Uncertainty Quantification for Sparse Estimation of Spectral Lines

Yi Han and Thomas C. M. Lee, *Senior Member, IEEE*

***Abstract*—Line spectral estimation is an important problem that finds many useful applications in signal processing. Many high-performance methods have been proposed for solving this problem: they select the number of spectral lines and provide point estimates of the frequencies and amplitudes of such spectral lines. This paper studies the line spectral estimation problem from a different and equally important angle: uncertainty quantification. More precisely, this paper develops a novel method that provides an uncertainty measure for the number of spectral lines and also offers point estimates and confidence intervals for other parameters of interest. The proposed method is based on the generalized fiducial inference framework and is shown to possess desirable theoretical and empirical properties. It has also been numerically compared with existing methods in the literature and applied for the detection of exoplanets.**

***Index Terms*—Confidence intervals, exoplanet detection, generalized fiducial inference, line spectral estimation, high-dimensional grid selection.**

I. Introduction

SPECTRAL analysis is an important topic that attracts much attention in the signal processing community. It has rich applications in areas like speech coding [1], radar and sonar signal processing [2], [3], and imaging system [4], to name a few.

This paper focuses on the sparse spectral line estimation problem as described, for example, in [5]. Let

$$\boldsymbol{Y} = [Y(t_1), Y(t_2), \ldots, Y(t_N)]^T \in \mathbb{C}^{N\times 1} \tag{1}$$

denote the complex-valued signal data vector, where the observed times $t_k \in \mathbb{R}^+$, $k \in 1, \ldots, N$, are not required to be regularly spaced. We shall focus on complex-valued signals, but our methodology can be naturally carried over to real-valued signals; see Section VI. We assume that $\boldsymbol{Y}$ satisfies the following model, sometimes known as the sinusoids-in-noise model [5], in which p represents the true number of significant frequencies:

$$\boldsymbol{Y} = \sum_{l=1}^{p} \alpha_l \boldsymbol{a}(f_l) + \boldsymbol{\epsilon}, \tag{2}$$

where $\alpha_l \in \mathbb{C}$ are the complex amplitudes of the p sinusoidal components, $f_l \in \mathbb{R}$ are the true frequencies, and

$$\boldsymbol{a}(f) = [e^{i2\pi f t_1}, e^{i2\pi f t_2}, e^{i2\pi f t_3}, \ldots, e^{i2\pi f t_N}]^T \in \mathbb{C}^{N\times 1}.$$

Also, $\boldsymbol{\epsilon} \in \mathbb{C}^{N\times 1}$ is the noise vector, and we assume that its elements are i.i.d. and follow the complex normal distribution with mean 0 and variance σ^2, denoted as $\mathcal{CN}(0, \sigma^2)$.

With this setup, the problem is to use the observation signal vector $\boldsymbol{Y}$ to estimate the true frequencies f_l and their amplitudes $|\alpha_l|$. This problem has been studied for a long time, and different methods have been proposed. An earlier set of methods are non-parametric, including conventional periodogram-based methods and variants like the Daniell method [6] and the Welch method [7]. There are also correlogram-based, temporal windowing, and lag windowing methods. However, these methods may show low performance, such as limited resolving power.

The second set of methods is parametric and models the time series data with auto-regressive or auto-regressive moving-average processes [8], [9]. They provide accurate spectral estimation if the assumed model is appropriate for the observed time series. However, a drawback of these methods is that they typically require prior knowledge of the number of true frequencies p, which is often not practical.

The third set of methods is semi-parametric, which mostly performs sparse estimation. The performances of these methods are similar to those of parametric methods despite not requiring prior knowledge of p. Some of these methods perform sparse data recovery using mixed norm approximation [10], or atomic norm denoising [11]. Also, there are other sparse estimation methods that need other prior information, such as the noise variance as in [12]. One notable exception is the LIKES method (LiKelihood-based Estimation of Sparse), which does not require prior information [5].

Lastly, Bayesian methods have also been proposed [13], [14]. In addition to offering point estimates, the latter work also provides uncertainty quantification for some parameters of interest.

We need more notations to proceed. Assume $f_{\max}$ to be the upper bound of all the true frequencies $\{f_l\}$; i.e., $f_{\max} \geq f_l, l = 1, \ldots, p$. Let Δ be the step size or the distance between two adjacent grid points of a uniform grid covering the interval $[0, f_{\max}]$. This paper only considers the positive frequencies for notation simplicity, but the discussion can be straightforwardly extended to negative frequencies. Finally, write

$$K = \left\lfloor \frac{f_{\max}}{\Delta} \right\rfloor \tag{3}$$

Manuscript received 8 May 2022; revised 28 October 2022 and 22 December 2022; accepted 4 January 2023. Date of current version 19 January 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Mr. Michael Muma. This work was supported by the National Science Foundation under Grants DMS-1916125, CCF-1934568, DMS-2113605, and DMS-2210388. *(Corresponding author: Thomas C. M. Lee.)*

The authors are with the Department of Statistics, University of California, Davis, CA 95616 USA (e-mail: yyihan@ucdavis.edu; tcmlee@ucdavis.edu).

Digital Object Identifier 10.1109/TSP.2023.3235662

and

$$\boldsymbol{A} = [\boldsymbol{a}(0), \boldsymbol{a}(\Delta), \ldots, \boldsymbol{a}((K-1)\Delta))] \in \mathbb{C}^{N\times K}.$$

Using these notations, we can approximately re-express (2) as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{4}$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^T$ is a sparse vector with mostly zero elements. Those non-zero elements of $\boldsymbol{\beta}$ equal to $\{\alpha_l\}$, while the indexes of these non-zero elements represent the corresponding frequencies in $\boldsymbol{A}$ that are equal to $\{f_l\}$.

The main idea of this so-called on-grid method is to use the grid that is closest to the true frequency to approximate it. Also, the problem of estimating $\{\alpha_l, f_l\}$ can be reformulated as estimating the sparse vector $\boldsymbol{\beta}$ in (4) and detecting the non-zero elements of the sparse vector. By choosing K to be sufficiently large and Δ to be sufficiently small, one can have the distance between the true frequencies and their closest grids be practically negligible. However, a very large value of K usually implies high-dimensional problems, so there is actually a trade-off between estimation accuracy and computational efficiency.

In practice, K is almost always much larger than N, which makes the estimation of $\boldsymbol{\beta}$ from $\boldsymbol{Y}$ in (4) a very challenging task. Different methods have been proposed to solve this problem, where some require additional prior knowledge such as noise variance or the number of non-zero frequencies; e.g., see [12], [15], [16].

It is fair to say that most existing methods focus on estimating the amplitudes α_l and noise variance σ^2. At the same time, very little treatment has been given to the issue of uncertainty quantification. The main goal of this paper is to construct confidence intervals for α_l, σ^2, as well as the number of the true frequencies, p. The proposed method is based on the relatively new methodology termed generalized fiducial inference (GFI) [17]. To the best of our knowledge, this is one of the first complete systematic analyses that capture these uncertainties in the line spectral estimation problem. It is also the first time that GFI is being applied to a complex-valued problem.

The rest of this paper is organized as follows. Section II provides some background material and usage on GFI. Section III applies the methodology to the sparse line spectral estimation problem, and one relatively simple and fast algorithm to generate fiducial samples is proposed. The theoretical properties of the proposed solution are examined in Section IV, while its empirical properties are illustrated in Sections V and VI by numerical simulations and a real data application. Lastly, concluding remarks are offered in Section VII, and technical details are provided in the appendix.

II. Methodology

A. A Brief History of Generalized Fiducial Inference

The idea of fiducial inference was first proposed by Fisher in 1930s [18] as an alternative to the Bayesian approach with the goal of constructing an appropriate statistical distribution on the estimator of an unknown parameter. One potential issue of the Bayesian approach is that, when inappropriate prior distributions are used, the performance and reliability of the approach could be affected. Fisher's fiducial method intends to avoid using the prior distribution; instead, it considers a switching mechanism between the model parameters and the observed data that is very similar to the idea of the method of maximum likelihood. In spite of Fisher's continuous endeavor to complete the framework of fiducial inference, it has not received much attention because it works well only for single-parameter problems but fails in the context of multiple parameters. Interested readers are referred to [17], where a more detailed introduction about the history of fiducial inference is given.

In recent years, there has been a resurgent interest in reformulating the fiducial concept. These modifications include Dempster-Shafer theory [19], [20] and inferential models [21]. Motivated by generalized confidence intervals [22], [23] and the surrogate variable method for obtaining confidence intervals for variance components [24], GFI was developed in a series of papers published around 2010 s, and summarized in [17]. It has been successfully applied to solve different uncertainty quantification problems, including wavelet regression [25], ultrahigh dimensional regression [26] and sparse additive models [27].

B. An Introduction to Generalized Fiducial Inference

As mentioned before, GFI utilizes the idea of a so-called switching principle that is similar to Fisher's celebrated maximum likelihood method. It first begins with expressing the relationship between the data $\boldsymbol{Y}$ and the parameter $\boldsymbol{\theta}$ with

$$\boldsymbol{Y} = \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{G}(\cdot,\cdot)$ is sometimes known as the "structural equation." Also, $\boldsymbol{U}$ is the random component of the problem whose distribution is **completely known** and is independent of $\boldsymbol{\theta}$. For example, for the problem of estimating μ from $\{X_i\}_{i=1}^n$ with X_i's as i.i.d. $\mathcal{N}(\mu, \sigma^2)$, we write $X_i = \mu + \sigma Z_i$ with Z_i as i.i.d. $\mathcal{N}(0,1)$, where the parameter $\boldsymbol{\theta} = \{\mu, \sigma\}$, data $\boldsymbol{Y} = \{X_i\}_{i=1}^n$ and random component $\boldsymbol{U} = \{Z_i\}_{i=1}^n$. Note that the distribution of $\boldsymbol{U}$ is completely known.

Similar to the main idea behind maximum likelihood estimation, with the switching principle, the roles of $\boldsymbol{\theta}$ and $\boldsymbol{Y}$ are switched in the GFI framework once the data are observed. That is, to treat the random data $\boldsymbol{Y}$ as deterministic and the deterministic parameter $\boldsymbol{\theta}$ as random. With this idea, we can define a set $\{\boldsymbol{\theta} : \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{U}^*, \boldsymbol{\theta})\}$ as the inverse mapping of $\boldsymbol{G}$, where $\boldsymbol{U}^*$ is an independent copy of $\boldsymbol{U}$ and $\boldsymbol{y}$ is an observation of $\boldsymbol{Y}$. A method is provided by [28] to ensure the existence and uniqueness of this inverse mapping.

With the above setup, we can build a distribution of $\boldsymbol{\theta}$ from (5) in the following manner. For any observed data $\boldsymbol{y}$ and $\boldsymbol{u}$, we can adopt the method from [28] to identify one $\boldsymbol{\theta}$ that guarantees the existence of the inverse

$$\boldsymbol{H}_{\boldsymbol{y}}(\boldsymbol{u}) = \{\boldsymbol{\theta} : \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{U}^*, \boldsymbol{\theta})\}. \tag{6}$$

Since the distribution of $\boldsymbol{U}$ is totally known and independent of $\boldsymbol{\theta}$, we can generate the random samples $\boldsymbol{U_1}, \boldsymbol{U_2}, \ldots$ and use (6) to obtain the random samples for $\boldsymbol{\theta}$ via

$$\boldsymbol{\theta}_1 = \boldsymbol{H}_{\boldsymbol{y}}(\boldsymbol{U}_1), \quad \boldsymbol{\theta}_2 = \boldsymbol{H}_{\boldsymbol{y}}(\boldsymbol{U}_2), \ldots$$

In other words, GFI transfers the randomness in $\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots$ to $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ via the inverse equation (6). We call these $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ *fiducial samples*, which can be used to calculate point estimates and construct confidence intervals of $\boldsymbol{\theta}$ in a way similar to posterior samples in the Bayesian context. Notice that an explicit expression for $\boldsymbol{H_y}$ may not exist for certain problems, but next, we describe how the fiducial samples can still be generated without calculating an expression for $\boldsymbol{H_y}$.

Through (6) one can see that a density function $r(\boldsymbol{\theta}|\boldsymbol{y})$ is implicitly defined for $\boldsymbol{\theta}$. We refer $r(\boldsymbol{\theta}|\boldsymbol{y})$ as the *generalized fiducial density* (GFD) of $\boldsymbol{\theta}$, which plays a similar role as the posterior density in the Bayesian context. It is shown in [17] that under some mild smoothness assumptions on the likelihood function $f(\boldsymbol{y}, \boldsymbol{\theta})$ of $\boldsymbol{y}$, the GFD $r(\boldsymbol{\theta}|\boldsymbol{y})$ admits the following expression

$$r(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}, \boldsymbol{\theta}) J(\boldsymbol{y}, \boldsymbol{\theta})}{\int_\Theta f\left(\boldsymbol{y}, \boldsymbol{\theta}'\right) J\left(\boldsymbol{y}, \boldsymbol{\theta}'\right) d\boldsymbol{\theta}'}, \tag{7}$$

where

$$J(\boldsymbol{y}, \boldsymbol{\theta}) = D\left(\frac{\boldsymbol{d}}{\boldsymbol{d\theta}} \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta})|_{\boldsymbol{U}=\boldsymbol{G}^{-1}(\boldsymbol{Y}, \boldsymbol{\theta})}\right) \tag{8}$$

with $D(\boldsymbol{A}) = |\det(\boldsymbol{A}^T \boldsymbol{A})|^{1/2}$ and $\boldsymbol{u} = \boldsymbol{G^{-1}}(\boldsymbol{y}, \boldsymbol{\theta})$ as the value of $\boldsymbol{u}$ such that $\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta})$.

We note that although (7) provides an explicit formula to calculate the GFD, it may not be as straightforward as it looks: the denominator requires the calculation of an integral that is intractable for some problems, and hence Monte Carlo or other numerical techniques are needed to sample from the GFD $r(\boldsymbol{\theta}|\boldsymbol{y})$.

C. Incorporating Model Selection in GFI

Up to now, our discussion on GFI assumes that the dimension of $\boldsymbol{\theta}$ is fixed and known. In other words, (7) cannot be used for model selection problems, where the size of $\boldsymbol{\theta}$ also needs to be chosen.

In the context of wavelet regression, [25] incorporated model selection in the GFI framework, which can be extended to more general situations. The idea is similar to penalized likelihood estimation, where a penalty term is added to the (log)-likelihood function to achieve a balanced trade-off between data fidelity and model complexity. Here we provide a brief description and refer the reader to [17] for further details.

Let $\mathcal{M}$ be the set of all possible models and $\boldsymbol{\theta}_M$ be the parameters of any model $M \in \mathcal{M}$. The GFD of $(\boldsymbol{\theta}_M, M)$ can be expressed as

$$r(\boldsymbol{\theta}_M, M|\boldsymbol{y}) = r(\boldsymbol{\theta}_M|\boldsymbol{y}, M) r(M|\boldsymbol{y}),$$

where the conditional GFD $r(\boldsymbol{\theta}_M|\boldsymbol{y}, M)$ of $\boldsymbol{\theta}_M$ (given M) can be calculated using (7), while the marginal GFD $r(M|\boldsymbol{y})$ of M admits the expression

$$r(M|\boldsymbol{y}) = \frac{\int r(\boldsymbol{\theta}_M|\boldsymbol{y}, M) e^{-q(M)} d\boldsymbol{\theta}_M}{\sum_{M' \in \mathcal{M}} \int r(\boldsymbol{\theta}_{M'}|\boldsymbol{y}, M') e^{-q(M')} d\boldsymbol{\theta}_{M'}}, \tag{9}$$

where $q(M)$ is the penalty associated with model M.

Different choices of $q(M)$ will lead to different penalty strengths, which will in turn affect the final results. In general, the stronger the penalty, the lesser the number of spectral lines we would expect to obtain. When $q(M)$ is suitably chosen, it leads to some well-known model selection methods commonly used in the signal processing and statistics communities. For example, if we set $q(M) = 2|M|$ with $|M|$ as the number of parameters in model M, we have the Akaike Information Criterion (AIC). Here we follow [26] and choose $q(M)$ as

$$q(M) = \frac{|M|}{2} \log N + \log_{e^{1/\gamma}} \binom{K}{|M|}, \tag{10}$$

where K is the number of parameters of the largest model in $\mathcal{M}$. Also, γ is a constant measuring the sparsity belief of the model. A natural choice is $\gamma = 1$, but other choices are also possible, and we note that there is not a universal choice of γ that is suitable for all different kinds of true models. In our work, we use $\gamma = 1$, which aligns (10) with the minimum description length principle [29] for high-dimensional problems [30]. This is a main reason behind our choice of $q(M)$, as the minimum description length principle is a well-studied model selection method that often produces excellent theoretical and empirical results.

III. GFI for Line Spectral Estimation

This section applies the above GFI methodology to the line spectral estimation problem represented by (4). We shall calculate the GFDs for this problem and devise a method for generating fiducial samples. To the best of our knowledge, this is the first time that GFI is being applied to a problem with complex-valued coefficients and responses.

Let M_0 be the true model and M be any candidate model such that $|M| < K$. Given M, the structural equation (5) for model (4) is:

$$\boldsymbol{y} = \boldsymbol{A}_M \boldsymbol{\beta}_M + \sigma \boldsymbol{U}, \tag{11}$$

where $\boldsymbol{A}_M$ and $\boldsymbol{\beta}_M$ represent, respectively, the design matrix and the parameter vector of model M. Also, σ is the standard deviation of the error term and $\boldsymbol{U}$ is a standard multivariate complex normal variable; i.e. $\boldsymbol{U} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$. To calculate the GFD of $\boldsymbol{\theta} = (\sigma, \boldsymbol{\beta})^T$ given M for (11), we first compute (8)

$$\begin{aligned}
J(\boldsymbol{y}, \boldsymbol{\theta}) &= D\left(\frac{\boldsymbol{d}}{\boldsymbol{d\theta}} \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta})|_{\boldsymbol{U}=\boldsymbol{G}^{-1}(\boldsymbol{y}, \boldsymbol{\theta})}\right) \\
&= D\left(\boldsymbol{A}_M, \frac{\boldsymbol{y} - \boldsymbol{A}_M \boldsymbol{\beta}_M}{\sigma}\right) \\
&= \left[\det\left\{\begin{pmatrix} \boldsymbol{A}_M^H \\ \frac{\boldsymbol{y}^H - \boldsymbol{\beta}_M^H \boldsymbol{A}_M^H}{\sigma} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}_M & \frac{\boldsymbol{y} - \boldsymbol{A}_M \boldsymbol{\beta}_M}{\sigma} \end{pmatrix}\right\}\right]^{\frac{1}{2}} \\
&= \sigma^{-1} |\det(\boldsymbol{A}_M^H \boldsymbol{A}_M)|^{\frac{1}{2}} \mathrm{RSS}_M^{\frac{1}{2}},
\end{aligned}$$

where $\boldsymbol{A}_M^H$ is the conjugate transpose of a $\boldsymbol{A}$ and RSS_M is the residual sum of squares of model M.

Next we calculate the GFD of $\boldsymbol{\theta}$ given M using (7):

$$r(\boldsymbol{\theta}|\boldsymbol{y},M)=\frac{c_N\sigma^{-1}\mathrm{RSS}_M^{\frac{1}{2}}(\frac{1}{\sigma^{-2n}})e^{\frac{-1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)^H(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)}}{\int_\Theta c_N\sigma^{-1}\mathrm{RSS}_M^{\frac{1}{2}}(\frac{1}{\sigma^{-2n}})e^{\frac{-1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)^H(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)}d\boldsymbol{\theta}'},$$

where $c_N=\frac{1}{\pi^N}|\det(\boldsymbol{A}_M^H\boldsymbol{A}_M)|^{\frac{1}{2}}$.

Let K be the length of $\boldsymbol{\beta}_M$. So the numerator of (9) can be calculated as:

$$\begin{aligned}r\left(M|\boldsymbol{y}\right)\propto&\int\sigma^{-1}|\det\left(\boldsymbol{A}_M^H\boldsymbol{A}_M\right)|^{1/2}\mathrm{RSS}_M^{1/2}\left(\frac{1}{\pi^N\sigma^{-2N}}\right)\\&\cdot e^{-\frac{1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)^H(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)}d\boldsymbol{\theta}e^{-q(M)}\\=&\,e^{-q(M)}\\&\cdot\int\pi^{-N}\sigma^{-2N+1}|\det\left(\boldsymbol{A}_M^H\boldsymbol{A}_M\right)|^{1/2}\mathrm{RSS}_M^{1/2}d\sigma\\&\cdot\int e^{-\frac{1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)^H(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)}d\boldsymbol{\beta}_M,\end{aligned}$$

where the last term is $\int e^{-\frac{1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)^H(\boldsymbol{y}-\boldsymbol{A}_M\boldsymbol{\beta}_M)}d\boldsymbol{\beta}_M=\pi^{|M|}\sigma^{2|M|}\det(\boldsymbol{A}_M^H\boldsymbol{A}_M)^{-1}\exp(-\frac{\mathrm{RSS}_M}{\sigma^2})$. Therefore

$$\begin{aligned}r\left(M|\boldsymbol{y}\right)\propto&\int\pi^{-N}\sigma^{-(2N+1)}|\det\left(\boldsymbol{A}_M^H\boldsymbol{A}_M\right)|^{1/2}\pi^{|M|}\sigma^{2|M|}\\&\cdot\frac{1}{\det\left(\boldsymbol{A}_M^H\boldsymbol{A}_M\right)}e^{-\frac{1}{\sigma^2}\mathrm{RSS}_M}d\sigma e^{-q(M)}\\=&\,\pi^{|M|-N}|\det\left(\boldsymbol{A}_M^H\boldsymbol{A}_M\right)|^{-1/2}\mathrm{RSS}_M^{1/2}e^{-q(M)}\\&\cdot\int\sigma^{2|M|-2N-1}e^{-\frac{1}{\sigma^2}\mathrm{RSS}_M}d\sigma\\=&\,\pi^{|M|-N}|\det\left(\boldsymbol{A}_M^H\boldsymbol{A}_M\right)|^{-1/2}\cdot\mathrm{RSS}_M^{\frac{1}{2}+|M|-N}\\&\cdot\Gamma\left(N-|M|\right)\cdot e^{-q(M)}.\end{aligned}\tag{12}$$

A. Generating Fiducial Samples

This subsection presents a method for generating fiducial samples for the line spectral estimation problem that this paper considers. The idea is to first generate a candidate model M, then given M, generate $\boldsymbol{\theta}=(\sigma,\boldsymbol{\beta})$.

First of all, due to the large number of columns of $\boldsymbol{A}$ in the line spectral estimation context, we are facing an extremely large number of potential models in the model set $\mathcal{M}$; i.e., the cardinality of $\mathcal{M}$ equals 2^K, which is often intractable. Therefore, for various practical considerations, we only consider models from a subset $\mathcal{M}^*$ of $\mathcal{M}$. We delay our discussion of how to choose $\mathcal{M}^*$ to Appendix A. In principle, an ideal $\mathcal{M}^*$ should include all the models that have a non-negligible value of $r(M|\boldsymbol{Y})$, while at the same time excluding other models that have a zero or near-zero $r(M|\boldsymbol{Y})$ value.

Suppose now we have a good $\mathcal{M}^*$. For each $M\in\mathcal{M}^*$, we compute (see (12))

$$\begin{aligned}R(M)=&\,\pi^{|M|-N}|\det(\boldsymbol{A}_M^H\boldsymbol{A}_M)|^{-1/2}\cdot\mathrm{RSS}_M^{\frac{1}{2}+|M|-N}\\&\cdot\Gamma(N-|M|)\cdot e^{-q(M)},\end{aligned}$$

where $e^{-q(M)}$ is given by (10). The generalized fiducial probability $r(M|\boldsymbol{y})$ (12) can then be well approximated by

$$r(M|\boldsymbol{y})\approx\frac{R(M)}{\sum_{M^*\in\mathcal{M}^*}R(M^*)}.\tag{13}$$

We can then sample a candidate model $M\in\mathcal{M}^*$ from (13).

Once a model M is generated, we set up the corresponding design matrix $\boldsymbol{A}_M$. Then we estimate the parameters $\boldsymbol{\beta}_M$ of the generated model M using maximum likelihood and obtain the estimate $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ and the corresponding residual sum of squares RSS_M. As $\boldsymbol{A}_M^H\boldsymbol{A}_M$ is of full rank (i.e., not in a high-dimensional setting), these two quantities can be calculated using classical regression formulae: $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}=(\boldsymbol{A}_M^H\boldsymbol{A}_M)^{-1}\boldsymbol{A}_M^H\boldsymbol{y}$ and $\mathrm{RSS}_M=\boldsymbol{y}^H(I-\boldsymbol{A}_M(\boldsymbol{A}_M^H\boldsymbol{A}_M)^{-1}\boldsymbol{A}_M^H)\boldsymbol{y}$. Then, using the properties of the complex normal distribution, σ and $\boldsymbol{\beta}$ can be sampled using the following distributional results:

$$\frac{2\mathrm{RSS}_M}{\sigma^2}\sim\chi^2_{2(N-|M|)}\tag{14}$$

and

$$\boldsymbol{\beta}\sim\mathcal{CN}\left(\hat{\boldsymbol{\beta}}_{\mathrm{ML}},\sigma^2(\boldsymbol{A}_M^H\boldsymbol{A}_M)^{-1}\right),\tag{15}$$

where $\chi^2_{2(N-|M|)}$ is the chi-square distribution with $2(N-|M|)$ degrees of freedom.

To sum up, a fiducial sample for $(M,\sigma,\boldsymbol{\beta})$ can be generated by the following steps.

1) Sample a model M from $\mathcal{M}^*$ using (13).
2) Fit M using maximum likelihood and obtain $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ and RSS_M.
3) Sample σ^2 using (14).
4) Sample $\boldsymbol{\beta}$ using (15), where the σ^2 obtained from the above step is used in the RHS of (15).

By repeating the above steps, one can generate enough samples of $(M,\sigma,\boldsymbol{\beta})$ for forming point estimates and constructing confidence intervals. Notice that (13) only needs to be calculated once, so it is fast to generate an M. Also, notice that no costly procedures are required to generate σ or $\boldsymbol{\beta}$ so overall the whole sample method is fast.

B. Point Estimates and Confidence Intervals

Repeating the above procedure, we obtain multiple fiducial samples for $(M,\sigma,\boldsymbol{\beta})$, which can be used to perform statistical inference in a similar manner as with posterior samples in the Bayesian context. For the case of σ, we can use the mean or the median of its fiducial samples as a point estimate, and the $(\alpha/2,1-\alpha/2)$ quantiles of the fiducial samples to be its $100(1-\alpha)\%$ confidence intervals.

The situation is less straightforward for M, as its domain $\mathcal{M}$ is a discrete space with 2^K elements and it is not entirely clear what would be a universally accepted definition for a "confidence interval" for a model. However, the fiducial samples of M could still provide valuable information on uncertainties. For example, for any M the samples can be used to approximate the generalized fiducial probability $r(M|\boldsymbol{y})$ as in (13), which is a numerical measure indicating how likely (or unlikely) M is the true model.

The fiducial samples can also provide uncertainty information for p, the number of significant frequencies. For example, the generalized fiducial probability for $p = l$ can be approximated by the sum of the generalized fiducial probabilities $r(M|\boldsymbol{y})$ of all models M with $p = l$.

One can also construct confidence intervals for $\boldsymbol{\beta}$ in the following manner. First, notice that, unless the GFD $r(M|\boldsymbol{y})$ has all its mass at one model, the fiducial samples will contain different models. In other words, for any $l = 1, \ldots, K$, β_l may be declared insignificant by some of the fiducial samples. These insignificant fiducial samples for β_l are zero, which will have an adverse effect when calculating averages or quantiles with those non-zero β_l values. We follow [26] to handle this issue: for each β_l, we count the percentage of non-zero fiducial sample values. If it is more than 50%, we claim that this specific β_l is significant and use all the non-zero fiducial sample values to obtain point estimates and confidence intervals, in the same manner as for σ.

IV. Theoretical Properties

This section investigates the theoretical properties of the above proposed GFI-based method under the situation that K is diverging and the size of the true model is fixed.

First, some notations. Recall M is any candidate model and M_0 is the true model. Let $\boldsymbol{P}_M$ be the projection matrix of $\boldsymbol{A}_M$; i.e., $\boldsymbol{P}_M = \boldsymbol{A}_M(\boldsymbol{A}_M^H\boldsymbol{A}_M)^{-1}\boldsymbol{A}_M^H$. Define $\Delta_M = \|\boldsymbol{\mu} - \boldsymbol{P}_M\boldsymbol{\mu}\|^2$, where $\boldsymbol{\mu} = E(\boldsymbol{Y}) = \boldsymbol{A}_{M_0}\boldsymbol{\beta}_{M_0}$.

Throughout this section, we assume that the following identifiability condition holds:

$$\lim_{n\to\infty} \min\left\{ \frac{\Delta_M}{|M_0|\log(K)} : M_0 \not\subset M, |M| \le b|M_0| \right\} = \infty \tag{16}$$

for some fixed constant $b > 1$. This b ensures that we only consider models whose size is comparable to the true model. This assumption is an identifiability condition because it guarantees the uniqueness of the true model among all the models that have a comparable size to the true model. To be more specific, this condition guarantees that if the true model $M_0 \not\subset M$, the residuals will become unbounded as $n \to \infty$. The restriction $|M| \le b|M_0|$ is imposed because in practice only those models with sizes comparable to the true model will be considered. Overall, this assumption means the true model is identifiable if no model other than the true model of comparable size can predict the response almost equally well, which ensures the true model can be differentiated from the other models.

Theorem 4.1: Assume condition (16). If $N \to \infty$, $K \to \infty$, $|M_0|\log(K) = o(N)$, $\frac{\log(|M_0|)}{\log(K)} \to \delta$ and $\frac{\log(N)}{\log(K)} \to \eta$, then there exists $\gamma > \frac{1+\delta}{1-\delta} - \frac{5\eta}{2(1-\delta)}$ such that

$$\max_{M\neq M_0, M\in\mathcal{M}^*} \frac{r(M)}{r(M_0)} \xrightarrow{P} 0. \tag{17}$$

Moreover, Suppose there exists a procedure for obtaining $\mathcal{M}^*$ that satisfies:

$$P\left(M_0 \in \mathcal{M}^*\right) \to 1 \text{ and } \log\left(\left|\mathcal{M}_j^*\right|\right) = o(j\log(N)), \tag{18}$$

where $\mathcal{M}_j^*$ denotes the set of all sub-models in $\mathcal{M}^*$ of size j, we have

$$r(M_0) \xrightarrow{P} 1.$$

Theorem 4.1 implies that, under some regularity conditions, the true model M_0 has the highest generalized fiducial probability amongst all the candidate models. Assumption (18) guarantees the true model in the candidate set and the candidate set not to be too large. The proof of this theorem is provided in the appendix.

V. Simulation Results

Two simulation experiments were conducted to evaluate the practical performance of the proposed GFI method under the line spectral model (1).

A. Confidence Intervals and Widths

In the first experiment, we follow the experimental setting of [5], where

- *Number of spectral lines:* $p = 3$.
- *Parameters:* $f_1 = 0.4230$, $f_2 = 0.6875$, $f_3 = f_2 + \delta_f$, $\alpha_1 = 5e^{i2\pi u_1}$, $\alpha_2 = 5e^{i2\pi u_2}$ and $\alpha_1 = 10e^{i2\pi u_3}$, where u_1, u_2, u_3 are randomly chosen from Unif$(0, 1)$. See below for a discussion on the values used for δ_f.
- *Number of observations:* $N = 50$.
- *Sampling times:* $t_1 = 0$ [sec], $t_N = 50$ [sec], and $\{t_k\}_{k=2}^{49}$ are uniformly randomly selected (real numbers) from the interval (0, 50).
- *Noise:* $\boldsymbol{\epsilon}$ is sampled from a complex normal distribution $\mathcal{CN}(0, \sigma^2\boldsymbol{I}_N)$.

The signal-to-noise ratio (SNR) is defined as

$$SNR = 10\log_{10}\left(\frac{|\alpha_1|^2+|\alpha_2|^2+|\alpha_3|^2}{\sigma^2}\right) = 10\log_{10}\left(\frac{150}{\sigma^2}\right).$$

As in here we have $\min_k(t_{k+1} - t_k) < 0.5$ [sec], we can set $f_{\max} = 1$ [Hz]. For the choices of K and Δ in (3), we adopted the suggestion by [5] and set

$$\Delta = \frac{1}{c(t_N - t_1)}$$

with $c = 20$, which gives $\Delta = 1 \times 10^{-3}$ [Hz]. As we chose $f_{\max} = 1$ [Hz], using (3) we have $K = 1000$.

For the frequency separation δ_f between f_2 and f_3, we considered three values: $\delta_f = \{0.01, 0.015, 0.1\}$. The first two values are considered "high-resolution" cases, and the last is a "normal" case. We also considered two SNRs $= \{5, 10\}$. Therefore, we have six different scenarios in this first simulation experiment. For each scenario, we generated 1,000 data sets and applied the proposed GFI method to each of them, where the number of fiducial samples for each data set was 10,000.

Recall that, unlike many traditional methods, the proposed GFI method also provides the generalized fiducial probabilities $r(M|\boldsymbol{Y})$ for all the candidate models, which in turn can be used to generate the corresponding generalized fiducial probabilities for the number of frequencies p; see Section III-B. These probabilities provide valuable information about how certain or

TABLE I
PERCENTAGES OF TIMES THAT DIFFERENT FREQUENCY NUMBERS p WERE SELECTED BY GFI IN THE SIX DIFFERENT SCENARIOS

estimated number of frequencies:		1	2	3	4
δ_f	SNR	percentages selected			
0.01	5	2%	12.5%	83.3%	2.2%
0.01	10	0%	7.5%	92.1%	0.4%
0.015	5	1.1%	11.2%	86.5%	1.2%
0.015	10	0%	6.2%	93.2%	0.6%
0.1	5	0.9%	10.3%	87.6%	1.2%
0.1	10	0%	4.5%	94.4%	0.1%

Recall that the true $p = 3$.

TABLE II
EMPIRICAL COVERAGE RATES OF THE CONFIDENCE INTERVALS FOR σ^2 OBTAINED BY THE PROPOSED GFI METHOD AND ORACLE

(δ_f, SNR)	method	90% CI	95% CI	99% CI
(0.1, 5)	GFI	91.0% (1.81)	93.8% (1.14)	98.4% (2.88)
	Oracle	88.7% (1.67)	94.4% (2.00)	99.0% (2.66)
(0.1, 10)	GFI	91.6% (0.97)	96.0% (1.16)	99.6% (1.53)
	Oracle	89.7% (0.94)	94.7% (1.12)	99.1% (1.49)
(0.015, 5)	GFI	90.8% (1.81)	96.0% (2.23)	98.5% (2.87)
	Oracle	88.3% (1.68)	94.5% (2.01)	99.3% (2.67)
(0.015, 10)	GFI	88.6% (0.95)	93.8% (1.14)	98.0% (1.51)
	Oracle	89.0% (0.95)	94.8% (1.13)	98.7% (1.50)
(0.01, 5)	GFI	90.6% (1.82)	95.8% (2.76)	98.8% (2.86)
	Oracle	90.5% (1.68)	95.0% (2.00)	98.9% (2.66)
(0.01, 10)	GFI	88.8% (0.96)	96.0% (1.15)	99.6% (1.52)
	Oracle	89.5% (0.94)	94.0% (1.13)	98.5% (1.49)

The numbers in the parentheses are the average widths of the intervals.

uncertain we are with the estimated results. For the six scenarios, Table I lists the percentages of times that different frequency numbers p were selected by the fiducial samples. As expected, the percentages for choosing the correct p are higher when the separation δ_f and/or the SNR are higher. Also, given the high percentages for selecting the true $p = 3$, one may conclude that the GFI estimation results are reliable.

For each data set, we also applied the LIKES method of [31] and the so-called Oracle method that has the knowledge of the true p and f_l's. Of course, the Oracle results cannot be obtained in practice as such knowledge is not available, but they are used here for benchmark comparisons. Table II provides the empirical coverage rates of the confidence intervals from Oracle and the GFI method for σ^2 (note that LIKES does not produce confidence intervals for σ^2). One can observe that the GFI results are comparable to those from Oracle.

We also constructed confidence intervals for the amplitudes α_l's using all three methods: GFI, LIKES, and Oracle. Note that for GFI and LIKES, we only used those results where the true number of frequencies was selected. The empirical coverage rates of these confidence intervals are reported in Table III. One can see that the GFI results are slightly worse than those from Oracle, but in general, are superior to those from LIKES. Also, very often, GFI produced higher empirical coverage rates with narrower confidence intervals.

B. Comparison With Bayesian Approach

In this second experiment, the simulation setting is similar to [14]:

- *Number of spectral lines:* $p = 3$.
- *Parameters:* $f_1 = 0.4230$, $f_2 = 0.6875$, $f_3 = 0.7875$, $\alpha_1 = 1 + 0.1e^{i2\pi u_1}$, $\alpha_2 = 1 + 0.1e^{i2\pi u_2}$ and $\alpha_3 = 5 + 0.1e^{i2\pi u_3}$, where u_1, u_2, u_3 are randomly chosen from Unif$(0, 1)$.
- Other quantities such as the sampling times are the same as in the first experiment.

As in [14] we use these two metrics to measure the quality of the estimation results: the normalized mean-squared-error (NMSE) of $\hat{\boldsymbol{A}}$ (only with those columns selected by the methods) and the mean-squared-error (MSE) of $\boldsymbol{f} = (f_1, f_2, f_3)$, defined respectively as

$$\text{NMSE}(\hat{\boldsymbol{A}}) = 20 \log \left(\|\boldsymbol{A}\boldsymbol{\beta} - \hat{\boldsymbol{A}}\hat{\boldsymbol{\beta}}\|_F / \|\hat{\boldsymbol{A}}\hat{\boldsymbol{\beta}}\|_F \right)$$

and

$$\text{MSE}\left(\hat{\boldsymbol{f}}\right) = 20 \log \left(\|\hat{\boldsymbol{f}} - \boldsymbol{f}\|_2 \right),$$

where $\|\cdot\|_F$ is the Frobenius norm for matrices and $\|\cdot\|_2$ is the L_2 norm for vectors. Following [14], MSE$(\hat{\boldsymbol{f}})$ is calculated only when both the model order p is correctly estimated and MSE$(\hat{\boldsymbol{f}}) \leq 0(dB)$. In addition, we also approximated the probability that the correct model order p is selected; i.e., $P(\hat{p} = 3)$.

For each simulated data set, we applied the GFI method and the MVALSE method of [14] and calculated the above metrics. Fig. 1 summarizes the results when the number of observations is fixed at $N = 75$ with changing SNRs $= \{-5, 0, 5, 10\}$. One can observe that when SNR $= 10$, both methods give comparable results, while GFI is better for the remaining SNRs. Similarly, Fig. 2 presents the results when SNR $= 2$ is fixed for different values of $N = \{25, 50, 75, 100, 125\}$. The results suggest that GFI is superior.

To sum up, results from these two sets of numerical experiments suggest that the proposed GFI method produces highly reliable results, and compares favorably with some of the leading methods in the literature. This agrees largely with the authors' experience in applying GFI to other problems. A thorough theoretical study is underway to identify those conditions under which GFI is expected to produce reliable results.

VI. REAL DATA EXAMPLE: RADIAL VELOCITY ANALYSIS

A. Background

The detection of extrasolar planets, also known as exoplanets, has always been a challenging and fascinating area in astronomy. Until the end of 2021, a total of 1274 exoplanets have been discovered. Popular techniques for exoplanet detection include radial velocity analysis, the transit method, direct imaging, gravitational microlensing, and astrometry minuscule movements; e.g., [32]. Among these techniques, radial velocity analysis is one of the most commonly used.

Radial velocity refers to the speed at which an object (in this case an exoplanet) moves away from Earth (or approaches it, with a negative radial velocity). Orbiting exoplanets cause the

TABLE III
EMPIRICAL COVERAGE RATES OF THE CONFIDENCE INTERVALS FOR THE FREQUENCY AMPLITUDES OBTAINED BY THE PROPOSED GFI METHOD, LIKES, AND ORACLE

	method	90% CI	95% CI	99% CI
(δ_f, SNR) = (0.1, 5)	GFI	90.2%, 87.9%, 90.1% (2.4, 2.4, 2.4)	95.5%, 93.8%, 94.6% (2.9, 2.8, 2.9)	98.4%, 98.4%, 99.2% (3.9, 3.8, 3.8)
	LIKES	84.9%, 84.9%. 85.8% (6.5, 6.4, 6.8)	90.2%, 92.0%, 89.3% (6.1, 6.1, 6.6)	94.9%, 95.6%, 94.5% (8.6, 8.6, 8.9)
	Oracle	88.3%, 89.4%, 90.2% (2.3, 2.3, 2.3)	94.6%, 95.4%, 95.7% (2.8, 2.8, 2.8)	99.2%, 98.3%, 99.1% (3.6, 3.6, 3.7)
(δ_f, SNR) = (0.1, 10)	GFI	91.0%, 91.5%, 91.7% (1.4 1.4, 1.4)	96.6%, 95.8%, 95.8% (1.7, 1.7, 1.7)	98.4%, 98.6%, 99.3% (2.3, 2.2, 2.3)
	LIKES	83.4%, 84.4%, 84.2% (2.2, 2.2, 2.2)	89.6%, 91.0%, 90.6% (2.6, 2.6, 2.6)	96.4%, 97.0%, 95.6% (3.4, 3.4, 3.5)
	Oracle	87.8%, 89.8%, 90.0% (1.3, 1.3, 1.3)	95.4%, 94.8%, 93.8% (1.6, 1.6, 1.6)	99.2%, 99.0%, 99.0% (2.1, 2.1, 2.1)
(δ_f, SNR) = (0.015, 5)	GFI	89.5%, 88.9%, 89.3% (2.4, 2.7, 2.6)	95.2%, 94.0%, 95.1% (2.8, 3.0, 2.8)	98.6%, 97.6%, 98.2% (3.7, 4.1, 3.9)
	LIKES	85.8%, 84.7%, 81.3% (6.3, 6.4, 6.6)	92.8%, 90.2%, 89.9% (6.6, 6.8, 6.8)	95.1%, 95.8%, 94.7% (8.8, 8.7, 8.9)
	Oracle	90.3%, 90.0%, 89.5% (2.3, 2.4, 2.4)	95.3%, 95.1%, 95.1% (2.8, 2.9, 2.9)	98.9%, 98.8%, 97.8% (3.6, 3.7, 3.8)
(δ_f, SNR) = (0.015, 10)	GFI	91.5%, 91.4%, 91.5% (1.5, 1.7, 1.6)	96.1%, 95.9%, 96.0% (1.8, 1.9, 1.9)	99.2%, 99.5%, 99.2% (2.2, 2.6, 2.4)
	LIKES	84.2%, 84.4%, 85.2% (2.2, 2.2, 2.2)	89.0%, 85.6%, 88.2% (2.6, 2.6, 2.6)	95.6%, 96.4%, 96.4% (3.4, 3.4, 3.5)
	Oracle	91.1%, 89.7%, 90.8% (1.3, 1.4, 1.4)	95.2%, 95.6%, 94.5% (1.8, 1.9, 1.9)	98.8%, 98.8%, 98.8% (2.1, 2.1, 2.1)
(δ_f, SNR) = (0.01, 5)	GFI	90.8%, 89.6%, 87.1% (2.4, 3.0, 3.1)	95.4%, 94.4%, 92.8% (2.9, 3.7, 3.7)	98.3%, 98.7%, 98.0% (3.9, 4.9, 4.8)
	LIKES	86.9%, 85.9%, 85.6% (6.3, 6.3, 6.1)	90.3%, 93.0%, 89.4% (6.9, 6.9, 6.4)	95.0%, 95.9%, 95.7% (8.8, 8.2, 8.4)
	Oracle	90.0%, 91.5%, 88.9% (2.3, 2.9, 2.9)	93.9%, 94.8%, 94.1% (2.8, 3.5, 3.5)	99.4%, 99.1%, 98.7% (3.6, 4.6, 4.6)
(δ_f, SNR) = (0.01, 10)	GFI	91.8%, 91.3%, 89.2% (1.5, 1.9, 1.9)	96.7%, 95.1%, 93.5% (1.7, 2.2, 2.2)	99.2%, 98.9%, 98.1% (2.3, 2.9, 2.9)
	LIKES	85.2%, 85.4%, 86.2% (2.2, 2.2, 2.2)	90.4%, 91.4%, 90.8% (2.6, 2.6, 2.6)	97.0%, 96.2%, 96.6% (3.4, 3.4, 3.5)
	Oracle	89.5%, 88.7%, 91.3% (1.3, 1.6, 1.6)	94.1%, 94.6%, 93.8% (1.6, 2.0, 2.0)	98.8%, 99.2%, 98.7% (1.7, 2.2, 2.2)

The numbers in the parentheses are the average widths of the intervals.

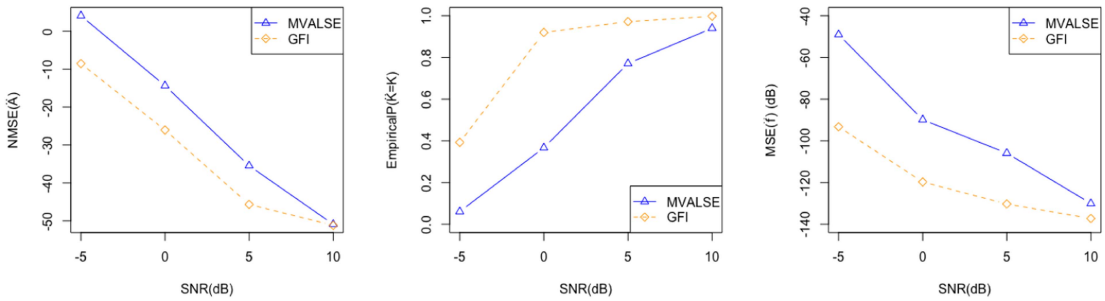


Fig. 1. Empirical performances of the MVALSE method [14] and the proposed GFI method with different SNRs and $N = 75$.

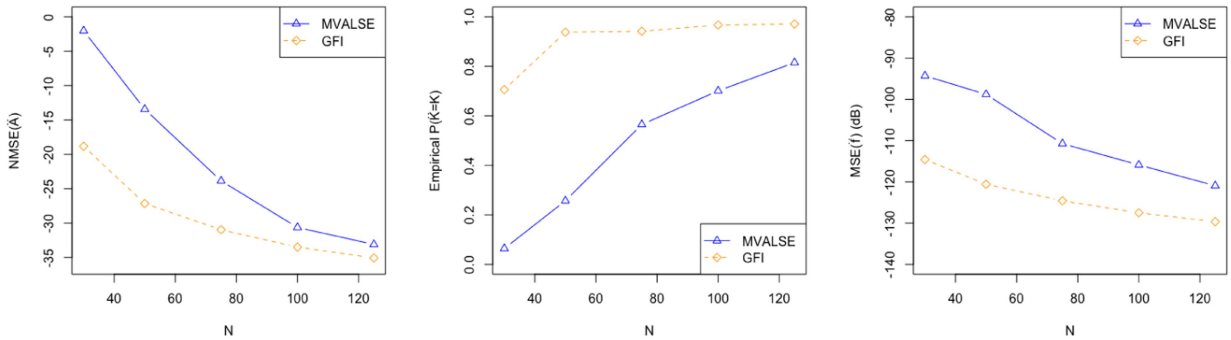


Fig. 2. Empirical performances of the MVALSE method [14] and the proposed GFI method with different N and SNR$= 2$.

TABLE IV
ESTIMATED RESULTS FOR STAR HD 63454 OBTAINED FROM DIFFERENT METHODS

	[32]		GFI	
Exoplanet No	$\hat{f}$	$\hat{\beta}$	$\hat{f}$	$\hat{\beta}$
1	0.3549	0.0634	0.3549	0.0482 (0.0255, 0.0708)

The GFI 95% confidence interval is given in parentheses.

TABLE V
PERCENTAGES THAT DIFFERENT NUMBERS OF EXOPLANETS WERE SELECTED FOR THE THREE STARS

Number of exoplanets detected	1	2	3
HD 63454	99.9%	0.1%	0
HD 208478	93.9%	6.1%	0
GJ 876	0.03%	15.02%	84.95%

stars to wobble in space, which in turn changes the color of the light astronomers observe. This permits an analysis of the Doppler shifts to confirm if there is any exoplanet revolving around a star. In order to do so, the radial velocity frequencies and amplitudes of the stars, need to be estimated. Notice that the radial velocity measurements are often obtained at non-uniformly spaced time intervals due to hardware and practical constraints, which limits the applications of many spectral analysis methods designed for equally-spaced data.

Here we apply the proposed GFI method to estimate the radial velocity frequencies and amplitudes of three different stars: HD 63454 [33], HD 208487 [34], and GJ 876 [35]. We note that the model we use (2) and (4) is simpler than those that are based on Keplerian's planetary motion, which also consider eccentricity and periastron parameters of the orbital planets thus more accurate; e.g., [32]. We also note that our model is complex-valued while the radial velocity data are real-valued. To circumvent this issue, we follow [32] and require both positive and negative frequencies in the model to represent a real-valued component. Below we compare our results with those reported in [32].

B. HD 63454

The radial velocity data set of star HD 63454 contains 26 samples spanning 350 days. The sampling pattern and the radial velocity measurements are shown in Fig. 3(a) and (b), respectively. The proposed GFI method was applied to the data set and the results are shown in Table IV, where f represents its orbital frequency (in cycles day^{-1}) and β represents the corresponding amplitude. As the results suggest, only one exoplanet was detected whose estimated frequency was 0.3549 cycles day^{-1} (i.e., an orbital period of 2.8176 days), which is the same as in [32]. The GFI estimated amplitude is smaller than the one reported by [32] but the corresponding GFI confidence interval does cover it, so overall the GFI results are consistent with those in [32] for HD 63454. Table V shows the percentages that different numbers of exoplanets were selected. One can see that for this star the proposed method is highly confident (99.9%) that there is only one exoplanet.

TABLE VI
SIMILAR TO TABLE IV BUT FOR STAR HD 208487

	[32]		GFI	
Exoplanet No	$\hat{f}$	$\hat{\beta}$	$\hat{f}$	$\hat{\beta}$
1	0.0078	19.9	0.0078	17.5 (13.7, 21.3)
2	0.0690	12.18	-	-
3	0.0408	4.96	-	-

TABLE VII
SIMILAR TO TABLE IV BUT FOR STAR GJ 876

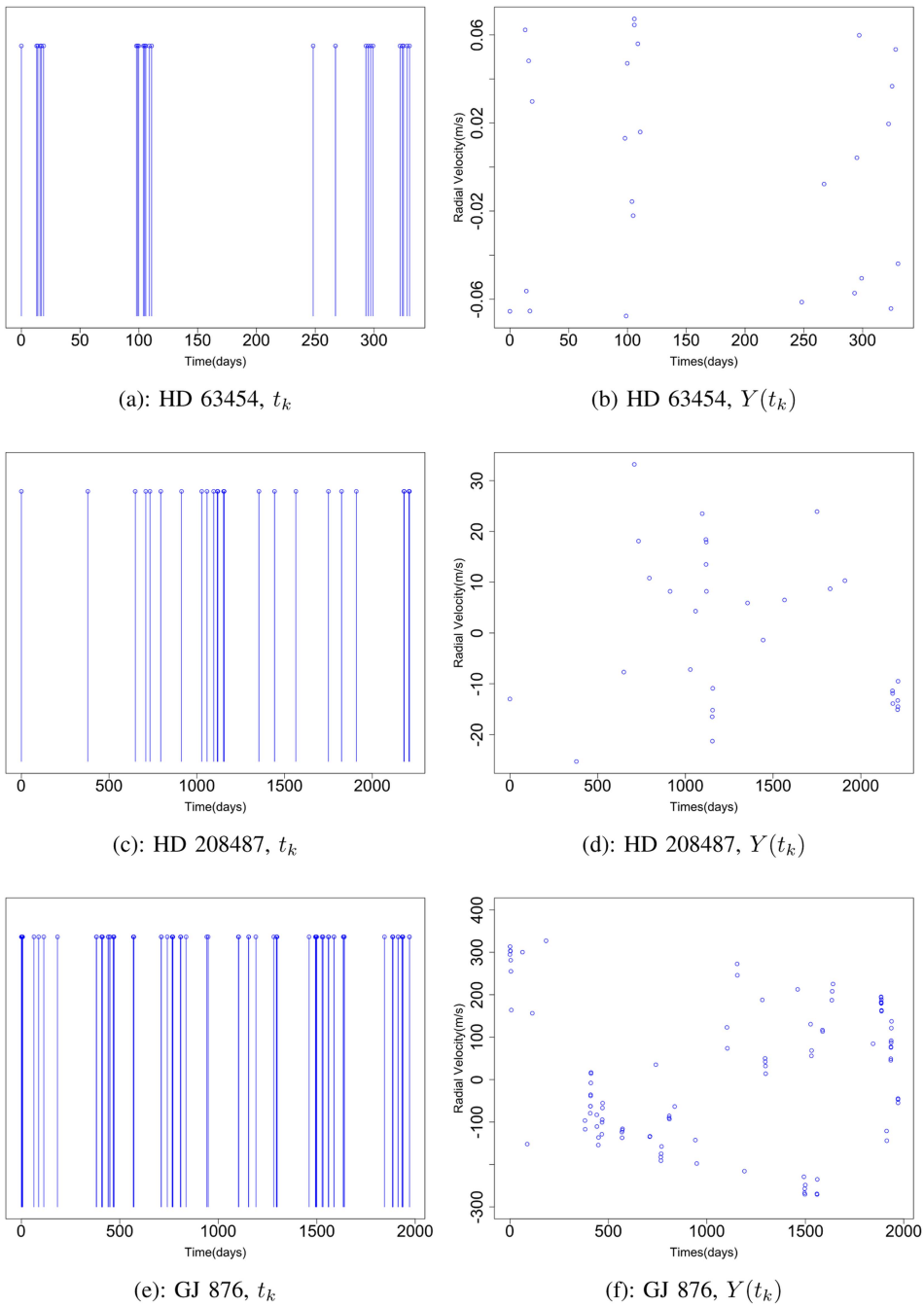
	[32]		[35]		GFI	
Exoplanet No	$\hat{f}$	$\hat{\beta}$	$\hat{f}$	$\hat{\beta}$	$\hat{f}$	$\hat{\beta}$
1	0.0164	215.09	0.0164	212.60	0.0165	201.03 (181.44, 220.61)
2	0.0331	82.62	0.0331	88.36	0.0332	69.74 (46.99, 92.49)
3	0.0011	9.61	0.516	6.40	0.0666	31.86 (10.44, 53.27)
4	0.0066	9.95	-	-	-	-
5	0.0168	23.57	-	-	-	-

C. HD 208487

The data set for star HD 208487 contains 31 samples spanning 2250 days. The sampling pattern and the radial velocity measurements are displayed in, respectively, Fig. 3(c) and (d). The GFI method only detected one exoplanet with an estimated orbital frequency of 0.0078 cycles day^{-1}, while 3 detected exoplanets were reported in [32]; see Table VI. However, as noted in both [32] and [34], there is no convincing evidence to support the claim of the existence of the two additional exoplanets for this star system, so the GFI method provided reasonable results for HD 208487. Table V also provides strong evidence (around 94%) that there is only one exoplanet for this star.

D. GJ 876

The last data set is for star GJ 876. It consists of 100 samples spanning 2000 days; see Fig. 3(e) and (f) for the sampling pattern and the radial velocity measurements, respectively. The results are shown in Table VII. The GFI method detected 3 exoplanets with orbital frequencies 0.0165, 0.0332 and 0.0666 cycles day^{-1}. A previous study by [35] also detected 3 exoplanets, but with a different orbital frequency (0.516 cycles day^{-1}) for the last one. The method of [32] detected 5 exoplanets. However, [32] also suggested that there is no concrete evidence to support the existence of the additional 2 exoplanets. In any case, all these methods agreed on the first 2 exoplanets in this star system. The uncertainty information in Table V also suggests that there are three exoplanets, but with lower confidence (around 85%). This indicates that for this star, the true number



(a): HD 63454, t_k (b) HD 63454, $Y(t_k)$

(c): HD 208487, t_k (d): HD 208487, $Y(t_k)$

(e): GJ 876, t_k (f): GJ 876, $Y(t_k)$

Fig. 3. Sampling times t_k's (left column) and radial velocity measurements $Y(t_k)$'s (right column) for stars HD 63454 (top row), HD 208487 (middle row), and GJ 876 (bottom row).

of exoplanets is more challenging to estimate, as can be seen from the very different results obtained from previous studies.

VII. Concluding Remarks

This paper developed a new method to perform statistical inference on the line spectral estimation problem. The proposed method is based on the approach of generalized fiducial inference. In greater detail, a procedure was developed to generate fiducial samples from a so-called generalized fiducial density for a set of candidate models. This generalized fiducial density plays a similar role as the posterior density in the Bayesian context. Its samples (i.e., fiducial samples) can be used to perform statistical inferences such as forming point estimates and confidence intervals. The proposed method was shown to enjoy desirable asymptotic properties under some regularity conditions. Through numerical experiments, it was also demonstrated that the proposed method possesses promising empirical properties and often outperforms existing methods in the literature. Lastly, the proposed method was applied to analyze three radial velocity

data sets in the context of exoplanet detection and yielded similar results as those reported in the astronomy literature.

Recall that our method is an example of semi-parametric method, which can be further divided into three categories [36]: on-grid, off-grid, and gridless. The on-grid methods require a pre-selected grid and the true frequencies to be one of the grid values. Our method can be classified as on-grid. However, as mentioned in Section I and demonstrated by the simulation experiments in Section V, our method can handle the situation when some of the true frequencies do not fall on the grid, by suitably choosing the values of K and Δ. For off-grid methods, they also require a grid, which is estimated jointly with the sparse signals. Consequently, more variables are needed to be estimated, which increases the dimension of the problem. The last category of gridless methods does not require any grid when compared with the first two categories. However, they are typically designed for equally spaced sampled data, which may restrict their applicability. We believe that GFI can be applied to these methods, but the form of the structural equation (5) will need to be formulated differently. Overall, we are confident that the GFI approach can be applied to these off-grid and gridless methods. The main challenge will be the development of a practical algorithm for generating the fiducial samples. These are left for future work.

Appendix A
Obtaining $\mathcal{M}^*$

This appendix presents our method for obtaining $\mathcal{M}^*$. Recall that an ideal $\mathcal{M}^*$ should only contain those models that have a non-negligible value of $r(M|\boldsymbol{Y})$. Our method consists of two stages. The first stage applies a fast algorithm to traverse the space of $\mathcal{M}$ to obtain a set of non-negligible models, where the true model will be included with high probability. In the second stage, we obtain more models by data perturbation and add these models to $\mathcal{M}^*$. Notice that we are not choosing the models by comparing their values of $r(M|\boldsymbol{Y})$ with a threshold.

Stage 1: In the context of ultra-high dimensional regression, the lasso algorithm [37] has been applied by [26] to obtain a $\mathcal{M}^*$. The idea is that, by changing the lasso tuning parameter, a sequence of models (also known as a solution path) will be generated, and all these models form $\mathcal{M}^*$. We shall follow this idea in the first stage of our method. However, due to the complex-valued coefficients, the lasso algorithm cannot be directly applied, as it does not guarantee to select both the real and imaginary parts of a complex coefficient simultaneously. To circumvent this, one can use for example the complex lasso [38]. Below, however, we shall re-express the problem and apply the group lasso algorithm of [39].

First, express the lasso problem as:

$$\min_{\boldsymbol{\beta}} \left(\frac{1}{2} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1^* \right), \tag{19}$$

where $\boldsymbol{X} = \Re(\boldsymbol{X}) + i\Im(\boldsymbol{X}) \in \mathbb{C}^{N\times K}$, $\boldsymbol{y} = \Re(\boldsymbol{y}) + i\Im(\boldsymbol{y}) \in \mathbb{C}^N$, $\boldsymbol{\beta} = \Re(\boldsymbol{\beta}) + i\Im(\boldsymbol{\beta}) \in \mathbb{C}^K$, and

$$\|\boldsymbol{\beta}\|_1^* = \sum_{j=1}^{n} \sqrt{\Re(\beta))_j^2 + \Im(\beta)_j^2},$$

with $\Re(\beta), \Im(\beta) \in \mathbb{R}$ and $j = 1, \ldots, n$. Minimizing (19) with different values of λ will give different models. However, as suggested before, there is no guarantee that all the resulting models are legitimate in the sense that the corresponding estimates in $\Re(\boldsymbol{\beta})$ and $\Im(\boldsymbol{\beta})$ are both zeros or non-zeros.

Now we can re-express

$$\begin{aligned}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 &= \|\Re(\boldsymbol{X})\Re(\boldsymbol{\beta}) - \Im(\boldsymbol{X})\Im(\boldsymbol{\beta}) - \Re(\boldsymbol{y})\|_2^2 \\ &\quad + \|\Re(\boldsymbol{X})\Im(\boldsymbol{\beta}) + \Im(\boldsymbol{X})\Re(\boldsymbol{\beta}) - \Im(\boldsymbol{y})\|_2^2 \\ &= \left\| \begin{pmatrix} \Re(\boldsymbol{X}) & -\Im(\boldsymbol{X}) \\ \Im(\boldsymbol{X}) & \Re(\boldsymbol{X}) \end{pmatrix} \begin{pmatrix} \Re(\boldsymbol{\beta}) \\ \Im(\boldsymbol{\beta}) \end{pmatrix} - \begin{pmatrix} \Re(\boldsymbol{y}) \\ \Im(\boldsymbol{y}) \end{pmatrix} \right\|_2^2\end{aligned}$$

and (19) becomes

$$\min_{\boldsymbol{\beta}} \left(\frac{1}{2} \|\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{y}}\|_2^2 + \lambda \|\tilde{\boldsymbol{\beta}}\|_{2,1} \right) \tag{20}$$

with $\tilde{\boldsymbol{X}} = \begin{pmatrix} \Re(\boldsymbol{X}) & -\Im(\boldsymbol{X}) \\ \Im(\boldsymbol{X}) & \Re(\boldsymbol{X}) \end{pmatrix}$, $\tilde{\boldsymbol{y}} = \begin{pmatrix} \Re(\boldsymbol{y}) \\ \Im(\boldsymbol{y}) \end{pmatrix}$, $\tilde{\boldsymbol{\beta}} = \begin{pmatrix} \Re(\boldsymbol{\beta}) \\ \Im(\boldsymbol{\beta}) \end{pmatrix}$ and $\|\boldsymbol{\beta}\|_{2,1} = \sum_{j=1}^{n} \sqrt{\Re(\beta)_j^2 + \Im(\beta)_j^2}$. With the above, we can apply the group lasso algorithm to (20) to generate different models with different values of λ. In practice, we observe that the true model was almost always included as one of the models generated by this algorithm.

Stage 2: To achieve theoretical guarantee, in the second stage, we apply the adaptive group lasso algorithm to generate more models, which was shown by [40] that the true model will be selected consistently. We can also obtain more models by applying the group lasso algorithm to various re-sampled data sets [41], so that in practice most non-negligible models are included in $\mathcal{M}^*$. We can yet further enrich $\mathcal{M}^*$ by adding solutions from other methods to $\mathcal{M}^*$, such as SPICE [42] and GIST [43]. By doing so, we expect the size of $\mathcal{M}^*$ to be much smaller than the size of $\mathcal{M}$ (which is 2^K) and yet $\sum_{M\in\mathcal{M}^*} r(M|\boldsymbol{Y})$ is close to 1.

Lastly, we note that before we generate the fiducial samples, the model parameters will be re-fitted using maximum likelihood, and therefore the parameter estimation bias from group lasso will not be carried over.

Appendix B
Proof and Technical Details

This appendix proves Theorem 4.1. When compared to earlier theoretical results in GFI, a major difference is that the current work considers complex-valued coefficients and responses. We begin by presenting three lemmas.

A. Lemmas

Lemma B.1: If $\log j/\log p \to \delta$ as $p \to \infty$, then $\log \binom{p}{j} = j \log p(1-\delta)(1+o(1))$.

Proof: First, calculate $\binom{p}{j} = \frac{p!}{j!(p-j)!} = \frac{p(p-1)\cdots(p-j+1)}{j!} = \frac{p^j(1-\frac{1}{p})(1-\frac{2}{p})\cdots(1-\frac{j-1}{p})}{j!}$ and we have

$$\left(1 - \frac{j-1}{p}\right)^{j-1} < \left(1 - \frac{1}{p}\right)\left(1 - \frac{2}{p}\right)\cdots$$

$$\left(1-\frac{j-1}{p}\right) < \left(1-\frac{1}{p}\right)^{j-1}.$$

By sterling's formula,

$$\sqrt{2\pi}j^{j+1/2}e^{\frac{-j+1}{12j+1}} < j! < \sqrt{2\pi}j^{j+1/2}e^{\frac{-j+1}{12j}}$$

so we have

$$\begin{aligned}
\log\tbinom{p}{j} &\le j\log p + (j-1)\log\left(1-\frac{1}{p}\right) - \log j! \\
&\le j\log p + (j-1)\log\left(1-\frac{1}{p}\right) \\
&\quad - \left(j+\frac{1}{2}\right)\log j + j - \frac{1}{12j+1} - \log\sqrt{2\pi} \\
&\le j\log p - \left(j+\frac{1}{2}\right)\log j + j \\
&= j\log p\left[1 - \frac{\left(j+\frac{1}{2}\right)\log j}{j\log p} + \frac{1}{\log p}\right] \\
&= j\log p(1-\delta)(1+o(1))
\end{aligned}$$

and

$$\begin{aligned}
\log\tbinom{p}{j} &\ge j\log p + (j-1)\log\left(1-\frac{j-1}{p}\right) \\
&\quad - \left(j+\frac{1}{2}\right)\log j + j - \frac{1}{12j} - \log\sqrt{2\pi} \\
&\ge j\log p + (j-1)\log\left(1-\frac{j-1}{p}\right) \\
&\quad - \left(j+\frac{1}{2}\right)\log j - \log\sqrt{2\pi} \\
&= j\log p\left(1 + \frac{(j-1)\log\left(1-\frac{j-1}{p}\right)}{j\log p}\right) \\
&\quad - j\log p\left(\frac{\left(j+\frac{1}{2}\right)\log j}{j\log p} - \frac{\log\sqrt{2\pi}}{j\log p}\right) \\
&= j\log p(1-\delta)(1+o(1)),
\end{aligned}$$

which completes the proof. ■

Lemma B.2: Let χ_j^2 be a chi-square random variable with j degrees of freedom. If $c\to\infty$ and $\frac{J}{c}\to 0$, then

$$P\left(\chi_j^2 > c\right) = \frac{1}{\Gamma\left(\frac{j}{2}\right)}\left(\frac{c}{2}\right)^{k/2-1}e^{-c/2}(1+o(1))$$

uniformly over $j\le J$.

Proof: The pdf of χ_j^2 is $f(x) = \frac{1}{2^{\frac{j}{2}}\Gamma(\frac{j}{2})}x^{\frac{j}{2}-1}e^{-\frac{x}{2}}$, so

$$\begin{aligned}
P\left(x_j^2 > c\right) &= \int_c^\infty \frac{\left(\frac{1}{2}\right)^{\frac{j}{2}}}{\Gamma\left(\frac{j}{2}\right)}x^{\frac{j}{2}-1}e^{-\frac{x}{2}}dx \\
&= \frac{\left(\frac{1}{2}\right)^{\frac{j}{2}}}{\Gamma\left(\frac{j}{2}\right)}\int_c^{+\infty} x^{\frac{j}{2}-1}e^{-\frac{x}{2}}dx.
\end{aligned}$$

Now calculate

$$\begin{aligned}
\int_c^\infty x^{\frac{j}{2}-1}e^{-\frac{x}{2}}dx &= \int_c^\infty x^{\frac{j}{2}-1}(-2)de^{-\frac{x}{2}} \\
&= (-2)x^{\frac{j}{2}} - e^{-\frac{x}{2}}\big|_c^\infty \\
&\quad - \int_c^\infty e^{-\frac{x}{2}}(-2)\left(\frac{j}{2}-1\right)x^{\frac{j}{2}-2}dx \\
&= (j-2)\int_c^{+\infty} x^{\frac{j}{2}-2}e^{-\frac{x}{2}}dx + 2c^{\frac{j}{2}-1}e^{-\frac{c}{2}}.
\end{aligned}$$

Therefore

$$\begin{aligned}
F_j(c) = P\left(X_j^2 > c\right) &= \frac{\left(\frac{1}{2}\right)^{\frac{j}{2}-1}}{\Gamma\left(\frac{j}{2}\right)}c^{\frac{j}{2}-1}e^{-\frac{c}{2}} \\
&\quad + \frac{\left(\frac{1}{2}\right)^{\frac{j}{2}}}{\Gamma\left(\frac{j}{2}\right)}(j-2)\int_c^\infty x^{\frac{j}{2}-2}e^{-\frac{x}{2}}dx \\
&= \frac{\left(\frac{1}{2}\right)^{\frac{j}{2}-1}}{\Gamma\left(\frac{j}{2}\right)}c^{\frac{j}{2}-1}e^{-\frac{c}{2}} + F_{j-2}(c).
\end{aligned}$$

So if j is even,

$$\begin{aligned}
F_j(c) &= \frac{1}{\Gamma\left(\frac{j}{2}\right)}\left(\frac{c}{2}\right)^{\frac{j}{2}-1} \\
&\quad \cdot e^{-\frac{c}{2}}\left[1 + \sum_{i=1}^{\frac{j-2}{2}}\left(\frac{\left(\frac{j}{2}-1\right)\cdots\left(\frac{j}{2}-i\right)}{(c/2)^i}\right)\right]
\end{aligned}$$

and if j is odd,

$$\begin{aligned}
F_j(c) &= \frac{1}{\Gamma\left(\frac{j}{2}\right)}\left(\frac{c}{2}\right)^{\frac{j}{2}-1} \\
&\quad \cdot e^{-\frac{c}{2}}\left[1 + \sum_{i=1}^{\frac{j-3}{2}}\left(\frac{\left(\frac{j}{2}-1\right)\cdots-\left(\frac{j}{2}-i\right)}{\left(\frac{c}{2}\right)^i}\right)\right], \\
&\quad + F_1(m),
\end{aligned}$$

where

$$\begin{aligned}
F_1(c) = P(\chi_1^2 \ge c) &\approx 2\frac{\exp\left(-\frac{c}{2}\right)}{\sqrt{2\pi c}} \\
&= \frac{1}{\Gamma\left(\frac{j}{2}\right)}\left(\frac{c}{2}\right)^{\frac{j}{2}-1}e^{-\frac{c}{2}}\frac{2T\left(\frac{j}{2}\right)}{\sqrt{2\pi}\left(\frac{c}{2}\right)^{\frac{j-1}{2}}}.
\end{aligned}$$

Now when $c\to\infty$, we have

$$F_j(c) = \frac{1}{\Gamma\left(\frac{j}{2}\right)}\left(\frac{c}{2}\right)^{\frac{j}{2}-1}e^{-\frac{c}{2}}[1+R(j,c)].$$

Finally, it is straightforward to see that $R(j,c)\le R(J,c)\to 0$ as $c\to\infty$, which completes the proof. ■

Lemma B.3: Let χ_j^2 be a chi-square random variable with j degree of freedom. Let $c_j = 2j\log p + \log(j\log p)$. If $p\to\infty$,

then for any $J \leq p$,

$$\sum_{j=1}^{J} \binom{p}{j} P(\chi_j^2 > c_j) \to 0.$$

Proof: Here we can directly apply Lemma B.2. Let $q_j = \sqrt{\frac{c_j}{(j \log p)^2}}$, by using $\binom{p}{j} \leq p^j$, we have

$$\begin{aligned}
\binom{p}{j} P\left(x_j^2 > c_j\right) &= \binom{p}{j} \frac{\left(\frac{c_j}{2}\right)^{\frac{j}{2}-1} e^{-\frac{c_j}{2}}}{\Gamma\left(\frac{j}{2}\right)} (1 + O(1)) \\
&\leq (c_j)^{\frac{j}{2}-1} \frac{p^j e^{-\frac{1}{2} \cdot 2j(\log p + \log(j \log p))}}{2^{\frac{j}{2}-1} \Gamma\left(\frac{j}{2}\right)} (1 + o(1)) \\
&= (c_j)^{\frac{j}{2}-1} (1 + o(1)) \cdot \frac{p^j e^{-j \log p - j \log(j \log p)}}{2^{\frac{j}{2}} - 1 \Gamma\left(\frac{j}{2}\right)} \\
&= (c_j)^{\frac{j}{2}-1} (1 + o(1)) \cdot \frac{(j \log p)^{-j}}{2^{\frac{j}{2}-1} \Gamma(\frac{j}{2})}.
\end{aligned}$$

Let

$$q_j = \sqrt{\frac{c_j}{(j \log p)^2}} \leq \frac{(c_j)^{\frac{j}{2}-1}}{(j \log p)^j} (1 + o(1)) \leq \frac{q_j^j}{c_j} (1 + o(1))$$

and therefore

$$\sum_{j=1}^{J} \binom{p}{j} P(\chi_j^2 > c_j) \leq \sum_{j=1}^{J} \frac{q_j^j}{c_j} (1 + o(1)) \overset{q_j \to 0}{\longrightarrow} 0,$$

which completes the proof. ■

B. Proof of Theorem 4.1

Denote $\mathcal{M}$ as the collection of models for which (16) holds. We shall prove that $\max_{\mathcal{M}} \frac{r(M)}{r(M_0)} \xrightarrow{P} 0$. Without loss of generality, we assume $\sigma^2 = 1$. We write $|M_0| = m_0, |M| = m$, where $m_0 = o(N)$ and $m = o(N)$. For simplicity, we can rewrite

$$\frac{r(M)}{r(M_0)} = \exp\{-T_1 - T_2 - T_3\},$$

where

$$T_1 = \left(N - m - \frac{1}{2}\right) \log \frac{\mathrm{RSS}_M}{\mathrm{RSS}_{M_0}},$$

$$\begin{aligned}
T_2 = &\frac{m - m_0}{2} \log n + (m - m_0) \log \pi \mathrm{RSS}_{M_0} \\
&+ \log \frac{\Gamma(N - m_0)}{\Gamma(N - m)} + \gamma \log \binom{K}{m} - \gamma \log \binom{K}{m_0}
\end{aligned}$$

and

$$T_3 = -\frac{1}{2} \log \frac{[\det(A_{M_0}^H A_{M_0})]}{[\det(A_M^H A_M)]}.$$

Next we consider the following two cases:
Case 1: $M_0 \not\subset M$.

Let $\mathcal{M}_{|} = \{M : |M| = j, M \in \mathcal{M}\}$. Notice that $\mathrm{RSS}_{M_0} = (N - m_0)(1 + o_p(1)) = N(1 + o_p(1))$,

$$\begin{aligned}
\mathrm{RSS}_M - \mathrm{RSS}_{M_0} = &\Delta_M + 2\boldsymbol{\mu}^H (\boldsymbol{I} - \boldsymbol{P}_M)\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^H \boldsymbol{P}_M \boldsymbol{\epsilon} \\
&+ \boldsymbol{\epsilon}^H (\boldsymbol{I} - \boldsymbol{P}_{M_0})\boldsymbol{\epsilon},
\end{aligned} \tag{21}$$

where $\boldsymbol{\mu} = \boldsymbol{A}_{M_0} \boldsymbol{\beta}_{M_0}$, $\Delta_M = ||(\boldsymbol{I} - \boldsymbol{P}_M)\boldsymbol{\mu}||^2$ and $\boldsymbol{\epsilon}^H \boldsymbol{P}_{M_0} \boldsymbol{\epsilon} = m_0(1 + o_p(1))$.

First consider the second term in (21) and denote $\boldsymbol{Z}_M = \boldsymbol{\mu}^H (\boldsymbol{I} - \boldsymbol{P}_M)\boldsymbol{\epsilon} / \sqrt{\Delta_M}$, we have

$$\boldsymbol{\mu}^H (\boldsymbol{I} - \boldsymbol{P}_M) \boldsymbol{\epsilon} = \sqrt{\Delta_M} \boldsymbol{Z}_M,$$

where $\boldsymbol{Z}_M \sim \mathcal{CN}(0, 1)$. Let $c_j = j\{\log K + \log j \log K\}$. For simplicity, we denote $c_{|M|}$ by c_m. Then by Lemma B.3

$$\begin{aligned}
P\left(\max_{\mathcal{M}} |\boldsymbol{Z}_M / \sqrt{c_m}| > 1\right) &\leq \sum_{j=1}^{bm_0} \sum_{\mathcal{M}_j} P(\boldsymbol{Z}_M^2 > c_j) \\
&= \sum_{j=1}^{bm_0} \binom{K}{j} P\left(\frac{\chi_2^2}{2} > c_j\right) \\
&\leq \sum_{j=1}^{bm_0} \binom{K}{j} P(\chi_j^2 > 2c_j) \\
&\to 0.
\end{aligned}$$

Therefore,

$$|\boldsymbol{\mu}^H (\boldsymbol{I} - \boldsymbol{P}_M)\boldsymbol{\epsilon}| \leq \sqrt{\Delta_M} |\boldsymbol{Z}_M| \leq \sqrt{\Delta_M} \sqrt{c_M} (1 + o_p(1))$$

uniformly over $\mathcal{M}$. Since $c_m = o(m_0 \log K)$ and the identifiability condition (16) states $m_0 \log(K) = o_p(\Delta_M)$ uniformly over $\mathcal{M}$ s.t. $M_0 \not\subset M$,

$$|\boldsymbol{\mu}^H (\boldsymbol{I} - \boldsymbol{P}_M)\boldsymbol{\epsilon}| = o_p(\Delta_M).$$

Next, we consider the third term in (21). By Lemma B.3 again, we have

$$\begin{aligned}
P\left(\max_{\mathcal{M}} \boldsymbol{\epsilon}^H \boldsymbol{P}_M \boldsymbol{\epsilon} / c_m > 1\right) &\leq \sum_{j=1}^{km_0} \sum_{\mathcal{M}_j} P\left(\boldsymbol{\epsilon}^H \boldsymbol{P}_M \boldsymbol{\epsilon} > c_j\right) \\
&\leq \sum_{j=1}^{km_0} \binom{K}{j} P\left(\chi_j^2 > 2c_j\right) \to 0.
\end{aligned}$$

So $\boldsymbol{\epsilon}^H \boldsymbol{P}_M \boldsymbol{\epsilon} \leq c_m(1 + o_p(1))$, and $\boldsymbol{\epsilon}^H \boldsymbol{P}_M \boldsymbol{\epsilon} = o_p(\Delta_M)$ uniformly over $\mathcal{M}$ s.t. $M_0 \not\subset M$. Therefore

$$\mathrm{RSS}_M - \mathrm{RSS}_{M_0} = \Delta(M)(1 + o_p(1))$$

and we have

$$\begin{aligned}
\log\left(\frac{\mathrm{RSS}_M}{\mathrm{RSS}_{M_0}}\right) &= \log\left(1 + \frac{\mathrm{RSS}_M - \mathrm{RSS}_{M_0}}{\mathrm{RSS}_{M_0}}\right) \\
&= \log\left(1 + \frac{\Delta(M)}{N}(1 + o_p(1))\right)
\end{aligned}$$

uniformly over all $M \in \mathcal{M}$ s.t. $M_0 \not\subset M$. Thus

$$T_1 = \left(N - m - \frac{1}{2}\right) \log\left(\frac{\mathrm{RSS}_M}{\mathrm{RSS}_{M_0}}\right)$$

$$= \left(N-m-\frac{1}{2}\right)\log\left(1+\frac{\Delta(M)}{N}(1+o_p(1))\right)$$

$$= \frac{2N\,(o_p(1)+1)}{2}\cdot\frac{\Delta(M)}{N}\,(1+o_p(1)) = \Delta(M)(1+o_p(1))$$

uniformly for $M \in \mathcal{M}$ such that $M_0 \not\subset M$.

Also,

$$(m-m_0)\log(\pi \mathrm{RSS}_{M_0}) + \log\frac{\Gamma(N-m_0)}{\Gamma(N-m)}$$

$$= (m-m_0)\log N(1+o_p(1)) + (m-m_0)\log N(1+o_p(1))$$

$$= 2(m-m_0)\log N(1+o_p(1)).$$

Finally, we have

$$T_2 = \frac{5}{2}(m-m_0)\log N(1+o_p(1)) - \gamma\log\binom{p}{m_0} + \gamma\log\binom{K}{m}$$

$$\geq \frac{-5}{2}m_0\log N(1+o_p(1)) - \gamma m_0\log K.$$

Case 2: Let $\mathcal{M}^* = \{M \in \mathcal{M}, M_0 \subset M, M \neq M_0\}$ and $\mathcal{M}_j^* = \{M, |M| = j, M_0 \subset M\}$. First notice that when $M_0 \subset M$, we have $\mathrm{RSS}_{M_0} - \mathrm{RSS}_M = \frac{1}{2}\chi^2_{2(m-m_0)}(M)$, where $\chi^2_{2(m-m_0)}(M)$ is a chi-square distribution with $2(m-m_0)$ degrees of freedom depending on M. Write $c_j = j\{\log K + \log(j\log K)\}$. Then by Lemma B.3 we have

$$P\left(\max_{1\leq j\leq bm_0-m_0}\max_{\mathcal{M}\in\mathcal{M}_j^*}\chi_j^2(M)/2c_j \geq 1\right)$$

$$\leq \sum_{j=1}^{bm_0-m_0} P\left(\max_{\mathcal{M}\in\mathcal{M}_j^*} \geq 2c_j\right)$$

$$= \sum_{j=1}^{bm_0-m_0}\binom{K-m_0}{j}P\left(\chi_j^2(M)\geq 2c_j\right)$$

$$\leq \sum_{j=1}^{bm_0-m_0}\binom{K}{j}P\left(\chi_j^2 \geq 2c_j\right) \to 0,$$

which implies

$$\chi^2_{2(m-m_0)}(M) \leq 2c_{2(m-m_0)}\,(1+o_p(1)).$$

Notice that $2c_{2(m-m_0)} = o(N)$, and therefore

$$\left(N-m-\frac{1}{2}\right)\log\left(\frac{\mathrm{RSS}_m}{\mathrm{RSS}_{M_0}}\right)$$

$$= -\left(N-m-\frac{1}{2}\right)\log\left(1+\frac{\frac{1}{2}\chi^2_{2(m-m_0)}(M)}{\mathrm{RSS}_{M_0}-\frac{1}{2}\chi_{2(m-m_0)^2}(M)}\right)$$

$$\geq \left(N-m-\frac{1}{2}\right)\log\left(\frac{\chi_{2(m-m_0)^2}(M)}{2\mathrm{RSS}_{M_0}-\chi^2_{2(m-m_0)}(M)}\right)$$

$$\geq -\frac{c_{2(m-m_0)}}{2}(1+o_p(1)) \geq -2(m-m_0) \times\left[1+\frac{\log\{(bm_0-m_0)\log K\}}{\log K}\right]\log K\,(1+o_p(1))$$

$$\geq -2(m-m_0)(1+\delta)\log K\,(1+o_p(1))$$

uniformly over $\mathcal{M}^*$. Consequently, we have shown that

$$T_1 \geq -2(m-m_0)(1+\delta)\log K\,(1+o_p(1))$$

uniformly over $\mathcal{M}^*$. By Lemma B.1, for $m_0 < m < bm_0$, we have

$$\log\binom{K}{m} = (1-\delta)m\log K(1+o(1))$$

uniformly over $\mathcal{M}^*$. So

$$T_2 = \frac{5}{2}(m-m_0)\log N(1-o_p(1)) + \gamma(1-\delta)(m-m_0)\log K(1+o(1))$$

uniformly over $\mathcal{M}^*$.

Finally, we have

$$\max_{M\neq M_0, M\in\mathcal{M}}\frac{r(M)}{r(M_0)} = \max\left\{\max_{M_0\not\subset M}\exp(-T_1-T_2-T_3), \max_{M_0\subset M}\exp(-T_1-T_2-T_3)\right\}.$$

As $T_3 = -\frac{1}{2}\log\frac{[\det(A_{M_0}^H A_{M_0})]}{[\det(A_M^H A_M)]}$ and under the identifiability condition (16), where we only consider $|M| \leq b|M_0|$, we have $T_3 > -\infty$. Together with the above analysis, we have

$$\max_{M_0\not\subset M}\exp(-T_1-T_2-T_3) \xrightarrow{P} 0, \tag{22}$$

since $\min_{M_0\not\subset M}\{T_1+T_2+T_3\} \to \infty$. Also,

$$\max_{M_0\subset M}\exp(-T_1-T_2-T_3) \to 0, \tag{23}$$

as $\min_{M_0\subset M} T_1+T_2+T_3 \to \infty$ if $\gamma > \frac{1+\delta}{1-\delta} - \frac{5\eta}{2(1-\delta)}$, which is guaranteed by the assumption.

So (22) and (23) together show that

$$\max_{M\neq M_0, M\in\mathcal{M}^*}\frac{r(M)}{r(M_0)} \xrightarrow{P} 0.$$

□

Moreover, if condition (18) holds, we have

$$\sum_{M\neq M_0, M\in\mathcal{M}^*}\frac{r(M)}{r(M_0)} \leq \sum_{j=1}^{km_0}\sum_{\mathcal{M}^*}\frac{r(M)}{r(M_0)} \leq km_0\max_{M\neq M_0, M\in\mathcal{M}^*}|M_j^*|\frac{r(M)}{r(M_0)} \xrightarrow{P} 0$$

which shows that $r(M_0) \xrightarrow{P} 0$ over the class $\mathcal{M}^*$.

ACKNOWLEDGMENT

They are most grateful to the reviewers and the associate editor for their most useful and constructive comments.

REFERENCES

[1] A. S. Spanias, "Speech coding: A tutorial review," *Proc. IEEE*, vol. 82, no. 10, pp. 1541–1582, Oct. 1994.

[2] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[3] R. Carriere and R. L. Moses, "High resolution radar target modeling using a modified Prony estimator," *IEEE Trans. Antennas Propag.*, vol. 40, no. 1, pp. 13–18, Jan. 1992.
[4] L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, "Imaging and time reversal in random media," *Inverse Problems*, vol. 18, no. 5, 2002, Art. no. 1247.
[5] P. Stoica and P. Babu, "Sparse estimation of spectral lines: Grid selection problems and their solutions," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 962–967, Feb. 2012.
[6] M. S. Bartlett, "On the theoretical specification and sampling properties of autocorrelated time-series," *Suppl. J. Roy. Stat. Soc.*, vol. 8, pp. 27–41, 1946.
[7] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.
[8] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.
[9] J. Cadzow, "High performance spectral estimation–a new ARMA method," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 5, pp. 524–529, Oct. 1980.
[10] M. M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $l_{2,0}$ norm approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4646–4655, Sep. 2010.
[11] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.
[12] A. Maleki and D. L. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 330–341, Apr. 2010.
[13] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.
[14] J. Zhu, Q. Zhang, P. Gerstoft, M.-A. Badiu, and Z. Xu, "Grid-less variational bayesian line spectral estimation with multiple measurement vectors," *Signal Process.*, vol. 161, pp. 155–164, 2019.
[15] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[16] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[17] J. Hannig, H. Iyer, R. C. Lai, and T. C. M. Lee, "Generalized fiducial inference: A review and new results," *J. Amer. Stat. Assoc.*, vol. 111, pp. 1346–1361, 2016.
[18] R. A. Fisher, "Inverse probability," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 26. Cambridge, U.K.: Cambridge Univ. Press, 1930, pp. 528–535.
[19] A. P. Dempster, "The Dempster–Shafer calculus for statisticians," *Int. J. Approx. Reasoning*, vol. 48, no. 2, pp. 365–377, 2008.
[20] P. T. Edlefsen et al., "Estimating limits from Poisson counting data using Dempster–Shafer analysis," *Ann. Appl. Statist.*, vol. 3, no. 2, pp. 764–790, 2009.
[21] J. Zhang and C. Liu, "Dempster-Shafer inference with weak beliefs," *Statistica Sinica*, vol. 21, pp. 475–494, 2011.
[22] K.-W. Tsui and S. Weerahandi, "Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters," *J. Amer. Stat. Assoc.*, vol. 84, no. 406, pp. 602–607, 1989.
[23] S. Weerahandi, "Generalized confidence intervals," in *Exact Statistical Methods for Data Analysis*. New York, NY: Springer, 1995, pp. 143–168.
[24] A. K. L. Chiang, "A simple general method for constructing confidence intervals for functions of variance components," *Technometrics*, vol. 43, no. 3, pp. 356–367, 2001.
[25] J. Hannig and T. C. M. Lee, "Generalized fiducial inference for wavelet regression," *Biometrika*, vol. 96, pp. 847–860, 2009.
[26] R. C. Lai, J. Hannig, and T. C. M. Lee, "Generalized fiducial inference for ultrahigh-dimensional regression," *J. Amer. Stat. Assoc.*, vol. 110, pp. 760–772, 2015.
[27] Q. Gao, R. C. Lai, T. C. M. Lee, and Y. Li, "Uncertainty quantification for high-dimensional sparse nonparametric additive models," *Technometrics*, vol. 62, no. 4, pp. 513–524, 2020.
[28] J. Hannig, "On generalized fiducial inference," *Statistica Sinica*, vol. 19, pp. 491–544, 2009.
[29] J. Rissanen, *Information and Complexity in Statistical Modeling*. Berlin/Heidelberg, Germany: Springer, 2007.
[30] Z. Wei, R. K. W. Wong, and T. C. M. Lee, "Extending the use of MDL for high-dimensional problems: Variable selection, robust fitting, and additive modeling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2022, pp. 5707–5711.
[31] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Process.*, vol. 92, no. 7, pp. 1580–1590, 2012.
[32] P. Babu, P. Stoica, J. Li, Z. Chen, and J. Ge, "Analysis of radial velocity data by a novel adaptive approach," *Astronomical J.*, vol. 139, 2010, Art. no. 783.
[33] C. Moutou et al., "The harps search for southern extra-solar planets-iv. three close-in planets around HD 2638, HD 27894 and HD 63454," *Astron. Astrophys.*, vol. 439, pp. 367–373, 2005.
[34] P. C. Gregory, "A Bayesian Kepler periodogram detects a second planet in HD 208487," *Monthly Notices Roy. Astronomical Soc.*, vol. 374, pp. 1321–1333, 2007.
[35] E. J. Rivera et al., "A 7.5 $M_{\oplus}$ planet orbiting the nearby star, GJ 876," *Astrophysical J.*, vol. 634, 2005, Art. no. 625.
[36] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing*, Volume 7. Cambridge, MA, USA: Academic Press Elsevier, 2018, pp. 509–581.
[37] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc.: Ser. B. (Methodological)*, vol. 58, pp. 267–288, 1996.
[38] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4290–4308, Jul. 2013.
[39] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc.: Ser. B (Statistical Methodol.)*, vol. 68, pp. 49–67, 2006.
[40] H. Liu and J. Zhang, "Estimation consistency of the group lasso and its applications," in *Proc. Artif. Intell. Statist.*, 2009, pp. 376–383.
[41] N. Meinshausen and P. Bühlmann, "Stability selection," *J. Roy. Stat. Society: Ser. B. (Stat. Methodol.)*, vol. 72, pp. 417–473, 2010.
[42] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 35–47, Jan. 2011.
[43] Y. She, J. Wang, H. Li, and D. Wu, "Group iterative spectrum thresholding for super-resolution sparse spectral selection," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6371–6386, Dec. 2013.

Yi Han received the B.A. degree in statistics from Wuhan University, Wuhan, China, in 2019. He is currently working toward the Ph.D. degree in statistics from the University of California, Davis, CA, USA. His research interests include change point detection, network analysis, and statistical applications in other disciplines.

Thomas C. M. Lee (Senior Member, IEEE) received the B.App.Sc. degree in math and the B.Sc. (Hons.) degree in math with University Medal from the University of Technology, Sydney, NSW, Australia, in 1992 and 1993, respectively, and the Ph.D. degree from Macquarie University, Sydney, and CSIRO Mathematical and Information Sciences, Sydney, in 1997. He is currently a Professor of statistics and the Associate Dean of the Faculty for Mathematical and Physical Sciences with the University of California, Davis, Davis, CA, USA. His research interests include inference methods, image and signal processing, and statistical applications in other scientific disciplines. He is an elected Fellow of the American Association for the Advancement of Science, the American Statistical Association, and the Institute of Mathematical Statistics. From 2013 to 2015, he was the Editor-in-Chief for the *Journal of Computational and Graphical Statistics*, and from 2015 to 2018, he was the Chair of the Department of Statistics at UC Davis.