sensors

MDPI

*Letter*

# Bandwidth Detection of Graph Signals with a Small Sample Size

**Xuan Xie, Hui Feng * and Bo Hu**

Research Center of Smart Networks and Systems, Fudan University, Shanghai 200433, China; xxie15@fudan.edu.cn (X.X.); bohu@fudan.edu.cn (B.H.)
* Correspondence: hfeng@fudan.edu.cn

**Abstract:** Bandwidth is the crucial knowledge to sampling, reconstruction or estimation of the graph signal (GS). However, it is typically unknown in practice. In this paper, we focus on detecting the bandwidth of bandlimited GS with a small sample size, where the number of spectral components of GS to be tested may greatly exceed the sample size. To control the significance of the result, the detection procedure is implemented by multi-stage testing. In each stage, a Bayesian score test, which introduces a prior to the spectral components, is adopted to face the high dimensional challenge. By setting different priors in each stage, we make the test more powerful against alternatives that have similar bandwidth to the null hypothesis. We prove that the Bayesian score test is locally most powerful in expectation against the alternatives following the given prior. Finally, numerical analysis shows that our method has a good performance in bandwidth detection and is robust to the noise.

**Keywords:** graph signals; bandwidth detection; bayesian score test

check for updates

**Citation:** Xie, X.; Feng, H.; Hu, B. Bandwidth Detection of Graph Signals with a Small Sample Size. *Sensors* **2021**, *21*, 146. https://dx.doi.org/10.3390/s21010146

Received: 29 October 2020
Accepted: 23 December 2020
Published: 28 December 2020

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Graph signal (GS) is a versatile model for describing information in irregular domains, which has been widely used in sensor networks [1,2], biological networks [3,4] and image and 3-D point cloud processing [5,6]. Graph signal processing (GSP) theory generalizes the classic discrete signal processing theory to irregular domains by introducing graph Fourier transform (GFT) [7,8]. In most real world networks, adjacent vertices usually have similar signals, which leads to bandlimitedness or approximate bandlimitedness in graph spectral domain. For example, the differences of temperature measured by sensors within short distances may not vary a lot; people tend to have similar interests or views with their friends in the social network.

Bandwidth is a widely used prior of GS in selecting the best sensor placements to monitor spatial phenomena [9–11], low-pass filter designed for denoising [7,8,12] and estimating the signals on all the sensors from partial observations [13,14]. However, the bandwidth of GS is typically unknown in practice. There are only a few works considering the problem of bandwidth detection. The bandwidth estimation in Reference [9] relies on noise-free samples. The scenario of noisy observations is considered in Reference [15], but it only applies to the spectral sparse case. Meanwhile, the estimation accuracy of the bandwidth in Reference [15] highly depends on the parameter controlling the sparsity of the GS, and the process of finding an optimal parameter is time-consuming.

The detection of bandwidth for a GS aims to detect the number of spectral components in a GS, which can be seen as a model selection problem in linear regression. The well-known model selection procedures like Akaike information criterion (AIC) and Bayesian information criterion (BIC) [16] pick the most likely model under some criteria. But they do not consider the significance of the chosen model. In contrast, hypothesis testing, such as likelihood ratio (LR) test and $F$-test can provide the significance of a model. If the bandwidth of a GS is $k$, the largest index of non-zero elements in frequency coefficients should be $k$. Let the vertex number of a GS be $N$, the regression coefficients of the last

$N-k$ frequency components are denoted by $\hat{\boldsymbol{f}}_{-k}$. To test whether the bandwidth of the GS is $k$, the hypothesis can be expressed as

$$H_0 : \hat{\boldsymbol{f}}_{-k} = \mathbf{0}, \quad H_A : \hat{\boldsymbol{f}}_{-k} \neq \mathbf{0}. \tag{1}$$

The null hypothesis will be rejected if the model with the assumed bandwidth $k$ is not significant enough. Therefore, the bandwidth of GS can be detected by applying tests over all the assumed bandwidths to see whether there is a model achieving the given significance.

For large-scale GS, such as social networks, it is impossible to sample the signal on every vertex due to huge data collection cost, which means we can only have the sample size $M < N-k$ sometimes. In this high dimensional situation, the conventional test for testing regression coefficients such as LR test and $F$-test is infeasible, which will be explained in Section 3.1. There are plenty of literature on testing the coefficients of high dimensional linear regression. For example, Reference [17,18] propose new test statistics by modifying the $F$-statistic, and Reference [19] introduces Bayesian priors to the parameters being tested. When $k \ll N$, the spectrum of the GS can be seen as sparse. Several approaches reconstruct the signal in compressive sensing which estimate the bandwidth as a by-product [20,21]. The frequency components detected by compressive sensing methods can help estimate the spectrum sparse signal well. However, the bandwidth obtained by the compressive sensing methods may be inaccurate since the choose frequency components in the whole space without considering the bandlimitedness. More accurate bandwidth is need in some applications like filter design and sampling set design of which the performance highly depend on the bandwidth prior.

In this letter, we try to detect the bandwidth of bandlimited GS with small sample size, which is not much larger than the bandwidth and there is no sparsity constraint to the spectrum. The bandwidth is detected by a forward multi-stage test. The bandwidth of model being tested is increasing over stages and the bandwidth is obtained when the null hypothesis is accepted. Since the samples are not adequate, we try to use the limited samples to better distinguish the assumed bandwidth from those close to it in each stage for a small detection bias. Therefore, we customize priors for $\hat{\boldsymbol{f}}_{-k}$ in each stage to describe our attention on testing whether some elements in $\hat{\boldsymbol{f}}_{-k}$ are non-zero. By doing so, we only use the limited samples to focus on testing a few alternatives in each stage. Bayesian score test [19] is adopted in this paper, but we do not give uniform prior to the parameters being tested as Reference [19] does. Since the bandwidth being tested is increasing over stages, our attentions on coefficients of different frequencies are also updated in each stage which makes our method able to locate the true bandwidth with a small bias even though the samples are insufficient.

The widely used multi-stage test for model selection, stepwise regression ([16] Chapter 10) does poorly with a small sample size since the estimation of the regression coefficients in it is biased [22]. In contrast, our method is proved to be locally most powerful (LMP) in expectation against alternatives following the given prior, which means it performs well in distinguishing among models with neighbouring bandwidth even using a small amount of samples in each stage.

## 2. Problem Formulation

Consider an $N$-vertex undirected connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{W})$, where $\mathcal{V}$ is the vertex set, $\mathcal{E}$ is the edge set and $\boldsymbol{W}$ is the $N \times N$ weighted adjacency matrix with $\boldsymbol{W}(i,j) = w_{ij}$. A GS defined on $\mathcal{V}$ can be represented as a vector $\boldsymbol{f} \in \mathbb{R}^N$, and its element $f_i$ represents the signal value at the $i$-th vertex in $\mathcal{V}$. Laplacian matrix of the undirected graph is given by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{D}$ is the diagonal degree matrix $\mathrm{diag}\{d_1, \cdots, d_N\}$ with $d_i = \sum_j w_{ij}$. Since $w_{ij} = w_{ji}$ for undirected graphs, Laplacian matrix is symmetric. Therefore, it has real non-negative eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ and an orthogonal set of eigenvectors $\boldsymbol{V}$. As a result, the spectral decomposition of graph Laplacian is given as $\boldsymbol{L} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues. The columns of $\boldsymbol{V}$ denoted by $\{\boldsymbol{v}_i\}_{1 \leq i \leq N}$ are

regarded as the graph Fourier bases and the eigenvalues are regarded as frequencies [7]. The expansion coefficients of $\boldsymbol{f}$ corresponding to eigenvectors are defined as $\hat{\boldsymbol{f}}$.

A GS is called *bandlimited* when there exists a $K \in \{0, 1, \cdots, N-1\}$ such that its GFT $\hat{\boldsymbol{f}}$ satisfies $\hat{f}_i = 0$ for all $i \geq K$ [10]. The smallest such $K$ is called the bandwidth of $\boldsymbol{f}$. If all the frequency coefficients are non-zero, the GS is not bandlimited. If $\boldsymbol{f}$ is a signal with bandwidth $K$, then it satisfies $\boldsymbol{f} = \boldsymbol{V}_K \hat{\boldsymbol{f}}_K$, where $\boldsymbol{V}_K$ denotes the first $K$ columns of $\boldsymbol{V}$ and $\hat{\boldsymbol{f}}_K$ denotes the first $K$ coefficients of $\hat{\boldsymbol{f}}$.

Suppose that $M (M < N)$ noisy observations $\boldsymbol{y} \in \mathbb{R}^M$ are sampled from $\boldsymbol{f} \in \mathbb{R}^N$ to detect the bandwidth, the observation model can be summarized as

$$\boldsymbol{y} = \boldsymbol{\Psi}(\boldsymbol{f} + \boldsymbol{w}), \tag{2}$$

where $\boldsymbol{w}$ is an $N \times 1$ vector with $i$th element representing the observation noise on the $i$th vertex and $\boldsymbol{\Psi} : \mathbb{R}^N \to \mathbb{R}^M$ is the sampling matrix, of which the element in the $i$th row and $j$th column is defined as

$$(\boldsymbol{\Psi})_{i,j} = \begin{cases} 1, & \text{if the } j\text{th vertex is the } i\text{th sample;} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

We use a simple example to illustrate the role of the sampling matrix. For a graph with 5 vertices and signal $\boldsymbol{f} = [f_1, \cdots, f_5]^{\mathrm{T}}$, if the 1st, 2nd and 4th vertices are sampled, then the sampling matrix is a matrix with 3 rows and 5 columns as follow

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{4}$$

The sampled graph signal is $\boldsymbol{\Psi}\boldsymbol{f} = [f_1, f_2, f_4]^{\mathrm{T}}$.

In this paper, we assume that the observation noise on each sample follows the Gaussian distribution $\mathcal{N}(0, \sigma^2)$ and the noise on all the samples are independently and identically distributed (*i.i.d.*), so the observation follows $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{N \times N})$, where $\boldsymbol{I}_{N \times N}$ is an $N \times N$ identity matrix.

By dividing the columns of $\boldsymbol{V}$ into two parts $\boldsymbol{V}_k$ and $\boldsymbol{V}_{-k}$ for any assumed bandwidth $k \in \{1, 2, \cdots, N-1\}$, the observation model (2) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{U}_k \hat{\boldsymbol{f}}_k + \boldsymbol{U}_{-k} \hat{\boldsymbol{f}}_{-k} + \boldsymbol{\Psi}\boldsymbol{w}, \tag{5}$$

where $\hat{\boldsymbol{f}}_k$ and $\hat{\boldsymbol{f}}_{-k}$ denote the first $k$ and the last $N-k$ elements of $\hat{\boldsymbol{f}}$, respectively, and $\boldsymbol{U}_k = \boldsymbol{\Psi}\boldsymbol{V}_k$, $\boldsymbol{U}_{-k} = \boldsymbol{\Psi}\boldsymbol{V}_{-k}$. If the actual bandwidth $K = k$, $H_0$ in (1) will be accepted. Otherwise, $H_0$ will be rejected.

For a GS with actual bandwidth $K$, the hypothesis with assumed bandwidth $k < K$ should all be rejected. Therefore, we detect the bandwidth of GS in a forward multi-stage way with assumed bandwidth increasing over stages. The bandwidth is obtained when the null hypothesis is accepted. By doing so, we reduce the number of models to be tested from $N$ to $K$.

## 3. Bandwidth Detection

### 3.1. High Dimensional Challenge

*F*-test and LR test are common approaches to test the hypothesis (1) in linear regression model. Let the maximum likelihood estimation (MLE) of $\hat{\boldsymbol{f}}$ under $H_0$ and $H_1$ be $\hat{\boldsymbol{f}}_{\mathrm{R}}$ and $\hat{\boldsymbol{f}}_{\mathrm{U}}$, respectively. The corresponding sum of squared residuals (SSR) are denoted by $\mathrm{SSR}(\hat{\boldsymbol{f}}_{\mathrm{R}})$ and $\mathrm{SSR}(\hat{\boldsymbol{f}}_{\mathrm{U}})$. Then, LR test statistic ([23] Chapter 10) equals $M \log(\mathrm{SSR}(\hat{\boldsymbol{f}}_{\mathrm{U}})/\mathrm{SSR}(\hat{\boldsymbol{f}}_{\mathrm{R}}))$, which is asymptotically distributed as $\chi^2(N-k)$ as the sample size approaches to infinity. When the sample size is small, the test statistic of LR test has no specific distribution which makes it hard to implement. *F*-test statistic ([23] Chapter 10) equals

$$\frac{\mathrm{SSR}(\hat{\boldsymbol{f}}_{\mathrm{U}}) - \mathrm{SSR}(\hat{\boldsymbol{f}}_{\mathrm{R}})/N-k}{\mathrm{SSR}(\hat{\boldsymbol{f}}_{\mathrm{R}})/M-k} = \frac{\boldsymbol{y}^T\boldsymbol{M}_k\boldsymbol{U}_{-k}\left(\boldsymbol{U}_{-k}^T\boldsymbol{M}_k\boldsymbol{U}_{-k}\right)^{-1}\boldsymbol{U}_{-k}^T\boldsymbol{M}_k\boldsymbol{y}/N-k}{\boldsymbol{y}^T\boldsymbol{M}_k\left(\boldsymbol{I}_{M\times M}-\boldsymbol{U}_{-k}\left(\boldsymbol{U}_{-k}^T\boldsymbol{M}_k\boldsymbol{U}_{-k}\right)^{-1}\boldsymbol{U}_{-k}^T\right)\boldsymbol{M}_k\boldsymbol{y}/M-k}, \tag{6}$$

where $\mathbf{M}_k = \boldsymbol{I}_{M\times M} - \boldsymbol{U}_k(\boldsymbol{U}_k^T\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^T$ is a projection matrix that projects signals to the complementary subspace of the column space of $\boldsymbol{U}_k$. $F$-test is exact in finite samples, however, to ensure that $\boldsymbol{U}_k^T\boldsymbol{U}_k$ in $\boldsymbol{M}_k$ is invertible, we have to let $M \geq k$. Similarly, to ensure that $\boldsymbol{U}_{-k}^T\boldsymbol{M}_k\boldsymbol{U}_{-k}$ is invertible, there must be $M \geq N-k$. Therefore, (6) requires that $M \geq \max(K, N-k)$, which means the sample size needs to be bigger than the assumed bandwidth, as well as the dimension of $\hat{\boldsymbol{f}}_{-k}$. For large-scale GS, the sampling cost for bandwidth detection using $F$-test or LR test is too high. Therefore, a test with small sample size is preferred. The sample size only needs to be larger than the assumed bandwidth, that is, $M > k$, in score test ([23] Chapter 10). However, when $N-k > M$, there may exist some alternatives which have $\hat{\boldsymbol{f}}_{-k} \neq \mathbf{0}$ but $\boldsymbol{U}_{-k}\hat{\boldsymbol{f}}_{-k} = \mathbf{0}$. We can never hope to have any power against these alternatives. Since the GS is known to be bandlimited, we should pay more attention on testing whether the low frequency coefficients in $\hat{\boldsymbol{f}}_{-k}$ are non-zero than the high frequency coefficients. Thus, we adopt Bayesian score test [19] to allocate our attention among the alternatives by designing a prior for $\hat{\boldsymbol{f}}_{-k}$. Different from Reference [19], which pays equal attention to all the alternatives, we customize a prior for $\hat{\boldsymbol{f}}_{-k}$ to fit our bandwidth detection purpose.

### *3.2. Design of the Prior*

In the test for each bandwidth, the alternative that $\hat{\boldsymbol{f}}_{-k} = \boldsymbol{\alpha}$ and $\hat{\boldsymbol{f}}_{-k} = -\boldsymbol{\alpha}$ for every $\boldsymbol{\alpha} \neq \mathbf{0}$ contributes equally in rejecting $H_0$, so we assign $\boldsymbol{E}(\hat{\boldsymbol{f}}_{-k}) = \mathbf{0}$ to make the prior unbiased, where $\boldsymbol{E}(\cdot)$ denotes the expectation of the given stochastic variable. The covariance matrix of $\hat{\boldsymbol{f}}_{-k}$ can be assigned as

$$\boldsymbol{E}\left(\hat{\boldsymbol{f}}_{-k}\hat{\boldsymbol{f}}_{-k}^T\right) = \tau^2\boldsymbol{\Sigma}_{-k}, \tag{7}$$

where $\boldsymbol{\Sigma}_{-k}$ is a positive semidefinite $k \times k$ matrix to be designed and $\tau^2$ is an unknown parameter. In this paper, we let $\boldsymbol{\Sigma}_{-k}$ be a diagonal matrix, then each of its diagonal element is the variance of the corresponding element in $\hat{\boldsymbol{f}}_{-k}$. Since $\boldsymbol{E}(\hat{\boldsymbol{f}}_{-k}) = \mathbf{0}$, a larger variance indicates the corresponding element in $\hat{\boldsymbol{f}}_{-k}$ is more likely to have a larger amplitude. To decide whether $H_0$ should be accepted, more attention should be payed to the element in $\hat{\boldsymbol{f}}_{-k}$ with larger variance. Thus, *the attention among elements in* $\hat{\boldsymbol{f}}_{-k}$ *in bandwidth detection is linked to their variances in the prior*. The attention is allocated by designing a $\boldsymbol{\Sigma}_{-k}$.

In our forward multi-stage test, the bandwidth being tested updates over stages, thus $\boldsymbol{\Sigma}_{-k}$ also needs to be updated. To avoid the the multi-stage test accepting the null hypothesis too early, the test should be more distinguishable from the model with bandwidth close to the null hypothesis in each stage. Therefore, we use the limited samples to make the test concentrate on distinguishing bandwidth close to $k$. The design of $\boldsymbol{\Sigma}_{-k}$ follows the guideline that as the frequency increases, element in $\hat{\boldsymbol{f}}_{-k}$ with higher frequency is paid less attention. Since we cannot determine a bandwidth larger than $M-1$ with sample size $M$, we set the attention on $\hat{\boldsymbol{f}}_{-k}$ with frequency larger than $M$ to a small constant $\delta$. For example, the $i$th diagonal element of $\boldsymbol{\Sigma}_{-k}$ can be designed to be

$$(\boldsymbol{\Sigma}_{-k})_{i,i} = \frac{1}{Z}\exp\left(-\frac{i^2}{2\sigma^2}\right) + \delta, \tag{8}$$

where $Z = \exp\left(-1/(2\sigma^2)\right)$ is the normalization factor. According to the three-standard-deviations property of normal distribution, $\sigma$ is set to be $(M-k)/3$ to make the attention on coefficients with frequencies larger than $M$ equals to $\delta$ with high probability. Equation (8)

is not the only choice of $\mathbf{\Sigma}_{-k}$, but this form is applicable. The details about how it affects the power of the test is analyzed later in Section 3.3.

As the multi-stage test moves forward, the attentions among frequency coefficients update, and in each stage, the test is distinguishable among bandwidth close to $k$, which makes the small-bias bandwidth detection with small sample size possible. An illustration of the attention update is shown in Figure 1.

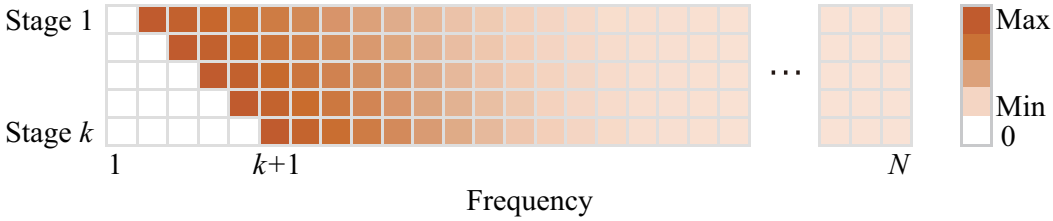



**Figure 1.** Illustration of the attentions on coefficients of different frequencies in different stages. Each line illustrates the attention distribution of one stage. Attention equals to 0 indicates that the corresponding frequency coefficient does not need to be tested.

We can complete the specification of the distribution of $\hat{\boldsymbol{f}}_{-k}$ by choosing a value for $\tau^2$ and a distribution shape. Let $L_k(\hat{\boldsymbol{f}}_{-k};\boldsymbol{y})$ be the likelihood of $\hat{\boldsymbol{f}}_{-k}$, for a specific prior distribution of $\hat{\boldsymbol{f}}_{-k}$, the likelihood of $\tau^2$ is

$$\bar{L}_k(\tau^2;\boldsymbol{y}) = \int L_k(\hat{\boldsymbol{f}}_{-k};\boldsymbol{y})\pi(\hat{\boldsymbol{f}}_{-k}|\tau^2)\mathrm{d}\hat{\boldsymbol{f}}_{-k}, \tag{9}$$

where $\pi(\hat{\boldsymbol{f}}_{-k}|\tau^2)$ is the distribution of $\hat{\boldsymbol{f}}_{-k}$ for a given $\tau^2$.

Given $\bar{L}_k(\tau^2;\boldsymbol{y})$, we convert hypothesis (1) to

$$\bar{H}_0 : \tau^2 = 0, \quad \bar{H}_A : \tau^2 \neq 0. \tag{10}$$

$\tau^2 = 0$ implies $\hat{\boldsymbol{f}}_{-k} = \mathbf{0}$, since they lead to the same distribution of $\boldsymbol{y}$. Thus $H_0 : \hat{\boldsymbol{f}}_{-k} = \mathbf{0}$ will be rejected if $\bar{H}_0$ is rejected. *The score test of $\bar{H}_0$ is named the Bayesian score test of $H_0$ with the given prior of $\hat{\boldsymbol{f}}_{-k}$.*

The score test of $\bar{H}_0$ in stage $k$ is a one-sided test for one parameter, the test statistic is

$$S_{\mathbf{\Sigma}_{-k}} = \frac{\mathrm{d}}{\mathrm{d}\tau^2}\log\bar{L}_k(0;\boldsymbol{y}), \tag{11}$$

$$= \frac{1}{2}\boldsymbol{g}_k^T\mathbf{\Sigma}_{-k}\boldsymbol{g}_k - \frac{1}{2}\mathrm{tr}(\mathbf{\Sigma}_{-k}\boldsymbol{F}_k), \tag{12}$$

where $\mathrm{tr}(\mathbf{\Sigma}_{-k}\boldsymbol{F}_k)$ denotes the trace of $\mathbf{\Sigma}_{-k}\boldsymbol{F}_k$. $\bar{H}_0$ will be rejected when $S_{\mathbf{\Sigma}_{-k}} \geq t$ for some threshold $t$. From (11), we can find that the test statistic of $\tau^2$ can be seen as the slope of log-likelihood function of $\tau^2$ under $\bar{H}_0$. The slope will equal to 0 when $\tau^2$ equals its MLE $\hat{\tau}^2$. The more closer $\hat{\tau}^2$ to 0, the more closer $S_{\mathbf{\Sigma}_{-k}}$ to 0. If $S_{\mathbf{\Sigma}_{-k}}$ is larger than a given $t$, which means $\hat{\tau}^2$ is much larger than 0, then $\bar{H}_0$ will be rejected. Equation (12) is given by Reference [19], where $\boldsymbol{g}_k = \sigma^{-2}\boldsymbol{U}_{-k}\boldsymbol{M}_k\boldsymbol{y}$ and $\boldsymbol{F}_k = \sigma^{-2}\boldsymbol{U}^T\boldsymbol{U}$ are the gradient and Fisher information matrix of $\log L_k(\mathbf{0};\boldsymbol{y})$, respectively.

### *3.3. Method*

Considering that the second part of (12) is not related to the observations, it is more convenient to work with the equivalent test statistic $\sigma^{-2}\boldsymbol{y}^T\boldsymbol{M}_k\boldsymbol{U}_{-k}\mathbf{\Sigma}_{-k}\boldsymbol{U}_{-k}^T\boldsymbol{M}_k\boldsymbol{y}$. Because $\sigma^2$ is not known, we plug in the MLE result $\hat{\sigma^2} \propto \boldsymbol{y}^T\boldsymbol{M}_k\boldsymbol{y}$ under the null hypothesis, the resulting test statistic is

$$S_{\mathbf{\Sigma}_{-k}} = \frac{\boldsymbol{y}^T \boldsymbol{M}_k \boldsymbol{U}_{-k} \mathbf{\Sigma}_{-k} \boldsymbol{U}_{-k}^T \boldsymbol{M}_k \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{M}_k \boldsymbol{y}}. \tag{13}$$

- *Interpretation of the test statistic*: $\boldsymbol{M}_k \boldsymbol{y}$ indicates the part of $\boldsymbol{y}$ out of the range of $\boldsymbol{U}_k$. The numerator of (13) can be rewritten as $Q = \sum_{i=1}^{M} (\mathbf{\Sigma}_{-k})_{i,i} Q_i$ with $Q_i = \left(\boldsymbol{y}^T \boldsymbol{M}_k \boldsymbol{U}_{k+i}\right)^2$. Larger $Q_i$ indicates a larger energy of $\boldsymbol{M}_k \boldsymbol{y}$ lying in the range of $\boldsymbol{U}_{k+i}$, which links to a larger amplitude of the $i$th element in $\hat{\boldsymbol{f}}_{-k}$. $S_{\mathbf{\Sigma}_{-k}}$ equals $Q$, which is a weighted sum of $Q_i$, normalized by the energy of $\boldsymbol{M}_k \boldsymbol{y}$. When the weight $(\mathbf{\Sigma}_{-k})_{i,i}$ decreases with $i$, $S_{\mathbf{\Sigma}_{-k}}$ will be larger if $Q_i$ is larger for smaller $i$. Therefore, the test is more powerful against $H_A$ that has non-zero frequency coefficients in low frequency than that has non-zero frequency coefficients in high frequency. This is in accordance with the purpose of our design that the test should be more distinguishable from the model with bandwidth close to $k$.

Finally, we decide whether the null hypothesis is accepted by a $p$-value threshold. For a given significance level $\alpha$, if $p \leq \alpha$, the null hypothesis is rejected. Let $S_{-k}$ be the observation value of $S_{\mathbf{\Sigma}_{-k}}$, then the $p$-value of the test is

$$\begin{aligned} p &= P_{H_0}(S_{\mathbf{\Sigma}_{-k}} > S_{-k}), \\ &= P_{H_0}(\boldsymbol{y}^T \boldsymbol{M}_k (\boldsymbol{U}_{-k} \mathbf{\Sigma}_{-k} \boldsymbol{U}_{-k}^T - S_{-k}) \boldsymbol{M}_k \boldsymbol{y} > 0). \end{aligned} \tag{14}$$

Since $\boldsymbol{M}_k \boldsymbol{y} = \boldsymbol{M}_k (\boldsymbol{y} - \boldsymbol{U}_k \hat{\boldsymbol{f}}_k) = \boldsymbol{M}_k \boldsymbol{w}$ under null hypothesis, it follows a multivariate normal distribution. Thus, $p$ is the probability that the quadratic form of normal variables is non-negative. The distribution of the quadratic forms in normal variables is approximated by a $\chi^2$ distribution in Reference [24] and $p$-value can be calculated approximately as,

$$p = \begin{cases} P(\chi^2_{d_k} > d_k - h_k), & \mathrm{tr}(\boldsymbol{X}_k^3) > 0; \\ P(\chi^2_{d_k} < d_k - h_k), & \mathrm{tr}(\boldsymbol{X}_k^3) < 0; \\ \Phi\left(\frac{\mathrm{tr}(\boldsymbol{X}_k)}{\sqrt{2\mathrm{tr}(\boldsymbol{X}_k^2)}}\right), & \mathrm{tr}(\boldsymbol{X}_k^3) = 0, \end{cases} \tag{15}$$

where $\boldsymbol{X}_k = \boldsymbol{M}_k \boldsymbol{U}_{-k} \mathbf{\Sigma}_{-k} \boldsymbol{U}_{-k}^T \boldsymbol{M}_k - S_{-k} \boldsymbol{M}_k$, $d_k = [\mathrm{tr}(\boldsymbol{X}_k^2)]^3 / [\mathrm{tr}(\boldsymbol{X}_k^3)]^2$, and $h_k = \mathrm{tr}(\boldsymbol{X}_k^2)\mathrm{tr}(\boldsymbol{X}_k) / \mathrm{tr}(\boldsymbol{X}_k^3)$.

Our algorithm for the multi-stage bandwidth detection is shown in Algorithm 1. If $p \leq \alpha$ for all the stages, the output of Algorithm 1 will be 'None', which means the bandwidth cannot be determined with the given sample size. Then we can say that the bandwidth is larger than $M - 1$.

**Algorithm 1:** Multi-stage Bandwidth Detection.

**Input** Samples $\boldsymbol{y}$ and significance level $\alpha$;
**Output** Detected bandwidth $K$ or None;
**for** $k = 1, 2, \cdots, M-1$ **do**
  Calculate test statistic according to (13);
  Calculate the $p$-value of the test according to (15);
  **if** $p > \alpha$ **then**
    Let bandwidth $K = k$ and **stop**;
  **else**
    $k = k + 1$;
  **end if**
**end for**

### 3.4. Power Analysis

In the situation $N - k > M$, there may exist some alternatives which have $\hat{\boldsymbol{f}}_{-k} \neq \mathbf{0}$ but $\boldsymbol{U}_{-k}\hat{\boldsymbol{f}}_{-k} = \mathbf{0}$. It is impossible to find a test which is optimal against all the alternatives, that is, uniformly most powerful. If the true $\hat{\boldsymbol{f}}_{-k}$ has large deviations from $H_0$, it is very easy to detect. However, if the deviations are small, the detection becomes harder. In the Bayesian score test of $H_0$, the deviations of the true $\hat{\boldsymbol{f}}_{-k}$ from $H_0$ is denoted by $\tau\boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is assumed to follow a prior distribution with $\boldsymbol{E}(\boldsymbol{\xi}) = \mathbf{0}$ and $\boldsymbol{E}(\boldsymbol{\xi}\boldsymbol{\xi}^{\boldsymbol{T}}) \propto \boldsymbol{\Sigma}_{-k}$ and $\tau^2$ indicates how large the deviation is. Next, we will analyze the power of the Bayesian score test of $H_0$ when $\tau^2$ is small.

**Definition 1** (Locally most powerful (LMP) [25]). *Consider the problem of testing the simple null hypothesis $H : \theta = \theta_0$ against the one-sided alternative $K : \theta > \theta_0$. A significance level $\alpha$ test $\Phi_0$ with power function $\beta_{\Phi_0}(\theta)$ is said to be LMP if given any other level $\alpha$ test $\Phi$ with power function $\beta_{\Phi}(\theta)$, there exists $\Delta > 0$ such that $\beta_{\Phi_0}(\theta) \geq \beta_{\Phi}(\theta)$ for all $\theta \in K$ with $0 \leq \theta - \theta_0 \leq \Delta$.*

An LMP test is one of the best tests for detecting small deviations from null hypothesis, though it is not good at detecting all kinds of alternatives. In each stage of Algorithm 1, we aim at deciding whether the bandwidth is $k$ or not, so small deviations in the coefficients of frequencies around $k$ should be detected as possible as we can. Therefore, a LMP test is preferred in each stage. In Theorem 1, we will show that the Bayesian score test in each stage of Algorithm 1 is LMP in expectation.

**Theorem 1.** *Let $\bar{w}(\hat{\boldsymbol{f}}_{-k})$ be the power function of the level $\alpha$ Bayesian score test of $H_0$ and let $w(\hat{\boldsymbol{f}}_{-k})$ be the power function of any level $\alpha$ test of $H_0$. The Bayesian score test of $H_0$ is LMP in expectation against all $\hat{\boldsymbol{f}}_{-k} \in H_A$ with $\hat{\boldsymbol{f}}_{-k} = \tau\boldsymbol{\xi}$, where $0 \leq \tau^2 \leq \Delta$, that is $\boldsymbol{E}_{\boldsymbol{\xi}}[w(\tau\boldsymbol{\xi})] \leq \boldsymbol{E}_{\boldsymbol{\xi}}[\bar{w}(\tau\boldsymbol{\xi})]$.*

**Proof of Theorem 1.** Let $\bar{\beta}(\tau^2)$ be the power function of the level $\alpha$ score test of $\bar{H}_0$. Since $\hat{\boldsymbol{f}}_{-k} = \mathbf{0}$ and $\tau^2 = 0$ lead to the same distribution of $\boldsymbol{y}$, the level $\alpha$ tests of $H_0$ and $\bar{H}_0$ lead to the same critical region. The same conclusion holds for any other level $\alpha$ tests of $H_0$ and $\bar{H}_0$ with power function $w(\hat{\boldsymbol{f}}_{-k})$ and $\beta(\tau^2)$. It has been proved in Reference [25] that the one-side score test for one dimensional parameter is LMP. Therefore, for the given level $\alpha$, there exists a $\Delta > 0$ such that for all $\tau^2 \in \bar{H}_A$ with $0 \leq \tau^2 \leq \Delta$, the power function of score test for (10) is larger than that of any other test for (10)

$$\beta(\tau^2) \leq \bar{\beta}(\tau^2). \tag{16}$$

Let the critical region of the level $\alpha$ Bayesian score test be $A$, we have

$$\beta(\tau^2) = \int_A \bar{L}_k(\tau^2; \boldsymbol{y}) \mathrm{d}\mu, \tag{17}$$

where $\mu$ is a dominating measure. According to (9), $\bar{L}_k(\tau^2; \boldsymbol{y})$ is the marginalized distribution of $L_k(\hat{\boldsymbol{f}}_{-k}; \boldsymbol{y})$. According to (12), every distribution shape of $\hat{\boldsymbol{f}}_{-k} = \tau\boldsymbol{\xi}$ with $\boldsymbol{E}(\boldsymbol{\xi}) = \mathbf{0}$ and $\boldsymbol{E}(\boldsymbol{\xi}\boldsymbol{\xi}^{\boldsymbol{T}}) \propto \boldsymbol{\Sigma}_{-k}$ leads to the same $S_{\boldsymbol{\Sigma}_{-k}}$, and therefore the same power function. So (17) can be written as

$$\beta(\tau^2) = \int_A \left( \int L_k(\tau\boldsymbol{\xi}; \boldsymbol{y}) \pi(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \right) \mathrm{d}\mu = \int \left( \int_A L_k(\tau\boldsymbol{\xi}; \boldsymbol{y}) \mathrm{d}\mu \right) \pi(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} = \boldsymbol{E}_{\boldsymbol{\xi}}[\bar{w}(\tau\boldsymbol{\xi})]. \tag{18}$$

According to (16) and (18), we have $\boldsymbol{E}_{\boldsymbol{\xi}}[w(\tau\boldsymbol{\xi})] \leq \boldsymbol{E}_{\boldsymbol{\xi}}[\bar{w}(\tau\boldsymbol{\xi})]$. □

To make Theorem 1 more intuitive, we give the following example.

*Example*: Suppose there are two alternatives to be tested in step $k$, $H_A^{\mathrm{l}} = \hat{\boldsymbol{f}}_{-k} = [0.1, 0, \cdots, 0]$, which has a small deviation 0.1 at a low frequency and $H_A^{\mathrm{h}} = \hat{\boldsymbol{f}}_{-k} = [0, \cdots, 0.1, \cdots, 0]$, which has the same deviation at high frequency. The prior of $\hat{\boldsymbol{f}}_{-k}$ is given as (8), which

means the test pays more attention to the low frequency components. Then, the power of testing $H_0$ against $H_A^{\mathrm{l}}$ will be larger than the power of testing $H_0$ against $H_A^{\mathrm{h}}$ at the same significance level. As a result, the average power of the Bayesian score test is increased since there are more alternatives having small deviations at the low frequencies in a bandlimited GS.

Furthermore, the proof of Theorem 1 does not rely on the specific form of $\mathbf{\Sigma}_{-k}$, which means that it will hold for different designs of $\mathbf{\Sigma}_{-k}$. We can design different $\mathbf{\Sigma}_{-k}$ to allocate attentions among frequency components to achieve various testing purposes with small sample size, for example, spectrum anomaly detection.

## 4. Numerical Analysis

### *4.1. Bandwidth Detection*

An accurate bandwidth is helpful in various applications. For example, an accurate bandwidth in low-pass filter design for graph signals can help remove the noise and keep the original signal well; An accurate bandwidth can also help in choosing the minimal sample size in the sampling set design for graph signals. So in this section, we first validate the performance of bandwidth detection on an Erdős-Rènyi random graph with $N = 250$ and the probability of edge presence being 0.25. The frequency coefficients of bandlimited GS are independently generated from a uniform distribution over the interval $[0, 1]$. Gaussian noise is added to the GS to produce the observations, the signal to noise ratio (SNR) is calculated by $\mathrm{SNR} = 10\log\left(\|\boldsymbol{f}\|_2^2/(N\sigma^2)\right)$. The performance of bandwidth detection is shown in bias of 1000 simulations, the SNR is set to 20 dB and vertices are sampled randomly in each simulation. In this simulation, we design two priors for the alternatives in our method follows the guideline in Section 3.2 and the significance level of the test is set to $\alpha = 0.05$. Prior 1 is given by (8) with $\delta = 0.01$. Prior 2 is given by

$$(\mathbf{\Sigma}_{-k})_{i,i} = \begin{cases} 1 - \frac{i-1}{M-k} + \delta & \text{if } i \leq M-k+1; \\ \delta & \text{if } i > M-k+1, \end{cases} \tag{19}$$

with $\delta = 0.01$. To show the effectiveness of our prior designed for bandwidth detection, we compare the mean bandwidth obtained by giving $\hat{\boldsymbol{f}}_{-k}$ our designed priors and uniform prior $\mathbf{\Sigma}_{-k} = \boldsymbol{I}_{(N-k)\times(N-k)}$, as shown in Figure 2. The test with uniform prior fails to detect the bandwidth when $M = 80$, while the test with designed prior 1 and prior 2 performs well under the same sample size. When the sample size increases to $2N$, the test with uniform prior turns out to have an acceptable performance, but still worse than the test with designed priors with $M = 80$. This implies that the prior we designed for bandwidth detection is very helpful for saving the sample size. We can also find that the bandwidth detection performance of our method with prior 1 and prior 2 are comparable, which indicates the guideline of prior design in our algorithm improves the detection performance instead of a specific form of prior.

The performance of our method is also compared with the bandwidth estimation method in Reference [15], in which a dictionary of 80 bandlimited kernels is constructed with $\beta = 10^3$ and the regularization parameter $\mu$ in it is set to 0.01 by cross validation. We first simulate the performance of bandwidth detection at different noise levels for bandwidth $K = \{10, 30, 50\}$, the sample size is set to $M = 80$. The results are shown in Figure 3. We can find that the bias and standard deviation (SD) of Algorithm 1 and sparsity based method [15] are similar for small bandwidth at low noise level. For large bandwidth, the bias and SD of our method are significantly smaller than those of Reference [15]. This is because the GS is no longer sparse for large bandwidth, which is out of the scope of sparsity based method [15]. At high noise level, Algorithm 1 is much more robust than Reference [15] especially when the bandwidth is close to the sample size.

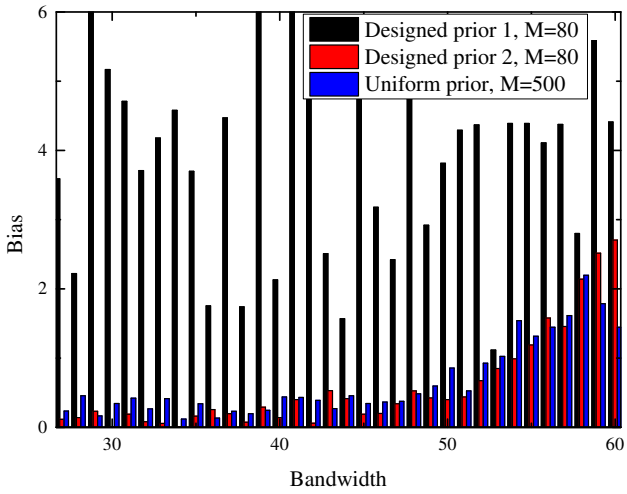



**Figure 2.** Bias of bandwidth detection under different priors.

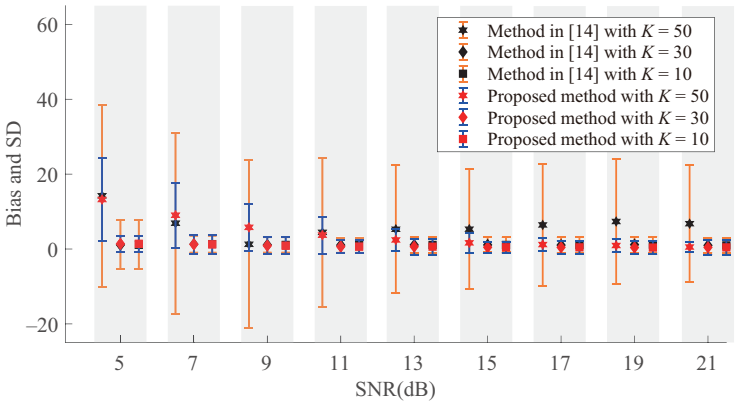



**Figure 3.** Bias and SD of bandwidth detection at different noise levels.

In Figure 4, we show that the increase of sample size can help decrease the bias and SD of Algorithm 1. However, when the sample size is abundant, the performance improvement caused by the increase of sample size can be ignored.

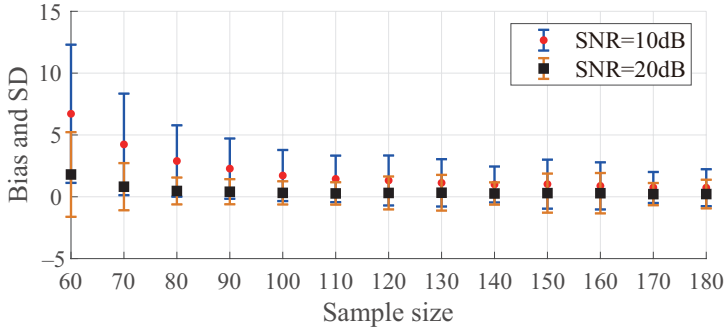



**Figure 4.** Bias and SD of bandwidth detection with different sample sizes when $K = 30$.

### *4.2. Signal Estimation*

Bandwidth is an useful prior when estimating signals from partial observations. In this section, we use the bandwidth detected by Algorithm 1 as a prior in signal estimation and a least squares (LS) estimator is used to estimate the signal with the form

$$f' = \boldsymbol{V}_K(\boldsymbol{V}_K^T\boldsymbol{V}_K)^{-1}\boldsymbol{V}_K^T\boldsymbol{y}. \tag{20}$$

The estimation performance is evaluated by the normalized mean square error (NMSE), which is

$$\text{NMSE} = \frac{\|(f' - f)\|_2^2}{\|f\|_2^2}. \tag{21}$$

We first show the estimation performance on the same graph signal with that in Section 4.1. The bandwidth of signal varies from 10 to 60, the sample size is set to $M = 80$ and the SNR is set to 20 dB. Our method is compared with two methods based on sparsity, namely BP [26] and BCS [21], which also select frequency components to estimate the signal. The result is show in Figure 5, we can find that as the bandwidth increases, the performance of the sparsity base methods degrade rapidly.

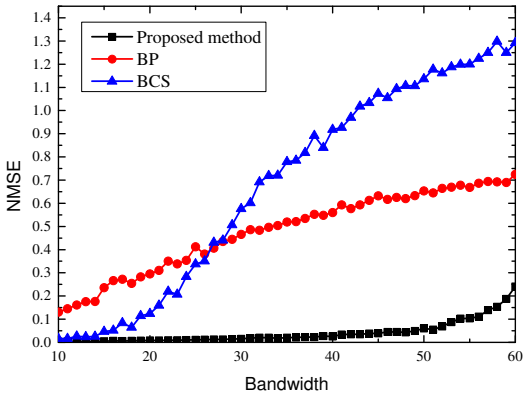
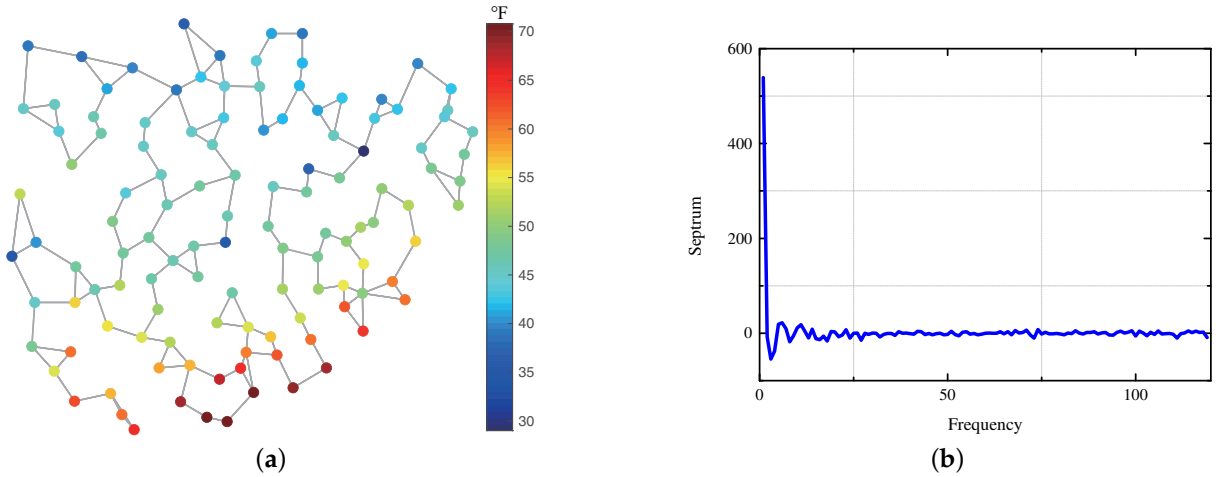



**Figure 5.** NMSE of graph signal (GS) estimation of different algorithms under different bandwidths.

This is because the sparsity based methods select frequency components in the whole space and the sparsity constraint makes the amount of frequency components they selected as small as possible. Therefore, they are not fit for the application when the frequency spectrum is not sparse.

We also compare the performance of the signal estimation in real-data set with BP and BCS. The data set comprises 22 signals corresponding to the average temperature on January 1st in the intervals 1998–2019 measured by 119 stations in China [27]. Each station is identified with a vertex and the distance between 2 vertices is calculated by the haversine formula. A 3-NN unweighted graph is constructed. The GS is shown in Figure 6a and the corresponding spectrum is shown in Figure 6b.




(**a**) (**b**)

**Figure 6.** GS constructed from the temperature measured by 119 stations on 1st January 1998 and its spectrum. (**a**) GS on the 3-NN graph. (**b**) Spectrum of the GS in (**a**).

*i.i.d*. Gaussian noise with SNR = 20 dB is added to the signal to generate observations. 200 simulations are implemented for each signal, the vertices are sampled randomly in each simulation. We can find that the temperatures distribute over the graph smoothly, so there are only a few low frequency components, which is sparse in spectrum. Since the GS is approximately bandlimited in real-data, the approximate bandwidth obtained by Algorithm 1 is used in (20). The average NMSE of all the simulations under different sample sizes is shown in Figure 7.

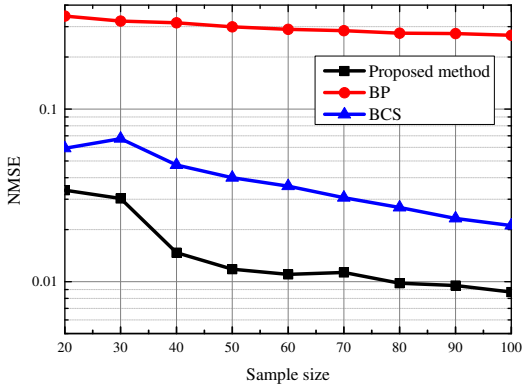



**Figure 7.** NMSE of GS estimation of different algorithms under different sample sizes.

We can find that the estimation performance of our method is also better than the sparsity based methods in real-data. Since BP ensures the sparsity of the signal by $l_1$-norm, it results in obtaining the same signal values on each vertex in this experiment. BCS selects 14 nonadjacent frequency components in all the frequency components, including high frequencies and low frequencies. The approximate bandwidth obtained by Algorithm 1 is 12, which means we estimate the GS using the first 12 frequency components. We can find that if the GS is known to be bandlimited, the signal estimation performance can be improved by detecting the bandwidth accurately first and using the bandwidth as a prior in estimation. If the knowledge that the GS is bandlimited is not take advantage of, more samples are needed to achieve the same estimation performance.

## 5. Conclusions

In this letter, we proposed a multi-stage Bayesian score test approach for the bandlimited GS bandwidth detection with a small sample size. By customizing a prior for frequency coefficients being tested, we made the test more distinguishable from the models with bandwidths close to the assumed one. In practice, our method may obtain a bandwidth smaller than the true bandwidth $K$ if there is a frequency band of which the coefficients are zeros in band, since we focus on testing the frequency coefficients close to the assumed one but ignore the others in the small sample size situation. We will try to improve the performance under this situation in the future. In addition, we would like to analyze how the different sampling sets affect the performance of bandwidth detection and design a sampling set to make the test more powerful.

**Author Contributions:** Conceptualization, X.X. and H.F.; methodology, X.X.; validation, X.X.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X., H.F. and B.H.; visualization, X.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Shanghai Municipal Natural Science Foundation number 19ZR1404700 and 2020 Okawa Foundation Research Grant.

**Acknowledgments:** We thank the Fudan-Zhuhai Innovation Institute for supporting this research.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References


1. Egilmez, H.E.; Ortega, A. Spectral anomaly detection using graph-based filtering for wireless sensor networks. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1085–1089.
2. Sakiyama, A.; Tanaka, Y.; Tanaka, T.; Ortega, A. Efficient sensor position selection using graph signal sampling theory. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 6225–6229.
3. Goldsberry, L.; Huang, W.; Wymbs, N.F.; Grafton, S.T.; Bassett, D.S.; Ribeiro, A. Brain signal analytics from graph signal processing perspective. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 851–855.
4. Hu, C.; Cheng, L.; Sepulcre, J.; Johnson, K.A.; Fakhri, G.E.; Lu, Y.M.; Li, Q. A spectral graph regression model for learning brain connectivity of Alzheimer's disease. *PLoS ONE* **2015**, *10*, e0128136. [CrossRef] [PubMed]
5. Hu, W.; Cheung, G.; Ortega, A.; Au, O.C. Multiresolution graph fourier transform for compression of piecewise smooth images. *IEEE Trans. Image Process.* **2014**, *24*, 419–433. [CrossRef] [PubMed]
6. Thanou, D.; Chou, P.A.; Frossard, P. Graph-based compression of dynamic 3D point cloud sequences. *IEEE Trans. Image Process.* **2016**, *25*, 1765–1778. [CrossRef] [PubMed]
7. Shuman, D.I.; Narang, S.K.; Frossard, P.; Ortega, A.; Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* **2013**, *30*, 83–98. [CrossRef]
8. Ortega, A.; Frossard, P.; Kovačević, J.; Moura, J.M.F.; Vandergheynst, P. Graph Signal Processing: Overview, Challenges, and Applications. *Proc. IEEE* **2018**, *106*, 808–828. [CrossRef]
9. Anis, A.; Gadde, A.; Ortega, A. Efficient sampling set selection for bandlimited graph signals using graph spectral proxies. *IEEE Trans. Signal Process.* **2016**, *64*, 3775–3789. [CrossRef]
10. Chen, S.; Varma, R.; Sandryhaila, A.; Kovačević, J. Discrete Signal Processing on Graphs: Sampling Theory. *IEEE Trans. Signal Process.* **2015**, *63*, 6510–6523. [CrossRef]
11. Wei, Z.; Li, B.; Guo, W. Optimal sampling for dynamic complex networks with graph-bandlimited initialization. *IEEE Access* **2019**, *7*, 150294–150305. [CrossRef]
12. Onuki, M.; Ono, S.; Yamagishi, M.; Tanaka, Y. Graph signal denoising via trilateral filter on graph spectral domain. *IEEE Trans. Signal Inf. Process. Over Netw.* **2016**, *2*, 137–148. [CrossRef]
13. Wang, X.; Chen, J.; Gu, Y. Local measurement and reconstruction for noisy bandlimited graph signals. *Signal Process.* **2016**, *129*, 119–129. [CrossRef]
14. Huang, C.; Zhang, Q.; Huang, J.; Yang, L. Reconstruction of bandlimited graph signals from measurements. *Digit. Signal Process.* **2020**, *101*, 102728. [CrossRef]
15. Romero, D.; Ma, M.; Giannakis, G.B. Kernel-based reconstruction of graph signals. *IEEE Trans. Signal Process.* **2016**, *65*, 764–778. [CrossRef]
16. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012; Volume 821.
17. Lan, W.; Wang, H.; Tsai, C.L. Testing covariates in high-dimensional regression. *Ann. Inst. Stat. Math.* **2013**, *66*, 279–301. [CrossRef]
18. Zhong, P.S.; Chen, S.X. Tests for High-Dimensional Regression Coefficients With Factorial Designs. *J. Am. Stat. Assoc.* **2011**, *106*, 260–274. [CrossRef]
19. Goeman, J.J.; Van De Geer, S.A.; Van Houwelingen, H.C. Testing against a high dimensional alternative. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2006**, *68*, 477–493. [CrossRef]
20. Tropp, J.; Gilbert, A.C. Signal recovery from partial information via orthogonal matching pursuit. *IEEE Trans. Inform. Theory* **2007**, *53*, 4655–4666. [CrossRef]
21. Ji, S.; Xue, Y.; Carin, L. Bayesian compressive sensing. *IEEE Trans. Signal Process.* **2008**, *56*, 2346–2356. [CrossRef]
22. Steyerberg, E. Stepwise Selection in Small Data Sets A Simulation Study of Bias in Logistic Regression Analysis. *J. Clin. Epidemiol.* **1999**, *52*, 935–942. [CrossRef]
23. Davidson, R.; MacKinnon, J.G. *Econom. Theory Methods*; Oxford University Press New York, NY, USA, 2004; Volume 5.
24. Imhof, J.P. Computing the distribution of quadratic forms in normal variables. *Biometrika* **1961**, *48*, 419–426. [CrossRef]
25. Omelka, M. The behavior of locally most powerful tests. *Kybernetika* **2005**, *41*, 699–712.
26. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic decomposition by basis pursuit. *SIAM Rev.* **2001**, *43*, 129–159. [CrossRef]
27. Federal Climate Complex Global Surface Summary of Day Data. Available online: http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html (accessed on 29 August 2019).