease when sample size decreases under LS with  $k$ -bandlimited noise, we demonstrate that this is not the case.

1) *Simplification*: For coherence with the full-band case, we use the single vertex simplification.

2) *Characterisation*: We first show that under  $k$ -bandlimited noise the choice of sample set  $\mathcal{S}$  only influences the MSE through  $\text{rank}([U]_{\mathcal{S},\mathcal{K}})$ .

**Lemma 2.**  $\text{MSE}_{\mathcal{S}} = k + (\sigma^2 - 1)\text{rank}([U]_{\mathcal{S},\mathcal{K}})$ .

Lemma 2 makes it easy to prove the following variants of Corollary 1.1 and Theorem 2 for bandlimited noise.

**Corollary 3.2.** Let  $v \in \mathcal{S}$ , then  $\mathcal{S} \setminus \{v\}$  is as good or better than  $\mathcal{S}$  if and only if

$$\text{SNR} \leq 1. \quad (48)$$

We can understand this as being like Corollary 1.1 with  $\tau_{LS\_bl} = 1$ . The criterion in Corollary 3.2 does not depend on vertex choice, unlike Corollary 1.1. Therefore under LS and bandlimited noise, whether  $\mathcal{S} \setminus \{v\}$  is as good or better than  $\mathcal{S}$  is not contingent on which vertices are in  $\mathcal{S}$ .

3) *Existence*: Next, our variant of Theorem 2 for bandlimited noise concerns when MSE changes at all, rather than when it reduces.

**Corollary 3.3.** Suppose we use a sequential noiseless-optimal scheme to select a vertex sample set  $\mathcal{S}_m$  of size  $m$ . For  $m \leq k$ :

$$\forall v \in \mathcal{S}_m : \tau(\mathcal{S}_m, v) = \tau_{LS\_bl} = 1, \quad (49)$$

i.e., for any  $v \in \mathcal{S}_m$ ,  $\mathcal{S}_m \setminus \{v\}$  is better than  $\mathcal{S}_m$  if and only if  $\text{SNR} < \tau_{LS\_bl}$ . For  $m > k$ :

$$\forall \text{SNR}, \forall v \in \mathcal{S}_m : \text{MSE}_{\mathcal{S}_m} = \text{MSE}_{\mathcal{S}_m \setminus \{v\}} \quad (50)$$

We prove Lemma 2 and Corollaries 3.2 & 3.3 in Appendix J.

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

**Remark 10.** As  $\tau_{LS\_bl}$  is not a function of the graph, we find that at a fixed  $\text{SNR} < \tau_{LS\_bl} = 1$ , for graphs of any size (even arbitrarily large), we can reduce sample size to improve MSE.

APPENDIX B  
THEORETICAL RESULTS FOR GLR UNDER  
 $k$ -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 to disprove this.

1) *Simplification*: As with the full-band GLR case, we consider when a set  $\mathcal{S} \subset \mathcal{N}$  is better than  $\mathcal{N}$ .

2) *Characterisation*: We first note that Corollary 2.1 also applies to the bandlimited case with the appropriate definition of  $\Delta_2$ , providing a full characterisation of when a set  $\mathcal{S}$  is better than  $\mathcal{N}$  under GLR reconstruction with bandlimited noise.

3) *Existence*: We now present a bandlimited variant of Lemma 1. It can be understood as Lemma 1 with  $\lambda_N$  replaced by  $\lambda_k$ .

**Lemma 3.** Let  $\lambda_i$  be the eigenvalues of  $\mathbf{L}$  and let  $\omega$  be defined as in Lemma 1. Let

$$r_{bl} = \omega \left( \frac{\lambda_k}{\lambda_2} \right) \quad (51)$$

$$B_k(m) = r_{bl} \frac{N}{m} + \sum_{i=2}^m \omega \left( \max \left[ 1, \frac{\lambda_{k+2-i}}{\lambda_i} \right] \right) \quad (52)$$

where the summation is 0 if  $m = 0$  or 1. Then

$$\xi_{2,bl}(\mathcal{S}) \leq 1 + B_k(m) \leq 1 + r_{bl} \left( \frac{N}{m} + m - 1 \right). \quad (53)$$

*Proof.* We prove a bandlimited variant of the Kantorovich inequality and use the same structure as the proof of Lemma 1. See Appendix U.  $\square$

A version of Theorem 3 follows in the same manner as with full-band noise:

**Theorem 4.** Let  $B_k(m)$  be defined as in Lemma 3. Let  $m_{opt\_bl}$  be the sample size minimising  $B_k(m)$ .

$$\text{If } B_k(m_{opt\_bl}) < k - 1 \quad (54)$$

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

$$\text{if } \text{SNR} < \tau_{GLR\_bl}(\mu) \quad (55)$$

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

*Proof.* Apply Lemma 3 and follow similar proof steps to Theorem 3. See Appendix V for details.  $\square$

Given the similarity of the bounds and proof approaches between Theorems 3 and 4, the same analysis of parameters applies to Theorem 4 as Theorem 3. Thus we expect the parameters in Theorem 4 –  $m_{opt\_bl}$ ,  $B_k(m_{opt\_bl})$ ,  $\mu_{ub\_bl}$  and  $\tau_{GLR\_bl}$  to behave similarly to their non-bandlimited counterparts. This is reinforced by the asymptotics, which we will see in Proposition 5.

We again present an MSE upper bound to link our experiments in Section IV to Fig 1:

**Corollary 4.1.** Let  $B_k(m)$  be defined as in Lemma 3. For a sample set  $\mathcal{S}$  of size  $m$ ,

$$\text{MSE}_{\mathcal{S}} \leq (k - 1) + (1 + \sigma^2) \cdot (1 + B_k(m)). \quad (56)$$

*Proof.* By Lemma 3,  $\xi_2(\mathcal{S}) \leq B(m)$ . By Lemma 9 in Appendix P,  $\xi_1(\mathcal{S}) \leq (k - 1) + \xi_{2,bl}(\mathcal{S}) \leq k + B_k(m)$ . Combining these using (16) gives the desired bound.  $\square$

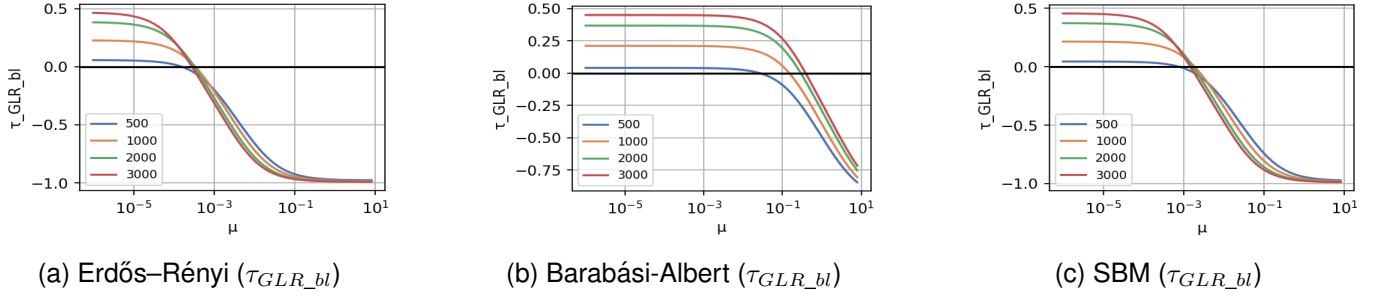


Fig. 10:  $\tau_{GLR\_bl}$  for different random graph models (#vertices = colour, bandwidth =  $\frac{\# \text{vertices}}{10}$ ).

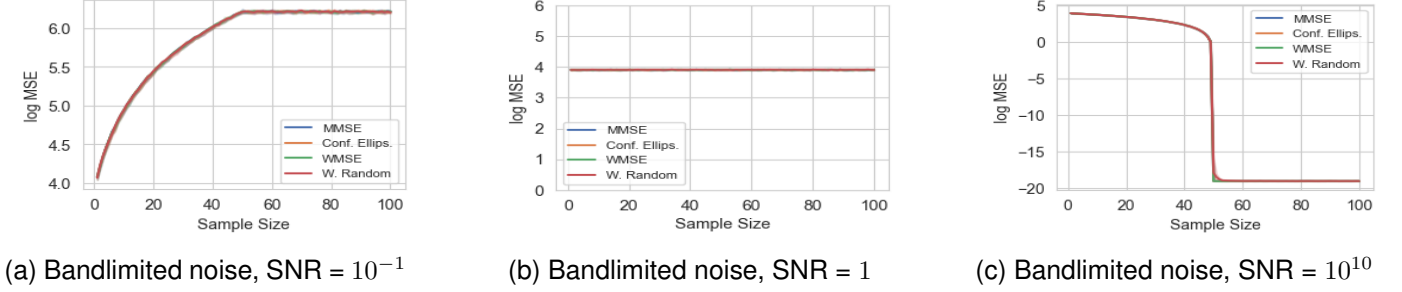


Fig. 11: Average MSE under LS on ER graphs (#vertices=500, bandwidth = 50).

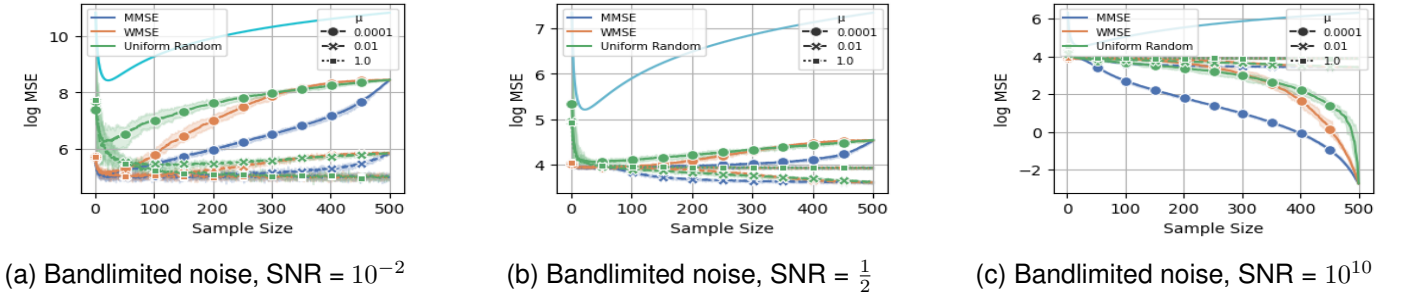


Fig. 12: Average MSE under GLR on ER graphs (#vertices=500, bandwidth = 50). Line without markers is an upper bound.

4) *Asymptotics*: Finally, we consider the case of large graphs. We prove a variant of Proposition 4.

**Proposition 5.** As  $N \rightarrow \infty$ , condition (54) in Theorem 4 holds w.h.p. . Furthermore,

$$r \xrightarrow{P} 1, \quad m_{opt\_bl} \cdot N^{-\frac{1}{2}} \xrightarrow{P} 1, \quad \mu_{ub\_bl} \lambda_2 \xrightarrow{P} +\infty. \quad (57)$$

Assume  $\frac{k}{N}$  is fixed and choose  $\mu = \frac{c}{\lambda_2}$ , or  $\frac{c}{\lambda_N}$ , or  $\frac{c}{\sqrt{\lambda_2 \lambda_N}}$  for optimal bias-variance trade-off at  $\mathcal{S} = \mathcal{N}$  [37], then

$$\tau_{GLR\_bl} \rightarrow (1 + 2c)^{-1}. \quad (58)$$

*Proof.* See Appendix W.  $\square$

We see that the parameters asymptotically behave the same as the non-bandlimited parameters for large Erdős-Rényi graphs.

## APPENDIX C

### EXPERIMENTAL RESULTS UNDER $k$ -BANDLIMITED NOISE

In this section we provide results on synthetic datasets under  $k$ -bandlimited noise.

#### A. $\tau$ and $\tau_{GLR}$ plots

1) *LS*: We do not provide a plot as in this case, as our theoretical results provide the magnitude of  $\tau$ : by Corollary 3.3,  $\tau(\mathcal{S}, v) = 1$  for  $|\mathcal{S}| \leq k$  under sequential noiseless-optimal sampling. Corollary 3.3 overall says something stronger, i.e. we can always reduce MSE by reducing sample size if  $\text{SNR} < 1$  at any sample size.

2) *GLR*: Our purpose in showing bandlimited variants is to show that reducing sample size can reduce MSE even under bandlimited noise. Figs. 10a- 10c show the same overall trend as Figs. 3a-3c, i.e.  $\tau_{GLR\_bl}$  can be positive for ER, BA and SBM graphs, giving evidence to our claim. As with the full-band case, Fig. 10a validates our asymptotic result for ER graphs (Proposition 5).

#### B. MSE plots

The MSE plots demonstrate the validity of our theoretical results linking MSE and sample size.

1) *LS*: Figs. 11a-11c show how MSE behaves as sample size increases. Fig. 11a shows that under high noise, MSE

increases with sample size when sample size is no larger than bandwidth, and is constant beyond that. Fig. 11b shows at an SNR of  $\tau_{LS,bl} = 1$ , MSE is constant with sample size. Finally, Fig. 11c shows that in the almost noiseless case MSE decreases with sample size. Fig. 11a demonstrates Corollary 3.2 by showing that for  $\text{SNR} < \tau_{LS,bl} = 1$  the MSE is increasing with sample size. In all cases, MSE remains unchanged for sample sizes exceeding the bandlimit  $k$ , demonstrating Corollary 3.3.

2) *GLR*: Our observations about Figs. 5a-5c and Theorem 3 also apply to Figs. 12a-12c and Theorem 4; specifically, for sufficiently small  $\mu$  and SNR, MSE is minimised at sample sizes well below  $N$  under bandlimited noise and  $\tau_{GLR,bl}$  is a lower bound for  $\tau(\mathcal{N}, \mathcal{S}^c)$  rather than an exact characterisation.

### C. Checking conditions

As in the full-band case, our theorems on GLR rely on conditions around graph invariants. We sample graphs from each random graph model to empirically show the probability the conditions of Theorem 4 are met at a sample size of  $m_{opt}$  for some  $\mu > 0$ :

TABLE VII: Probability theorem conditions are met.

	ER	SBM	BA
Theorem 4 conditions met	100%	100%	99.4%

As with full-band noise, Proposition 5 shows the conditions in Theorem 4 hold w.h.p. for ER graphs as  $N \rightarrow \infty$ , and Table VII shows empirically that the conditions hold under  $k$ -bandlimited noise (Theorem 3) for all ER and SBM graphs tested with 500 vertices, and most BA graphs tested.

### APPENDIX D BIAS-VARIANCE DECOMPOSITION

We present a proof of the Bias-Variance decomposition we use. We assume that  $\mathbf{x}$  and  $\epsilon$  are independent. Let  $\hat{\mathbf{x}}$  be any reconstruction of the signal  $\mathbf{x}$ , then

$$\text{MSE}_{\mathcal{S}} = \mathbb{E}_{\mathbf{x}} \left[ \mathbb{E}_{\epsilon} \left[ \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \right] \right] \quad (59)$$

$$= \mathbb{E}_{\mathbf{x}} \left[ \mathbb{E}_{\epsilon} \left[ \|\mathbf{x} - \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] + \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \hat{\mathbf{x}}\|_2^2 \right] \right] \quad (60)$$

$$= \mathbb{E}_{\mathbf{x}} \left[ 2\mathbb{E}_{\epsilon} \left[ (\mathbf{x} - \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}])^T (\mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \hat{\mathbf{x}}) \right] \right] \quad (61)$$

$$+ \mathbb{E}_{\mathbf{x}} \left[ \|\mathbf{x} - \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}]\|_2^2 + \mathbb{E}_{\epsilon} \left[ \|\mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \hat{\mathbf{x}}\|_2^2 \right] \right]. \quad (62)$$

Note that by the properties of expectation,

$$\mathbb{E}_{\epsilon} \left[ \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}]^T \hat{\mathbf{x}} \right] = \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}]^T \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] = \mathbb{E}_{\epsilon} \left[ \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}]^T \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] \right] \quad (63)$$

$$\text{so} \quad \mathbb{E}_{\epsilon} \left[ \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}]^T (\mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \hat{\mathbf{x}}) \right] = 0 \quad (64)$$

and as the value of the signal  $\mathbf{x}$  is independent to the value of  $\epsilon$ ,

$$\mathbb{E}_{\epsilon} \left[ \mathbf{x}^T (\mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \hat{\mathbf{x}}) \right] = \mathbf{x}^T \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \mathbf{x}^T \hat{\mathbf{x}} = 0 \quad (65)$$

$$\mathbb{E}_{\epsilon} \left[ (\mathbf{x} - \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}])^T (\mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \hat{\mathbf{x}}) \right] = 0 \quad (66)$$

for any value of  $\mathbf{x}$ . Therefore

$$\text{MSE}_{\mathcal{S}} = \mathbb{E}_{\mathbf{x}} \left[ \underbrace{\|\mathbf{x} - \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}]\|_2^2}_{\text{Bias}(\hat{\mathbf{x}}, \mathbf{x})^2} + \underbrace{\mathbb{E}_{\epsilon} \left[ \|\mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \hat{\mathbf{x}}\|_2^2 \right]}_{\text{Var}(\hat{\mathbf{x}})} \right]. \quad (67)$$

Note that if  $\mathbf{x}$  was fixed (which can be understood as being drawn from a distribution with one possible value),  $\epsilon$  and  $\mathbf{x}$  must be independent, and therefore

$$\text{MSE}_{\mathcal{S}}|_{\mathbf{x}} = \mathbb{E}_{\epsilon} \left[ \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \mid \mathcal{S} \text{ observed} \right] \quad (68)$$

$$= \underbrace{\|\mathbf{x} - \mathbb{E}_{\epsilon}[\hat{\mathbf{x}}]\|_2^2}_{\text{Bias}(\hat{\mathbf{x}}, \mathbf{x})^2} + \underbrace{\mathbb{E}_{\epsilon} \left[ \|\mathbb{E}_{\epsilon}[\hat{\mathbf{x}}] - \hat{\mathbf{x}}\|_2^2 \right]}_{\text{Var}(\hat{\mathbf{x}})}. \quad (69)$$

### APPENDIX E EXTENDING RESULTS TO OTHER SETTINGS - PROOF OF PROPOSITION 1

In this proof we refer to the distribution of  $\mathbf{x}$  as the signal model and the distribution of  $\epsilon$  as the noise model.

We apply Theorem 1; we note that all results in Section III-B don't require  $\mathbb{E}[\mathbf{x}] = 0$ , so apply when  $\mathbf{x}$  is a fixed nonzero signal. By (15),

$$\Delta_2(\mathcal{S}, \mathcal{T}) = \xi_2(\mathcal{S}) - \xi_2(\mathcal{S} \setminus \mathcal{T}) \quad (70)$$

$$= \|\mathbf{R}_{\mathcal{S}}[\epsilon]_{\mathcal{S}}\|_F^2 - \|\mathbf{R}_{\mathcal{S}}[\epsilon]_{\mathcal{S} \setminus \mathcal{T}}\|_F^2 \quad (71)$$

therefore  $\Delta_2(\mathcal{S}, \mathcal{T})$  is not a function of the signal model.

Pick a signal model  $\mathbf{x}$  with  $\mathbb{E}_{\mathbf{x}} \left[ \|\mathbf{x}\|_2^2 \right] < \infty$ . Then let

$$\tau'(\mathcal{S}, \mathcal{T}) = \begin{cases} \frac{\mathbb{E} \left[ \|\mathbf{x}\|_2^2 \right]}{\mathbb{E} \left[ \|\epsilon\|_2^2 \right]} \cdot \frac{\Delta_2(\mathcal{S}, \mathcal{T})}{-\Delta_1(\mathcal{S}, \mathcal{T})} & \text{if } \Delta_1 < 0 \\ \infty & \text{otherwise} \end{cases} \quad (72)$$

As  $\Delta_2(\mathcal{S}, \mathcal{T}) > 0$ ,  $\tau'(\mathcal{S}, \mathcal{T}) > 0$ .

We now apply Theorem 1. We look at the three cases

1)  $\Delta_1 < 0$ : In this case,  $\tau'$  is  $\tau$  in Theorem 1, and the Theorem directly gives the proposition.

2)  $\Delta_1 > 0$ : This case in Theorem 1 results in negative  $\tau$ , and thus  $\text{SNR} \geq 0 > \tau$  is always fulfilled, and is equivalent to  $\text{SNR} < \infty$ . Thus we set  $\tau' = \infty$ .

3)  $\Delta_1 = 0$ : As  $\Delta_2 > 0$  by assumption, this case always holds and reduces to the case where  $\Delta_1 > 0$ .

Therefore under this new signal model if  $\text{SNR} < \tau'(\mathcal{S}, \mathcal{T})$ ,  $\text{MSE}_{\mathcal{S}} > \text{MSE}_{\mathcal{S} \setminus \mathcal{T}}$ .

### APPENDIX F PROOF OF TABLE IV

For GLR reconstruction, Table IVb does not rule out any option. In experiments with Erdős-Rényi graphs, all options marked as  $\checkmark$  in Figure IVb can be observed at some sample size without too much difficulty under random or WMSE sampling.

The options marked  $\sim$  do not turn up in such experiments, but may happen for very specific values of  $\mu$  dependent on the graph and sampling scheme.

For LS reconstruction, we decompose the pattern in Table IVa into the following statements:

- $\Delta_1 \leq 0$
- $\Delta_1 < 0$  if and only if  $\Delta_2 > 0$

A. Under LS reconstruction,  $\Delta_1 \leq 0$

For LS we have:

$$\mathbf{R}_S = [\mathbf{U}]_{\mathcal{N}, \mathcal{K}} [\mathbf{U}]_{S, \mathcal{K}}^\dagger.$$

**Lemma 4.** For LS,

$$\xi_1(\mathcal{S}) = k - \text{rank}([\mathbf{U}]_{S, \mathcal{K}}), \quad (73)$$

$$\Delta_1(\mathcal{S}, v) \in \{0, -1\}. \quad (74)$$

*Proof.* As  $\|[\mathbf{U}]_{\mathcal{N}, \mathcal{K}} \mathbf{A}\|_F^2 = \|\mathbf{A}\|_F^2$  for any matrix  $\mathbf{A} \in \mathbb{R}^{k \times k}$ ,

$$\xi_1(\mathcal{S}) = \|[\mathbf{U}]_{\mathcal{N}, \mathcal{K}} - \mathbf{R}_S [\mathbf{U}]_{S, \mathcal{K}}\|_F^2 \quad (75)$$

$$= \|[\mathbf{U}]_{\mathcal{N}, \mathcal{K}} - [\mathbf{U}]_{\mathcal{N}, \mathcal{K}} ([\mathbf{U}]_{S, \mathcal{K}})^\dagger [\mathbf{U}]_{S, \mathcal{K}}\|_F^2 \quad (76)$$

$$= \|\mathbf{I}_k - ([\mathbf{U}]_{S, \mathcal{K}})^\dagger [\mathbf{U}]_{S, \mathcal{K}}\|_F^2 \quad (77)$$

Let  $\mathbf{\Pi} = ([\mathbf{U}]_{S, \mathcal{K}})^\dagger [\mathbf{U}]_{S, \mathcal{K}}$ .  $\mathbf{\Pi}$  is of the form  $\mathbf{A}^\dagger \mathbf{A}$ , so is a symmetric orthogonal projection onto the range of  $([\mathbf{U}]_{S, \mathcal{K}})^T$  [30, p. 290]. Orthogonal projections are idempotent ( $\mathbf{\Pi} = \mathbf{\Pi}^2$ ) hence have eigenvalues which are 0 or 1, and therefore  $\text{tr}(\mathbf{\Pi}) = \text{rank}([\mathbf{U}]_{S, \mathcal{K}}) = \text{rank}([\mathbf{U}]_{S, \mathcal{K}})$ . We then have:

$$\xi_1(\mathcal{S}) = \|\mathbf{I}_k - \mathbf{\Pi}\|_F^2 \quad (78)$$

$$= \text{tr}((\mathbf{I}_k - \mathbf{\Pi})(\mathbf{I}_k - \mathbf{\Pi})^T) \quad (79)$$

$$= \text{tr}((\mathbf{I}_k - \mathbf{\Pi})(\mathbf{I}_k - \mathbf{\Pi})) \quad (80)$$

$$= \text{tr}(\mathbf{I}_k - 2\mathbf{\Pi} + \mathbf{\Pi}^2) \quad (81)$$

$$= \text{tr}(\mathbf{I}_k - \mathbf{\Pi}) \quad (82)$$

$$= \text{tr}(\mathbf{I}_k) - \text{tr}(\mathbf{\Pi}) \quad (83)$$

$$= k - \text{rank}([\mathbf{U}]_{S, \mathcal{K}}) \quad (84)$$

proving (73). We now prove (74). Removing a vertex from  $\mathcal{S}$  removes a row from  $[\mathbf{U}]_{S, \mathcal{K}}$ , reducing the rank by 0 or 1, so

$$\Delta_1(\mathcal{S}, v) = \xi_1(\mathcal{S}) - \xi_1(\mathcal{S} \setminus \{v\}) \quad (85)$$

$$= -\text{rank}([\mathbf{U}]_{S, \mathcal{K}}) + \text{rank}([\mathbf{U}]_{\mathcal{S} \setminus \{v\}, \mathcal{K}}) \quad (86)$$

$$\in \{0, -1\}. \quad (87)$$

□

Non-positivity of  $\Delta_1$  immediately follows from Lemma 4

B. Under LS reconstruction,  $\Delta_1 < 0$  if and only if  $\Delta_2 > 0$

We first need the following lemmas.

**Lemma 5.**

$$\xi_2(\mathcal{S}) = \sum_{\lambda_i^S \neq 0} \frac{1}{\lambda_i^S} \quad (88)$$

where  $\lambda_i^S$  is the  $i^{\text{th}}$  eigenvalue of  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_S$ .

*Proof.* As

$$\xi_2(\mathcal{S}) = \|\mathbf{R}_S\|_F^2 = \|[\mathbf{U}]_{\mathcal{N}, \mathcal{K}} [\mathbf{U}]_{S, \mathcal{K}}^\dagger\|_F^2 = \|[\mathbf{U}]_{S, \mathcal{K}}^\dagger\|_F^2 \quad (89)$$

$\xi_2(\mathcal{S})$  is the sum of the squares of the singular values of  $([\mathbf{U}]_{S, \mathcal{K}})^\dagger$  [30, Corollary 2.4.3]. The pseudoinverse maps

the singular values of  $[\mathbf{U}]_{S, \mathcal{K}}$  onto the singular values of  $([\mathbf{U}]_{S, \mathcal{K}})^\dagger$  in the following way [30, Section 5.5.2]:

$$\sigma_i(([\mathbf{U}]_{S, \mathcal{K}})^\dagger) = \begin{cases} 0 & \text{if } \sigma_i([\mathbf{U}]_{S, \mathcal{K}}) = 0 \\ \sigma_i([\mathbf{U}]_{S, \mathcal{K}})^{-1} & \text{otherwise} \end{cases} \quad (90)$$

and the squares of the singular values of  $[\mathbf{U}]_{S, \mathcal{K}}$  are  $\lambda_i$  [30, Eq. (8.6.1)]. As  $[\mathbf{U}]_{S, \mathcal{K}} [\mathbf{U}]_{S, \mathcal{K}}^T = [\mathbf{\Pi}_{bl(\mathcal{K})}]_S$ , summing the singular values gives the result. □

**Lemma 6.**

$$\text{rank}([\mathbf{\Pi}_{bl(\mathcal{K})}]_S) = \text{rank}([\mathbf{U}]_{S, \mathcal{K}}) \leq k. \quad (91)$$

*Proof.* Remember that  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_S = [\mathbf{U}]_{S, \mathcal{K}} [\mathbf{U}]_{S, \mathcal{K}}^T$ .

The equality: The number of strictly positive singular values of a matrix is its rank [30, Corollary 2.4.6] and both  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_S = [\mathbf{U}]_{S, \mathcal{K}} [\mathbf{U}]_{S, \mathcal{K}}^T$  and  $[\mathbf{U}]_{S, \mathcal{K}}$  have the same number of strictly positive singular values [30, Eq. (8.6.2)].

The inequality:  $[\mathbf{U}]_{S, \mathcal{K}}$  has  $k$  columns and so column  $\text{rank}([\mathbf{U}]_{S, \mathcal{K}}) \leq k$  and rank equals column rank. □

We can now prove the overall result:

**Lemma 7.** For LS,  $\Delta_1 < 0$  if and only if  $\Delta_2 > 0$ .

*Proof.* As  $\Delta_1 \in \{0, -1\}$  (Lemma 4), we instead prove that  $\Delta_1 = 0$  if and only if  $\Delta_2 \leq 0$ .

Write the eigenvalues of  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{S \setminus \{v\}}$  as  $\lambda_1, \dots, \lambda_n$  and the eigenvalues of  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_S$  as  $\mu_1, \dots, \mu_{n+1}$ . As  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{S \setminus \{v\}}$  is a principal submatrix of  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_S$ , by Cauchy's Interlacing Theorem [45, p. 59],

$$0 \leq \mu_1 \leq \lambda_1 \leq \dots \leq \lambda_n \leq \mu_{n+1} \leq 1 \quad (92)$$

where the outer bounds come from the fact that both matrices are principal submatrices of  $\mathbf{\Pi}_{bl(\mathcal{K})}$ , an orthogonal projection matrix.

1)  $\Delta_1 = 0 \implies \Delta_2 \leq 0$ :  $\Delta_1 = 0$  implies  $\text{rank}([\mathbf{U}]_{S, \mathcal{K}}) = \text{rank}([\mathbf{U}]_{S \setminus \{v\}, \mathcal{K}})$ , so  $\text{rank}([\mathbf{\Pi}_{bl(\mathcal{K})}]_S) = \text{rank}([\mathbf{\Pi}_{bl(\mathcal{K})}]_{S \setminus \{v\}})$ . As the rank is unchanged,  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_S$  has one more zero-eigenvalue than  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{S \setminus \{v\}}$ . This means:

$$\mu_1 = 0 \quad (93)$$

$$\lambda_i = 0 \iff \mu_{i+1} = 0 \quad (94)$$

By Cauchy's Interlacing Theorem,  $\lambda_i \leq \mu_{i+1}$  and so

$$\frac{1}{\lambda_i} \geq \frac{1}{\mu_{i+1}} \text{ if } \lambda_i \neq 0 \text{ and } \mu_{i+1} \neq 0. \quad (95)$$

Therefore

$$\sum_{\lambda_i^S \neq 0} \frac{1}{\lambda_i^S} \geq \sum_{\mu_i^S \neq 0} \frac{1}{\mu_i^S} \quad (96)$$

as we have the same number of non-zero terms in each of these terms by (93) and (94), and the inequality is proved by summing over the non-zero terms using (95). Equation (96) is exactly

$$\xi_2(\mathcal{S} \setminus \{v\}) \geq \xi_2(\mathcal{S}). \quad (97)$$

Rearranging gives  $\Delta_2 \leq 0$ .

2)  $\Delta_1 = 0 \iff \Delta_2 \leq 0$ : We prove the equivalent statement

$$\Delta_1 \neq 0 \implies \Delta_2 > 0. \quad (98)$$

By Lemma 4, if  $\Delta_1 \neq 0$  then  $\Delta_1 = -1$ . This means that  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) - 1 = \text{rank}([\mathbf{U}]_{\mathcal{S} \setminus \{v\}, \mathcal{K}})$ , therefore  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$  has one more non-zero eigenvalue than  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S} \setminus \{v\}}$ . This means:

$$\mu_{n+1} > 0 \quad (99)$$

$$\lambda_i \neq 0 \iff \mu_i \neq 0 \quad (100)$$

By Cauchy's interlacing theorem,  $\lambda_i \geq \mu_i$  and so

$$\frac{1}{\lambda_i} \leq \frac{1}{\mu_i} \text{ if } \lambda_i \neq 0 \text{ and } \mu_i \neq 0. \quad (101)$$

Let  $I$  be the number of zero eigenvalues of  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$ . Then

$$\sum_{I \leq i \leq n} \frac{1}{\lambda_i^{\mathcal{S}}} \leq \sum_{I \leq i \leq n} \frac{1}{\mu_i^{\mathcal{S}}} < \sum_{I \leq i \leq n+1} \frac{1}{\mu_i^{\mathcal{S}}}. \quad (102)$$

With the left inequality by matching terms via (100) and then summing over (101), and the right inequality because (99) means  $\frac{1}{\mu_{n+1}} > 0$ . We then note the left and the right terms in this inequality say:

$$\sum_{\lambda_i^{\mathcal{S}} \neq 0} \frac{1}{\lambda_i^{\mathcal{S}}} < \sum_{\mu_i^{\mathcal{S}} \neq 0} \frac{1}{\mu_i^{\mathcal{S}}} \quad (103)$$

or equivalently,

$$\xi_2(\mathcal{S} \setminus \{v\}) < \xi_2(\mathcal{S}). \quad (104)$$

Rearranging gives  $\Delta_2 > 0$ .  $\square$

#### APPENDIX G PROOF OF THEOREM 1

*Proof.*  $\mathcal{S} \setminus \mathcal{T}$  is better than  $\mathcal{S}$  if and only if  $\text{MSE}_{\mathcal{S} \setminus \mathcal{T}} < \text{MSE}_{\mathcal{S}}$ . By (20) this happens if and only if

$$\Delta_1(\mathcal{S}, \mathcal{T}) + \sigma^2 \cdot \Delta_2(\mathcal{S}, \mathcal{T}) > 0. \quad (105)$$

By substituting in  $\sigma^2 = \frac{\mathbb{E}[\|\mathbf{x}\|_2^2]}{\mathbb{E}[\|\boldsymbol{\epsilon}\|_2^2]} \cdot \frac{1}{\text{SNR}}$  and multiplying both sides by SNR (which does not change the direction of the inequality, as  $\text{SNR} > 0$ ),  $\mathcal{S} \setminus \mathcal{T}$  is better than  $\mathcal{S}$  if and only if

$$\frac{\mathbb{E}[\|\mathbf{x}\|_2^2]}{\mathbb{E}[\|\boldsymbol{\epsilon}\|_2^2]} \cdot \Delta_2(\mathcal{S}, \mathcal{T}) > -\Delta_1(\mathcal{S}, \mathcal{T}) \cdot \text{SNR}. \quad (106)$$

We consider the conditions of (23):

(23a):  $\Delta_1(\mathcal{S}, \mathcal{T}) < 0$

We can divide both sides of (106) by  $-\Delta_1(\mathcal{S}, \mathcal{T})$  without changing the inequality, so (106) holds if and only if

$$\frac{\mathbb{E}[\|\mathbf{x}\|_2^2]}{\mathbb{E}[\|\boldsymbol{\epsilon}\|_2^2]} \cdot \frac{\Delta_2(\mathcal{S}, \mathcal{T})}{-\Delta_1(\mathcal{S}, \mathcal{T})} > \text{SNR}. \quad (107)$$

(23b):  $\Delta_1(\mathcal{S}, \mathcal{T}) > 0$

Dividing both sides of (106) by  $-\Delta_1(\mathcal{S}, \mathcal{T})$  flips the inequality, so (106) holds if and only if

$$\frac{\mathbb{E}[\|\mathbf{x}\|_2^2]}{\mathbb{E}[\|\boldsymbol{\epsilon}\|_2^2]} \cdot \frac{\Delta_2(\mathcal{S}, \mathcal{T})}{-\Delta_1(\mathcal{S}, \mathcal{T})} < \text{SNR}. \quad (108)$$

(23c):  $\Delta_1(\mathcal{S}, \mathcal{T}) = 0$

$-\Delta_1(\mathcal{S}, \mathcal{T}) \cdot \text{SNR} = 0$  so (106) holds if and only if  $\frac{\mathbb{E}[\|\mathbf{x}\|_2^2]}{\mathbb{E}[\|\boldsymbol{\epsilon}\|_2^2]} \cdot \Delta_2(\mathcal{S}, \mathcal{T}) > 0$ , if and only if

$$\Delta_2(\mathcal{S}, \mathcal{T}) > 0. \quad (109)$$

$\square$

#### APPENDIX H PROOF OF COROLLARY 1.1

*Proof.* For brevity, we fix  $\mathcal{S}$  and  $v$  and write  $\Delta_1 = \Delta_1(\mathcal{S}, v)$  and  $\Delta_2 = \Delta_2(\mathcal{S}, v)$ .

Rearranging (20) gives us that  $\mathcal{S} \setminus \{v\}$  is better than  $\mathcal{S}$  if and only if

$$\Delta_1 + \sigma^2 \cdot \Delta_2 > 0 \quad (110)$$

or equivalently if and only if

$$\Delta_1 > -\sigma^2 \cdot \Delta_2. \quad (111)$$

By definition,  $\sigma^2 = \frac{k}{N \cdot \text{SNR}}$ , so this condition is equivalent to

$$\Delta_1 > -\frac{k}{N \cdot \text{SNR}} \Delta_2 \quad (112)$$

and as SNR is strictly positive, this is equivalent to

$$\text{SNR} \cdot \Delta_1 > -\frac{k}{N} \Delta_2. \quad (113)$$

We can now use the major lemmas from the previous appendices. By Lemma 4, we have two possible values of  $\Delta_1(\mathcal{S}, v)$ :

$\Delta_1 = 0$ :

Lemma 7 means  $\Delta_2 < 0$ , so

$$\Delta_1 + \sigma^2 \cdot \Delta_2 = \sigma^2 \cdot \Delta_2 < 0 \quad (114)$$

and so  $\mathcal{S} \setminus \{v\}$  is not better than  $\mathcal{S}$ .

$\Delta_1 = -1$ :

Eq. (113) simplifies to:

$$-\text{SNR} > -\frac{k}{N} \Delta_2 \quad (115)$$

which is equivalent to

$$\text{SNR} < \frac{k}{N} \Delta_2. \quad (116)$$

On the one hand,  $v$  improves  $\mathcal{S}$  implies  $\Delta_1 = -1$ , which implies (116). On the other hand, (116) implies  $\Delta_2 > 0$  which in turn implies  $\Delta_1 = -1$ , which means (116) implies (113), which implies  $\mathcal{S} \setminus \{v\}$  is better than  $\mathcal{S}$ .

Note that the right-hand side of (116) is  $\tau(\mathcal{S}, v)$ ; this completes the proof.  $\square$

APPENDIX I  
PROOF OF PROPOSITION 2

We reframe the Proposition to the following equivalent statement:

Consider any sequence of vertices  $v_1, \dots, v_N$  with no repeated vertices, and let  $\mathcal{S}_i = \{v_1, \dots, v_i\}$ . Then there are exactly  $k$  indices  $I_1, \dots, I_k$  such that under LS reconstruction of a noisy  $k$ -bandlimited signal,

$$\forall 1 \leq j \leq k : \tau(\mathcal{S}_{I_j}, v_{I_j}) > 0 \quad (117)$$

and so for some  $\text{SNR} > 0$ ,  $\mathcal{S}_{I_j} \setminus \{v_{I_j}\}$  is better than  $\mathcal{S}_{I_j}$ .

*Proof.* By Lemma 4 in Appendix H:

$$\xi_1(\mathcal{S}_i) = k - \text{rank}([\mathbf{U}]_{\mathcal{S}_i, \mathcal{K}}), \quad (118)$$

$$\Delta_1 \in \{0, -1\} \quad (119)$$

and as  $\text{rank}([\mathbf{U}]_{\mathcal{N}, \mathcal{K}}) = k$ ,  $\xi_1(\mathcal{S}_N) = 0$ . As  $\xi_1(\mathcal{S}_0) = k$ , we must have exactly  $k$  indices for which  $\Delta_1(\mathcal{S}_i, v_i) = -1$ , and by Lemma 7 in Appendix H we have exactly  $k$  indices for which  $\Delta_2(\mathcal{S}_i, v_i) > 0$ . As  $\tau(\mathcal{S}_i, v_i) = \frac{k}{N} \Delta_2(\mathcal{S}_i, v_i)$ , we're done.  $\square$

APPENDIX J  
PROOFS FOR LS RECONSTRUCTION WITH BANDLIMITED NOISE

A. *Proof of Lemma 2*

*Proof.* By Appendix F, Lemma 4, under LS reconstruction,

$$\xi_1(\mathcal{S}) = k - \text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}). \quad (120)$$

Assuming LS reconstruction,

$$\xi_2(\mathcal{S}) = \|\|[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}^+[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}\|\|_F^2 \quad (121)$$

$$= \|\|[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}^+[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}\|\|_F^2. \quad (122)$$

As  $\mathbf{A}^+ \mathbf{A}$  is an orthogonal projection matrix, its eigenvalues are 0 or 1. Therefore

$$\|\|[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}^+[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}\|\|_F^2 = \text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}). \quad (123)$$

Add this times  $\sigma^2$  to  $\xi_1(\mathcal{S})$  to get the result.  $\square$

B. *Proof of Corollary 3.2*

*Proof.* As sample size increases,  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}})$  is increasing. If  $\text{SNR} < 1$ , then  $1 < \sigma^2$  and the MSE increases with sample size by Lemma 2.  $\square$

C. *Proof of Corollary 3.3*

*Proof.* Under a noiseless-optimal sampling scheme, after sampling  $k$  vertices we have perfect reconstruction of any clean  $k$ -bandlimited signal, and so  $\xi_1(\mathcal{S}_k) = k - \text{rank}([\mathbf{U}]_{\mathcal{S}_k, \mathcal{K}}) = 0$ .

Let  $m \leq k$ . As  $[\mathbf{U}]_{\mathcal{S}_k, \mathcal{K}}$  is of full rank, for any  $\mathcal{S}_m \subseteq \mathcal{S}_k$ ,  $[\mathbf{U}]_{\mathcal{S}_m, \mathcal{K}}$  must also be full rank. Therefore

$$\text{MSE}_{\mathcal{S}_m} - \text{MSE}_{\mathcal{S}_m \setminus \{v\}} = (\sigma^2 - 1)(m - (m - 1)) \quad (124)$$

$$= \sigma^2 - 1 \quad (125)$$

so  $\mathcal{S}_m \setminus \{v\}$  is better than  $\mathcal{S} \iff \sigma^2 > 1 \iff \text{SNR} < 1$ .

In the case where  $m > k$ :

$$|\mathcal{S}| \geq k \implies \text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = k \quad (126)$$

$$\implies \text{MSE}_{\mathcal{S}} = \sigma^2 k. \quad (127)$$

which is constant as sample size increases.  $\square$

APPENDIX K  
PROOF OF THEOREM 2

A. *Proof of (34)*

We first show that if  $m \leq k$  then

$$\forall m \leq k : \forall v \in \mathcal{S}_m : \Delta_1(\mathcal{S}_m, v) = -1. \quad (128)$$

*Proof.* By Appendix H, Lemma 4, the noiseless error

$$\xi_1(\mathcal{S}) = k - \text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) \quad (129)$$

must be 0, as we can perfectly reconstruct any  $k$ -bandlimited signal. Therefore,  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = k$ .

$[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}$  is a  $k \times k$  matrix of full rank, so its rows must be linearly independent. Any subset of linearly independent rows is linearly independent, so for any non-empty  $\mathcal{R} \subset \mathcal{S}$ ,  $[\mathbf{U}]_{\mathcal{R}, \mathcal{K}}$  has linearly independent rows.

Greedy schemes pick increasing sample sets: that is, if asked to pick a vertex sample set  $\mathcal{S}_m$  of size  $m$  for  $m < k$  and a sample set  $\mathcal{S}$  of size  $k$ ,  $\mathcal{S}_m \subset \mathcal{S}$ . Therefore for any sample set  $\mathcal{S}_m$  of size  $m \leq k$  picked by the scheme,  $[\mathbf{U}]_{\mathcal{S}_m, \mathcal{K}}$  has independent rows.

If  $[\mathbf{U}]_{\mathcal{S}_m, \mathcal{K}}$  has independent rows, then removal of any row (corresponding to removing any vertex) reduces its rank by 1; which is (128).  $\square$

We now show that for  $m \leq k$ ,

$$\forall m \leq k : \forall v \in \mathcal{S}_m : \Delta_2(\mathcal{S}_m, v) \geq 1. \quad (130)$$

*Proof.* By the previous section, we know that  $[\mathbf{U}]_{\mathcal{R}, \mathcal{K}}$  is full rank for  $\mathcal{R} \subseteq \mathcal{S}_m$ , so  $[\mathbf{U}]_{\mathcal{R}, \mathcal{K}}[\mathbf{U}]_{\mathcal{R}, \mathcal{K}}^T = [\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{R}}$  is invertible. By Appendix F, (89),  $\xi_2(\mathcal{S}_m) = \text{tr}([\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}_m}^{-1})$ . We have

$$\xi_2(\mathcal{S}_m) \quad (131)$$

$$= \text{tr}([\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}_m}^{-1}) \quad (132)$$

$$= [\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}_m \setminus \{v\}}^{-1} + \text{tr}([\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}_m \setminus \{v\}}^{-1}) \quad (133)$$

$$\geq [\mathbf{\Pi}_{bl(\mathcal{K})}]_{\{v\}}^{-1} + \text{tr}([\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}_m \setminus \{v\}}^{-1}) \quad (134)$$

$$= \frac{1}{[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\{v\}}} + \xi_2(\mathcal{S}_m \setminus \{v\}) \quad (135)$$

$$\geq 1 + \xi_2(\mathcal{S}_m \setminus \{v\}) \quad (136)$$

where the inequality in (134) is by [25, Eq. 5], and the final inequality is because the diagonal elements of  $\mathbf{\Pi}_{bl(\mathcal{K})}$  are bounded above by its maximum eigenvalue, which is 1 as  $\mathbf{\Pi}_{bl(\mathcal{K})}$  is a projection. Therefore, for all  $v \in \mathcal{S}_m$ ,

$$\Delta_2(\mathcal{S}_m, v) = \xi_2(\mathcal{S}_m) - \xi_2(\mathcal{S}_m \setminus \{v\}) \geq 1. \quad (137)$$

$\square$

Finally as  $\tau(\mathcal{S}_m, v) = \frac{k}{N} \Delta_2(\mathcal{S}_m, v)$ ,

$$\forall m \leq k : \forall v \in \mathcal{S}_m : \tau(\mathcal{S}_m, v) \geq \frac{k}{N}. \quad (138)$$

### B. Proof of (35)

*Proof.* As  $[\mathbf{U}]_{\mathcal{S}_k, \mathcal{K}}$  has  $k$  independent rows, it is of rank  $k$ . Adding further rows cannot decrease its rank, so for  $m' > k$ ,  $\text{rank}([\mathbf{U}]_{\mathcal{S}_{m'}, \mathcal{K}}) \geq k$ . As  $[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}$  is of rank  $k$ ,  $\text{rank}([\mathbf{U}]_{\mathcal{S}_{m'}, \mathcal{K}}) \leq k$ . This means for all samples sizes  $m' > k$ ,  $\text{rank}([\mathbf{U}]_{\mathcal{S}_{m'}, \mathcal{K}}) = k$ . This says that further additions of rows do not change rank; that is:

$$\forall m' > k : \forall v \in \mathcal{S}_{m'} \setminus \mathcal{S}_k : \Delta_1(\mathcal{S}_{m'}, v) = 0 \quad (139)$$

Then, by Appendix H, Lemma 7,

$$\forall m' > k : \forall v \in \mathcal{S}_{m'} \setminus \mathcal{S}_k : \Delta_2(\mathcal{S}_{m'}, v) \leq 0 \quad (140)$$

and, like for (34), as  $\tau(\mathcal{S}_m, v) = \frac{k}{N} \Delta_2(\mathcal{S}_m, v)$  and  $\frac{k}{N} > 0$ ,

$$\forall m' > k : \forall v \in \mathcal{S}_{m'} \setminus \mathcal{S}_k : \tau(\mathcal{S}_{m'}, v) \leq 0. \quad (141)$$

□

## APPENDIX L PROOF OF REMARK 5

### A-optimality

A-optimality depends on the existence of the inverse of  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$  existing, which requires it to be of full rank (at  $|\mathcal{S}| = k$  our definition of  $\mathbf{P}$  requires  $[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}^T [\mathbf{U}]_{\mathcal{S}, \mathcal{K}}$  invertible; this implies  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$  invertible as  $0 < \det([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}^T [\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = \det([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}^T) \det([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = \det([\mathbf{U}]_{\mathcal{S}, \mathcal{K}} [\mathbf{U}]_{\mathcal{S}, \mathcal{K}}^T)$ ). By Appendix H, Lemma 6, if an A-optimal scheme picks a set  $\mathcal{S}$  of size  $k$ , then  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = k$ . Therefore,  $\mathcal{S}$  is a uniqueness set [17] and can perfectly reconstruct any  $k$ -bandlimited signal.

### D- and E-optimality

We show that for sample sizes less than  $k$  we can always pick a row which keeps  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$  full rank (of rank  $|\mathcal{S}|$ ), and that D- and E-optimal schemes do so.

By Appendix H, Lemma 6,  $\text{rank}([\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}) = \text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}})$ , so we only need to ensure  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = |\mathcal{S}|$ .

We proceed by induction: given  $\mathcal{S}_1$  with  $|\mathcal{S}_1| = 1$ ,  $\text{rank}([\mathbf{U}]_{\mathcal{S}_1, \mathcal{K}}) = 1$ . Assume that for  $\mathcal{S}_i$  with  $|\mathcal{S}_i| = i < k$ ,  $\text{rank}([\mathbf{U}]_{\mathcal{S}_i, \mathcal{K}}) = i$ . As  $\text{rank}([\mathbf{U}]_{\mathcal{N}, \mathcal{K}}) = k$  and  $i < k$ , we can find a row to add to  $[\mathbf{U}]_{\mathcal{S}_i, \mathcal{K}}$  which will increase its rank (else all other rows would lie in the  $i$ -dimensional space spanned by the rows of  $[\mathbf{U}]_{\mathcal{S}_i, \mathcal{K}}$ , which would imply  $\text{rank}([\mathbf{U}]_{\mathcal{N}, \mathcal{K}}) = i$ , which is a contradiction as  $i < k$ ). Adding the vertex which corresponds to the row to  $\mathcal{S}_i$  gives  $\mathcal{S}_{i+1}$  with  $\text{rank}([\mathbf{U}]_{\mathcal{S}_{i+1}, \mathcal{K}}) = i + 1$ .

We have shown that we can greedily choose to keep  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = |\mathcal{S}|$ . We now show that D- and E-optimal schemes do so. The eigenvalues of  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$  are non-negative (see Appendix H, Eq. (92)), so any invertible  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$  will have a strictly positive determinant and minimum eigenvalue, which are preferable under the D- and E- optimality criterion respectively to a non-invertible  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$ , which has a determinant and minimum eigenvalue of 0. Therefore, greedy D- and E- optimal sampling schemes will make sure  $[\mathbf{\Pi}_{bl(\mathcal{K})}]_{\mathcal{S}}$  is invertible, and thus keep  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = |\mathcal{S}|$  for  $|\mathcal{S}| \leq k$ . Therefore when D- and E- optimal schemes pick a set  $\mathcal{S}$  of size  $k$ ,  $\text{rank}([\mathbf{U}]_{\mathcal{S}, \mathcal{K}}) = k$ . Therefore,  $\mathcal{S}$  is a uniqueness set [17] and can perfectly reconstruct any  $k$ -bandlimited signal.

## APPENDIX M PROOF OF COROLLARY 2.1

We first simplify  $\xi_1(\mathcal{S})$ :

$$[\mathbf{U}]_{\mathcal{N}, \mathcal{K}} - \mathbf{R}_{\mathcal{S}} [\mathbf{U}]_{\mathcal{S}, \mathcal{K}} \quad (142)$$

$$= \left( \mathbf{I} - (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^{-1} \mathbf{\Pi}_{\mathcal{S}} \right) [\mathbf{U}]_{\mathcal{N}, \mathcal{K}} \quad (143)$$

$$= (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^{-1} \mu \mathbf{L} [\mathbf{U}]_{\mathcal{N}, \mathcal{K}} \quad (144)$$

$$= (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^{-1} [\mathbf{U}]_{\mathcal{N}, \mathcal{K}} (\mu \mathbf{\Lambda}_k) \quad (145)$$

where  $\mathbf{\Lambda}_k$  is a  $k \times k$  diagonal matrix with the corresponding graph frequencies to  $[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}$  as its diagonal. Write  $u_i$  for the  $i^{\text{th}}$  column of  $[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}$ , so  $\mathbf{u}_i$  is an eigenvector of  $\mathbf{L}$ .

$$\xi_1(\mathcal{S}) = \|[ [\mathbf{U}]_{\mathcal{N}, \mathcal{K}} - \mathbf{R}_{\mathcal{S}} [\mathbf{U}]_{\mathcal{S}, \mathcal{K}} ]\|_F^2 \quad (146)$$

$$= \sum_{i=2}^k (\mu \lambda_i)^2 \mathbf{u}_i^T (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^{-2} \mathbf{u}_i \quad (147)$$

Note that the condition is equivalent to the following:

$$[\mathbf{U}]_{\mathcal{S}^c, \{2, \dots, k\}} = \mathbf{0} \iff \forall i \in [2, k] : \mathbf{\Pi}_{\mathcal{S}} \mathbf{u}_i = \mathbf{u}_i \quad (148)$$

that is, the projection is idempotent on each of the  $k - 1$  non-constant eigenvectors. We consider the cases where this is and is not true and correlate them to cases in Theorem 1.

### A. The projection is idempotent

As  $\mathbf{\Pi}_{\mathcal{S}} \mathbf{u}_i = \mathbf{u}_i$ ,

$$(\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L}) \mathbf{u}_i = (1 + \mu \lambda_i) \mathbf{u}_i \quad (149)$$

therefore  $\mathbf{u}_i$  is an eigenvector of  $(\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})$  with eigenvalue  $1 + \mu \lambda_i$  and

$$\mathbf{u}_i^T (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^{-2} \mathbf{u}_i = (1 + \mu \lambda_i)^{-2}. \quad (150)$$

By Lemma 8, in this case  $\xi_1(\mathcal{S}) = \xi_1(\mathcal{N})$ , i.e. that  $\Delta_1(\mathcal{N}, \mathcal{S}^c) = 0$ . This corresponds to condition (23c), and gives us condition (37b) in our Corollary.

### B. The projection is not idempotent

Applying Cauchy-Schwartz to  $\mathbf{x} = (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^{-1} \mathbf{u}_i$  and  $\mathbf{y} = (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L}) \mathbf{u}_i$  gives, as  $\mathbf{u}_i^T \mathbf{u}_i = 1$ ,

$$1 \leq \mathbf{u}_i^T (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^{-2} \mathbf{u}_i \mathbf{u}_i^T (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^2 \mathbf{u}_i. \quad (151)$$

We note that

$$\| \mathbf{\Pi}_{\mathcal{S}} \mathbf{u}_i \|_2^2 = \sum_{j \in \mathcal{S}} (\mathbf{u}_i)_j^2 \leq \sum_{j=1}^N (\mathbf{u}_i)_j^2 = \| \mathbf{u}_i \|_2^2 = 1. \quad (152)$$

and see that (152) must be strict for at least one  $i \in [2, k]$ , because some component of  $\mathbf{u}_i$  in  $\mathcal{S}^c$  is nonzero for at least one  $i \in [2, k]$ . Therefore

$$\mathbf{u}_i^T (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^2 \mathbf{u}_i = (\mu \lambda_i)^2 + (1 + 2\mu \lambda_i) \| \mathbf{\Pi}_{\mathcal{S}} \mathbf{u}_i \|_2^2 \quad (153)$$

$$\leq (1 + \mu \lambda_i)^2. \quad (154)$$

and substituting into (151):

$$\left( \frac{1}{1 + \mu \lambda_i} \right)^2 \leq \mathbf{u}_i^T (\mathbf{\Pi}_{\mathcal{S}} + \mu \mathbf{L})^{-2} \mathbf{u}_i \quad (155)$$

with these inequality being strict for at least one  $i \in [2, k]$ . Therefore, by Lemma 8 and (147),

$$\xi_1(\mathcal{S}) > \xi_1(\mathcal{N}) \quad (156)$$

so  $\Delta_1(\mathcal{N}, \mathcal{S}^c) < 0$ .

### C. Simplifying Theorem 1

We see that the projection is idempotent on  $(\mathbf{u}_i)_{i=2}^k$  implies  $\Delta_1(\mathcal{N}, \mathcal{S}^c) = 0$ , and the projection is not idempotent implies  $\Delta_1(\mathcal{N}, \mathcal{S}^c) < 0$ . As the projection must either be idempotent or not idempotent, these implications must be ‘if and only if’ statements. We therefore rule out (23b) in Theorem 1.

## APPENDIX N PROOF OF THEOREM 3

See Appendix R for a proof that  $m_{opt} < \frac{N+1}{2}$  if  $B(m) < N$ .

We restate the following definitions from Lemmas 1 and 3:

$$r = \omega \left( \frac{\lambda_N}{\lambda_2} \right) \quad (157)$$

$$B(m) = r \frac{N}{m} + \sum_{i=1}^m \omega \left( \max \left[ 1, \frac{\lambda_{N+2-i}}{\lambda_i} \right] \right) \quad (158)$$

$$r_{bl} = \omega \left( \frac{\lambda_k}{\lambda_2} \right) \quad (159)$$

$$B_k(m) = r_{bl} \frac{N}{m} + \sum_{i=1}^m \omega \left( \max \left[ 1, \frac{\lambda_{k+2-i}}{\lambda_i} \right] \right) \quad (160)$$

We define the following:

$$\bar{\lambda} = \frac{1}{N} \sum_{i=1}^N \lambda_i \quad (161)$$

$$\mu_{ub} = \bar{\lambda}^{-1} \left( \sqrt{\frac{N}{B(m_{opt})}} - 1 \right) \quad (162)$$

$$\tau_{GLR}(\mu) = \frac{k}{N} \cdot \frac{\left( \sum_{i=1}^N (1 + \mu \lambda_i)^{-2} \right) - B(m_{opt})}{k + B_k(m_{opt}) - \sum_{i=1}^k (1 - (1 + \mu \lambda_i)^{-1})^2} \quad (163)$$

We start by calculating  $\xi_i(\mathcal{N})$ .

**Lemma 8.** Under GLR reconstruction with parameter  $\mu$ ,

$$\xi_1(\mathcal{N}) = \sum_{i=1}^k \left( 1 - \frac{1}{1 + \mu \lambda_i} \right)^2 \quad (164)$$

$$\xi_2(\mathcal{N}) = \sum_{i=1}^N \left( \frac{1}{1 + \mu \lambda_i} \right)^2 \quad (165)$$

*Proof.* Set  $\mathbf{R}_{\mathcal{N}} = (\mathbf{I} + \mu \mathbf{L})^{-1}$  in (28) and (29), noting  $[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}$  are eigenvectors for  $\mathbf{R}_{\mathcal{N}}$ .  $\square$

Let  $\mathcal{S}$  be any sample set of size  $m$ . We show that under our conditions  $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$  and then apply Corollary 2.1. To do so, we use the following bounds:

$$\xi_2(\mathcal{S}) \leq B(m) \quad (166)$$

$$\xi_1(\mathcal{S}) \leq k + B_k(m) \quad (167)$$

$$\xi_1(\mathcal{N}) = \sum_{i=1}^k \left( 1 - \frac{1}{1 + \mu \lambda_i} \right)^2 \quad (168)$$

$$\xi_2(\mathcal{N}) = \sum_{i=1}^N \left( \frac{1}{1 + \mu \lambda_i} \right)^2 \quad (169)$$

These are proven in Lemma 1, Appendix P Lemma 9, and Lemma 8. We therefore see that, as  $\Delta_i(\mathcal{N}, \mathcal{S}^c) = \xi_i(\mathcal{N}) - \xi_i(\mathcal{S})$ ,

$$\Delta_2(\mathcal{N}, \mathcal{S}^c) \geq \sum_{i=1}^N \left( \frac{1}{1 + \mu \lambda_i} \right)^2 - B(m) \quad (170)$$

$$\Delta_1(\mathcal{N}, \mathcal{S}^c) \geq \sum_{i=1}^k \left( 1 - \frac{1}{1 + \mu \lambda_i} \right)^2 - (k + B_k(m)) \quad (171)$$

We now consider a sets  $\mathcal{S}$  of size  $m_{opt}$  and show that  $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ . We have that  $r > 0$  so  $B(m) > 0$  and by assumption  $B(m_{opt}) < N$ . Therefore  $\mu_{ub} > 0$  and is real and it therefore possible to pick  $0 < \mu < \mu_{ub}$ . By assumption  $0 < \mu < \mu_{ub}$ , so by Jensen’s Inequality,

$$\sum_{i=1}^N \left( \frac{1}{1 + \mu \lambda_i} \right)^2 \geq \frac{N}{\left( 1 + \mu \frac{\text{tr}(\mathbf{L})}{N} \right)^2} \quad (172)$$

$$> \frac{N}{\left( 1 + \mu_{ub} \frac{\text{tr}(\mathbf{L})}{N} \right)^2} \quad (173)$$

$$= B(m_{opt}) \quad (174)$$

And therefore by (170),  $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ .

We now apply Corollary 2.1. We case-split on whether  $\mathbf{\Pi}_{\mathcal{S}^c} [\mathbf{U}]_{\mathcal{N}, \{2, \dots, k\}}$  is or is not  $\mathbf{0}$ , and show  $\mathcal{S}$  is better than  $\mathcal{N}$  in both cases.

1) (37a) - is not 0: We assume  $\mathbf{\Pi}_{\mathcal{S}^c} [\mathbf{U}]_{\mathcal{N}, \{2, \dots, k\}} \neq \mathbf{0}$ . By (170) and (171),

$$\tau(\mathcal{N}, \mathcal{S}^c) = \frac{k}{N} \cdot \frac{\Delta_2(\mathcal{N}, \mathcal{S}^c)}{-\Delta_1(\mathcal{N}, \mathcal{S}^c)} \geq \tau_{GLR}(\mu) > \text{SNR} \quad (175)$$

where the last inequality is by our assumption. Therefore Corollary 2.1 (37a) holds and  $\mathcal{S}$  is better than  $\mathcal{N}$ .

2) (37b) - is 0: We assume  $\mathbf{\Pi}_{\mathcal{S}^c} [\mathbf{U}]_{\mathcal{N}, \{2, \dots, k\}} = \mathbf{0}$ . We have that  $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ , so Corollary 2.1 (37b) holds and  $\mathcal{S}$  is better than  $\mathcal{N}$ .

Therefore  $\mathcal{S}$  is better than  $\mathcal{N}$  regardless of whether  $\mathbf{\Pi}_{\mathcal{S}^c} [\mathbf{U}]_{\mathcal{N}, \{2, \dots, k\}}$  is or is not  $\mathbf{0}$  and we are done.

## APPENDIX O

BOUNDING  $\xi_2(\mathcal{S})$  UNDER GLR – PROOF OF LEMMA 1

In this Appendix, we aim to prove Lemma 1 which states that under GLR

$$\xi_2(\mathcal{S}) \leq r \frac{N}{m} + \sum_{i=2}^m \omega \left( \max \left[ 1, \frac{\lambda_{N+2-i}}{\lambda_i} \right] \right) \quad (176)$$

$$\leq r \frac{N}{m} + r(m-1). \quad (177)$$

## A. Preliminaries and notation

We assume that  $\mathbf{L}$  is the combinatorial Laplacian. We write the eigenvalues of  $\mathbf{L}$  as  $0 = \lambda_1 \leq \dots \leq \lambda_N$ . We note, for any  $\mathbf{X} \in \mathbb{R}^{x \times N}$ ,  $\mathbf{Y} \in \mathbb{R}^{N \times y}$ ,

$$[\mathbf{X}]_{a,N} [\mathbf{Y}]_{N,c} = [\mathbf{XY}]_{a,c}. \quad (178)$$

We pick a basis suited to our proof. Let the standard basis for  $\mathbb{R}^N$  be  $(\mathbf{e}_i)_{i=1}^N$ . Let

$$\mathbf{v}_1 = \frac{1}{\sqrt{m}} \sum_{i \in \mathcal{S}} \mathbf{e}_i = \frac{1}{\sqrt{m}} \mathbf{\Pi}_{\mathcal{S}} \mathbf{1}_N. \quad (179)$$

so  $\|\mathbf{v}_1\|_2 = 1$ . Pick  $\{\mathbf{v}_2, \dots, \mathbf{v}_m\}$  so that  $(\mathbf{v}_i)_{i=1}^m$  is an orthonormal basis for  $(\mathbf{e}_i)_{i \in \mathcal{S}}$ . Finally let  $(\mathbf{v}_i)_{i=m+1}^N = (\mathbf{e}_i)_{i \in \mathcal{S}^c}$ . Now  $(\mathbf{v}_i)_{i=1}^N$  is a basis for  $\mathbb{R}^N$ .

For the rest of this Appendix, we will write out matrices in this new basis. In our new basis, the top left entry when writing out  $\mathbf{L}^\dagger$  is

$$\mathbf{L}_{1,1}^\dagger = \frac{1}{m} \mathbf{1}_m^T [\mathbf{L}^\dagger]_{\mathcal{S}} \mathbf{1}_m \in \mathbb{R} \quad (180)$$

and our frequently used projection  $\mathbf{\Pi}_{\mathcal{S}}$  looks like:

$$\mathbf{\Pi}_{\mathcal{S}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \quad (181)$$

We define the set  $\Theta = \{2, \dots, m\}$ , and note that

$$\mathbf{\Pi}_{\Theta} = \mathbf{I}_m - \frac{1}{m} \mathbf{1}_{m \times m}. \quad (182)$$

We have  $\{1\} \cup \Theta = \mathcal{S}$  and  $\{1\} \cup \Theta \cup \mathcal{S}^c = \mathcal{N}$ .

Finally, we define a useful matrix:

$$\mathbf{P} = \mathbf{I} - \frac{1}{m} [\mathbf{I}]_{\mathcal{N}, \mathcal{S}} \mathbf{1}_{m \times N}. \quad (183)$$

## B. Proof overview

We decompose  $\xi_2(\mathcal{S}) = \|\mathbf{R}_{\mathcal{S}}\|_F^2$  row-wise in our new basis (Subsection O-C).

$$\|\mathbf{R}_{\mathcal{S}}\|_F^2 = \|\mathbf{R}_{\mathcal{S}}\|_{\mathcal{N}, \{1\}}^2 + \|\mathbf{R}_{\mathcal{S}}\|_{\mathcal{N}, \Theta}^2 \quad (184)$$

$$= \frac{N}{m} + \|\mathbf{R}_{\mathcal{S}}\|_{\mathcal{N}, \Theta}^2 \quad (185)$$

We explicitly write out  $[\mathbf{R}_{\mathcal{S}}]_{\mathcal{N}, \Theta}$  (Subsection O-D),

$$[\mathbf{R}_{\mathcal{S}}]_{\mathcal{N}, \Theta} = \mathbf{P}^T \begin{bmatrix} \mathbf{1} \\ \mu \mathbf{L}^\dagger \end{bmatrix}_{\mathcal{N}, \Theta} \begin{bmatrix} \mathbf{I} + \frac{1}{\mu} \mathbf{L}^\dagger \\ \mathbf{0} \end{bmatrix}_{\Theta}^{-1} \quad (186)$$

$$= \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mu \mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-1} \quad (187)$$

and use this to remove the dependence on  $\mu$  in our bound (Subsection O-E):

$$\|\mathbf{R}_{\mathcal{S}}\|_{\mathcal{N}, \Theta}^2 \leq \left\| \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (188)$$

We exactly calculate the effects of  $\mathbf{P}^T$  on the Frobenius norm (which yields the  $\frac{N}{m}$  term in (190))(Subsection O-F):

$$\left\| \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (189)$$

$$= \left( \frac{N}{m} \right) \left\| [\mathbf{L}^\dagger]_{\{1\}, \Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (190)$$

$$+ \left\| [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (191)$$

As

$$\left\| [\mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 = \|\mathbf{I}_{m-1}\|_F^2 = m-1, \quad (192)$$

we get that

$$\xi_2(\mathcal{S}) = \left( \frac{N}{m} + m-1 \right) + \text{error} \quad (193)$$

where the error term is quantified in (190) and (191).

Finally, we bound (190) and (191) using variants of the Kantorovich Inequality. We have that [40, Eq. 20-23] gives (Subsection O-G)

$$\left\| [\mathbf{L}^\dagger]_{\{1\}, \Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \leq (r-1) \quad (194)$$

We use another variant to show that (Subsection O-H)

$$\left\| [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \leq \sum_{i=2}^m \omega \left( \max \left[ 1, \frac{\lambda_{N+2-i}}{\lambda_i} \right] \right) \quad (195)$$

Combine these bounds with (188 - 191) to prove the first inequality in the Lemma.

The second, weaker bound follows as  $\omega$  is increasing so

$$\omega \left( \max \left[ 1, \frac{\lambda_{N+2-i}}{\lambda_i} \right] \right) \leq \omega \left( \frac{\lambda_N}{\lambda_2} \right) = r. \quad (196)$$

## C. Row decomposition

As  $(\mathbf{v}_i)_{i=1}^m$  are orthogonal and span  $(\mathbf{e}_i)_{i \in \mathcal{S}}$ , and  $\mathbf{R}_{\mathcal{S}} \mathbf{1}_m = \mathbf{1}_N$ ,

$$\|\mathbf{R}_{\mathcal{S}}\|_F^2 = \sum_{i=1}^m \|\mathbf{R}_{\mathcal{S}} \mathbf{v}_i\|_2^2 \quad (197)$$

$$= \sum_{i=1}^m \left\| [\mathbf{R}_{\mathcal{S}}]_{\mathcal{N}, \{i\}} \right\|_2^2 \quad (198)$$

$$= \left\| \mathbf{R}_{\mathcal{S}} \frac{1}{\sqrt{m}} \mathbf{1}_m \right\|_2^2 + \|\mathbf{R}_{\mathcal{S}}\|_{\mathcal{N}, \Theta}^2 \quad (199)$$

$$= \frac{N}{m} + \|\mathbf{R}_{\mathcal{S}}\|_{\mathcal{N}, \Theta}^2. \quad (200)$$

#### D. Explicit submatrix computation

We explicitly compute that

$$[\mathbf{R}_S]_{\mathcal{N},\Theta} = (\mathbf{\Pi}_S + \mu\mathbf{L})^{-1} [\mathbf{I}]_{\mathcal{N},\Theta} \quad (201)$$

$$= \mathbf{P}^T \begin{bmatrix} \mathbf{1} \\ \mu \mathbf{L}^\dagger \end{bmatrix}_{\mathcal{N},\Theta} \begin{bmatrix} \mathbf{I} + \frac{1}{\mu} \mathbf{L}^\dagger \\ \mu \end{bmatrix}_{\Theta}^{-1} \quad (202)$$

*Proof.* We show the equivalent statement,

$$[\mathbf{I}]_{\mathcal{N},\Theta} \begin{bmatrix} \mathbf{I} + \frac{1}{\mu} \mathbf{L}^\dagger \\ \mu \end{bmatrix}_{\Theta} = (\mathbf{\Pi}_S + \mu\mathbf{L}) \mathbf{P}^T \begin{bmatrix} \mathbf{1} \\ \mu \mathbf{L}^\dagger \end{bmatrix}_{\mathcal{N},\Theta}. \quad (203)$$

We have that

$$\mu\mathbf{L}\mathbf{P}^T = \mu\mathbf{L} \quad (204)$$

$$\mathbf{\Pi}_S\mathbf{P}^T = [\mathbf{I}]_{\mathcal{N},S} \left( \mathbf{I} - \frac{1}{m} \mathbf{1}_{m \times m} \right) [\mathbf{I}]_{S,\mathcal{N}} \quad (205)$$

$$= [\mathbf{I}]_{\mathcal{N},S} [\mathbf{\Pi}_\Theta]_S [\mathbf{I}]_{S,\mathcal{N}} \quad (206)$$

$$= [\mathbf{I}]_{\mathcal{N},\Theta} [\mathbf{I}]_{\Theta,\mathcal{N}} \quad (207)$$

so

$$\mu\mathbf{L}\mathbf{P}^T \begin{bmatrix} \mathbf{1} \\ \mu \mathbf{L}^\dagger \end{bmatrix}_{\mathcal{N},\Theta} = \mu\mathbf{L} \begin{bmatrix} \mathbf{1} \\ \mu \mathbf{L}^\dagger \end{bmatrix}_{\mathcal{N},\Theta} \quad (208)$$

$$= \mathbf{L}\mathbf{L}^\dagger [\mathbf{I}]_{\mathcal{N},\Theta} \quad (209)$$

$$= \left( \mathbf{I} - \frac{1}{N} \mathbf{1}_{N \times N} \right) [\mathbf{I}]_{\mathcal{N},\Theta} \quad (210)$$

$$= [\mathbf{I}]_{\mathcal{N},\Theta} \quad (211)$$

and

$$\mathbf{\Pi}_S\mathbf{P}^T \begin{bmatrix} \mathbf{1} \\ \mu \mathbf{L}^\dagger \end{bmatrix}_{\mathcal{N},\Theta} = [\mathbf{I}]_{\mathcal{N},\Theta} [\mathbf{I}]_{\Theta,\mathcal{N}} \begin{bmatrix} \mathbf{1} \\ \mu \mathbf{L}^\dagger \end{bmatrix}_{\mathcal{N},\Theta} \quad (212)$$

$$= [\mathbf{I}]_{\mathcal{N},\Theta} \begin{bmatrix} \mathbf{1} \\ \mu \mathbf{L}^\dagger \end{bmatrix}_{\Theta} \quad (213)$$

Sum these two terms for the result.  $\square$

#### E. Removing the dependency on $\mu$

By multiplying out the constant in 202, we see that

$$[\mathbf{R}_S]_{\mathcal{N},\Theta} = \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mu\mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-1} \quad (214)$$

We prove that

$$\forall \mu > 0 : \left\| [\mathbf{R}_S]_{\mathcal{N},\Theta} \right\|_F^2 \leq \left\| \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (215)$$

*Proof.* In this proof, all matrix inequalities are in the standard Loewner ordering [45, pg. 112]. We first show  $[\mathbf{L}^\dagger]_{\Theta}^{-1}$  is well defined. Principal submatrices of positive definite matrices are positive definite [45, Corollary III.1.5], so

$$\mathbf{L}^\dagger + \mathbf{1}_{N \times N} > 0 \implies [\mathbf{L}^\dagger + \mathbf{1}_{N \times N}]_{\Theta} = [\mathbf{L}^\dagger]_{\Theta} > 0. \quad (216)$$

Therefore  $[\mathbf{L}^\dagger]_{\Theta}$  is invertible and  $[\mathbf{L}^\dagger]_{\Theta}^{-1} > 0$ . Next, as  $\mu > 0$ ,  $\mu^2[\mathbf{I}]_{\Theta} > 0$ . Therefore, by expanding the quadratic,

$$[\mu\mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^2 > [\mathbf{L}^\dagger]_{\Theta}^2. \quad (217)$$

By [45, Proposition V.1.6, page 114],

$$[\mu\mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-2} \leq [\mathbf{L}^\dagger]_{\Theta}^{-2}. \quad (218)$$

As trace is the sum of inner products, by the definition of the Loewner ordering,  $\forall \mathbf{H} \in \mathbb{R}^{N \times m-1}$ ,  $\forall \mu > 0$ ,

$$\text{tr} \left( \mathbf{H} [\mu\mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-2} \mathbf{H}^T \right) \leq \text{tr} \left( \mathbf{H} [\mathbf{L}^\dagger]_{\Theta}^{-2} \mathbf{H}^T \right) \quad (219)$$

which is the same as

$$\left\| \mathbf{H} [\mu\mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_2^2 \leq \left\| \mathbf{H} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_2^2. \quad (220)$$

Substitute  $\mathbf{H} = \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N},\Theta}$  to finish.  $\square$

#### F. Column-wise decomposition

We first show that

$$\left\| \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 = \frac{N}{m} \left\| [\mathbf{L}^\dagger]_{\{1\},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (221)$$

$$+ \left\| [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (222)$$

*Proof.* Let  $\mathbf{K} = [\mathbf{I}]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1}$ , then

$$\left\| \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 = \text{tr} (\mathbf{K}^T \mathbf{L}^\dagger \mathbf{P} \mathbf{P}^T \mathbf{L}^\dagger \mathbf{K}) \quad (223)$$

As  $\mathbf{1}_{m \times N} \mathbf{L}^\dagger = 0$  cross terms in  $\mathbf{P} \mathbf{P}^T$  disappear in  $\mathbf{L}^\dagger \mathbf{P} \mathbf{P}^T \mathbf{L}^\dagger$ ,

$$\mathbf{L}^\dagger \mathbf{P} \mathbf{P}^T \mathbf{L}^\dagger = \mathbf{L}^\dagger \left( \mathbf{I} + \frac{N}{m} \begin{pmatrix} \frac{1}{m} \mathbf{1}_{m \times m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right) \mathbf{L}^\dagger \quad (224)$$

$$= (\mathbf{L}^\dagger)^2 + \frac{N}{m} \mathbf{L}^\dagger \mathbf{v}_1 \mathbf{v}_1^T \mathbf{L}^\dagger. \quad (225)$$

Therefore

$$\left\| \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (226)$$

$$= \text{tr} (\mathbf{K}^T \mathbf{L}^\dagger \mathbf{L}^\dagger \mathbf{K}) + \frac{N}{m} \text{tr} (\mathbf{K}^T \mathbf{L}^\dagger \mathbf{v}_1 \mathbf{v}_1^T \mathbf{L}^\dagger \mathbf{K}) \quad (227)$$

$$= \left\| [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 + \frac{N}{m} \left\| [\mathbf{L}^\dagger]_{\{1\},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2. \quad (228)$$

$\square$

#### G. Bounding (190)

We follow the idea in [40, Eq. 20-23], but use newer work to account for  $\mathbf{L}^\dagger$  not being positive definite.

$$1 + \left\| [\mathbf{L}^\dagger]_{\{1\},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right\|_F^2 \quad (229)$$

$$= 1 + \lambda_{\max} \left( [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger \mathbf{v}_1 \mathbf{v}_1^T \mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right) \quad (230)$$

$$= \lambda_{\max} \left( [\mathbf{I}]_{\Theta} + [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger \mathbf{v}_1 \mathbf{v}_1^T \mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right) \quad (231)$$

$$= \lambda_{\max} \left( [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right) \quad (232)$$

$$+ [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger \mathbf{v}_1 \mathbf{v}_1^T \mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \quad (233)$$

$$\leq \lambda_{\max} \left( [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger \mathbf{\Pi}_\Theta \mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right) \quad (234)$$

$$+ [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger \mathbf{v}_1 \mathbf{v}_1^T \mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \quad (235)$$

$$+ [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger \mathbf{\Pi}_{S^c} \mathbf{L}^\dagger]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \quad (236)$$

$$= \lambda_{\max} \left( [\mathbf{L}^\dagger]_{\Theta}^{-1} \left[ (\mathbf{L}^\dagger)^2 \right]_{\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} \right) \quad (237)$$

where in (234) we use  $[\mathbf{L}^\dagger]_\Theta [\mathbf{L}^\dagger]_\Theta = [\mathbf{L}^\dagger \mathbf{\Pi}_\Theta \mathbf{L}^\dagger]_\Theta$ . We weaken the inequality in step (236) by adding a positive semi-definite matrix ( $\mathbf{\Pi}_{S^c}$  is positive semidefinite so for any  $\mathbf{X}$ ,  $\mathbf{X} \mathbf{\Pi}_{S^c} \mathbf{X}^T$  is positive semidefinite) – this is valid by Weyl’s inequality [45, Corollary III.2.2, page 63]. The final step comes from  $\{1\} \cup \Theta \cup S^c = \mathcal{N}$  so  $\mathbf{X} \mathbf{v}_1 \mathbf{v}_1^T \mathbf{X}^T + \mathbf{X} \mathbf{\Pi}_\Theta \mathbf{X}^T + \mathbf{X} \mathbf{\Pi}_{S^c} \mathbf{X}^T = \mathbf{X} \mathbf{X}^T$  for any matrix  $\mathbf{X}$ .

We now apply a matrix variant of Kantorovich which holds for positive semi-definite matrices – [46, Theorem 3] with  $\mathbf{X} = [\mathbf{I}]_{\mathcal{N}, \Theta}$ ,  $\mathbf{H} = \mathbf{L} \mathbf{L}^\dagger = \mathbf{I}_N - \frac{1}{N} \mathbf{1}_{N \times N}$  and  $\mathbf{G} = \mathbf{L}^\dagger$ .

This version has extra constraints on the ranges of  $\mathbf{X}$ ,  $\mathbf{G}$  and  $\mathbf{H}$  to handle the semi-definiteness, which we now check. First,  $\mathbf{G} \mathbf{H} = \mathbf{H} \mathbf{G} = \mathbf{L}^\dagger \geq 0$ . Now  $\text{range}(\mathbf{G}) = \text{range}(\mathbf{L}^\dagger) = \text{range}(\mathbf{L} \mathbf{L}^\dagger) = \text{range}(\mathbf{I} - \frac{1}{N} \mathbf{1}_{N \times N}) = \text{range}(\mathbf{H})$ . As  $[\mathbf{1}_N]_\Theta = \mathbf{X}^T \mathbf{1}_N = \mathbf{0}$ , so  $\text{range}(\mathbf{1}_N \mathbf{1}_N^T) \subseteq \ker(\mathbf{X}^T)$ , and by the Fundamental Theorem of Linear Algebra  $\ker(\mathbf{X}^T)^\perp = \text{range}(\mathbf{X})$ , so we see that  $\text{range}(\mathbf{X}) \subseteq \text{range}(\mathbf{G})$  and  $\text{range}(\mathbf{X}) \subseteq \text{range}(\mathbf{H})$ .

Applying [46, Theorem 3] gives

$$[\mathbf{L}^\dagger]_\Theta^{-1} \left[ (\mathbf{L}^\dagger)^2 \right]_\Theta [\mathbf{L}^\dagger]_\Theta^{-1} \leq r \mathbf{\Pi}_\Theta \quad (238)$$

and therefore

$$\lambda_{\max} \left( [\mathbf{L}^\dagger]_\Theta^{-1} \left[ (\mathbf{L}^\dagger)^2 \right]_\Theta [\mathbf{L}^\dagger]_\Theta^{-1} \right) \leq r \quad (239)$$

By (229-237) we have

$$\left\| [\mathbf{L}^\dagger]_{\{1\}, \Theta}^{-1} [\mathbf{L}^\dagger]_\Theta^{-1} \right\|_F^2 \leq r - 1. \quad (240)$$

therefore,

$$\left( \frac{N}{m} \right) \left\| [\mathbf{L}^\dagger]_{\{1\}, \Theta}^{-1} [\mathbf{L}^\dagger]_\Theta^{-1} \right\|_2^2 \leq \left( \frac{N}{m} \right) (r - 1). \quad (241)$$

#### H. Bounding (191)

We use the Bloomfield-Watson-Knott extension of Kantorovich’s inequality [39, Thm 5]. Let  $\mathbf{X} = \mathbf{Y} = [\mathbf{I}]_{\mathcal{N}, \Theta}$ ,  $\mathbf{G} = \mathbf{L}^\dagger$  and  $\mathbf{H} = \mathbf{L} \mathbf{L}^\dagger = \mathbf{I}_N - \frac{1}{N} \mathbf{1}_{N \times N}$  in [39, Thm 5]. As  $[\mathbf{H}^2]_\Theta = [\mathbf{I}]_\Theta$ ,

$$\left\| [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}^\dagger]_\Theta^{-1} \right\|_F^2 \leq \sum_{i=2}^m \omega \left( \max \left[ 1, \frac{\lambda_{N+2-i}}{\lambda_i} \right] \right) \quad (242)$$

The same checks we ran in Section O-G apply to [39, Thm 5], and so the constraints hold.

In [39, Thm 5], in the first part of (2.14),  $\omega_i = \omega \left( \frac{\lambda_{N+2-i}}{\lambda_i} \right)$  – these correspond to the case where  $\frac{\lambda_{N+2-i}}{\lambda_i} \geq 1$ . We have weakened the second part of (2.14) to be increasing by 1 each time (as  $\omega(1) = 1$ ) for ease of presentation; this does not affect the downstream effectiveness of the bound, and lets us have the same form as the bandlimited bound.

#### APPENDIX P BOUNDING $\xi_1(\mathcal{S})$

**Lemma 9.** Under GLR reconstruction,

$$\xi_1(\mathcal{S}) \leq k + B_k(m) \quad (243)$$

where  $m = |\mathcal{S}|$  and  $B_k(m)$  is as defined in Lemma 3.

*Proof.* Note that

$$\mathbf{R}_S[\mathbf{I}]_{S, \mathcal{N}} = (\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{\Pi}_S \quad (244)$$

$$= \mathbf{I} - (\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mu \mathbf{L} \quad (245)$$

we have

$$\xi_1(\mathcal{S}) = \left\| [\mathbf{U}]_{\mathcal{N}, \mathcal{K}} - \mathbf{R}_S[\mathbf{U}]_{S, \mathcal{K}} \right\|_F^2 \quad (246)$$

$$= \text{tr}(\mathbf{I}_k - 2[\mathbf{U}]_{\mathcal{N}, \mathcal{K}}^T \mathbf{R}_S[\mathbf{U}]_{S, \mathcal{K}}) + \left\| \mathbf{R}_S[\mathbf{U}]_{S, \mathcal{K}} \right\|_F^2 \quad (247)$$

$$= \text{tr} \left( 2 \left( [\mathbf{U}]_{\mathcal{N}, \mathcal{K}}^T (\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mu \mathbf{L} [\mathbf{U}]_{\mathcal{N}, \mathcal{K}} \right) - \mathbf{I}_k \right) \quad (248)$$

$$+ \left\| \mathbf{R}_S[\mathbf{U}]_{S, \mathcal{K}} \right\|_F^2 \quad (249)$$

$$= 2 \left( \sum_{i=1}^k \mathbf{u}_i^T (\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{u}_i \mu \lambda_i \right) - k \quad (250)$$

$$+ \left\| \mathbf{R}_S[\mathbf{U}]_{S, \mathcal{K}} \right\|_F^2 \quad (251)$$

By [47, Eq. (1.7)] for  $i > 1$ ,  $\mathbf{u}_i^T (\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{u}_i \leq \mathbf{u}_i^T (\mu \mathbf{L})^{-1} \mathbf{u}_i = (\mu \lambda_i)^{-1}$ , and for  $i = 1$ ,  $\lambda_1 = 0$ , so

$$\left( \sum_{i=1}^k \mathbf{u}_i^T (\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{u}_i \mu \lambda_i \right) \leq k - 1 \quad (252)$$

and therefore

$$\xi_1(\mathcal{S}) \leq (k - 1) + \left\| \mathbf{R}_S[\mathbf{U}]_{S, \mathcal{K}} \right\|_F^2. \quad (253)$$

As  $\left\| \mathbf{R}_S[\mathbf{U}]_{S, \mathcal{K}} \right\|_F^2 = \xi_{2, bl}$ , by Lemma 3,

$$\xi_1(\mathcal{S}) \leq k + B_k(m). \quad (254)$$

□

#### APPENDIX Q PROOF OF PROPOSITION 3

As in the proof of Theorem 3, we show that  $\Delta_2 > 0$  and then apply Corollary 2.1. We use the following bounds and equalities:

$$B_{weak}(m) = r \left( \frac{N}{m} + m - 1 \right) \quad (255)$$

$$\xi_2(\mathcal{S}) \leq B_{weak}(m) \quad (256)$$

$$\xi_1(\mathcal{S}) \leq k + B_{weak}(m) \quad (257)$$

$$\xi_1(\mathcal{N}) = \sum_{i=1}^k \left( 1 - \frac{1}{1 + \mu \lambda_i} \right)^2 \quad (258)$$

$$\xi_2(\mathcal{N}) = \sum_{i=1}^N \left( \frac{1}{1 + \mu \lambda_i} \right)^2 \quad (259)$$

These are proven in Lemma 1, Appendix P Lemma 9, and Lemma 8. We also use that  $\xi_{2, bl}(\mathcal{S}) = \left\| \mathbf{R}_S[\mathbf{U}]_{S, \mathcal{K}} \right\|_F^2 \leq \left\| \mathbf{R}_S \right\|_F^2 = \xi_2(\mathcal{S})$ , which follows because  $[\mathbf{U}]_{\mathcal{N}, \mathcal{K}} [\mathbf{U}]_{\mathcal{N}, \mathcal{K}}^T = \mathbf{\Pi}_{bl(\mathcal{K})}$  is a projection and thus reduces the Frobenius norm. We in fact use the following lower bounds:

$$\xi_2(N) \geq N(1 + \mu \bar{\lambda})^{-2} \quad (260)$$

$$\xi_1(N) \geq 0 \quad (261)$$

where the first comes via applying Jensen's inequality, and the second by positivity of quadratics.

We therefore see that, as  $\Delta_i(\mathcal{N}, \mathcal{S}^c) = \xi_i(\mathcal{N}) - \xi_i(\mathcal{S})$ ,

$$\Delta_2(\mathcal{N}, \mathcal{S}^c) \geq N(1 + \mu\bar{\lambda})^{-2} - B_{weak}(m) \quad (262)$$

$$\Delta_1(\mathcal{N}, \mathcal{S}^c) \geq 0 - (k + B_{weak}(m)). \quad (263)$$

We now consider a sets  $\mathcal{S}$  of size  $\lceil \sqrt{N} \rceil$  and show that  $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ . By assumption  $2r\sqrt{N} < N$  so  $\frac{4\sqrt{N}}{(2r)^{-\frac{1}{2}}} > 1$  and therefore  $\mu_{ub\_weak} > 0$  and is real and it therefore possible to pick  $0 < \mu < \mu_{ub\_weak}$ . By assumption  $0 < \mu < \mu_{ub\_weak}$ , so

$$\frac{N}{(1 + \mu\bar{\lambda})^2} > \frac{N}{(1 + \mu_{ub\_weak}\bar{\lambda})^2} \quad (264)$$

$$= 2r\sqrt{N} \quad (265)$$

$$= r \left( \frac{N}{\sqrt{N}} + \sqrt{N} \right) \quad (266)$$

$$\geq r \left( \frac{N}{\lceil \sqrt{N} \rceil} + \lceil \sqrt{N} \rceil - 1 \right) = B_{weak}(\lceil \sqrt{N} \rceil) \quad (267)$$

And therefore by (262) (170). We then applying Corollary 2.1 in the same was as in the proof of Theorem 3, which mostly finishes the proof.

The remainder is computing  $\tau_{GLR\_weak}$ ; this is done in the same manner as in Theorem 3, except using the bounds for  $\Delta_1$  and  $\Delta_2$  presented earlier in this proof, along with  $B_{weak}(\lceil \sqrt{N} \rceil) \leq 2r\sqrt{N}$ . which was presented above.

#### APPENDIX R PROOF OF REMARK 9

We first show  $m_{opt} \leq \lceil \frac{N+1}{2} \rceil$ .

First note that if  $x > 0$  then  $\frac{1}{2} \left( \sqrt{x} + \frac{1}{\sqrt{x}} \right) \geq \sqrt{\sqrt{x} \frac{1}{\sqrt{x}}} = 1$  by the AM-GM inequality, and so  $\forall x, \omega(x) \geq 1$ . Secondly,  $\omega$  is increasing. Therefore

$$r \frac{N}{m} + m - 1 \leq B(m) \leq r \left( \frac{N}{m} + m - 1 \right). \quad (268)$$

If  $f(m) < g(m)$  then  $\min f(m) < \min g(m)$  so

$$2\sqrt{rN} - 1 \leq B(m_{opt}) < N \quad (269)$$

the LHS has a minimum at  $m = \sqrt{rN}$  with value  $2\sqrt{rN} - 1$  and so because  $\min(\text{LHS}) < \min B(m)$ , under the assumption  $B(m_{opt}) < N$  we have

$$2\sqrt{rN} - 1 \leq N \quad (270)$$

For  $m \geq \lfloor \frac{N}{2} \rfloor$ ,

$$B(m) = r \frac{N}{m} + m - 1 + c \quad (271)$$

for a constant  $c \geq 0$  not dependent on  $m$ . The function  $r \frac{N}{m} + m - 1 + c$  is increasing for  $m \geq \sqrt{rN}$  and by 270  $\lceil \sqrt{rN} \rceil \leq \lfloor \frac{N+1}{2} \rfloor$ . Therefore, considering that  $m$  is discrete,  $m_{opt} \leq \lfloor \frac{N+1}{2} \rfloor$ .

We now show that  $m_{opt} \in \left[ \lfloor \sqrt{N} \rfloor, \lceil \sqrt{rN} \rceil \right]$ . Write  $r_i = \omega \left( \max \left[ 1, \frac{N+2-i}{\lambda_i} \right] \right)$ . Then

$$B(m) = r \frac{N}{m} + \sum_{i=2}^m r_i \quad (272)$$

and

$$B(m+1) - B(m) = r_{m+1} - \frac{rN}{m(m+1)} \quad (273)$$

As  $B(m_{opt})$  is a global minimum it is a local minimum, so  $B(m_{opt}+1) \geq B(m_{opt})$  and  $B(m_{opt}-1) \geq B(m_{opt})$ . This can be written:

$$r_{m_{opt}} m_{opt} (m_{opt} + 1) \geq rN \quad (274)$$

$$r_{m_{opt}-1} m_{opt} (m_{opt} - 1) \leq rN \quad (275)$$

As  $1 \leq r_i \leq r$  for all  $i$ , we get

$$m_{opt} (m_{opt} + 1) \geq N \quad (276)$$

$$m_{opt} (m_{opt} - 1) \leq rN \quad (277)$$

and as  $(m+1)^2 > m(m+1)$  and  $(m-1)^2 < m(m-1)$ ,

$$\sqrt{N} - 1 < m_{opt} < \sqrt{rN} + 1 \quad (278)$$

as these inequalities are strict and  $m_{opt}$  is an integer,

$$\lfloor \sqrt{N} \rfloor \leq m_{opt} \leq \lceil \sqrt{rN} \rceil. \quad (279)$$

#### APPENDIX S SENSITIVITY ANALYSIS OF THEOREM 3 AND PROPOSITION 3

Proposition 3 is centered around the graph property  $r^3$  which measures the spread of the eigenvalues of  $L$ ;  $r$  is increasing in  $\frac{\lambda_N}{\lambda_2}$ .  $r$  can also be understood as a bound on degree spread; let  $d_{min}$  and  $d_{max}$  be the minimum and maximum node degrees on the graph, then  $4r \geq \frac{d_{max}}{d_{min}}$  (via the min-max theorem). From condition (43) we require  $r$  to be small for the theorem to hold – which means the degree spread needs to be narrow.

To better understand  $\mu_{ub\_weak}$  and  $\tau_{ub\_weak}$ , we present a sensitivity analysis; this analysis varies one parameter of  $\bar{\lambda}$ ,  $N$ ,  $r$  and  $\mu$  while holding the others constant.

TABLE VIII: Sensitivity analysis for Proposition 3 .

	$\bar{\lambda}$	$N$	$r$	$\mu$
$\mu_{ub\_weak}$	$\mathcal{O}(\bar{\lambda}^{-1})$	$\mathcal{O}(\sqrt[4]{N})$	$\mathcal{O}(r^{-\frac{1}{2}})$	N/A
$\tau_{GLR\_weak}$	$\mathcal{O}(\bar{\lambda}^{-2})$	$\mathcal{O}(N^{-\frac{1}{2}})$	$\mathcal{O}(r^{-1})$	$\mathcal{O}(\mu^{-2})$

We see both  $\mu_{ub\_weak}$  and  $\tau_{GLR\_weak}$  are decreasing in  $r$ , and furthermore the constraint (43) also requires low  $r$ . This shows that graphs with low  $r$  are more amenable to reconstruction with fewer observations via GLR.

Proposition 3 summarises the full spectrum of  $L$  via  $\lambda_2$ ,  $\bar{\lambda}$  and  $\lambda_N$  and thus can fail to show that reducing sample size

<sup>3</sup>We give some intuition for  $r$ : low  $r$ , equivalent to low  $\frac{\lambda_N}{\lambda_2}$ , is a known condition in the Network Synchronisation literature which allows for dynamic oscillators on a network to synchronise [48].

reduces MSE for graph models like the Barabasi-Albert graph model which have a few very large eigenvalues with an otherwise generally concentrated spectrum. Theorem 3 overcomes this by using the full spectrum; its broader applicability makes it worth understanding.

We briefly talk directly about the parameters of to Theorem 3; as  $r$  decreases,  $m_{opt}$  decreases to  $\sqrt{N}$  while  $\mu_{ub}$  and  $\tau_{GLR}$  increase, which is in line with our suggestion of  $r$  as a measure of amenability to reconstruction.

#### APPENDIX T ASYMPTOTIC RESULTS FOR GLR UNDER FULL-BAND NOISE

**Proposition 6.** Fix  $p \in (0, 1]$ . Consider graphs drawn from the distribution of random connected Erdős-Rényi graphs with  $N$  vertices and edge probability  $p$ . Then as  $N \rightarrow \infty$ , condition (41) in Theorem 3 holds w.h.p. . Furthermore, as  $N \rightarrow \infty$ ,

$$r \xrightarrow{p} 1, \quad m_{opt} \cdot N^{-\frac{1}{2}} \xrightarrow{p} 1, \quad \mu_{ub} \lambda_2 \xrightarrow{p} +\infty. \quad (280)$$

Assume  $\frac{k}{N}$  is fixed and choose  $\mu = \frac{c}{\lambda_2}$ , or  $\frac{c}{\lambda_N}$ , or  $\frac{c}{\sqrt{\lambda_2 \lambda_N}}$  for optimal bias-variance trade-off at  $\mathcal{S} = \mathcal{N}$  [37], then

$$\tau_{GLR} \rightarrow (1 + 2c)^{-1}. \quad (281)$$

*Proof Sketch.* By [49, Theorem 1],  $\lambda_2 \approx Np - \sqrt{2N \log N}$  and  $\lambda_N \approx Np + \sqrt{2N \log N}$  as  $N \rightarrow \infty$  so  $\frac{\lambda_N}{\lambda_2} \xrightarrow{p} 1$  and  $r \xrightarrow{p} 1$ . Approximately,  $B(m) \rightarrow \frac{N}{m} + m - 1$ , which lets us bound  $\mu_{ub}$  and  $\tau_{GLR}$ . See Appendix T for a full proof.  $\square$

*Proof.* We prove the proposition by proving some bounds relating terms to  $r$ ,  $\lambda_2$  and  $\lambda_N$ , then showing that  $r \rightarrow 1$ .

##### A. Notation and terms

While we use standard terminology in probability theory, for convenience we define some of it here.

An event  $E_n$  happens *with high probability* (abbreviated w.h.p.) if  $\lim_{n \rightarrow \infty} \mathbb{P}(E_n) = 1$ . A sequence of random variables  $X_n$  *converges in probability* to a random variable or constant  $X$  if  $\forall \epsilon > 0 : \mathbb{P}(|X_n - X| > \epsilon) \rightarrow 0$ . We write this as

$$X_n \xrightarrow{p} X. \quad (282)$$

Similarly, we write

$$X_n \xrightarrow{p} +\infty. \quad (283)$$

to mean that  $\forall c > 0 : \mathbb{P}(X < c) \rightarrow 0$

##### B. Limits in probability

In this section, we show  $r \xrightarrow{p} 1$  in our setting. Morally our argument is that  $\frac{\lambda_N}{\lambda_2} \approx \frac{Np + \sqrt{2N \log N}}{Np - \sqrt{2N \log N}} \xrightarrow{p} 1$ . We now prove this formally, starting with statements including disconnected graphs, and using them to derive results about connected

graphs. By [49, Theorem 1 (i) & (ii)], across all Erdős-Rényi graphs with edge probability  $p$

$$\frac{Np - \lambda_2}{\sqrt{N \log N}} \xrightarrow{p} \sqrt{2} \quad (284)$$

$$\frac{\lambda_N - Np}{\sqrt{N \log N}} \xrightarrow{p} \sqrt{2}. \quad (285)$$

By Slutsky's Theorem, and as convergence in distribution to a constant implies convergence in probability, we add and square the ratios to get

$$\frac{(\lambda_N - \lambda_2)^2}{N \log N} \xrightarrow{p} 8. \quad (286)$$

We now bound  $\frac{N \log N}{\lambda_2 \lambda_N}$ . By the definition of convergence in probability, because  $\frac{Np}{N \log N} \rightarrow \infty$  and by the triangle inequality,

$$\forall \epsilon > 0, \lim_{N \rightarrow \infty} \mathbb{P} \left( \left| \frac{Np - \lambda_2}{\sqrt{N \log N}} - \sqrt{2} \right| > \epsilon \right) = 0 \quad (287)$$

$$\implies \forall c > 0, \lim_{N \rightarrow \infty} \mathbb{P} \left( \left| \frac{\lambda_2}{\sqrt{N \log N}} \right| < c \right) = 0 \quad (288)$$

$$\implies \forall \epsilon > 0, \lim_{N \rightarrow \infty} \mathbb{P} \left( \left| \frac{\sqrt{N \log N}}{\lambda_2} \mathbb{1}\{\lambda_2 > 0\} \right| > \epsilon \right) = 0 \quad (289)$$

therefore

$$\frac{\sqrt{N \log N}}{\lambda_2} \mathbb{1}\{\lambda_2 > 0\} \xrightarrow{p} 0 \quad (290)$$

As  $0 < \frac{\sqrt{N \log N}}{\lambda_N} < \frac{\sqrt{N \log N}}{\lambda_2}$ ,

$$\frac{N \log N}{\lambda_2 \lambda_N} \mathbb{1}\{\lambda_2 > 0\} \xrightarrow{p} 0. \quad (291)$$

we use Slutsky to multiply this with (286) to get

$$(r - 1) \mathbb{1}\{\lambda_2 > 0\} = \frac{(\lambda_N - \lambda_2)^2}{4\lambda_2 \lambda_N} \mathbb{1}\{\lambda_2 > 0\} \xrightarrow{p} 0 \quad (292)$$

By (288) we know  $\lambda_2 > 0$  w.h.p. . Therefore, for any  $\epsilon > 0$

$$\mathbb{P}(|r - 1| > \epsilon \mid \lambda_2 > 0) \quad (293)$$

$$= \frac{\mathbb{P}(|r - 1| > \epsilon \cap \lambda_2 > 0)}{\mathbb{P}(\lambda_2 > 0)} \quad (294)$$

$$= \frac{\mathbb{P}(|r - 1| \mathbb{1}\{\lambda_2 > 0\} > \epsilon)}{\mathbb{P}(\lambda_2 > 0)} \rightarrow \frac{0}{1} \quad (295)$$

Therefore under our setting of considering only connected graphs (i.e. that  $\lambda_2 > 0$ ),  $r \xrightarrow{p} 1$ .

##### C. Parameter bounds

We first bound  $B(m)$  in terms of  $r$ . Note that  $\omega$  is increasing and therefore for  $2 \leq i \leq N$ ,

$$\omega \left( \max \left[ 1, \frac{\lambda_{N+2-i}}{\lambda_i} \right] \right) \leq \omega \left( \frac{\lambda_N}{\lambda_2} \right) = r. \quad (296)$$

If  $x > 0$  then  $\frac{1}{2} \left( \sqrt{x} + \frac{1}{\sqrt{x}} \right) \geq \sqrt{\sqrt{x} \frac{1}{\sqrt{x}}} = 1$  by the AM-GM inequality, and so  $\forall x, \omega(x) \geq 1$ . Therefore

$$1 \leq \omega \left( \max \left[ 1, \frac{\lambda_{N+2-i}}{\lambda_i} \right] \right) \leq r \quad (297)$$

$$\frac{N}{m} + m - 1 \leq B(m) \leq r \left( \frac{N}{m} + m - 1 \right). \quad (298)$$

We also have that

$$\frac{1}{\lambda_N} \left( \sqrt{\frac{N}{B(m)}} - 1 \right) \leq \mu_{ub}(m) \leq \frac{1}{\lambda_2} \left( \sqrt{\frac{N}{B(m)}} - 1 \right). \quad (299)$$

#### D. Pulling it together

We already have that  $r \xrightarrow{P} 1$ . As  $\sqrt{N} \leq m_{opt} \leq \lceil \sqrt{rN} \rceil$ ,  $\frac{m_{opt}}{\sqrt{N}} \xrightarrow{P} 1$ . By the squeeze theorem, for a fixed  $m$ ,

$$\frac{B(m)}{N} \xrightarrow{P} \frac{1}{m}. \quad (300)$$

Because  $\frac{1}{m} < 1$  for  $m > 1$ , condition (41) in Theorem 3 holds w.h.p. as  $N \rightarrow \infty$ . As  $B(m) > 0$ ,

$$\sqrt{\frac{N}{B(m)}} - 1 \xrightarrow{P} \sqrt{m} - 1 \quad (301)$$

Conditioning on  $\lambda_2 > 0$  does not change (288), so it applies when we only consider the set of connected Erdős-Rényi graphs as well. By setting  $c = \sqrt{m} - 1$  in (288), we get

$$\frac{1}{\lambda_2} \left( \sqrt{\frac{N}{B(m)}} - 1 \right) \xrightarrow{P} 0. \quad (302)$$

and therefore for fixed  $m$ ,  $\mu_{ub}(m) \rightarrow 0$ . We now calculate  $\mu_{ub}(m_{opt})$ . First note that as  $x \mapsto x + \frac{1}{x}$  invertible and monotone on  $x \in (0, 1)$ , so its inverse is continuous. Consider that  $\frac{\lambda_2}{\lambda_N} \in (0, 1)$ . As the inverse is continuous, we have that  $\forall \epsilon > 0 \exists \delta > 0 : \left| \left( \frac{\lambda_2}{\lambda_N} + \frac{\lambda_N}{\lambda_2} \right) - 2 \right| < \epsilon \implies \left| \frac{\lambda_2}{\lambda_N} - 1 \right| < \delta$ . Using this with the definition of convergence in probability and that  $r \xrightarrow{P} 1$  gives

$$\frac{\lambda_2}{\lambda_N} \xrightarrow{P} 1. \quad (303)$$

Next, note that

$$\frac{\sqrt{N}}{m_{opt}} + \frac{m_{opt}}{\sqrt{N}} - \frac{1}{\sqrt{N}} \leq \frac{B(m_{opt})}{\sqrt{N}} \leq r \left( \frac{\sqrt{N}}{m_{opt}} + \frac{m_{opt}}{\sqrt{N}} \right). \quad (304)$$

so

$$\frac{m_{opt}}{\sqrt{N}} \rightarrow 1 \implies \frac{B(m_{opt})}{\sqrt{N}} \xrightarrow{P} 2. \quad (305)$$

Therefore (41) in Theorem 3 holds for  $m_{opt}$  w.h.p. as  $N \rightarrow \infty$ . Also

$$\frac{\mu_{ub}(m_{opt}) \lambda_2}{\sqrt[4]{N}} \geq \frac{\lambda_2}{\lambda_N} \left( \sqrt{\frac{\sqrt{N}}{B(m_{opt})}} - \frac{1}{\sqrt[4]{N}} \right) \quad (306)$$

$$\xrightarrow{P} \frac{1}{\sqrt{2}} \quad (307)$$

and therefore  $\mu_{ub}(m_{opt}) \lambda_2 \xrightarrow{P} \infty$ .

Finally, we bound  $\tau_{GLR}$ . We first take limits of  $\left( \frac{1}{1+\mu\lambda_i} \right)^2$ . Note that

$$1 \leq \frac{\lambda_i}{\lambda_2} \leq \frac{\lambda_N}{\lambda_2} \xrightarrow{P} 1, \quad (308)$$

$$1 \geq \frac{\lambda_i}{\lambda_N} \geq \frac{\lambda_2}{\lambda_N} \xrightarrow{P} 1 \quad (309)$$

and  $\frac{\lambda_i}{\sqrt{\lambda_2 \lambda_N}} \in \left[ \frac{\lambda_i}{\lambda_N}, \frac{\lambda_i}{\lambda_2} \right]$  and so  $\frac{\lambda_i}{\sqrt{\lambda_2 \lambda_N}} \xrightarrow{P} 1$ . Therefore, for all three choices of  $\mu$ ,  $\mu\lambda_i \xrightarrow{P} c$ . As

$$\frac{1}{N} + \frac{N-1}{N} \left( \frac{1}{1+\mu\lambda_N} \right)^2 \leq \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{1+\mu\lambda_i} \right)^2 \quad (310)$$

$$\frac{1}{N} + \frac{N-1}{N} \left( \frac{1}{1+\mu\lambda_2} \right)^2 \geq \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{1+\mu\lambda_i} \right)^2 \quad (311)$$

For all three choices of  $\mu$ ,

$$\frac{1}{N} \sum_{i=1}^N \left( \frac{1}{1+\mu\lambda_i} \right)^2 \rightarrow \left( \frac{1}{1+c} \right)^2. \quad (312)$$

Similarly,

$$\frac{1}{N} \sum_{i=1}^N \left( 1 - \frac{1}{1+\mu\lambda_i} \right)^2 \rightarrow \left( 1 - \frac{1}{1+c} \right)^2. \quad (313)$$

As  $\frac{B(m_{opt})}{\sqrt{N}} \xrightarrow{P} 2$ , we must have that  $\frac{B(m_{opt})}{N} \xrightarrow{P} 0$  and as  $\frac{k}{N}$  is constant,  $\frac{B(m_{opt})}{k} \xrightarrow{P} 0$ . Therefore, by Slutsky, for the three choices of optimal  $\mu$ ,

$$\tau_{GLR}(\mu, m_{opt}) \rightarrow \frac{\left( \frac{1}{1+c} \right)^2}{1 - \left( 1 - \frac{1}{1+c} \right)^2} = \frac{1}{1+2c}. \quad (314)$$

□

#### APPENDIX U PROOF OF LEMMA 3

We follow the structure and notation of Appendix 1 and follow the same broad outline. However, there are two additional complexities: firstly, we cannot remove the dependence on  $\mu$  at the same stage as we did in the proof of Lemma 1 because the terms are not of the form  $\mathbf{X} [\mu \mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-2} \mathbf{X}^T$ , as  $[\mathbf{U}]_{\mathcal{S}, \mathcal{K}}$  breaks the symmetry we were using. We overcome this by only removing the dependence on  $\mu$  at the end of the proof.

Secondly, we cannot directly apply the matrix versions of Kantorovich's inequality, because  $\mathbf{\Pi}_{\Theta}$  and  $\mathbf{\Pi}_{bl(\mathcal{K})}$  do not commute, thus breaking the assumptions needed. We prove two bandlimited variants of Kantorovich instead, one which works by directly building the proof from the ground up, and the second of which works by finding a matrix that commutes with  $\mathbf{L}\mathbf{L}^\dagger$  and upper bounds  $\mathbf{\Pi}_{bl(\mathcal{K})}$  without weakening the proof too much.

### A. Notation and useful matrices

While we follow the notation described in Appendix V, we also define the following matrices:

$$\mathbf{L}_\mu^\dagger = \mu \mathbf{I} + \mathbf{L}^\dagger \quad (315)$$

$$\mathbf{\Pi}_L = \mathbf{I} - \frac{1}{N} \mathbf{1}_{N \times N} = \mathbf{L}^\dagger \mathbf{L} = \mathbf{L} \mathbf{L}^\dagger \quad (316)$$

where the latter can be seen to be a projection operator as  $\mathbf{\Pi}_L = \mathbf{\Pi}_L^T = \mathbf{\Pi}_L^2$ . Furthermore,  $\mathbf{\Pi}_L \mathbf{L}^\dagger = \mathbf{L}^\dagger$  and because  $\mathbf{1}_N [\mathbf{I}]_{\mathcal{N}, \Theta} = 0$ ,  $\mathbf{\Pi}_L [\mathbf{I}]_{\mathcal{N}, \Theta} = [\mathbf{I}]_{\mathcal{N}, \Theta}$ .

### B. Proof overview

Overall we show that

$$\xi_{2,bl}(\mathcal{S}) = \|\mathbf{R}_S [\mathbf{U}]_{\mathcal{S}, \mathcal{K}}\|_F^2 \quad (317)$$

$$= 1 + \left\| \mathbf{R}_S [\mathbf{U}]_{\mathcal{S}, 2:k} \right\|_F^2 \quad (318)$$

$$= 1 + \left\| [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mu \mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{U}]_{\Theta, 2:k} \right\|_F^2 \quad (319)$$

$$+ \frac{N}{m} \left\| \mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta, \mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N}, 2:k} \right\|_2^2 \quad (320)$$

$$\leq 1 + B_k(m) \quad (321)$$

by proving that

$$\frac{N}{m} \left\| \mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta, \mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N}, 2:k} \right\|_2^2 \quad (322)$$

$$\leq \frac{N}{m} r_{bl} \quad (323)$$

and

$$\left\| [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mu \mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{U}]_{\Theta, 2:k} \right\|_F^2 \quad (324)$$

$$\leq \sum_{i=1}^m \omega \left( \max \left[ 1, \frac{\lambda_{k+2-i}}{\lambda_i} \right] \right). \quad (325)$$

### C. Row decomposition

Noting that  $\mathbf{R}_S [\mathbf{u}_1]_S = \frac{1}{\sqrt{N}} \mathbf{R}_S \mathbf{1}_S = \frac{1}{\sqrt{N}} \mathbf{1}_N$ ,

$$\xi_{2,bl}(\mathcal{S}) = \|\mathbf{R}_S [\mathbf{U}]_{\mathcal{S}, \mathcal{K}}\|_F^2 \quad (326)$$

$$= \|\mathbf{R}_S [\mathbf{u}_1]_S\|_2^2 + \left\| \mathbf{R}_S [\mathbf{U}]_{\mathcal{S}, 2:k} \right\|_F^2 \quad (327)$$

$$= 1 + \left\| \mathbf{R}_S [\mathbf{U}]_{\mathcal{S}, 2:k} \right\|_F^2. \quad (328)$$

### D. Explicit submatrix computation

We first show

$$(\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{\Pi}_L \quad (329)$$

$$= \frac{1}{\mu} \mathbf{P}^T \left( \mathbf{L}^\dagger - [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger]_{\Theta, \mathcal{N}} \right) \quad (330)$$

*Proof.* From Appendix V,  $(\mathbf{\Pi}_S + \mu \mathbf{L}) \mathbf{P}^T = \mathbf{\Pi}_\Theta + \mu \mathbf{L}$ , so

$$(\mathbf{\Pi}_S + \mu \mathbf{L}) \frac{1}{\mu} \mathbf{P}^T \mathbf{L}^\dagger = \left( \frac{1}{\mu} \mathbf{\Pi}_\Theta \mathbf{L}^\dagger + \mathbf{\Pi}_L \right) \quad (331)$$

as  $[\mathbf{\Pi}_L]_{\mathcal{N}, \Theta} = [\mathbf{I}]_{\mathcal{N}, \Theta}$ ,

$$(\mathbf{\Pi}_S + \mu \mathbf{L}) \frac{1}{\mu} \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} \quad (332)$$

$$= \left( \frac{1}{\mu} \mathbf{\Pi}_\Theta \mathbf{L}^\dagger + \mathbf{\Pi}_L \right) [\mathbf{I}]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} \quad (333)$$

$$= \frac{1}{\mu} [\mathbf{I}]_{\mathcal{N}, \Theta} \left( [\mathbf{L}^\dagger]_{\Theta} + \mu [\mathbf{I}]_{\Theta} \right) [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} \quad (334)$$

$$= \frac{1}{\mu} [\mathbf{I}]_{\mathcal{N}, \Theta} \quad (335)$$

Then

$$(\mathbf{\Pi}_S + \mu \mathbf{L}) \frac{1}{\mu} \mathbf{P}^T [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger]_{\Theta, \mathcal{N}} \quad (336)$$

$$= \frac{1}{\mu} [\mathbf{I}]_{\mathcal{N}, \Theta} [\mathbf{L}^\dagger]_{\Theta, \mathcal{N}} = \frac{1}{\mu} \mathbf{\Pi}_\Theta \mathbf{L}^\dagger \quad (337)$$

Therefore

$$(\mathbf{\Pi}_S + \mu \mathbf{L}) \frac{1}{\mu} \mathbf{P}^T \left( \mathbf{L}^\dagger - [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger]_{\Theta, \mathcal{N}} \right) \quad (338)$$

$$= \left( \frac{1}{\mu} \mathbf{\Pi}_\Theta \mathbf{L}^\dagger + \mathbf{\Pi}_L \right) - \frac{1}{\mu} \mathbf{\Pi}_\Theta \mathbf{L}^\dagger \quad (339)$$

$$= \mathbf{\Pi}_L. \quad (340)$$

Multiplying both sides by  $(\mathbf{\Pi}_S + \mu \mathbf{L})$  finishes the proof.  $\square$

We use this to compute  $(\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{\Pi}_S$ .

$$(\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} (\mathbf{\Pi}_S + \mu \mathbf{L}) = \mathbf{I} \quad (341)$$

means

$$(\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{\Pi}_S \quad (342)$$

$$= \mathbf{I} - \mu (\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{L} \quad (343)$$

$$= \mathbf{I} - \mu (\mathbf{\Pi}_S + \mu \mathbf{L})^{-1} \mathbf{\Pi}_L \mathbf{L} \quad (344)$$

$$= \mathbf{I} - \mathbf{P}^T \left( \mathbf{L}^\dagger - [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} [\mathbf{L}^\dagger]_{\Theta, \mathcal{N}} \right) \mathbf{L} \quad (345)$$

$$= \mathbf{I} - \mathbf{P}^T \left( \mathbf{\Pi}_L - [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta, \mathcal{N}} \right) \quad (346)$$

As

$$\mathbf{\Pi}_L [\mathbf{U}]_{\mathcal{N}, 2:k} = [\mathbf{U}]_{\mathcal{N}, 2:k}, \quad (347)$$

$$\mathbf{P}^T = \mathbf{I} - \frac{1}{m} \mathbf{1}_{N \times m} [\mathbf{I}]_{\mathcal{S}, \mathcal{N}} = \mathbf{I} - \sqrt{\frac{N}{m}} \mathbf{u}_1 \mathbf{v}_1^T \quad (348)$$

where  $\mathbf{u}_1 = \frac{1}{\sqrt{N}} \mathbf{1}_N$  is the null eigenvector of  $\mathbf{L}$  and  $\mathbf{v}_1 = \frac{1}{\sqrt{m}} \mathbf{1}_S$ . We have

$$\mathbf{R}_S [\mathbf{U}]_{\mathcal{S}, 2:k} \quad (349)$$

$$= [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mu \mathbf{I} + \mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{U}]_{\Theta, 2:k} \quad (350)$$

$$+ \sqrt{\frac{N}{m}} \mathbf{u}_1 \mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}_\mu^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta, \mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N}, 2:k} \quad (351)$$

Because  $\mathbf{X}^T \mathbf{L}^\dagger \mathbf{u}_1 \mathbf{y}^T = \mathbf{0}$ , we have  $\|\mathbf{L}^\dagger \mathbf{X} + \mathbf{u}_1 \mathbf{y}\|_F^2 = \|\mathbf{L}^\dagger \mathbf{X}\|_F^2 + \|\mathbf{y}\|_2^2$  for any matrix  $\mathbf{X}$  and any vector  $\mathbf{y}$ . This gives that

$$\left\| \mathbf{R}_S [\mathbf{U}]_{S,2:k} \right\|_F^2 \quad (352)$$

$$= \left\| [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger_\mu]_\Theta^{-1} [\mathbf{U}]_{\Theta,2:k} \right\|_F^2 \quad (353)$$

$$+ \frac{N}{m} \left\| \mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger_\mu]_\Theta^{-1} [\mathbf{I}]_{\Theta,\mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N},2:k} \right\|_F^2 \quad (354)$$

### E. Bounding (353)

We show that

$$\left\| [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger_\mu]_\Theta^{-1} [\mathbf{U}]_{\Theta,2:k} \right\|_F^2 \leq \sum_{i=2}^m \omega \left( \max \left[ 1, \frac{\lambda_{k+2-i}}{\lambda_i} \right] \right). \quad (355)$$

*Proof.* We seek to apply [39, Theorem 5].

For this, we first construct some matrices which are diagonal in the eigenbasis of  $\mathbf{L}$ . We present them by their values on  $\mathbf{u}_1$ ,  $[\mathbf{U}]_{\mathcal{N},2:k}$  and  $\mathbf{I} - \mathbf{\Pi}_{bl(\mathcal{K})}$  in Table IX. We write

$$\mathbf{\Pi}_{bl(2:k)} = [\mathbf{U}]_{\mathcal{N},2:k} [\mathbf{U}]_{\mathcal{N},2:k}^T.$$

TABLE IX: Matrices diagonal in the eigenbasis of  $\mathbf{L}$ .

	$\mathbf{u}_1$	$\mathbf{u}_2, \dots, \mathbf{u}_k$	$\mathbf{u}_{k+1}, \dots, \mathbf{u}_N$
$\mathbf{G}$	0	$\sqrt{\lambda_i(\mu + \frac{1}{\lambda_i})}$	$\sqrt{\lambda_{\lceil \frac{k}{2} \rceil}(\mu + \frac{1}{\lambda_i})}$
$\mathbf{H}$	0	$\sqrt{\frac{1}{\lambda_i}(\mu + \frac{1}{\lambda_i})}$	$\sqrt{\frac{1}{\lambda_{\lceil \frac{k}{2} \rceil}}(\mu + \frac{1}{\lambda_i})}$
$\mathbf{GH}$	0	$\mu + \frac{1}{\lambda_i}$	$\mu + \frac{1}{\lambda_i}$
$\mathbf{H}^\dagger \mathbf{G}$	0	$\lambda_i$	$\lambda_{\lceil \frac{k}{2} \rceil}$
$\mathbf{G}^2$	0	$1 + \mu \lambda_i$	$\lambda_{\lceil \frac{k}{2} \rceil}(\mu + \frac{1}{\lambda_i})$
$\mathbf{\Pi}_{bl(2:k)}$	0	1	0
$\mathbf{H}^2$	0	$\frac{1}{\lambda_i}(\mu + \frac{1}{\lambda_i})$	$\frac{1}{\lambda_{\lceil \frac{k}{2} \rceil}}(\mu + \frac{1}{\lambda_i})$
$(\mathbf{L}^\dagger)^2$	0	$\frac{1}{\lambda_i}^2$	$\frac{1}{\lambda_i}^2$

From Table IX we have constructed  $\mathbf{G}^2$  and  $\mathbf{H}^2$  such that

$$\mathbf{\Pi}_{bl(2:k)} \leq \mathbf{G}^2 \quad (356)$$

$$(\mathbf{L}^\dagger)^2 \leq \mathbf{H}^2 \quad (357)$$

By the way we constructed  $\Theta$ ,

$$[\mathbf{GH}]_\Theta = \left[ \left( \mathbf{I} - \frac{1}{N} \mathbf{1}_{N \times N} \right) (\mu \mathbf{I} + \mathbf{L}^\dagger) \right]_\Theta = [\mu \mathbf{I} + \mathbf{L}^\dagger]_\Theta. \quad (358)$$

We now note that  $\mathbf{A}^2 \leq \mathbf{B}^2$  means  $\forall \mathbf{x} : \mathbf{x}^T \mathbf{A}^2 \mathbf{x} \leq \mathbf{x}^T \mathbf{B}^2 \mathbf{x}$  so for all compatible  $\mathbf{V}$ ,  $\text{tr}(\mathbf{V}^T \mathbf{A}^2 \mathbf{V}) \leq \text{tr}(\mathbf{V}^T \mathbf{B}^2 \mathbf{V})$ . Therefore

$$\left\| [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mu \mathbf{I} + \mathbf{L}^\dagger]_\Theta^{-1} [\mathbf{U}]_{\Theta,2:k} \right\|_F^2 \quad (359)$$

$$= \left\| [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{GH}]_\Theta^{-1} [\mathbf{U}]_{\Theta,2:k} \right\|_F^2 \quad (360)$$

$$= \text{tr} \left( [\mathbf{U}]_{\Theta,2:k}^T [\mathbf{GH}]_\Theta^{-1} [(\mathbf{L}^\dagger)^2]_\Theta [\mathbf{GH}]_\Theta^{-1} [\mathbf{U}]_{\Theta,2:k} \right) \quad (361)$$

$$\leq \text{tr} \left( [\mathbf{U}]_{\Theta,2:k}^T [\mathbf{GH}]_\Theta^{-1} [\mathbf{H}^2]_\Theta [\mathbf{GH}]_\Theta^{-1} [\mathbf{U}]_{\Theta,2:k} \right) \quad (362)$$

$$= \text{tr} \left( [\mathbf{GH}]_\Theta^{-1} [\mathbf{H}^2]_\Theta [\mathbf{GH}]_\Theta^{-1} [\mathbf{\Pi}_{bl(2:k)}]_\Theta \right) \quad (363)$$

$$= \text{tr} \left( [\mathbf{H}]_{\mathcal{N},\Theta} [\mathbf{GH}]_\Theta^{-1} [\mathbf{\Pi}_{bl(2:k)}]_\Theta [\mathbf{GH}]_\Theta^{-1} [\mathbf{H}]_{\Theta,\mathcal{N}} \right) \quad (364)$$

$$\leq \text{tr} \left( [\mathbf{H}]_{\mathcal{N},\Theta} [\mathbf{GH}]_\Theta^{-1} [\mathbf{G}^2]_\Theta [\mathbf{GH}]_\Theta^{-1} [\mathbf{H}]_{\Theta,\mathcal{N}} \right) \quad (365)$$

$$= \text{tr} \left( [\mathbf{GH}]_\Theta^{-1} [\mathbf{G}^2]_\Theta [\mathbf{GH}]_\Theta^{-1} [\mathbf{H}^2]_\Theta \right) \quad (366)$$

$$\leq \sum_{i=2}^m \omega \left( \max \left[ 1, \frac{\lambda_{k+2-i}}{\lambda_i} \right] \right). \quad (367)$$

where the equalities are via the cyclic properties or via (358).

The final inequality is by applying [39, Theorem 5] with  $\mathbf{X} = \mathbf{Y} = [\mathbf{I}]_{\mathcal{N},\Theta}$ . One can see that the conditions of the Theorem are fulfilled by the way we constructed  $\mathbf{G}$  and  $\mathbf{H}$  in Table IX, that  $\omega_i = \omega \left( \max \left[ 1, \frac{\lambda_{k+2-i}}{\lambda_i} \right] \right)$  in the  $t \geq k + s$  case and the  $t < k + s$  case can be weakened to our result.  $\square$

### F. A bandlimited Kantorovich lemma

We require the following lemma to bound (354).

**Lemma 10.** Let  $\mathbf{A}, \mathbf{B}$  be positive semi-definite, and let  $\mathbf{\Pi}_k$  be a projection to the eigenbasis of  $\mathbf{A}$  s.t.

$$\exists 0 < m \leq M : m \mathbf{\Pi}_k \leq \mathbf{A} \mathbf{\Pi}_k \leq M \mathbf{\Pi}_k. \quad (368)$$

i.e. it bandlimits  $\mathbf{A}$  to between  $m$  and  $M$ . Furthermore, assume  $\mathbf{A}$  has no eigenvalues less than  $m$ .

Assume  $\mathbf{B} \mathbf{A} = \mathbf{A} \mathbf{B}$  and  $\mathbf{B} \leq \mathbf{I}$ . Then for any projection  $\mathbf{\Pi}_{\mathcal{X}}$  where  $\mathbf{\Pi}_{\mathcal{X}} \mathbf{A} \mathbf{A}^\dagger \mathbf{\Pi}_{\mathcal{X}}$  is idempotent,

$$\sigma_{max} \left( \sqrt{\mathbf{B} \mathbf{A}^\dagger \mathbf{\Pi}_{\mathcal{X}}} \sqrt{\mathbf{A}} \sqrt{\mathbf{A} \mathbf{\Pi}_k} \right) \leq \sqrt{\frac{(M+m)^2}{4Mm}}. \quad (369)$$

*Proof.* We abuse our notation to write  $\mathbf{\Pi}_{\mathcal{X}} = [\mathbf{I}]_{\mathcal{N},\mathcal{X}} [\mathbf{I}]_{\mathcal{X},\mathcal{N}}$ . Then

$$\sqrt{\mathbf{A}^\dagger \mathbf{\Pi}_{\mathcal{X}}} \sqrt{\mathbf{A}} = \left[ \sqrt{\mathbf{A}^\dagger} \right]_{\mathcal{N},\mathcal{X}} \left[ \sqrt{\mathbf{A}} \right]_{\mathcal{N},\mathcal{X}}. \quad (370)$$

Throughout this proof we will use the fact that  $\mathbf{\Pi}_{\mathcal{A}} = \mathbf{A} \mathbf{A}^\dagger$  is a projection to the full range of  $\mathbf{A}$  and thus  $\mathbf{\Pi}_{\mathcal{A}} \geq \mathbf{\Pi}_k$  (consider diagonalizing both, and note all eigenvalues are 1 or 0). Note that

$$m \mathbf{A}^\dagger (\mathbf{I} - \mathbf{\Pi}_k) \leq (\mathbf{I} - \mathbf{\Pi}_k) \quad (371)$$

because  $\mathbf{A}$  has no non-zero eigenvalues less than  $m$ , so  $\mathbf{A}^\dagger$  has no non-zero eigenvalues greater than  $\frac{1}{m}$ . Thus

$$M m \mathbf{A}^\dagger (\mathbf{I} - \mathbf{\Pi}_k) \leq M (\mathbf{I} - \mathbf{\Pi}_k) \quad \text{and so} \quad (372)$$

$$(M+m)(\mathbf{I} - \mathbf{\Pi}_k) - M m \mathbf{A}^\dagger (\mathbf{I} - \mathbf{\Pi}_k) \geq 0. \quad (373)$$

From [50, Theorem 2, (3.3)],

$$\mathbf{A}\mathbf{\Pi}_k \leq (m + M)\mathbf{\Pi}_k - Mm\mathbf{A}^\dagger\mathbf{\Pi}_k \quad (374)$$

$$\leq (m + M)\mathbf{\Pi}_{\mathcal{A}} - Mm\mathbf{A}^\dagger \quad (375)$$

where the second inequality comes from adding on (373)

Conjugating by  $[\mathbf{I}]_{\mathcal{N},\mathcal{X}}$  and  $[\mathbf{I}]_{\mathcal{X},\mathcal{N}}$  gives:

$$[\mathbf{A}\mathbf{\Pi}_k]_{\mathcal{X}} \leq (M + m)[\mathbf{\Pi}_{\mathcal{A}}]_{\mathcal{X}} - Mm[\mathbf{A}^\dagger]_{\mathcal{X}} \quad (376)$$

We assumed  $\mathbf{\Pi}_{\mathcal{X}}\mathbf{\Pi}_{\mathcal{A}}\mathbf{\Pi}_{\mathcal{X}}$  is idempotent, and it implies that  $[\mathbf{\Pi}_{\mathcal{A}}]_{\mathcal{X}}$  is idempotent so we can complete the square giving:

$$[\mathbf{A}\mathbf{\Pi}_k]_{\mathcal{X}} \leq \frac{(M + m)^2}{4Mm} [\mathbf{A}^\dagger]_{\mathcal{X}}^\dagger. \quad (377)$$

In more detail: this is the final step of the proof of [50, Theorem 2] along with the observation that the constant is invariant to the transformation  $(m, M) \mapsto (\frac{1}{m}, \frac{1}{M})$ .

Conjugating (377) by  $[\sqrt{\mathbf{A}^\dagger}]_{\mathcal{N},\mathcal{X}}$  and  $[\sqrt{\mathbf{A}^\dagger}]_{\mathcal{N},\mathcal{X}}^T$ :

$$[\sqrt{\mathbf{A}^\dagger}]_{\mathcal{N},\mathcal{X}} [\mathbf{A}\mathbf{\Pi}_k]_{\mathcal{X}} [\sqrt{\mathbf{A}^\dagger}]_{\mathcal{X},\mathcal{N}} \quad (378)$$

$$\leq \frac{(M + m)^2}{4Mm} [\sqrt{\mathbf{A}^\dagger}]_{\mathcal{N},\mathcal{X}} [\mathbf{A}^\dagger]_{\mathcal{X}}^\dagger [\sqrt{\mathbf{A}^\dagger}]_{\mathcal{N},\mathcal{X}}. \quad (379)$$

Note that

$$\mathbf{\Pi}_\dagger = [\sqrt{\mathbf{A}^\dagger}]_{\mathcal{N},\mathcal{X}} [\mathbf{A}^\dagger]_{\mathcal{X}}^\dagger [\sqrt{\mathbf{A}^\dagger}]_{\mathcal{N},\mathcal{X}} \quad (380)$$

is a projection because  $\mathbf{\Pi}_\dagger = \mathbf{\Pi}_\dagger^T = \mathbf{\Pi}_\dagger^2$ , and thus its eigenvalues are 0 or 1 (i.e the real roots of  $\lambda = \lambda^2$ ). Thus, for any vector  $\mathbf{v}$  with  $\|\mathbf{v}\|_2 = 1$ ,

$$\left\| \mathbf{v}^T \sqrt{\mathbf{B}\mathbf{A}^\dagger} \mathbf{\Pi}_{\mathcal{X}} \sqrt{\mathbf{A}\mathbf{\Pi}_k} \right\|_2^2 \quad (381)$$

$$= \mathbf{v}^T [\sqrt{\mathbf{B}\mathbf{A}^\dagger}]_{\mathcal{N},\mathcal{X}} [\mathbf{A}\mathbf{\Pi}_k]_{\mathcal{X}} [\sqrt{\mathbf{A}^\dagger\mathbf{B}}]_{\mathcal{X},\mathcal{N}} \mathbf{v} \quad (382)$$

$$\leq \frac{(M + m)^2}{4Mm} \mathbf{v}^T \sqrt{\mathbf{B}\mathbf{\Pi}_\dagger} \sqrt{\mathbf{B}\mathbf{v}} \quad (383)$$

$$\leq \frac{(M + m)^2}{4Mm} \left\| \sqrt{\mathbf{B}\mathbf{v}} \right\|_2^2 \quad (384)$$

$$\leq \frac{(M + m)^2}{4Mm} \quad (385)$$

where we have  $\sqrt{\mathbf{A}\mathbf{\Pi}_k}\sqrt{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}_\dagger$  because  $\sqrt{\mathbf{A}}$  commutes with  $\mathbf{\Pi}_k$  (as they are both diagonal in the same basis).

As (381 - 385) holds true for all normalised  $\mathbf{v}$ , taking square roots gives the proof.  $\square$

### G. Bounding (354)

We show that

$$\frac{N}{m} \left\| \mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta,\mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N},2:k} \right\|_2^2 \leq \frac{N}{m} r_{bl}. \quad (386)$$

*Proof.* First note that  $\mathbf{v}_1^T [\mathbf{I}]_{\mathcal{N},\Theta} = \mathbf{0}$ ,  $\mathbf{\Pi}_{\mathcal{L}} [\mathbf{I}]_{\mathcal{N},\Theta} = [\mathbf{I}]_{\mathcal{N},\Theta}$  and so  $\mathbf{v}_1^T [\mu\mathbf{\Pi}_{\mathcal{L}}]_{\mathcal{N},\Theta} = \mu\mathbf{v}_1^T [\mathbf{I}]_{\mathcal{N},\Theta} = 0$  so

$$\mathbf{v}_1^T [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} = \mathbf{v}_1^T [\mathbf{\Pi}_{\mathcal{L}} \mathbf{L}^\dagger]_{\mathcal{N},\Theta} \quad (387)$$

$$= \mathbf{v}_1^T [\mu\mathbf{\Pi}_{\mathcal{L}} + \mathbf{\Pi}_{\mathcal{L}} \mathbf{L}^\dagger \mathbf{\Pi}_{\mathcal{L}}]_{\mathcal{N},\Theta} \quad (388)$$

$$= \mathbf{v}_1^T [\mathbf{\Pi}_{\mathcal{L}} \mathbf{L}^\dagger \mathbf{\Pi}_{\mathcal{L}}]_{\mathcal{N},\Theta} \quad (389)$$

Write  $\mathbf{\Pi}_{\mathcal{L}} \mathbf{L}^\dagger \mathbf{\Pi}_{\mathcal{L}} = \mathbf{L}_{\mu-}^\dagger$ . This zeros the smallest eigenvalue of  $\mathbf{L}_{\mu-}^\dagger$  which is  $\mu$ , meaning the eigenvalues of  $\mathbf{L}_{\mu-}^\dagger$  are between  $\mu + \frac{1}{\lambda_{\mathcal{N}}}$  and  $\mu + \frac{1}{\lambda_2}$ . Then

$$\mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}_{\mu-}^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta,\mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N},2:k} \quad (390)$$

$$= \mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}_{\mu-}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}_{\mu-}^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta,\mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N},2:k}. \quad (391)$$

Note that because  $\mathbf{\Pi}_{\mathcal{L}} [\mathbf{I}]_{\mathcal{N},\Theta} = [\mathbf{I}]_{\mathcal{N},\Theta}$ ,

$$[\mathbf{L}_{\mu-}^\dagger]_{\Theta} = [\mathbf{L}_{\mu-}^\dagger]_{\Theta} \quad (392)$$

We note that the following are projection operators:

$$\mathbf{\Pi}_{\mu} = \left[ \sqrt{\mathbf{L}_{\mu-}^\dagger} \right]_{\mathcal{N},\Theta} [\mathbf{L}_{\mu-}^\dagger]_{\Theta}^{-1} \left[ \sqrt{\mathbf{L}_{\mu-}^\dagger} \right]_{\Theta,\mathcal{N}} \quad (393)$$

$$\mathbf{\Pi}_{\mu^c} = \mathbf{I} - \mathbf{\Pi}_{\mu} \quad (394)$$

which can be checked by seeing that  $\mathbf{\Pi}_{\mu} = \mathbf{\Pi}_{\mu}^T = \mathbf{\Pi}_{\mu}^2$ . As  $\mathbf{\Pi}_{\mathcal{L}} [\mathbf{U}]_{\mathcal{N},2:k} = [\mathbf{U}]_{\mathcal{N},2:k}$ ,

$$\mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}_{\mu-}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}_{\mu-}^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta,\mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N},2:k} \quad (395)$$

$$= \mathbf{v}_1^T \sqrt{\mathbf{L}_{\mu-}^\dagger} \mathbf{\Pi}_{\mu^c} \sqrt{\left( \mathbf{L}_{\mu-}^\dagger \right)^\dagger} [\mathbf{U}]_{\mathcal{N},2:k} \quad (396)$$

Therefore

$$\left\| \mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}^\dagger]_{\mathcal{N},\Theta} [\mathbf{L}_{\mu-}^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta,\mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N},2:k} \right\|_2^2 \quad (397)$$

$$= \left\| \mathbf{v}_1^T \sqrt{\mathbf{L}_{\mu-}^\dagger} \mathbf{\Pi}_{\mu^c} \sqrt{\left( \mathbf{L}_{\mu-}^\dagger \right)^\dagger} [\mathbf{U}]_{\mathcal{N},2:k} \right\|_2^2 \quad (398)$$

$$\leq \sigma_{max}^2 \left( \sqrt{\mathbf{L}_{\mu-}^\dagger} \mathbf{\Pi}_{\mu^c} \sqrt{\left( \mathbf{L}_{\mu-}^\dagger \right)^\dagger} [\mathbf{U}]_{\mathcal{N},2:k} \right) \quad (399)$$

We now apply Lemma 10 with  $\mathbf{A} = \left( \mathbf{L}_{\mu-}^\dagger \right)^\dagger$ ,  $\mathbf{B} = \mathbf{I}$  and  $\mathbf{\Pi}_k = [\mathbf{U}]_{\mathcal{N},2:k} [\mathbf{U}]_{\mathcal{N},2:k}^T$ . Then

$$M = \frac{1}{\mu + \frac{1}{\lambda_k}} \quad (400)$$

$$m = \frac{1}{\mu + \frac{1}{\lambda_2}} \quad (401)$$

so

$$\sigma_{max}^2 \left( \sqrt{\mathbf{L}_{\mu-}^\dagger} \mathbf{\Pi}_{\mu^c} \sqrt{\left( \mathbf{L}_{\mu-}^\dagger \right)^\dagger} [\mathbf{U}]_{\mathcal{N},2:k} \right) \leq \frac{(M + m)^2}{4Mm} \quad (402)$$

Note that

$$\frac{(M + m)^2}{4Mm} = \omega \left( \frac{M}{m} \right) = \omega \left( \frac{\mu + \frac{1}{\lambda_2}}{\mu + \frac{1}{\lambda_k}} \right) \quad (403)$$

as  $\frac{1}{\lambda_2} \geq \frac{1}{\lambda_k}$  we know that  $\frac{\mu + \frac{1}{\lambda_2}}{\mu + \frac{1}{\lambda_k}}$  is decreasing as a function of  $\mu$  so

$$\frac{\mu + \frac{1}{\lambda_2}}{\mu + \frac{1}{\lambda_k}} \leq \frac{1}{\lambda_2} \leq \frac{\lambda_k}{\lambda_2} \quad (404)$$

as  $\omega$  is increasing this means that

$$\frac{(M+m)^2}{4Mm} \leq \omega \left( \frac{\lambda_k}{\lambda_2} \right) = r_{bl} \quad (405)$$

combining these results gives

$$\left\| \mathbf{v}_1^T \left( \mathbf{I} - [\mathbf{L}^\dagger]_{\mathcal{N}, \Theta} [\mathbf{L}^\dagger]_{\Theta}^{-1} [\mathbf{I}]_{\Theta, \mathcal{N}} \right) [\mathbf{U}]_{\mathcal{N}, 2:k} \right\|_2^2 \leq r_{bl} \quad (406)$$

and multiplying both sides by  $\frac{N}{m}$  gives the result.  $\square$

#### APPENDIX V

##### PROOF OF THEOREM 4

*Proof.* To get the bounds on  $m_{opt\_bl}$  given  $B_k(m_{opt\_bl}) < k - 1$ , we can follow Appendix R replacing  $r$  with  $r_{bl}$ . Let

$$\mu_{ub\_bl} = \lambda_k^{-1} \left( \sqrt{\frac{k}{1 + B_k(m_{opt\_bl})}} - 1 \right) \quad (407)$$

$$\tau_{GLR\_bl}(\mu) = \frac{\left( \sum_{i=1}^k (1 + \mu \lambda_i)^{-2} \right) - B_k(m_{opt\_bl})}{k + B_k(m_{opt\_bl}) - \sum_{i=1}^k (1 - (1 + \mu \lambda_i)^{-1})^2} \quad (408)$$

We follow the structure of Appendix N. Firstly, we note that by the same arguments as for Lemma 8,

$$\xi_2(\mathcal{N}) = \sum_{i=1}^k \left( \frac{1}{1 + \mu \lambda_i} \right)^2. \quad (409)$$

Using 3 to get bounds on  $\xi_2(\mathcal{S})$ , we continue to follow Appendix N. Note that  $\xi_1$  is not changed by changing our noise model. Therefore

$$\Delta_2(\mathcal{N}, \mathcal{S}^c) \geq \sum_{i=1}^k \left( \frac{1}{1 + \mu \lambda_i} \right)^2 - (1 + B_k(m)) \quad (410)$$

$$\Delta_1(\mathcal{N}, \mathcal{S}^c) \geq \sum_{i=1}^N \left( 1 - \frac{1}{1 + \mu \lambda_i} \right)^2 - (k + B_k(m)). \quad (411)$$

We now show that  $\Delta_2 > 0$ . We have that  $\omega(x) > 0$  so  $B(m) > 0$  and by assumption  $B(m_{opt\_bl}) < k$ . Therefore  $\mu_{ub\_bl} > 0$  and is real and it is therefore possible to pick  $0 < \mu < \mu_{ub\_bl}$ . By assumption  $0 < \mu < \mu_{ub\_bl}$  and so

$$\sum_{i=1}^k \left( \frac{1}{1 + \mu \lambda_i} \right)^2 \geq \frac{k}{\left( 1 + \mu \frac{1}{\lambda_k} \right)^2} \quad (412)$$

$$> \frac{k}{\left( 1 + \mu_{ub\_bl} \frac{1}{\lambda_k} \right)^2} \quad (413)$$

$$= 1 + B_k(m_{opt\_bl}) \quad (414)$$

And therefore by (410),  $\Delta_2(\mathcal{N}, \mathcal{S}^c) > 0$ . Note that  $\text{SNR} = \frac{1}{\sigma^2}$  and apply Corollary 2.1 in the same manner as Appendix N and we are done.  $\square$

#### APPENDIX W

##### PROOF OF PROPOSITION 5

The proof in Appendix T adapts immediately to the bandlimited case, by noting that  $1 \leq r_{bl} \leq r$  and  $B_k(m) \leq B(m)$  and using the squeeze theorem.

#### APPENDIX X ADDITIONAL RESULTS

Under LS reconstruction, we show the threshold  $\tau(\mathcal{S}, v)$  for the BA and SBM graphs with 1000, 2000 and 3000 vertices (Figs. 13 and 14). We also present MSE plots for BA and SBM graphs under LS with full-band noise (Figs. 15 and 19), LS with bandlimited noise (Figs. 16 and 20), GLR with full-band noise (Figs. 17 and 21) and GLR with bandlimited noise (Figs. 18 and 22).

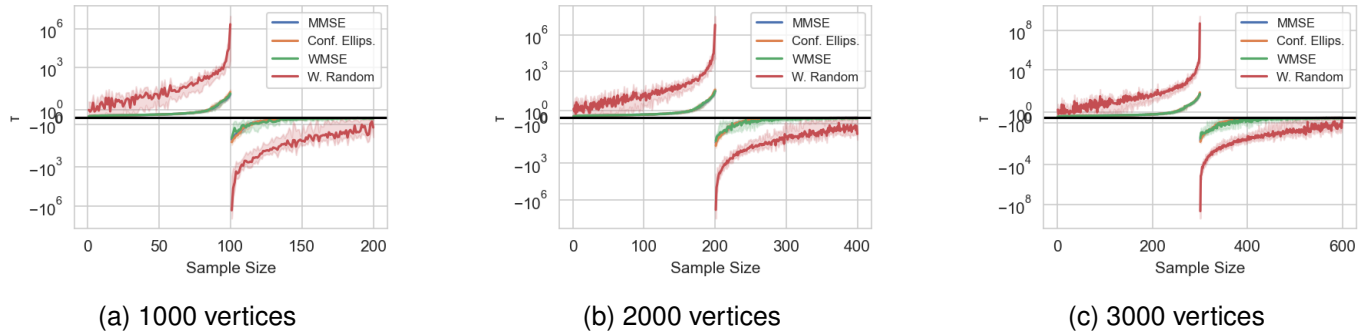


Fig. 13:  $\tau(\mathcal{S}, v)$  for different sized BA graphs under LS reconstruction (bandwidth =  $\frac{\# \text{ vertices}}{10}$ ).

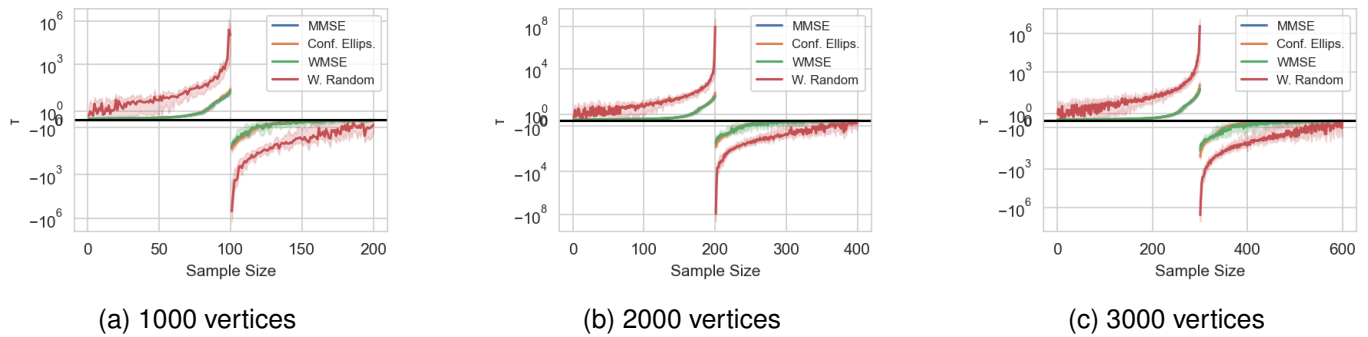


Fig. 14:  $\tau(\mathcal{S}, v)$  for different sized SBM graphs under LS reconstruction (bandwidth =  $\frac{\# \text{ vertices}}{10}$ ).

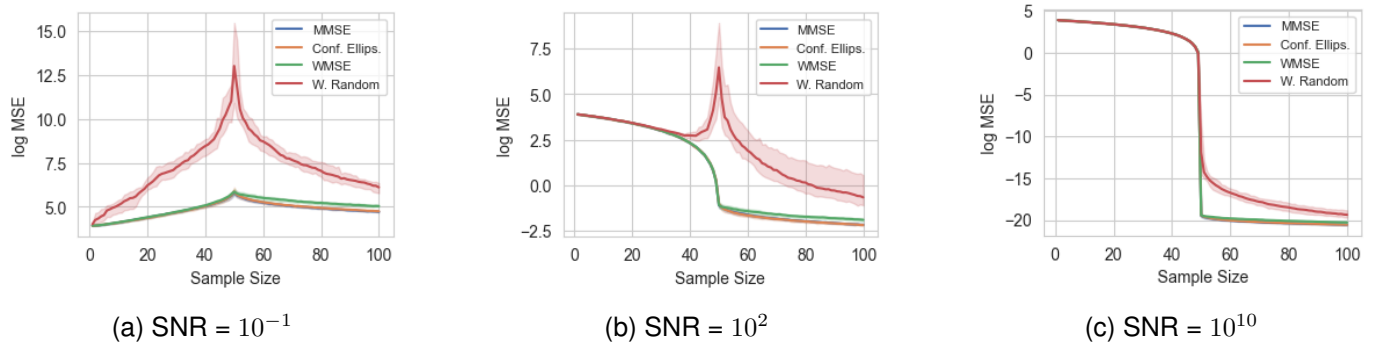


Fig. 15: Average MSE for LS reconstruction on BA graphs ( $\# \text{ vertices} = 500$ , bandwidth = 50) with different SNRs.

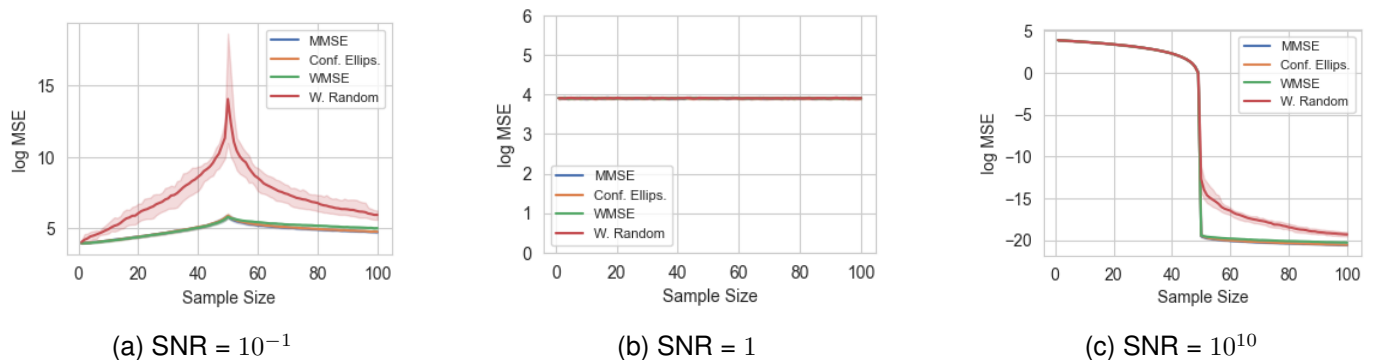


Fig. 16: Average MSE for LS reconstruction on BA graphs with bandlimited noise ( $\# \text{ vertices} = 500$ , bandwidth = 50) with different SNRs.

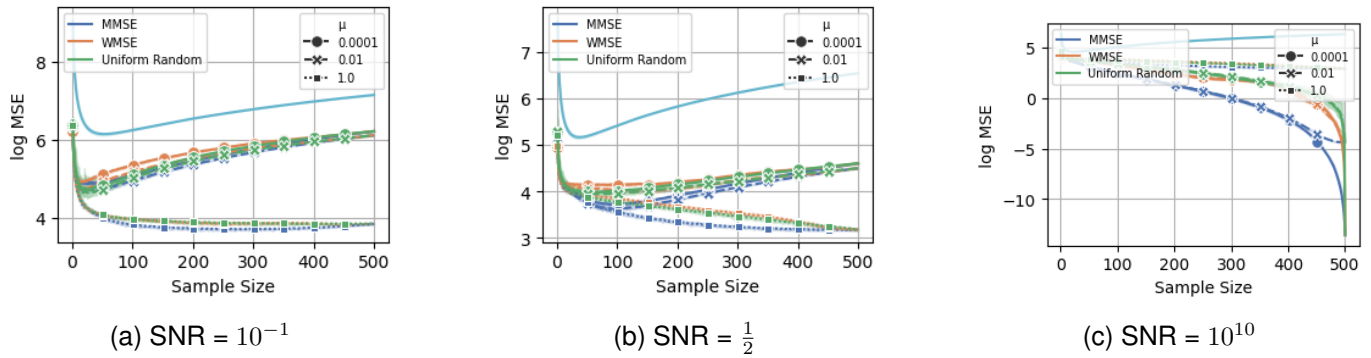


Fig. 17: Average MSE for GLR reconstruction on BA graphs (#vertices=500, bandwidth = 50) with different SNRs. Line without markers is an upper bound.

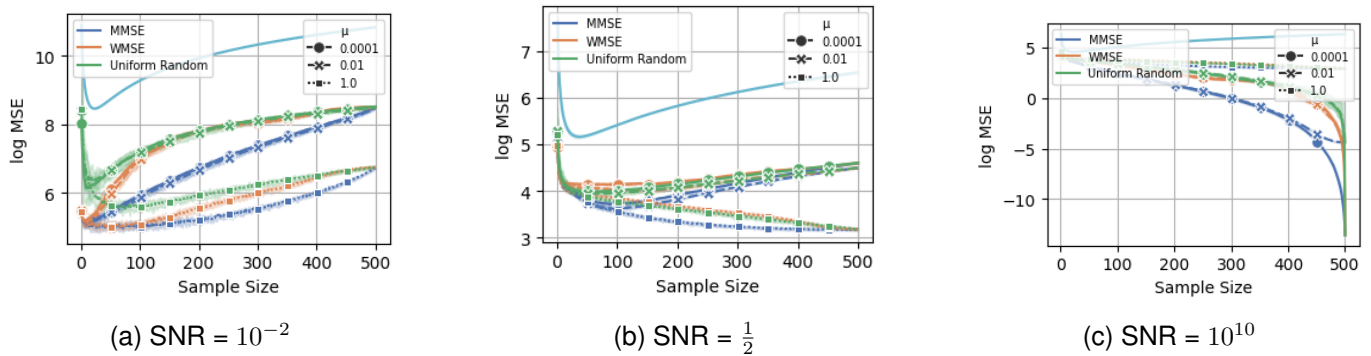


Fig. 18: Average MSE for GLR reconstruction on BA graphs under bandlimited noise (#vertices=500, bandwidth = 50) with different SNRs. Line without markers is an upper bound.

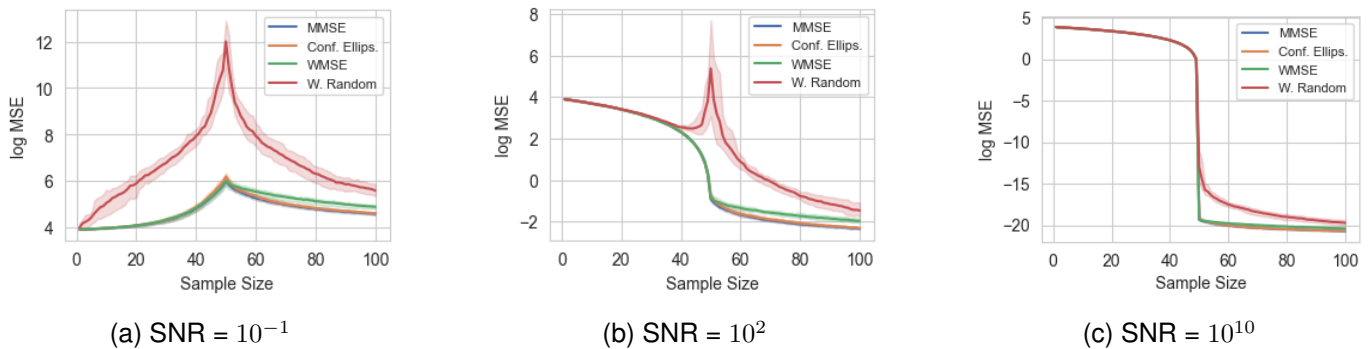


Fig. 19: Average MSE for LS reconstruction on SBM graphs (#vertices=500, bandwidth = 50) with different SNRs.

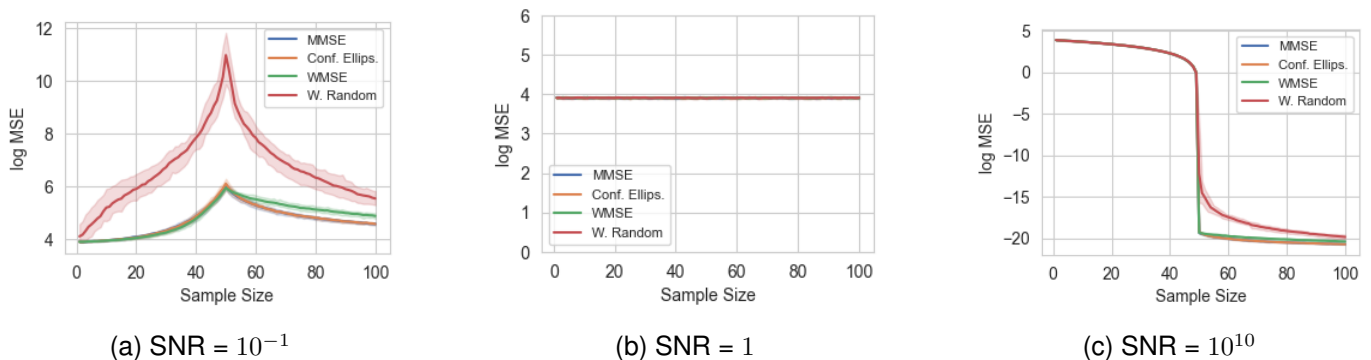
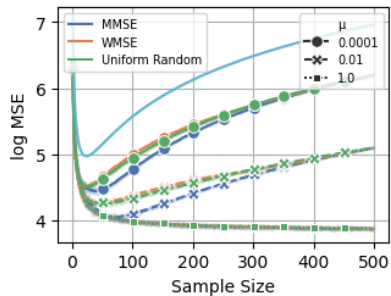
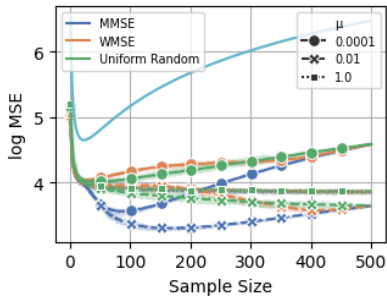


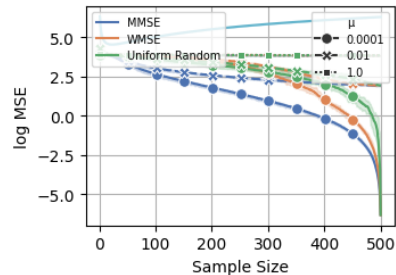
Fig. 20: Average MSE for LS reconstruction on SBM graphs with bandlimited noise (#vertices=500, bandwidth = 50) with different SNRs.



(a)  $\text{SNR} = 10^{-1}$

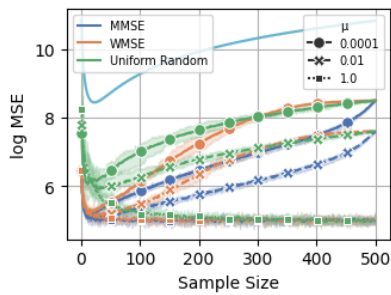


(b)  $\text{SNR} = \frac{1}{2}$

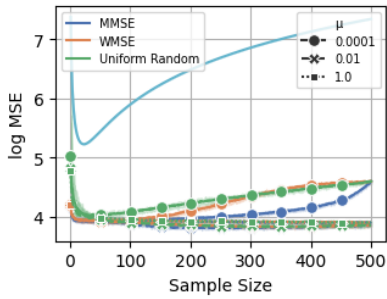


(c)  $\text{SNR} = 10^{10}$

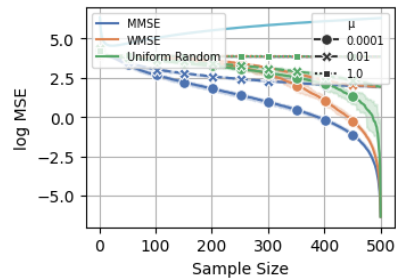
Fig. 21: Average MSE for GLR reconstruction on SBM graphs (#vertices=500, bandwidth = 50) with different SNRs. Line without markers is an upper bound.



(a)  $\text{SNR} = 10^{-2}$



(b)  $\text{SNR} = \frac{1}{2}$



(c)  $\text{SNR} = 10^{10}$

Fig. 22: Average MSE for GLR reconstruction on SBM graphs under bandlimited noise (#vertices=500, bandwidth = 50) with different SNRs. Line without markers is an upper bound.