

# Spectral Analysis of Nonuniformly Sampled Data: A New Approach Versus the Periodogram

Petre Stoica, *Fellow, IEEE*, Jian Li, *Fellow, IEEE*, and Hao He, *Student Member, IEEE*

**Abstract**—We begin by revisiting the periodogram to explain why arguably the plain least-squares periodogram (LSP) is preferable to the “classical” Fourier periodogram, from a data-fitting viewpoint, as well as to the frequently-used form of LSP due to Lomb and Scargle, from a computational standpoint. Then we go on to introduce a new enhanced method for spectral analysis of nonuniformly sampled data sequences. The new method can be interpreted as an iteratively weighted LSP that makes use of a data-dependent weighting matrix built from the most recent spectral estimate. Because this method is derived for the case of real-valued data (which is typically more complicated to deal with in spectral analysis than the complex-valued data case), it is iterative and it makes use of an adaptive (i.e., data-dependent) weighting, we refer to it as the real-valued iterative adaptive approach (RIAA). LSP and RIAA are nonparametric methods that can be used for the spectral analysis of general data sequences with both continuous and discrete spectra. However, they are most suitable for data sequences with discrete spectra (i.e., sinusoidal data), which is the case we emphasize in this paper.

**Index Terms**—BIC, iterative adaptive approach, least-squares method, nonuniformly sampled data, periodogram, spectral analysis.

## I. INTRODUCTORY DISCUSSION AND THE PERIODOGRAM

### A. The Data

LET  $\{y(t_n)\}_{n=1}^N$  denote the sequence of observations whose spectral analysis is our main goal. We assume that the observation times  $\{t_n\}_{n=1}^N$  are given, that  $y(t_n) \in \mathbb{R}$  ( $n = 1, \dots, N$ ), and that a possible nonzero mean has been removed from  $\{y(t_n)\}_{n=1}^N$ , so that  $\sum_{n=1}^N y(t_n) = 0$ . We will also assume throughout this paper that the data sequence consists of a finite number of sinusoidal components and of noise, which is a case of interest in many applications. Note that, while this assumption is not strictly necessary for the nonparametric spectral analysis methods discussed in this paper, these methods perform most satisfactorily when it is satisfied.

Manuscript received April 23, 2008; revised October 10, 2008. First published November 07, 2008; current version published February 13, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Petr Tichavsky. This work was supported in part by the Swedish Research Council (VR) and the National Science Foundation under Grant CCF-0634786 and ECCS-0729727.

P. Stoica is with the Department of Information Technology, Uppsala University, SE-75105 Uppsala, Sweden (e-mail: ps@it.uu.se).

J. Li and H. He are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130 USA (e-mail: li@dsp.ufl.edu; haohe@ufl.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2008.2008973

### B. The Fourier Periodogram

The “classical” Fourier transform-based periodogram (FP) associated with  $\{y(t_n)\}_{n=1}^N$  is given by

$$P_F(\omega) = \frac{1}{N^2} \left| \sum_{n=1}^N y(t_n) e^{-j\omega t_n} \right|^2 \quad (1)$$

where  $\omega$  is the frequency variable and where, depending on the application, the normalization factor might be different from  $1/N^2$  (such as  $1/N$ , see, e.g., [1] and [2]). It can be readily verified that  $P_F(\omega)$  can be obtained from the solution to the following least-squares (LS) data fitting problem:

$$P_F(\omega) = |\hat{\beta}(\omega)|^2, \quad \hat{\beta}(\omega) = \arg \min_{\beta(\omega)} \sum_{n=1}^N |y(t_n) - \beta(\omega) e^{j\omega t_n}|^2. \quad (2)$$

Because  $\{y(t_n)\}_{n=1}^N \in \mathbb{R}$ , the LS criterion above can be rewritten as (below  $\beta(\omega) \triangleq |\beta(\omega)| e^{j\phi(\omega)}$ )

$$\sum_{n=1}^N [y(t_n) - |\beta(\omega)| \cos(\omega t_n + \phi(\omega))]^2 + |\beta(\omega)|^2 \sum_{n=1}^N \sin^2(\omega t_n + \phi(\omega)). \quad (3)$$

Minimization of the first term in (3) makes sense, given the sinusoidal data assumption made previously. However, the same cannot be said about the second term in (3), which has no data fitting interpretation and hence only acts as an additive data-independent perturbation on the first term.

### C. The LS Periodogram

It follows from the discussion in the previous subsection that in the case of real-valued (sinusoidal) data, considered in this paper, the use of FP is not completely suitable, and that a more satisfactory spectral estimate should be obtained by solving the following LS fitting problem:

$$\min_{\substack{\alpha \geq 0 \\ \phi \in [0, 2\pi]}} \sum_{n=1}^N [y(t_n) - \alpha \cos(\omega t_n + \phi)]^2 \quad (4)$$

(we omit the dependence of  $\alpha$  and  $\phi$  on  $\omega$ , for notational simplicity). Using

$$a = \alpha \cos \phi \text{ and } b = -\alpha \sin \phi \quad (5)$$

we can reparameterize the LS criterion in terms of  $a$  and  $b$ :

$$\min_{a,b} \sum_{n=1}^N [y(t_n) - a \cos(\omega t_n) - b \sin(\omega t_n)]^2. \quad (6)$$

The solution to the minimization problem in (6) is well known to be

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \mathbf{R}^{-1} \mathbf{r} \quad (7)$$

where

$$\mathbf{R} = \sum_{n=1}^N \begin{bmatrix} \cos(\omega t_n) \\ \sin(\omega t_n) \end{bmatrix} \begin{bmatrix} \cos(\omega t_n) & \sin(\omega t_n) \end{bmatrix} \quad (8)$$

and

$$\mathbf{r} = \sum_{n=1}^N \begin{bmatrix} \cos(\omega t_n) \\ \sin(\omega t_n) \end{bmatrix} y(t_n). \quad (9)$$

The power of the sinusoidal component with frequency  $\omega$ , corresponding to  $(\hat{a}, \hat{b})$ , is given by

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \left( \begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix} \begin{bmatrix} \cos(\omega t_n) \\ \sin(\omega t_n) \end{bmatrix} \right)^2 \\ &= \frac{1}{N} \begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix} \mathbf{R} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \\ &= \frac{1}{N} \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r}. \end{aligned} \quad (10)$$

The LS periodogram is accordingly given by

$$P_{\text{LS}}(\omega) = \frac{1}{N} \mathbf{r}^T(\omega) \mathbf{R}^{-1}(\omega) \mathbf{r}(\omega) \quad (11)$$

(where we have reinstated the dependence on  $\omega$ , for clarity).

The LSP has been discussed, for example, in [3]–[8], under different forms and including various generalized versions. In particular, the papers [6] and [8] introduced a special case of LSP that has received significant attention in the subsequent literature. To describe this special instance of LSP, let us reparameterize (4) as follows:

$$\min_{\substack{\alpha \geq 0 \\ \tilde{\phi} \in [0, 2\pi]}} \sum_{n=1}^N [y(t_n) - \alpha \cos(\omega t_n + \tilde{\phi} + \bar{\phi})]^2 \quad (12)$$

where  $\tilde{\phi}$  is chosen as the solution to the equation

$$\sum_{n=1}^N \cos(\omega t_n + \tilde{\phi}) \sin(\omega t_n + \tilde{\phi}) = 0 \quad (13)$$

( $\tilde{\phi}$  can be obtained in an **explicit** form, see [6] and [8]). Reparameterizing (12) in terms of  $\tilde{a} = \alpha \cos(\tilde{\phi})$  and  $\tilde{b} = -\alpha \sin(\tilde{\phi})$ , and making use of (13), we obtain (similarity to (7))

$$\begin{bmatrix} \hat{\tilde{a}} \\ \hat{\tilde{b}} \end{bmatrix} = \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{r}} \quad (14)$$

where

$$\tilde{\mathbf{R}} = \sum_{n=1}^N \begin{bmatrix} \cos^2(\omega t_n + \tilde{\phi}) & 0 \\ 0 & \sin^2(\omega t_n + \tilde{\phi}) \end{bmatrix} \quad (15)$$

and

$$\tilde{\mathbf{r}} = \sum_{n=1}^N \begin{bmatrix} \cos(\omega t_n + \tilde{\phi}) \\ \sin(\omega t_n + \tilde{\phi}) \end{bmatrix} y(t_n). \quad (16)$$

The power of the sinusoidal component with frequency  $\omega$ , corresponding to  $\hat{\tilde{a}}$  and  $\hat{\tilde{b}}$  above, of course is identical to that associated with  $\hat{a}$  and  $\hat{b}$  in (7) (i.e., it is given by (11)), as the best LS sinusoidal fit to the data is unique. Then the question is whether the form of LSP obtained from (14), i.e.,  $(1/N) \tilde{\mathbf{r}}^T(\omega) \tilde{\mathbf{R}}^{-1}(\omega) \tilde{\mathbf{r}}(\omega)$ , is preferable to the plain LSP in (11) on computational or any other grounds.

In the words of [6], for numerical work it is simpler to use (7) than (14). We basically agree, as we briefly explain next. Equation (7) requires the inversion of a  $2 \times 2$  matrix, whereas this is not needed in (14) where the matrix is diagonal; however, (14) requires the computation of  $\tilde{\phi}$  (note that  $\tilde{\phi}$  depends on  $\omega$ ), which may be slightly trickier than the inversion of the matrix in (7). More importantly, the elements of the vector  $\mathbf{r}$  in (7) are equal to the real and imaginary parts of  $\sum_{n=1}^N y(t_n) e^{j\omega t_n}$ . There are fast algorithms, similar to the FFT, termed nonuniform FFT (NFFT) that can be used to compute the latter quantity and therefore the vector  $\mathbf{r}$  (see, e.g., [9]–[11]). The same algorithms can be used to evaluate the vector  $\tilde{\mathbf{r}}$  appearing in (14), which is given by the real and imaginary parts of

$$e^{j\tilde{\phi}(\omega)} \sum_{n=1}^N y(t_n) e^{j\omega t_n} \quad (17)$$

but at a computational cost a bit larger than that required for  $\mathbf{r}$ . To conclude this point, (7) may indeed be slightly more preferable than (14) from a computational standpoint (despite the fact that (14) avoids the  $2 \times 2$  matrix inversion operation).

The suggestion in both [6] and [8] was that, again in the words of [6], the use of (14) would facilitate the statistical description of the LSP. We do not find this suggestion to be justified. Indeed, the statistical analysis in either [6] or [8], which used (14) to show that  $NP_{\text{LS}}(\omega) \sim \chi^2(2)$  (i.e., that  $NP_{\text{LS}}(\omega)$  is distributed as a chi-square random variable with 2 degrees of freedom) under the assumption that  $\{y(t_n)\}_{n=1}^N$  is a normal white noise sequence with unit variance, is not simpler than the following proof of the same result based on (7). To prove the result under discussion, first observe that under the assumption made on  $\{y(t_n)\}_{n=1}^N$ , the vector  $\mathbf{r}$  is normally distributed with mean zero and covariance matrix equal to  $\mathbf{R}$ :  $\mathbf{r} \sim \mathcal{N}(0, \mathbf{R})$ . This observation implies that (here  $\mathbf{R}^{-(1/2)}$  is a symmetric square root of  $\mathbf{R}^{-1}$ )

$$\mathbf{R}^{-\frac{1}{2}} \mathbf{r} \sim \mathcal{N}(0, \mathbf{I}) \quad (18)$$

from which it follows at once that

$$NP_{LS}(\omega) = \left\| \mathbf{R}^{-\frac{1}{2}} \mathbf{r} \right\|^2 \sim \chi^2(2) \quad (19)$$

(for any given  $\omega$ , and under the normal white data assumption).

In view of the above arguments, we tend to recommend the use of the plain LSP in lieu of its version in [6] and [8]; this is in contrast to what has been usually done so far in the literature where the latter periodogram was favored on unclear grounds, sometimes even missing the fact that the final result is the same as that obtained with the plain LSP.

#### D. Spectral Window and Frequency Range

Consider a sinusoidal component with frequency  $\bar{\omega}$ , different from  $\omega$ , that might be present in the data:

$$\bar{a} \cos(\bar{\omega} t_n) + \bar{b} \sin(\bar{\omega} t_n), \quad n = 1, \dots, N \quad (20)$$

(if no such component exists in the data, then  $\bar{a} = \bar{b} = 0$ ). The effect of (20) on the LS estimate in (7) is to **introduce** an error term, or **leakage**, given by 泄露 介绍

$$\mathbf{R}^{-1} \sum_{n=1}^N \begin{bmatrix} \cos(\omega t_n) \\ \sin(\omega t_n) \end{bmatrix} [\cos(\bar{\omega} t_n) \quad \sin(\bar{\omega} t_n)] \begin{bmatrix} \bar{a} \\ \bar{b} \end{bmatrix}. \quad (21)$$

In general, (21) is a function of both  $\omega$  and  $\bar{\omega}$  that is relatively difficult to deal with. To simplify, we make use of the fact that, asymptotically in  $N$  and for “nonpathological” sampling patterns (see the Appendix for an explanation of this terminology)

$$\mathbf{R} \approx \frac{N}{2} \mathbf{I} \quad (22)$$

and

$$\begin{aligned} & \sum_{n=1}^N \begin{bmatrix} \cos(\omega t_n) \\ \sin(\omega t_n) \end{bmatrix} [\cos(\bar{\omega} t_n) \quad \sin(\bar{\omega} t_n)] \begin{bmatrix} \bar{a} \\ \bar{b} \end{bmatrix} \\ & \approx \sum_{n=1}^N \begin{bmatrix} \bar{a} \cos(\omega t_n) \cos(\bar{\omega} t_n) \\ \bar{b} \sin(\omega t_n) \sin(\bar{\omega} t_n) \end{bmatrix}. \end{aligned} \quad (23)$$

Next, assuming that  $\omega$  and  $\bar{\omega}$  are not both very close to zero, we have that  $\sum_{n=1}^N \cos(\omega + \bar{\omega}) t_n \approx 0$ , which implies that

$$\begin{aligned} & \sum_{n=1}^N \cos(\omega t_n) \cos(\bar{\omega} t_n) \\ & = \frac{1}{2} \sum_{n=1}^N [\cos((\omega - \bar{\omega}) t_n) + \cos((\omega + \bar{\omega}) t_n)] \\ & \approx \frac{1}{2} \sum_{n=1}^N \cos((\omega - \bar{\omega}) t_n) \end{aligned} \quad (24)$$

and

$$\begin{aligned} & \sum_{n=1}^N \sin(\omega t_n) \sin(\bar{\omega} t_n) \\ & = \frac{1}{2} \sum_{n=1}^N [\cos((\omega - \bar{\omega}) t_n) - \cos((\omega + \bar{\omega}) t_n)] \\ & \approx \frac{1}{2} \sum_{n=1}^N \cos((\omega - \bar{\omega}) t_n). \end{aligned} \quad (25)$$

The right-hand side in the above two equations can be further approximated as

$$\begin{aligned} \frac{1}{2} \sum_{n=1}^N \cos((\omega - \bar{\omega}) t_n) & = \frac{1}{2} \operatorname{Re} \left[ \sum_{n=1}^N e^{j(\omega - \bar{\omega}) t_n} \right] \\ & \approx \frac{1}{2} \sum_{n=1}^N e^{j(\omega - \bar{\omega}) t_n}. \end{aligned} \quad (26)$$

Inserting (24)–(26) into (23) leads to the following approximation of the error term in (21):

$$\frac{1}{N} \begin{bmatrix} \bar{a} \\ \bar{b} \end{bmatrix} \sum_{n=1}^N e^{j(\omega - \bar{\omega}) t_n}. \quad (27)$$

We can express the “size” of this error as the squared Euclidean norm of the vector in (27):

$$\frac{1}{N^2} P(\bar{\omega}) \left| \sum_{n=1}^N e^{j(\omega - \bar{\omega}) t_n} \right|^2 \quad (28)$$

where  $P(\bar{\omega}) = \bar{a}^2(\omega) + \bar{b}^2(\omega)$ . As might have been expected, (28) is the same as the error term that would be caused by a complex sinusoid,  $(\bar{a} + j\bar{b})e^{j\bar{\omega}t}$ , on the Fourier periodogram in (1) evaluated at  $\omega$ . By analogy with the uniformly sampled data case, the part of (28) that depends on the sampling pattern, written as a function of a single frequency variable viz.

$$W(\omega) = \left| \sum_{n=1}^N e^{j\omega t_n} \right|^2 \quad (29)$$

is called the spectral window (as a quick motivation for this name, note that  $W(\omega)$  above is proportional to the squared magnitude of the Fourier transform of the sampling window  $\sum_{n=1}^N \delta(t - t_n)$ ).

Next we discuss, following [12], the use of the spectral window to define the maximum frequency interval, let us say  $[0, \omega_{\max}]$ , that can be dealt with unambiguously, for a given set of sampling times  $\{t_n\}$ . The spectral window attains its maximum value of  $N^2$  at  $\omega = 0$ :  $W(0) = N^2 \geq W(\omega), \forall \omega$ . We evaluate  $W(\omega)$  for  $\omega > 0$  to find the smallest frequency value, let us say  $\omega_0$ , at which the spectral window has a peak whose height is close to  $N^2$  (if we cannot find any such peak even for  $\omega \gg 1$ , then we set  $\omega_0 = \infty$ ). Theoretically, we should consider  $W(\omega_0) = N^2$ . However, in applications with relatively low signal-to-noise ratios, we may choose  $W(\omega_0) < N^2$ , perhaps even as low as  $0.8N^2$ . In any case, once  $\omega_0$  was computed, then the claim is that

$$\omega_{\max} = \frac{\omega_0}{2} \quad (30)$$

should be the largest frequency of the data sequence in question. To motivate this claim, we must verify that no aliasing can take place if and only if (30) is satisfied. However, this follows easily from the fact that if (30) is true, then  $|\omega - \bar{\omega}| < \omega_0$  for any  $\omega, \bar{\omega} \in (-(\omega_0/2), (\omega_0/2)), \omega \neq \bar{\omega}$ ; and that if (30) does not hold, then

the frequency  $\bar{\omega} = (\omega_0/2) + \Delta$ , for any  $\Delta \in (0, (\omega_0/2))$ , is aliased as  $\omega = -(\omega_0/2) + \Delta \neq \bar{\omega}$ ,  $\omega \in [-(\omega_0/2), (\omega_0/2)]$ .

Note that often in the literature,  $\omega_{\max}$  is chosen as  $(\pi N)/(t_N - t_1)$  or as  $(\pi)/(\min_{k \neq p} |t_k - t_p|)$ . As noted in [12],  $\omega_{\max}$  in (30) is typically much larger than either of these choices.

Regarding the smallest frequency separation that can be safely detected, let us say  $\Delta\omega$ , this can be approximated by the requirement that

$$W(\Delta\omega) = \frac{N^2}{2}; \quad W(\omega) > \frac{N^2}{2} \text{ for } \omega \in [0, \Delta\omega]. \quad (31)$$

For **nonpathological** sampling patterns we typically have, with a reasonable approximation **非病态的**

$$\Delta\omega = \frac{2\pi}{t_N - t_1}. \quad (32)$$

Because  $\Delta\omega$  above is essentially the resolution limit of the LSP, we may want to evaluate the periodogram function on a grid with a step (much) smaller than (32), such as  $\Delta\omega/10$ .

## II. CONTRIBUTIONS AND OUTLOOK

As is well known, the spectral estimates obtained with either FP or LSP suffer from both local and global (or distant) leakage problems. Local leakage is due to the width of the main beam of the spectral window, and it is what limits the resolution capability of the periodogram. Global leakage is due to the sidelobes of the spectral window, and is what causes spurious peaks to occur (which leads to “false alarms”) and small peaks to drown in the leakage from large peaks (which leads to “misses”).

Additionally, there is no satisfactory procedure for testing the significance of the periodogram peaks. In the uniformly sampled data case, there is a relatively well-established test for the significance of the most dominant peak of the periodogram; see [1], [2], and [13] and the references therein. In the nonuniformly sampled data case, [8] (see also [14] for a more recent account) has proposed a test that mimics the uniform data case test mentioned above. However, it appears that the said test is not readily applicable to the nonuniform data case; see [13] and the references therein. As a matter of fact, even if the test were applicable, it would only be able to decide whether  $\{y(t_n)\}$  are white noise samples, and not whether the data sequence contains one or several sinusoidal components (we remark in passing on the fact that, even in the uniform data case, testing the existence of multiple sinusoidal components, i.e., the significance of the second largest peak of the periodogram, and so forth, is rather intricate [1], [2]). The only way of correcting the test, to make it applicable to nonuniform data, appears to be via Monte Carlo simulations, which may be a rather computationally intensive task (see the lucid discussion in [13] on this aspect).

The main contribution of the present paper is the introduction of a new method for spectral estimation and detection in the nonuniformly sampled data case, that does not suffer from the above drawbacks of the periodogram (i.e., poor resolution due to local leakage through the mainlobe of the spectral window, significant global leakage through the sidelobes, and lack of satisfactory tests for the significance of the dominant peaks). A preview of what the paper contains is as follows.

- The amplitude and phase estimation (APES) method, proposed in [15] for uniformly sampled data, has significantly less leakage (both local and global) than the periodogram. We follow here the ideas in [16]–[18] to extend APES to the nonuniformly sampled data case. The so-obtained generalized method is referred to as RIAA for reasons explained in the Abstract.
- Both LSP and RIAA provide nonparametric spectral estimates in the form of an estimated amplitude spectrum (or periodogram  $P(\omega)$ ). We use the frequencies and amplitudes corresponding to the dominant peaks of  $P(\omega)$  (first the largest one, then the second largest, and so on) in a Bayesian information criterion (BIC), see, e.g., [19] and the references therein, to decide which peaks we should retain and which ones we can discard. The combined methods, viz. LSP+BIC and RIAA+BIC, provide parametric spectral estimates in the form of a number of estimated sinusoidal components that are deemed to fit the data well. Therefore, the use of BIC in the outlined manner not only bypasses the need for testing the significance of the periodogram peaks in the manner of [8] (which would be an intractable problem for RIAA, and almost an intractable one for LSP as well—see [13]), but it also provides additional information in the form of an estimated number of sinusoidal components, which no periodogram test of the type discussed in the cited references can really provide.
- Finally, we present a method for designing an optimal sampling pattern that minimizes an objective function based on the spectral window. In doing so, we assume that a sufficient number of observations are already available, from which we can get a reasonably accurate spectral estimate. We make use of this spectral estimate to design the sampling times when future measurements should be performed. The literature is relatively scarce in papers that approach the sampling pattern design problem (see, e.g., [8] and [20]). One reason for this may be that, as explained later on, spectral window-based criteria are relatively insensitive to the sampling pattern, unless prior information (such as a spectral estimate) is assumed to be available—as in this paper. Another reason may be the fact that measurement plans might be difficult to realize in some applications, due to factors that are beyond the control of the experimenter. However, this is not a serious problem for the sampling pattern design strategy proposed here which is flexible enough to tackle cases with missed measurements by revising the measurement plan on the fly.

## III. RIAA AND BIC

### A. RIAA

Let  $\Delta\omega$  denote the step size of the grid **后续的** considered for the frequency variable, see (32) and the **subsequent** comment, and let  $K = \lfloor \omega_{\max}/\Delta\omega \rfloor$  denote the number of the grid points needed to cover the frequency interval  $[0, \omega_{\max}]$ , where  $\lfloor x \rfloor$  denotes the largest integer less than or equal to  $x$ ; also, let

$$\omega_k = k\Delta\omega, \quad k = 1, \dots, K. \quad (33)$$

The uniform frequency grid in (33) will be the one used in most applications, and this is why we have mentioned it here, but we should stress that the spectral analysis methods discussed in this paper can be used with any type of frequency grid without any conceptual modification (only the implementation code will depend on whether the grid is uniform or nonuniform).

The following additional notation will be needed as well in this section:

$$\mathbf{y} = \begin{bmatrix} y(t_1) \\ \vdots \\ y(t_N) \end{bmatrix}, \quad \boldsymbol{\theta}_k = \begin{bmatrix} a(\omega_k) \\ b(\omega_k) \end{bmatrix}, \quad \mathbf{A}_k = [\mathbf{c}_k \quad \mathbf{s}_k],$$

$$\mathbf{c}_k = \begin{bmatrix} \cos(\omega_k t_1) \\ \vdots \\ \cos(\omega_k t_N) \end{bmatrix}, \quad \mathbf{s}_k = \begin{bmatrix} \sin(\omega_k t_1) \\ \vdots \\ \sin(\omega_k t_N) \end{bmatrix}. \quad (34)$$

Using this notation, we can rewrite the LS fitting criterion in (6) in the following vector form (for  $\omega = \omega_k$ ):

$$\|\mathbf{y} - \mathbf{A}_k \boldsymbol{\theta}_k\|^2, \quad k = 1, \dots, K \quad (35)$$

where  $\|\cdot\|$  denotes the Euclidean norm. The LS estimate of  $\boldsymbol{\theta}_k$  in (7) can be rewritten as

$$\hat{\boldsymbol{\theta}}_k = (\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T \mathbf{y}. \quad (36)$$

Besides a possible sinusoidal component with frequency  $\omega_k$ , as considered in (35), the data  $\mathbf{y}$  might contain other sinusoidal components with frequencies different from  $\omega_k$ , as well as noise. Regarding the latter, we do not consider a noise component of  $\mathbf{y}$  explicitly, but rather implicitly via its contributions to the data spectrum at  $\{\omega_k\}_{k=1}^K$ ; for typical values of the signal-to-noise ratio, these noise contributions to the spectrum are comparatively small. Let us define

$$\mathbf{Q}_k = \sum_{p=1, p \neq k}^K \mathbf{A}_p \mathbf{D}_p \mathbf{A}_p^T,$$

$$\mathbf{D}_p = \frac{a^2(\omega_p) + b^2(\omega_p)}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (37)$$

which can be thought of as the covariance matrix of the other possible components in  $\mathbf{y}$ , besides the sinusoidal component with frequency  $\omega_k$  considered in (35).

*Remark:* In some applications, the covariance matrix of the noise component of  $\mathbf{y}$  is known (or, rather, can be assumed with a good approximation) to be

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^2 \end{bmatrix}; \quad \text{with } \{\sigma_n^2\}_{n=1}^N \text{ given.} \quad (38)$$

In such cases, we can simply add  $\boldsymbol{\Sigma}$  to the matrix  $\mathbf{Q}_k$  in (37). ■

Assuming that  $\mathbf{Q}_k$  is available, and that it is invertible, it would make sense to consider the following weighted LS (WLS) criterion, instead of (35),

$$[\mathbf{y} - \mathbf{A}_k \boldsymbol{\theta}_k]^T \mathbf{Q}_k^{-1} [\mathbf{y} - \mathbf{A}_k \boldsymbol{\theta}_k]. \quad (39)$$

Indeed, it is well known that the estimate of  $\boldsymbol{\theta}_k$  obtained by minimizing (39) is more accurate, under quite general conditions, than the LS estimate obtained from (35). Note that a necessary condition for  $\mathbf{Q}_k^{-1}$  to exist is that  $2(K-1) \geq N$ , which is easily satisfied in general.

The vector  $\boldsymbol{\theta}_k$  that minimizes (39) is given by

$$\hat{\boldsymbol{\theta}}_k = (\mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{A}_k)^{-1} (\mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{y}). \quad (40)$$

Similarly to (10), the WLS periodogram, corresponding to (40), can be defined as

$$P_{\text{WLS}}(\omega_k) = \frac{1}{N} \hat{\boldsymbol{\theta}}_k^T (\mathbf{A}_k^T \mathbf{A}_k) \hat{\boldsymbol{\theta}}_k. \quad (41)$$

The WLS estimate in (40) appears to require the inversion of an  $N \times N$  matrix  $\mathbf{Q}_k$ , for each value of  $k = 1, \dots, K$ . For  $N \gg 1$ , this would be computationally a rather intensive task. To show how we can simply reduce the computational complexity of (40), let us introduce the matrix

$$\boldsymbol{\Gamma} = \sum_{p=1}^K \mathbf{A}_p \mathbf{D}_p \mathbf{A}_p^T = \mathbf{Q}_k + \mathbf{A}_k \mathbf{D}_k \mathbf{A}_k^T. \quad (42)$$

A simple calculation shows that

$$\mathbf{Q}_k^{-1} \mathbf{A}_k = \boldsymbol{\Gamma}^{-1} \mathbf{A}_k (\mathbf{I} + \mathbf{D}_k \mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{A}_k). \quad (43)$$

To verify this equation, **premultiply** it with the  $\boldsymbol{\Gamma}$  in (42) and observe that

$$\begin{aligned} \boldsymbol{\Gamma} \mathbf{Q}_k^{-1} \mathbf{A}_k &= \mathbf{A}_k + \mathbf{A}_k \mathbf{D}_k \mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{A}_k \\ &= \mathbf{A}_k (\mathbf{I} + \mathbf{D}_k \mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{A}_k). \end{aligned} \quad (44)$$

Inserting (43) into (40) yields the following alternative expression for the WLS estimate:

$$\hat{\boldsymbol{\theta}}_k = (\mathbf{A}_k^T \boldsymbol{\Gamma}^{-1} \mathbf{A}_k)^{-1} (\mathbf{A}_k^T \boldsymbol{\Gamma}^{-1} \mathbf{y}) \quad (45)$$

which is computationally more appealing than (40), as  $\boldsymbol{\Gamma}^{-1}$  in (45) needs to be computed only once for all values of  $k = 1, \dots, K$ .

Next, we explain how to circumvent the problem that  $\boldsymbol{\Gamma}$  in (45) depends on the very quantities that we want to estimate, viz. the  $\{\boldsymbol{\theta}_k\}_{k=1}^K$ , and consequently that  $\hat{\boldsymbol{\theta}}_k$  cannot be implemented directly. The only apparent solution to this problem is some form of iterative process, for example the RIAA algorithm, containing equations (46) and (47), outlined in Table I.

In most applications, the RIAA algorithm is expected to require no more than 10–20 iterations (see, e.g., the numerical examples in Section V).

Note that we might think of replacing  $\hat{\boldsymbol{\Gamma}}^i$  in (46) by

$$\hat{\boldsymbol{\Gamma}}_k^{i,i+1} = \sum_{p=1}^{k-1} \mathbf{A}_p \hat{\mathbf{D}}_p^{i+1} \mathbf{A}_p^T + \sum_{p=k}^K \mathbf{A}_p \hat{\mathbf{D}}_p^i \mathbf{A}_p^T \quad (48)$$

which makes use not only of the  $i$ th iteration estimates of  $\{\boldsymbol{\theta}_k\}$  but also of the  $(i+1)$ th iteration estimates as they become available. However, the estimate of  $\boldsymbol{\Gamma}$  in (48) depends on  $k$  (as indicated by notation), which would complicate the computation of  $\hat{\boldsymbol{\Gamma}}_k^{i,i+1}$ : indeed, while the matrix inversion lemma could be used

TABLE I  
THE RIAA ALGORITHM

**Initialization**

Use the LS method (see Eq. (36)) to obtain initial estimates of  $\{\theta_k\}$ , denoted by  $\{\hat{\theta}_k^0\}$ .

**Iteration**

Let  $\{\hat{\theta}_k^i\}$  denote the estimates of  $\{\theta_k\}$  at the  $i^{\text{th}}$  iteration ( $i = 0, 1, 2, \dots$ ), and let  $\hat{\Gamma}^i$  denote the estimate of  $\Gamma$  obtained from  $\{\hat{\theta}_k^i\}$ . For  $i = 0, 1, 2, \dots$ , compute:

$$\hat{\theta}_k^{i+1} = [\mathbf{A}_k^T (\hat{\Gamma}^i)^{-1} \mathbf{A}_k]^{-1} [\mathbf{A}_k^T (\hat{\Gamma}^i)^{-1} \mathbf{y}], \quad k = 1, \dots, K, \quad (46)$$

until a given number of iterations is performed.

**Periodogram Calculation**

Let  $\{\hat{\theta}_k^I\}$  denote the estimates of  $\{\theta_k\}$  obtained by the iterative process outlined above (here  $I$  denotes the iteration number at which the iteration process is stopped). Use  $\{\hat{\theta}_k^I\}$  to compute the RIAA periodogram (as see Eq. (41)):

$$P_{\text{RIAA}}(\omega_k) = \frac{1}{N} (\hat{\theta}_k^I)^T (\mathbf{A}_k^T \mathbf{A}_k) (\hat{\theta}_k^I), \quad k = 1, \dots, K. \quad (47)$$

to compute  $(\Gamma_k^{i,i+1})^{-1}$  from  $(\Gamma_k^{i,i+1})^{-1}$ , the computation cost compared with that of (46) would increase, and this without any guarantee for better performance. Consequently, we will use the RIAA with the simpler estimate of  $\Gamma$  in (46).

**B. RIAA and APES**

In what follows, we make use of the analogy between RIAA and the amplitude and phase estimate (APES) periodogram (see, e.g., [15] and [21]) to provide some additional insights into the expected behavior of RIAA. In particular, we use this analogy to explain intuitively why RIAA is expected to have less (both local and global) leakage than LSP.

We begin by noting that there is an essential difference between RIAA and APES: the latter uses an estimate of the residual matrix  $\mathbf{Q}_k$  obtained from multiple realizations of the data vector  $\mathbf{y}$ , whereas RIAA estimates  $\mathbf{Q}_k$  using a theoretical formula for this matrix, along with the most recent spectral estimate available; note that in the spectral analysis problem considered in this paper, we dispose of only one realization of  $\mathbf{y}$ , so APES would not be applicable directly. However, once this difference is realized (as well as the fact that APES was derived in [15] and [21] for complex-valued data, whereas RIAA assumes real-valued data, which complicates its formulas to some extent), (40) used by RIAA to estimate  $\theta_k$  looks similar to the corresponding equation of APES (see [15] and [21]). The implication of this observation is that we should be able to obtain the estimate in (40) in the manner in which the APES estimate is usually derived (see the cited papers). To describe this alternative manner, let  $\hat{\mathbf{H}}_k$  denote an  $N \times 2$  matrix that is the solution to the following constrained minimization problem:

$$\min_{\mathbf{H}_k} f(\mathbf{H}_k^T \mathbf{Q}_k \mathbf{H}_k) \quad (49)$$

$$\text{s.t. } \mathbf{H}_k^T \mathbf{A}_k = \mathbf{I} \quad (50)$$

where  $f$  is a monotonically increasing function on the domain of positive definite matrices (which means that  $\mathbf{X} \geq \mathbf{Y}$ , i.e., the difference matrix  $\mathbf{X} - \mathbf{Y}$  is positive semi-definite, implies  $f(\mathbf{X}) \geq f(\mathbf{Y})$ ), for example the trace or the **determinant**. We use  $\hat{\mathbf{H}}_k$ , in the manner of APES, to obtain an estimate of  $\theta_k$  as

$$\hat{\theta}_k = \hat{\mathbf{H}}_k^T \mathbf{y}. \quad \text{行列式} \quad (51)$$

To prove the equivalence between (51) and (40), we need to show that the solution to (49) is given by

$$\hat{\mathbf{H}}_k^T = (\mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{A}_k)^{-1} \mathbf{A}_k^T \mathbf{Q}_k^{-1} \quad (52)$$

or, equivalently, that

$$\mathbf{H}_k^T \mathbf{Q}_k \mathbf{H}_k \geq \hat{\mathbf{H}}_k^T \mathbf{Q}_k \hat{\mathbf{H}}_k = (\mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{A}_k)^{-1} \quad (53)$$

for any  $\mathbf{H}_k$  that satisfies the constraint in (50). Using the fact that  $\mathbf{H}_k^T \mathbf{A}_k = \mathbf{I}$ , we can write (53) in the following form:

$$\mathbf{H}_k^T \mathbf{Q}_k \mathbf{H}_k - \mathbf{H}_k^T \mathbf{A}_k (\mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{A}_k)^{-1} \mathbf{A}_k^T \mathbf{H}_k \geq 0 \quad (54)$$

which is equivalent (by standard Schur complement results, see, e.g., [2]) to

$$\begin{bmatrix} \mathbf{H}_k^T \mathbf{Q}_k \mathbf{H}_k & \mathbf{H}_k^T \mathbf{A}_k \\ \mathbf{A}_k^T \mathbf{H}_k & \mathbf{A}_k^T \mathbf{Q}_k^{-1} \mathbf{A}_k \end{bmatrix} \geq 0 \quad (55)$$

and therefore to

$$\begin{bmatrix} \mathbf{H}_k^T \mathbf{Q}_k^{\frac{1}{2}} \\ \mathbf{A}_k^T \mathbf{Q}_k^{-\frac{1}{2}} \end{bmatrix} [\mathbf{Q}_k^{\frac{1}{2}} \mathbf{H}_k \quad \mathbf{Q}_k^{-\frac{1}{2}} \mathbf{A}_k] \geq 0. \quad (56)$$

Because the matrix in the left-hand side of (56) is evidently positive semi-definite, the proof that (52) is the solution to (49), and therefore that  $\hat{\theta}_k$  in (51) coincides with the WLS estimate of  $\theta_k$ , is concluded.

*Remark:* Making use of the constraint, we can reformulate the minimization problem in (49) as

$$\min_{\mathbf{H}_k} f(\mathbf{H}_k^T \Gamma \mathbf{H}_k) \quad (57)$$

$$\text{s.t. } \mathbf{H}_k^T \mathbf{A}_k = \mathbf{I} \quad (58)$$

whose solution, viz.

$$\hat{\mathbf{H}}_k^T = (\mathbf{A}_k^T \Gamma \mathbf{A}_k)^{-1} \mathbf{A}_k^T \Gamma^{-1} \quad (59)$$

must **coincide** with (52). This observation provides another proof of the equivalence between the WLS estimates in (40) and (45). **直觉**

The **intuition** afforded by the above APES-like derivation of WLS is interesting. The matrix  $\hat{\mathbf{H}}_k^T$  that solves the optimization problem in (49) can be viewed as a “filter” that passes the sinusoidal component of current interest (with frequency  $\omega_k$ ) without any distortion (by (50)) and attenuates all the other components in  $\mathbf{y}$  as much as possible (by (49)). To illustrate the latter property of  $\hat{\mathbf{H}}_k$ , we use the fact that, by assumption, the data contains a finite (usually small) number of sinusoidal components. This means that there are only a limited number of frequencies which contribute significant terms to  $\mathbf{Q}_k$ . Let  $\omega_p$  be

one of these frequencies. Then, in view of (49),  $\hat{\mathbf{H}}_k^T$  should be nearly orthogonal to  $\mathbf{A}_p$ , i.e.,  $\hat{\mathbf{H}}_k^T \mathbf{A}_p \approx 0$ , which means that  $\hat{\mathbf{H}}_k^T$  acts as an annihilating filter for any strong sinusoidal component in  $\mathbf{y}$  whose frequency is different from  $\omega_k$ . This observation explains why RIAA can be expected to have significantly reduced leakage problems compared with LSP: note that the “filtering” matrix used by LSP, viz.  $(\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T$ , is data-independent and as such it cannot eliminate the sinusoidal components in  $\mathbf{y}$ , with frequencies different from  $\omega_k$ , as well as  $\hat{\mathbf{H}}_k^T$  does.

C. LSP+BIC and RIAA+BIC

As already mentioned in the previous section, the use of the standard  $\chi^2$ -test (borrowed from the uniform data case, see [1], [2] and the references therein) to decide whether or not the most dominant peak of LSP is significant, as suggested in [8], faces a number of problems in the nonuniformly sampled data case. **First**, in the latter case, the test is in general invalid (see, e.g., [13]). **Second**, and more important in our opinion, even in the uniformly sampled data case the test can hardly be used to decide if there are sinusoidal components in the data (and, if yes, how many of them), but rather only to check whether or not the data consist of white noise alone. An additional problem is that for RIAA the distribution of the corresponding spectral estimate is unknown (even in the uniform data case) and hence a  $\chi^2$ -type of test would not be applicable even approximately.

The BIC rule is an alternative to standard statistical tests of hypothesis testing, such as the  $\chi^2$  test discussed above, which can be used in the problem of interest here, **albeit** in an approximate manner (see below for details). To explain how this can be done, we let  $\{P(\omega_k)\}_{k=1}^K$  denote the values taken by either the LSP or the RIAA periodogram at the points of the frequency grid, and let

$$\{\check{\omega}_m, \check{a}_m, \check{b}_m\}_{m=1}^M \quad (60)$$

denote the frequencies and the amplitude and phase related parameters corresponding to the  $M$  largest peaks of  $\{P(\omega_k)\}_{k=1}^K$ , arranged in a decreasing order of their associated periodogram values:

$$P(\check{\omega}_1) \geq P(\check{\omega}_2) \geq \dots \geq P(\check{\omega}_M) \quad (61)$$

for  $M \geq 1$ . Under the idealizing assumptions that the data sequence consists of a finite number of sinusoidal components and of normal white noise, and that  $\{\check{\omega}_m, \check{a}_m, \check{b}_m\}_{m=1}^M$  are the maximum likelihood (ML) estimates of the frequencies and the amplitude and phase related parameters of  $M$  such sinusoidal components, the BIC rule estimates  $M$  as follows:

$$\hat{M} = \arg \min_{M=0,1,2,\dots} \text{BIC}(M) \quad (62)$$

where

$$\text{BIC}(M) = N \ln \left\{ \sum_{n=1}^N \left[ y(t_n) - \sum_{m=1}^M (\check{a}_m \cos(\check{\omega}_m t_n) + \check{b}_m \sin(\check{\omega}_m t_n)) \right]^2 \right\} + \rho M \ln N \quad (63)$$

and where

$$\rho = \rho_{\text{ML}} = 5 \quad (64)$$

(see, e.g., [2] and [19] and the references therein). Note that BIC is made of two terms: a LS data fitting term that decreases as  $M$  increases, and a complexity **penalization** term which increases with increasing  $M$ ; therefore, the BIC estimate of  $M$ , i.e.,  $\hat{M}$  in (62), is obtained by a **tradeoff** between in-sample fitting accuracy and complexity of the sinusoidal data description. Also note that if  $\hat{M} = 0$  is selected then, according to BIC, the data consist of only white noise.

In applications, the idealizing assumptions made above do not hold exactly. Indeed, usually the noise in the data is not white and normally distributed. Moreover, the  $\{\check{\omega}_m, \check{a}_m, \check{b}_m\}_{m=1}^M$  obtained from the peaks of the periodogram are not ML estimates (unless  $M = 1$  and the noise is white and normal). Despite these problems, BIC can still be used for estimating  $M$  (i.e., essentially for testing the significance of the peaks of LSP or RIAA periodogram), **albeit** only in an approximate manner, as explained next.

It follows from the derivation of BIC (see, e.g., the cited works) that the large complexity **penalty** corresponding to the  $\rho$  in (64) is mainly due to the high accuracy of the ML estimates. Clearly, as  $\rho$  increases, the probability of “false alarms” decreases, whereas the probability of “misses” increases; an accurate estimation method, such as the ML, should allow a detection rule to operate at quite a small probability of false alarm without more than an insignificant increase in the probability of miss—this explains intuitively why the value of  $\rho$  in BIC, corresponding to the ML, is relatively high.

By comparison, the frequencies and amplitude and phase related parameters estimated with RIAA and LSP are expected to be less accurate and, respectively, significantly less accurate than the ML estimates. Consequently, we suggest the use of BIC with smaller values of  $\rho$  in the case of LSP and RIAA than the  $\rho_{\text{ML}} = 5$  used in the case of ML, namely

$$\rho_{\text{LSP}} = 1 \text{ and } \rho_{\text{RIAA}} = 4. \quad (65)$$

Admittedly, these choices are somewhat **ad-hoc**, but the corresponding BIC rules are simple to use and they appear to provide accurate estimates of  $M$  (see Section V for details).

IV. SAMPLING PATTERN DESIGN

We consider the following situation: the data samples  $\{y(t_n)\}_{n=1}^N$  have already been collected, and a spectral estimate,  $\{P(\omega_k)\}_{k=1}^K$ , has been obtained based on them; we want to choose the sampling times for future measurements,  $\{x_f\}_{f=1}^F$ , in an **optimal manner** according to some **objective function**. Let  $\{\tau_g\}_{g=1}^G$  denote a set of possible times when future measurements of  $y(t)$  could **in principle** be performed. Naturally, we assume that  $G > F$  and that

$$\{x_f\}_{f=1}^F \subset \{\tau_g\}_{g=1}^G \triangleq \mathcal{S}. \quad (66)$$

Note that any known **constraints** on the set of possible future sampling times, such as **black out dates** and the like, can be taken into account by properly selecting  $\mathcal{S}$ .

↑  
假设

解释

均衡

← 尽管

← 惩罚

↘ 点对点

← 最佳模式

← 限制

← 类似于数据丢失等问题

The discussion in the previous sections (see also [20]) has evidenced quite clearly the fact that the performance of LSP depends, to a large extent, on the shape of the spectral window (its mainlobe width and sidelobe heights), and so does the performance of RIAA, but to a smaller extent. Consequently, a spectral window-based function should be a logical choice of an objective to optimize with respect to  $\{x_f\}_{f=1}^F$ .

The spectral window corresponding to the set of past sampling times  $\{t_n\}_{n=1}^N$  and possible future sampling times  $\{\tau_g\}_{g=1}^G$  is given by (see (29))

$$W(\omega_k) = \left| \sum_{n=1}^N e^{j\omega_k t_n} + \sum_{g=1}^G e^{j\omega_k \tau_g} \right|^2, \quad k = 1, \dots, K \quad (67)$$

and the **integrated** spectral window by

$$\sum_{k=1}^K W(\omega_k) = \sum_{k=1}^K \left| \sum_{n=1}^N e^{j\omega_k t_n} + \sum_{g=1}^G e^{j\omega_k \tau_g} \right|^2. \quad (68)$$

We might think of minimizing (68) with respect to  $\{x_f\}_{f=1}^F \subset \mathcal{S}$ . **Furthermore**, because (68) does not depend on any **prior information**, we might even think of minimizing (68) with respect to  $\{x_f\}_{f=1}^F$  for  $N = 0$  (in which case we would design the **entire set** of sampling times). However, doing so is not advisable. The reason is that the integrated spectral window is rather insensitive to the choice of  $\{x_f\}$ . To see why this is so, observe that (68) is the **discrete approximation** of the following integral function (to within a multiplication constant):

$$u^T \left[ \int_0^{\omega_{\max}} v(\omega) v^*(\omega) d\omega \right] u \quad (69)$$

where  $(\cdot)^*$  denotes the conjugate transpose, and

$$u = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad v(\omega) = \begin{bmatrix} e^{j\omega t_1} \\ \vdots \\ e^{j\omega t_N} \\ e^{j\omega \tau_1} \\ \vdots \\ e^{j\omega \tau_G} \end{bmatrix}. \quad (70)$$

For **nonpathological** choices of  $\omega_{\max}$ ,  $\{t_n\}$  and  $\{\tau_g\}$ ,

$$\int_0^{\omega_{\max}} v(\omega) v^*(\omega) d\omega \sim \mathbf{I} \quad (71)$$

so that (69), and **therefore** (68), is approximately constant. This observation **extends** the well-known fact that in the uniform sampling case the integrated spectral window is constant (a fact that in particular implies that if the sidelobe heights are reduced, then the mainlobe width must increase, and vice versa).

It follows from the previous discussion that in order to get a meaningful spectral window-based objective function for the design of  $\{x_f\}_{f=1}^F$ , we must **assume** that prior information, e.g., in the form of a spectral estimate  $\{P(\omega_k)\}$ , is available. Under

this condition, the spectral window (which now is a two-dimensional function) is given at  $\omega = \omega_p$  by (see (28))

$$P(\omega_k) \left| \sum_{n=1}^N e^{j(\omega_p - \omega_k) t_n} + \sum_{g=1}^G e^{j(\omega_p - \omega_k) \tau_g} \right|^2, \quad p, k = 1, \dots, K \quad (72)$$

and therefore the integrated spectral window (at  $\omega = \omega_p$ ) is

$$\sum_{k=1}^K P(\omega_k) \left| \sum_{n=1}^N e^{j(\omega_p - \omega_k) t_n} + \sum_{g=1}^G e^{j(\omega_p - \omega_k) \tau_g} \right|^2. \quad (73)$$

Compared with (68) (which can be obtained from (73) by setting  $P(\omega_k) = 1$ ), the above function has a more appealing character: for a given  $\omega_p$ , the sidelobe corresponding to  $\omega_k$  is weighted by the power in the data sequence at  $\omega_k$ ; therefore, the larger the said power, the more emphasis will be put on minimizing the level of the corresponding sidelobe of the spectral window.

*Remark:* Instead of the weighted 2-norm type of function in (73), we might want to consider a weighted  $\infty$ -norm objective:

$$\max_{k \in [1, K]} P(\omega_k) \left| \sum_{n=1}^N e^{j(\omega_p - \omega_k) t_n} + \sum_{g=1}^G e^{j(\omega_p - \omega_k) \tau_g} \right|^2. \quad (74)$$

The largest (weighted) sidelobe that will result from the use of (74) will of course be smaller than the largest (weighted) sidelobe that is obtained by using (73). However, many other (weighted) sidelobes of the former design may be larger than those of the latter, a fact that suggests that (73) may be preferable in many cases. ■

The function in (73) still depends on  $\omega_p$ . Because the sampling times cannot be frequency dependent, we must eliminate the dependence of the objective function on the frequency. This can be done by “integrating” (73), weighted by  $P(\omega_p)$ , with respect to  $\omega_p$  as well:

$$\sum_{p=1}^K P(\omega_p) \sum_{k=1}^K P(\omega_k) \left| \sum_{n=1}^N e^{j(\omega_p - \omega_k) t_n} + \sum_{g=1}^G e^{j(\omega_p - \omega_k) \tau_g} \right|^2. \quad (75)$$

In the discussion so far, we have been able to define what appears to be a reasonable objective function that has the form of an integrated weighted spectral window. The problem that remains is how to minimize this function with respect to  $\{x_f\}_{f=1}^F \subset \mathcal{S}$ . For relatively small values of  $G$  (which is equal to the cardinality of  $\mathcal{S}$ ), the minimization problem in question can be solved by checking all possible combinations of  $\{x_f\}_{f=1}^F$  in  $\mathcal{S}$ . However, such an exhaustive combinatorial search might not be feasible for larger values of  $G$  (and of  $F$ ). In the latter case, we can proceed in the following approximate but simpler manner.

Let  $\{\mu_g\}_{g=1}^G$  be Boolean variables:  $\mu_g \in \{0, 1\}$ . Making use of these variables, we can write the problem of minimizing (75) with respect to  $\{x_f\}_{f=1}^F \subset \mathbb{S}$  in the following way:

$$\begin{aligned} \min_{\{\mu_g\}_{g=1}^G} & \sum_{p=1}^K \sum_{k=1}^K P(\omega_p) P(\omega_k) \left| \sum_{n=1}^N e^{j(\omega_p - \omega_k)t_n} \right. \\ & \left. + \sum_{g=1}^G \mu_g e^{j(\omega_p - \omega_k)\tau_g} \right|^2 \\ \text{s.t. } & \mu_g \in \{0, 1\} \\ & \sum_{g=1}^G \mu_g = F. \end{aligned} \quad (76)$$

Because the Boolean constraint is what makes the problem in (76) hard, we relax it to

$$0 \leq \mu_g \leq 1, \quad g = 1, \dots, G. \quad (77)$$

The problem in (76) with the constraint  $\mu_g \in \{0, 1\}$  replaced by (77) is a linearly constrained **quadratic program** that can be solved efficiently, for instance, by means of the MATLAB<sup>1</sup> function *quadprog*. Once the solution  $\{\mu_g\}$  to the said program has been found, we choose its  $F$  largest elements and set them to 1 (and the rest to 0). Letting  $g_1 < g_2 < \dots < g_F$  denote the indexes of these  $F$  largest elements of  $\{\mu_g\}_{g=1}^G$ , we use the following sampling times

$$x_f = \tau_{g_f}, \quad f = 1, \dots, F \quad (78)$$

as an approximation of the solution to the minimization problem of (76). (See, e.g., [22] for a general discussion on this type of approximation of the solution to a Boolean optimization problem.)

Finally, we remark on the possible use of a **data taper**, for a given sampling pattern. LSP with a data taper can be viewed as a special case of WLS corresponding to a diagonal weighting matrix (whose diagonal elements are determined by the taper). Consequently, a well chosen taper may improve the performance of LSP. However, for RIAA this can hardly be the case, as this method is based on a general and theoretically optimal weighting matrix. Therefore, because the main emphasis here is on RIAA, we do not consider data tapers in this paper.

## V. NUMERICAL ILLUSTRATIONS

### A. Methods Not Included in the Present Study

The main goal of this section is to illustrate the differences in performance between LSP and RIAA, as well as between LSP+BIC and RIAA+BIC. The literature on the spectral analysis of nonuniform data contains several other methods that we do not consider in this performance study, and we explain here why.

<sup>1</sup>MATLAB is a registered trademark of The MathWorks, Inc., Natick, MA.

First, note that we have emphasized the case of sinusoidal data in this paper, and therefore methods designed for nonuniform data with continuous spectra, such as ARMA sequences, fall outside the scope of the paper.

Of the existing methods for nonuniform sinusoidal data, the CAP, MUSIC and ESPRIT methods introduced in [16] appear to be the closest in spirit to the RIAA proposed here (see the cited paper for explanations of the acronyms used to designate these methods). Indeed, all these methods make use of the estimated covariance matrix that is computed in the first iteration of RIAA from LSP. In fact, CAP (when used with the same covariance matrix dimension as RIAA) is essentially identical to the first iteration of RIAA. MUSIC and ESPRIT, on the other hand, are parametric methods that require a guess of the number of sinusoidal components present in the data, otherwise they cannot be used; furthermore, they make the assumption that the noise in the data sequence is white, and they may perform poorly if this assumption is not satisfied.

Several different versions of a cyclic algorithm for estimating the parameters of sinusoidal sequences by minimizing a nonlinear LS fitting criterion have been discussed, apparently in an independent manner, in [3], [4], [7], and [23]–[25] (see also [26] for a recent account). The method of this type introduced in [24] was called RELAX, **because cyclic algorithms are sometimes referred to as relaxation algorithms in the optimization literature**, and we will use this name in the following discussion to designate a generic algorithm in this class. RELAX is a powerful and usually quite accurate parametric method; in particular, note that RELAX coincides with the ML under the assumption that the noise is white and normally distributed, and that RELAX appears to be relatively insensitive to the violation of the said noise assumption. Being a parametric method, RELAX is thus comparable with the LSP+BIC and RIAA+BIC methods discussed in this paper. However, unlike the latter methods, RELAX does not provide any nonparametric (intermediate) estimate. In our opinion, this is a drawback: indeed, a parametric spectral model can never provide an exact description of real-world data and therefore it may suffer from robustness issues; on the other hand, a nonparametric spectral estimate (such as the one obtained with RIAA) is much more robust, as it does not suffer from the mis-modeling problems of a parametric method, and therefore it is often good to have a nonparametric estimate at least as a back-up for and cross-checking of a parametric estimation method.

Another class of methods, that have been recently proposed for solving the spectral analysis problem dealt with in this paper, make use of sparsity enforcing-based techniques (see, e.g., [16] and [27]). While these methods have a seemingly nonparametric flavor, in actuality they are essentially parametric and do not provide any (intermediate) nonparametric spectral estimate, unlike LSP+BIC and RIAA+BIC. Consequently, they suffer from the same drawback as the RELAX approach discussed in the previous paragraph.

### B. Simulated Data

We consider a data sequence consisting of  $M = 3$  sinusoidal components with frequencies 0.1, 0.4 and 0.41 Hz, and amplitudes 2, 4 and 5, respectively. The phases of the three sinusoids are independently and uniformly distributed over  $[0, 2\pi]$  and the

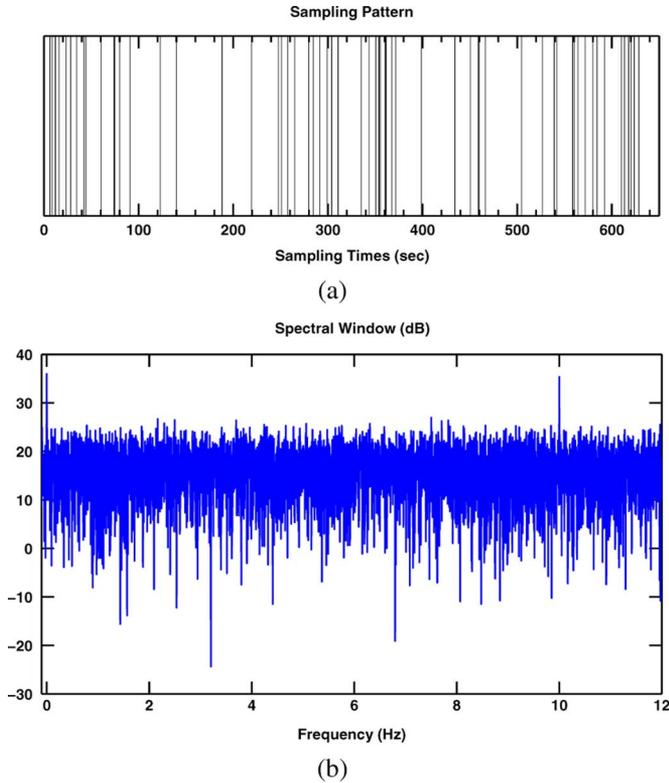


Fig. 1. Sampling pattern and spectral window for the simulated data case. (a) The sampling pattern used for all Monte Carlo trials in Figs. 2–4. The distance between two consecutive bars represents the sampling interval. (b) The corresponding spectral window.

← 加性噪声  
 additive noise is white and normally distributed with mean 0 and variance  $\sigma^2 = 0.01$ . We define the signal-to-noise ratio (SNR) of each sinusoid as

$$\text{SNR}_m = 10 \log_{10} \left( \frac{\alpha_m^2/2}{\sigma^2} \right) \text{ dB}, \quad m = 1, 2, 3 \quad (79)$$

where  $\alpha_m$  is the amplitude of the  $m$ th sinusoidal component. Hence, we have  $\text{SNR}_1 = 23$  dB,  $\text{SNR}_2 = 29$  dB and  $\text{SNR}_3 = 31$  dB in this example.

去尾 指数的  
 The sampling pattern follows a Poisson process with parameter  $\lambda = 0.1 \text{ s}^{-1}$ , that is, the sampling intervals are exponentially distributed with mean  $\mu = 1/\lambda = 10$  s (we round off the so-generated sampling times to ten decimals, see [12]). We choose  $N = 64$  and show the sampling pattern (which we will fix in all Monte Carlo trials) in Fig. 1(a). Note the highly irregular sampling intervals, which range from 0.2 to 51.2 s with mean value 9.3 s. Fig. 1(b) shows the spectral window corresponding to Fig. 1(a). The smallest frequency  $f_0 > 0$  at which the spectral window has a peak close to  $N^2$  is approximately 10 Hz. Therefore, according to (30),  $f_{\max} = f_0/2 \approx 5$  Hz. In the following plots, however, we zoom onto the interval  $f \in [0, 0.6]$  Hz that contains the three sinusoidal components of interest. The step of the frequency grid  $\Delta f$  is chosen as 0.005 Hz.

Fig. 2 presents the spectral estimates averaged over 100 independent realizations while Fig. 3 shows the overlapped estimates from the first 15 Monte Carlo trials. LSP nearly misses the

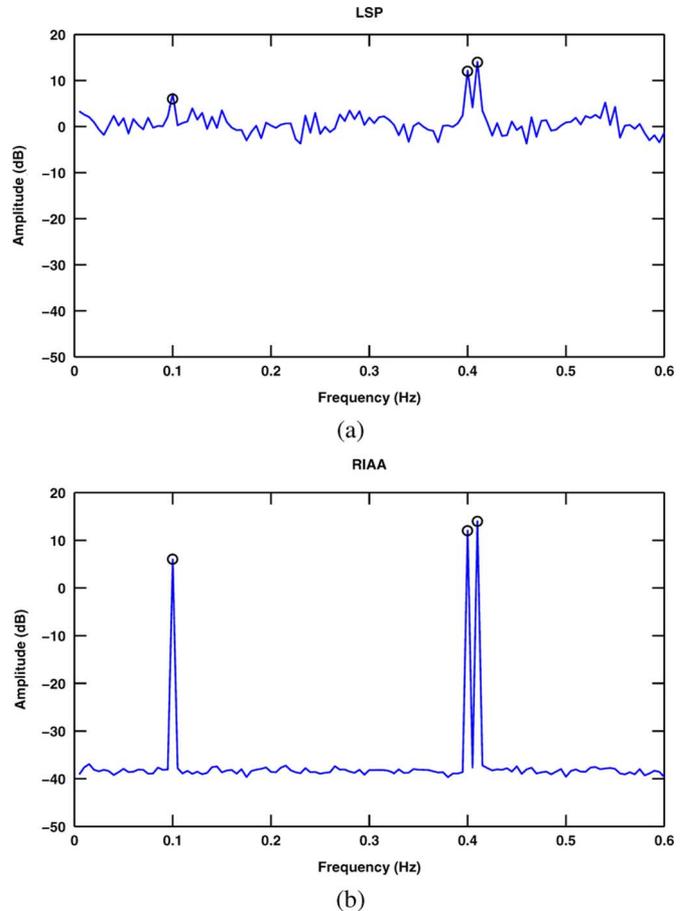


Fig. 2. Average spectral estimates from 100 Monte Carlo trials. The solid line is the estimated spectrum and the circles represent the true frequencies and amplitudes of the three sinusoids. (a) LSP and (b) RIAA.

smallest sinusoid while RIAA successfully resolves all three sinusoids. Note that RIAA suffers from much less variability than LSP from one trial to another.

Next, we use BIC to select the significant peaks of the spectra obtained by LSP and RIAA. We let  $M_{\text{BIC}}$  denote the number of sinusoids picked up by BIC. Then the probabilities of “correct detection,” of “false alarm” and of “miss” are defined as  $P_D = \text{Prob.}(M_{\text{BIC}} = 3)$ ,  $P_{\text{FA}} = \text{Prob.}(M_{\text{BIC}} > 3)$  and  $P_M = \text{Prob.}(M_{\text{BIC}} < 3)$ . Fig. 4 shows the scatter plots of the estimates for the dominant sinusoids obtained via using LSP+BIC and RIAA+BIC. The LSP amplitude estimates of the two closely-spaced sinusoids are biased and the smallest sinusoid is frequently missed. For LSP,  $P_D$  is only 0.03,  $P_{\text{FA}}$  is 0.00, and the missing probability  $P_M$  is as high as 0.97. Compared to LSP, RIAA shows much better stability and accuracy, as illustrated in Fig. 4(b). For RIAA,  $P_D = 0.96$ ,  $P_{\text{FA}} = 0.04$ , and  $P_M = 0.00$ . We also note that usually after 15 iterations RIAA’s performance does not improve visibly. So in the above and all subsequent examples, we terminate RIAA after 15 iterations.

Remark: In the case of a single sinusoidal signal in white Gaussian noise, the LSP is equivalent to the method of maximum likelihood and therefore it is asymptotically statistically efficient. Consequently, in this case, of academic rather than

渐近的 学术性的

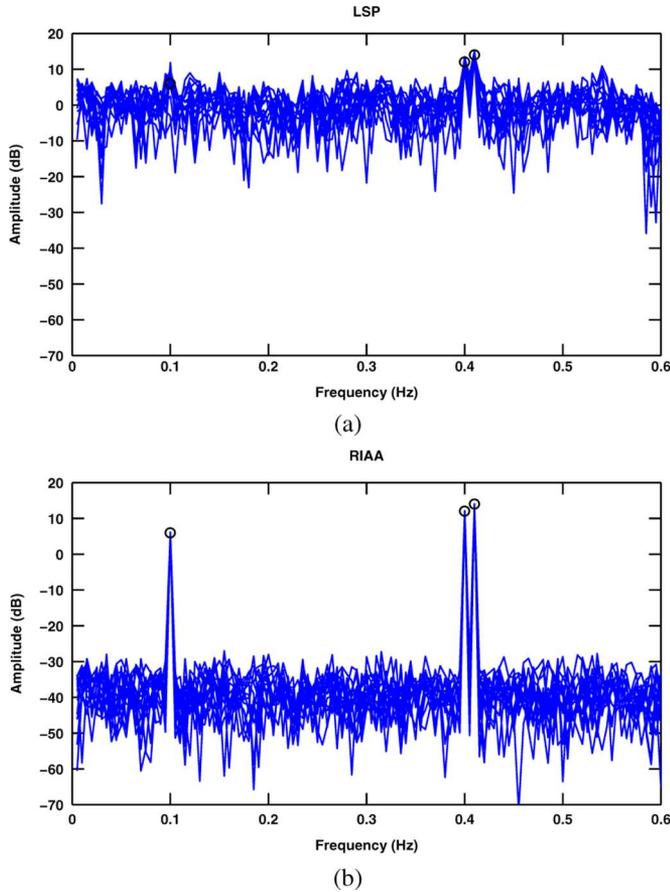


Fig. 3. Spectral estimates from 15 Monte Carlo trials. The solid lines are the estimated spectra and the circles represent the true frequencies and amplitudes of the three sinusoids. (a) LSP and (b) RIAA.

practical interest, LSP can be expected to outperform RIAA. In numerical examples not shown here (in the interest of brevity) we have observed that indeed in such a case LSP tends to be somewhat better than RIAA for relatively large values of  $N$  or SNR; however, we have also observed that, even under these conditions that are ideal for LSP, the performance of RIAA in terms of MSE (mean squared error) is slightly better (by a fraction of a dB) than that of LSP when  $N$  or SNR becomes smaller than a certain threshold. ■

### C. Sampling Pattern Design

Suppose we have already collected data at  $\{t_n\}_{n=1}^{N=38}$ , which are the sampling times of the astronomical data that will be analyzed in the next subsection (the unit has been changed here from ‘day’ to ‘second’ for simplicity). The sampling pattern and the corresponding spectral window are plotted in Fig. 5. Similarly to the discussion in the last subsection we get  $f_0 \approx 1$  Hz and therefore  $f_{\max} = f_0/2 \approx 0.5$  Hz. The set of sampling times of possible future measurements  $\{\tau_g\}_{g=1}^{G=200}$  are randomly selected from the interval  $[0, 200]$  s, which covers approximately the same period of time as  $\{t_n\}_{n=1}^{N=38}$ .

Consider the same data sequence as in the previous subsection, except that the amplitudes of the two closely-spaced sinusoids (at 0.4 and 0.41 Hz) are now set to 2 and 5, respectively.

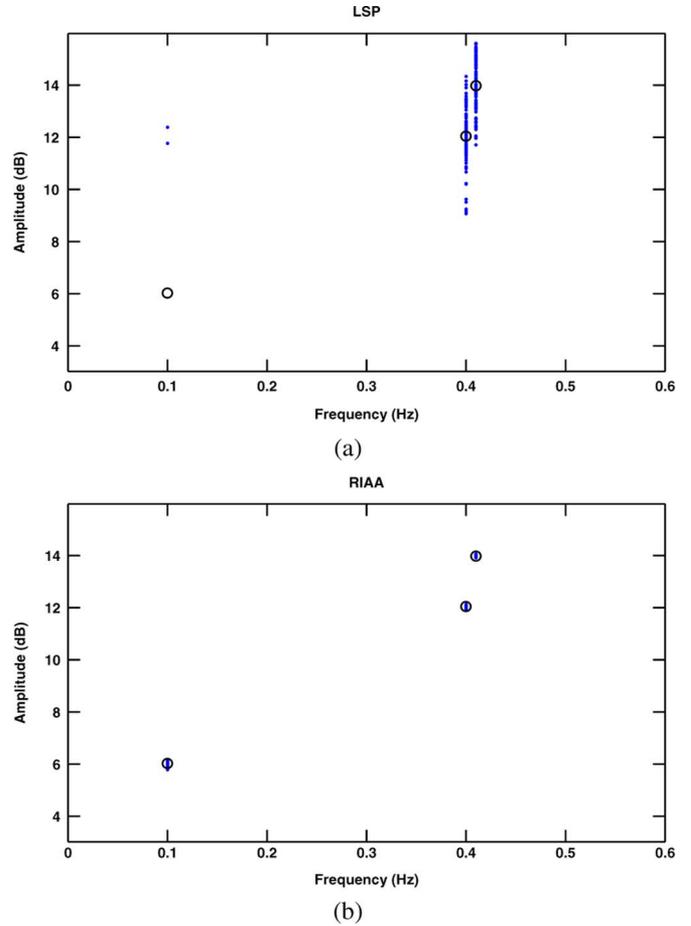


Fig. 4. Scatter plot of the dominant sinusoidal components selected by BIC in 100 Monte Carlo trials. Dots symbolize the estimates while circles represent the true frequencies and amplitudes of the three sinusoids. (a) LSP+BIC, and (b) RIAA+BIC.

实验 幅度估计有偏差

Fig. 6 shows the average estimated spectra and Fig. 7 shows the scatter plot. In most trials LSP+BIC finds only the strongest sinusoid and yields rather biased amplitude estimates. RIAA, on the other hand, has a much better detection rate, although the amplitude estimates are somewhat biased. The corresponding probabilities  $P_D$ ,  $P_{FA}$  and  $P_M$  are 0.01, 0.00 and 0.99 for LSP and 0.73, 0.01 and 0.26 for RIAA, respectively.

Next we use the optimal sampling design technique in Section IV to choose  $\{x_f\}_{f=1}^{F=38}$  from  $\{\tau_g\}_{g=1}^{G=200}$ . In (76) the  $\{P(\omega_k)\}_{k=1}^K$  are equal to the RIAA spectrum estimates obtained from the given 38 samples and  $N$  is chosen as 0 because we want to achieve better estimation from only the designed samples (for comparison with the accuracy achieved from the given samples). Fig. 8(a) shows the designed sampling pattern. Fig. 8(b) shows the corresponding spectral window, from which we can infer a much larger  $f_{\max}$  than 0.5 Hz [cf. Fig. 5(b)]. In the subsequent simulation with the designed sampling pattern, however, we still choose  $f_{\max} = 0.5$  Hz for comparison with the results obtained using the given sampling pattern.

With the data sequence sampled at  $\{x_f\}_{f=1}^{F=38}$ , we obtain the estimated spectra shown in Fig. 9 and the scatter plots in Fig. 10. The new probabilities  $P_D$ ,  $P_{FA}$ , and  $P_M$  are 0.05, 0.00, and 0.95 for LSP; and 0.94, 0.06, and 0.00 for RIAA. All three sinusoidal

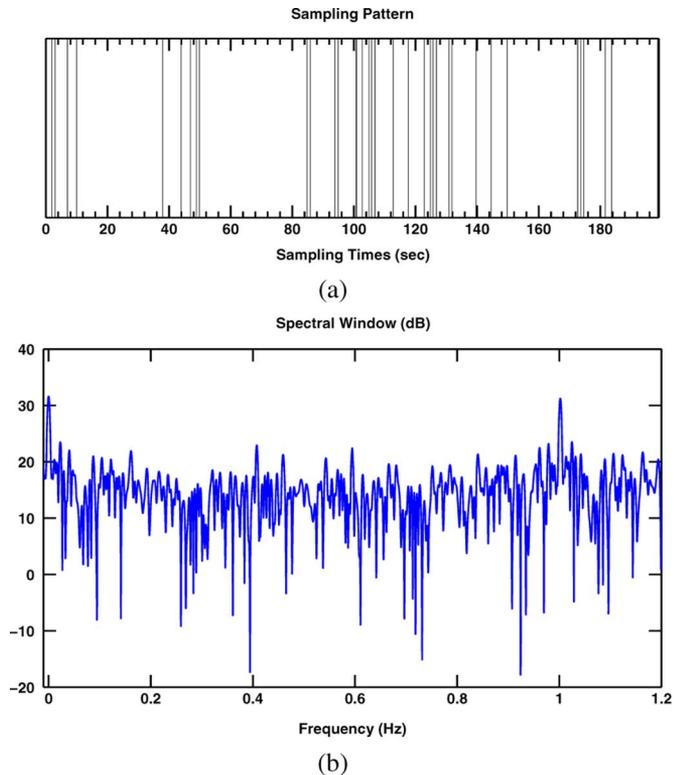


Fig. 5. Sampling pattern and spectral window for the given sampling times in the sampling design example. (a) Sampling pattern used for all Monte Carlo trials in Figs. 6 and 7. The distance between two consecutive bars represents the sampling interval. (b) Corresponding spectral window.

components are now clearly resolved by RIAA and the amplitude estimates are also quite accurate.

#### D. Radial Velocity Data

Finally, we present an example based on measured astronomical data, whose original analysis led to the discovery of an **extrasolar planet** named ET-1 (see [28] and the references therein). **Fig. 11** shows the data. The sampling pattern and the corresponding spectral window have been shown in Fig. 5 (with the unit changed from “day” to “second”). Note the highly-irregular sampling intervals which range from 0.01 to 34.91 days during a total of 198.61 days of observation. Because  $f_{\max} \approx 0.5$  cycles/day, we will be concerned with the frequency interval  $[0.01, 0.5]$  cycles/day (corresponding to periods between 2 and 100 days).

Fig. 12 shows the spectra estimated by LSP and RIAA. Although LSP and RIAA give almost the same **dominant peak**, the relatively large sidelobe levels of LSP are not desirable. The **dominant peak** selected by RIAA+BIC is located at 0.2439 Hz (4.1 days), which is **consistent** with the result obtained in [28] by a **relatively involved procedure**. Next we **synthesize** the data using the sinusoidal component picked up by BIC. Fig. 13 compares the original data sequence and the synthesized waveform. The synthesized sinusoid fits the original samples well with a normalized mean-squared error less than 6%.

To demonstrate the advantages of RIAA over LSP, we next use only the first 16 of the 38 available samples in the data set.

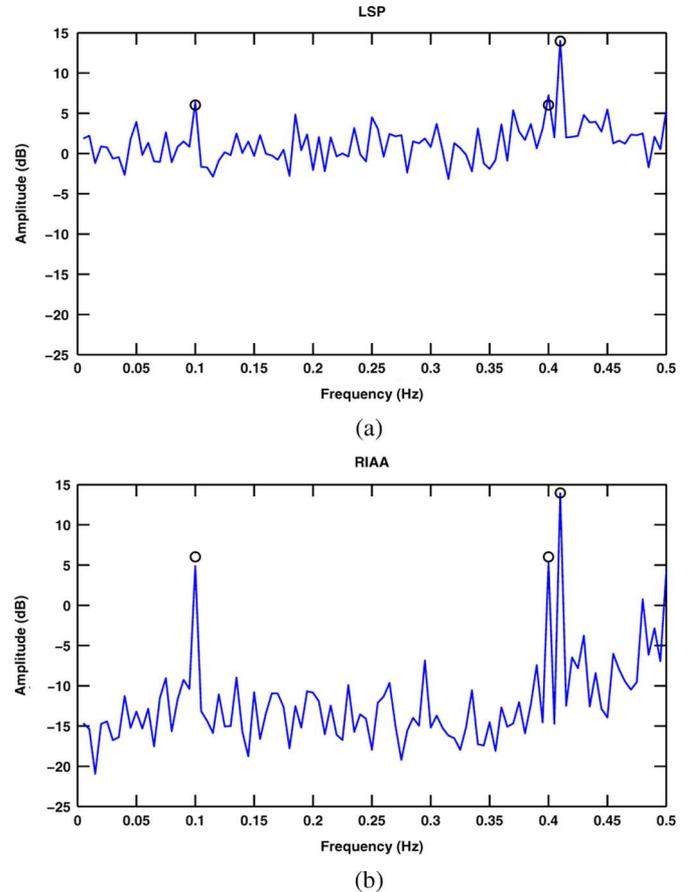


Fig. 6. Average spectral estimates from 100 Monte Carlo trials, using given 38 sampling times. Solid line is the estimated spectrum and the circles represent the true frequencies and amplitudes of the three sinusoids. (a) LSP and (b) RIAA.

The estimated spectra are plotted in Fig. 14. In this case, the LSP spectral estimate has large spurious peaks and the dominant peak occurs at about 0.05 Hz. In contrast with this, RIAA still succeeds to find the same dominant peak as before. It is worth mentioning that the first 16 samples cover an observation period of 100.87 days, which is about one half of the total observation time. This implies that the observation time can be significantly shortened via the use of RIAA.

*Remark:* We have also computed the LSP and RIAA spectral estimates for other segments of the radial velocity data set, not only for the segment made from the first 16 samples. While we omit the corresponding plots, for brevity, we note that in all cases we have considered the dominant peak of the RIAA spectrum is located close to the frequency of 0.24 cycles/day [exactly as in Figs. 12(b) and 14(b)], whereas the location of the largest peak of the LSP spectrum varies significantly and is rather often quite far from 0.24 cycles/day. ■

Finally we note that the sampling times of the radial velocity data, used in this subsection, are an example of what we call “pathological sampling patterns.” For such patterns, the matrix  $\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_2 \ \cdots \ \mathbf{A}_K]$  (where  $\{\mathbf{A}_k\}_{k=1}^K$  are given by (34)) is nearly rank-deficient, which leads to an ill-conditioned matrix  $\mathbf{\Gamma}$  (see (42)) regardless of the values of elements in  $\{\mathbf{D}_k\}_{k=1}^K$ . To avoid the problems induced by an ill-conditioned  $\mathbf{\Gamma}$ , we have modified RIAA as follows (only in this subsection).

一致的

太阳系外行星

合成

缺陷

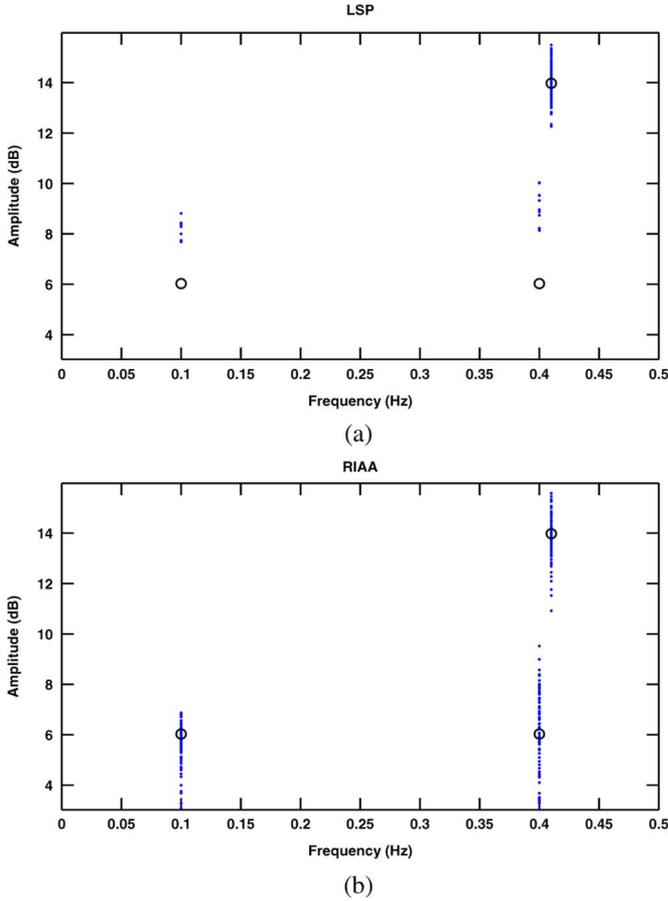


Fig. 7. Scatter plot of the dominant sinusoidal components selected by BIC in 100 Monte Carlo trials, using the given 38 sampling times. Dots symbolize the estimates while circles represent the true frequencies and amplitudes of the three sinusoids. (a) LSP+BIC, and (b) RIAA+BIC.

Let  $\lambda_1, \lambda_2, \dots, \lambda_N$  denote the **eigenvalues** of  $\mathbf{A}\mathbf{A}^T$  (arranged in a decreasing order), and let  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N$  denote the corresponding **eigenvectors**. Then we define a **transformation** matrix  $\mathbf{U}$  as  $[\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_n]$  where  $n$  is the largest integer in  $[1, N]$  for which  $\lambda_1/\lambda_n \leq 50$ . Finally, we replace the  $\mathbf{\Gamma}$ ,  $\mathbf{A}_k$  and  $\mathbf{y}$  in (45) by  $\mathbf{U}^T \mathbf{\Gamma} \mathbf{U}$ ,  $\mathbf{U}^T \mathbf{A}_k$  and  $\mathbf{U}^T \mathbf{y}$ , respectively. The above modification reduces the condition number of  $\mathbf{A}\mathbf{A}^T$  and therefore ensures the **stability** of RIAA.

## VI. CONCLUSION

We began by revisiting the periodogram to explain why arguably the plain least-squares periodogram (LSP) is preferable to the “classical” Fourier periodogram, from a data-fitting viewpoint, as well as to the frequently-used form of LSP due to Lomb and Scargle, from a computational standpoint. Then we went on to introduce a new enhanced method for spectral analysis of nonuniformly sampled data sequences. The new method can be interpreted as an iteratively weighted LSP that makes use of a data-dependent weighting matrix built from the most recent spectral estimate. Because this method was derived for the case of real-valued data (which is typically more complicated to deal

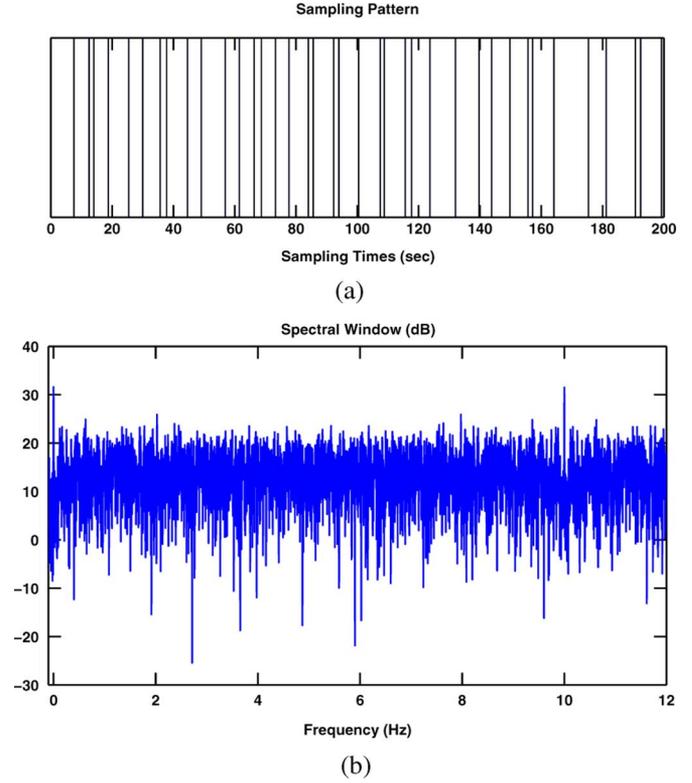


Fig. 8. Sampling pattern and spectral window for the designed sampling times in the sampling design example. (a) Sampling pattern used for all Monte Carlo trials in Figs. 9 and 10. The distance between two consecutive bars represents the sampling interval. (b) Corresponding spectral window.

with in spectral analysis than the complex-valued data case), it is iterative and it makes use of an adaptive (i.e., data-dependent) weighting, we referred to it as the real-valued iterative adaptive approach (RIAA).

LSP and RIAA are nonparametric methods that can be used for the spectral analysis of general data sequences with both continuous and discrete spectra. However, they are most suitable for data sequences with discrete spectra (i.e., sinusoidal data), which is the case we emphasized in this paper. For the latter type of data, we presented a procedure for obtaining a parametric spectral estimate, from the LSP or RIAA nonparametric estimate, by means of a Bayesian information criterion (BIC). The use of BIC for the said purpose can be viewed as a way of testing the significance of the dominant peaks of the LS or RIAA periodograms, a problem for which there was hardly any satisfactory solution available.

Finally, we discussed a possible strategy for designing the sampling pattern of future measurements, based on the spectral estimate obtained from the already available observations. To compare the performances of LSP and RIAA, as well as of their parametric versions proposed in the paper, and to illustrate the type of results that can be obtained with the sampling pattern design strategy presented here, we used both simulated and measured data (the latter were radial velocity data collected for extrasolar planet search).

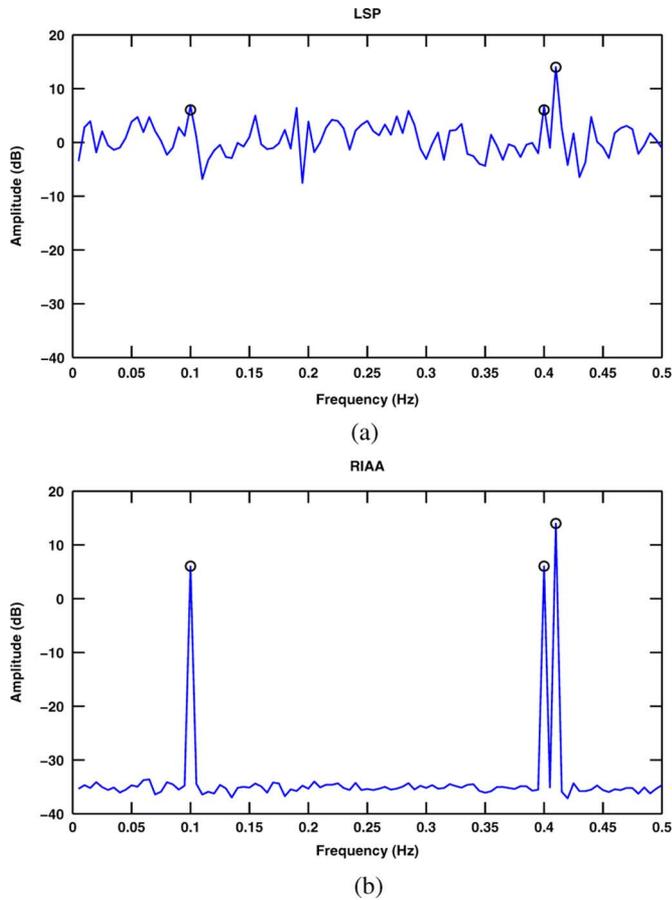


Fig. 9. Average spectral estimates from 100 Monte Carlo trials, using the designed 38 sampling times. The solid line is the estimated spectrum and the circles represent the true frequencies and amplitudes of the three sinusoids. (a) LSP and (b) RIAA.

#### APPENDIX I EXPLANATION OF SOME TERMINOLOGY

We say that a sampling pattern  $\{t_n\}_{n=1}^N$  is nonpathological if the matrix

$$\int_0^{\omega_{\max}} \mathbf{A}(\omega) \mathbf{A}^T(\omega) d\omega \quad (80)$$

where

$$\mathbf{A}(\omega) = \begin{bmatrix} \cos(\omega t_1) & \sin(\omega t_1) \\ \vdots & \vdots \\ \cos(\omega t_N) & \sin(\omega t_N) \end{bmatrix} \quad (81)$$

is reasonably well conditioned (otherwise the sampling pattern is said to be pathological). We define the conditioning number of a matrix  $\mathbf{\Omega}$  in the following standard way:

$$\text{cond}(\mathbf{\Omega}) = \frac{\lambda_{\max}(\mathbf{\Omega})}{\lambda_{\min}(\mathbf{\Omega})} \quad (82)$$

where  $\lambda_{\max}(\mathbf{\Omega})$  and  $\lambda_{\min}(\mathbf{\Omega})$  denote the maximum eigenvalue and, respectively, the minimum eigenvalue of  $\mathbf{\Omega}$ . Condition numbers smaller than  $10^3$  are well tolerated by common linear algebra software, and the corresponding matrices can be considered to be relatively well conditioned.

代数学

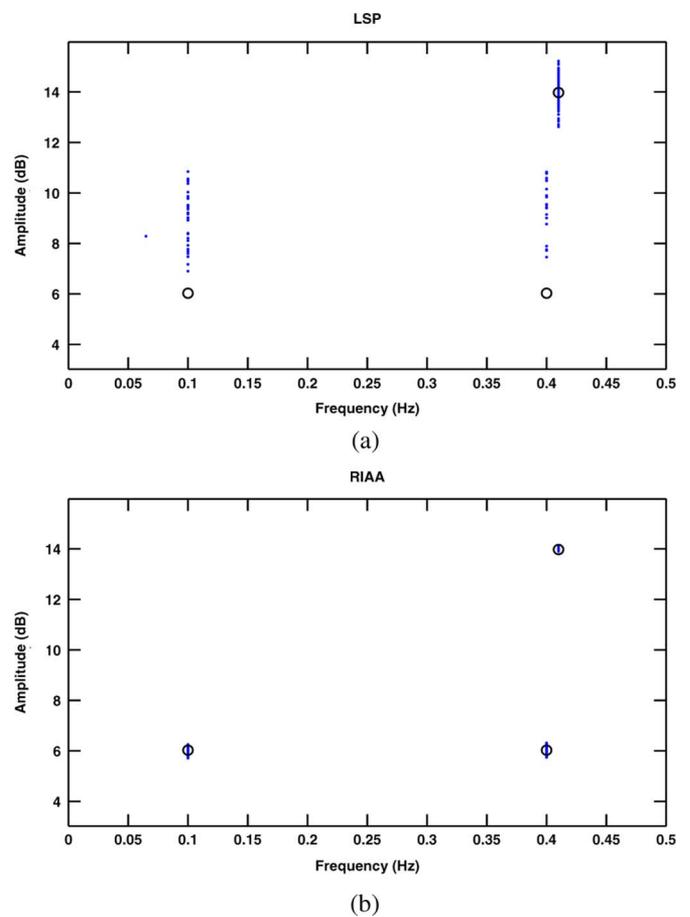


Fig. 10. Scatter plot of the dominant sinusoidal components selected by BIC in 100 Monte Carlo trials, using the designed 38 sampling times. Dots symbolize the estimates while circles represent the true frequencies and amplitudes of the three sinusoids. (a) LSP+BIC, and (b) RIAA+BIC.

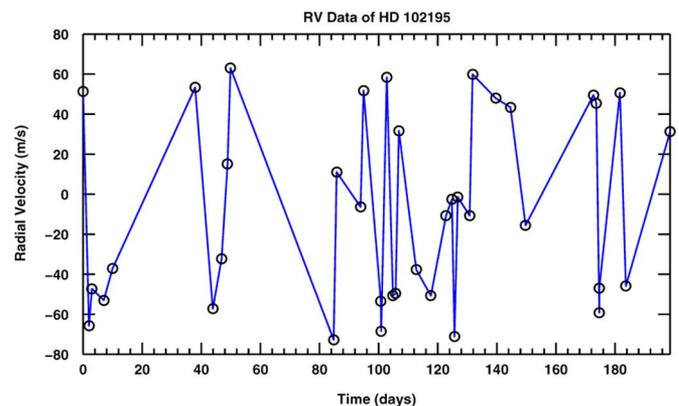


Fig. 11. Radial velocity data for HD 102195, which is the star that ET-1 orbits. Circles represent the samples.

乘法的, 增加

Note that the matrix in (80) is well approximated (to within a multiplicative constant) by the matrix  $\mathbf{\Gamma}$  in (42) with  $\mathbf{D}_p = \mathbf{I}$  ( $p = 1, \dots, K$ ), that is:  $\sum_{p=1}^K \mathbf{A}_p \mathbf{A}_p^T$ . This observation relates the present definition of a nonpathological sampling pattern to that mentioned briefly in Section V-D (indeed,  $\text{cond}(\sum_{p=1}^K \mathbf{A}_p \mathbf{A}_p^T) = \text{cond}(\mathbf{A} \mathbf{A}^T) = [\text{cond}(\mathbf{A})]^2$ , where

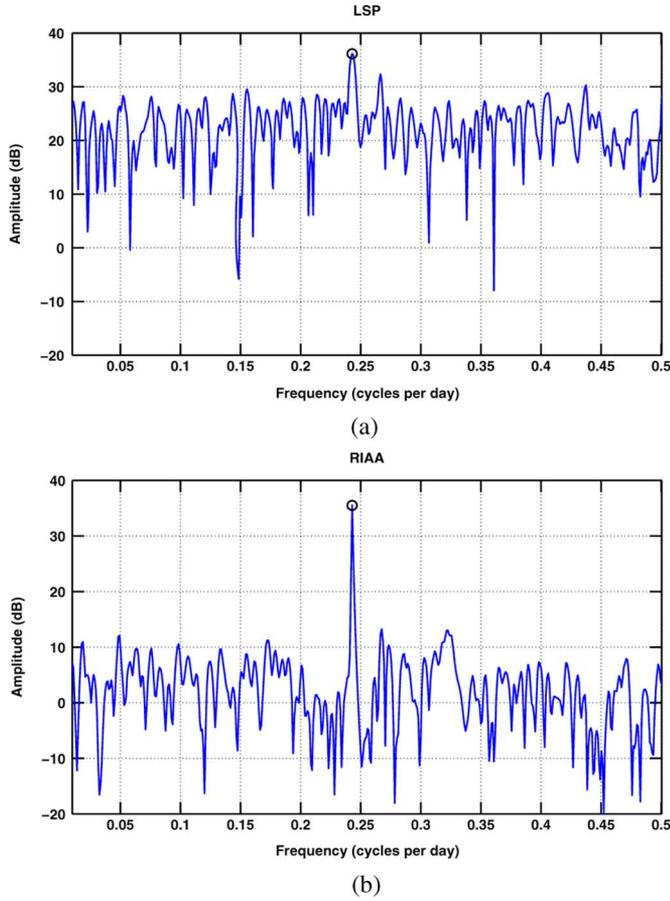


Fig. 12. Spectral estimates for the radial velocity data of HD 102195. The solid line is the estimated spectrum and the circle is the dominant peak selected by BIC ( $M_{BIC} = 1$  for both LSP and RIAA). (a) LSP and (b) RIAA.

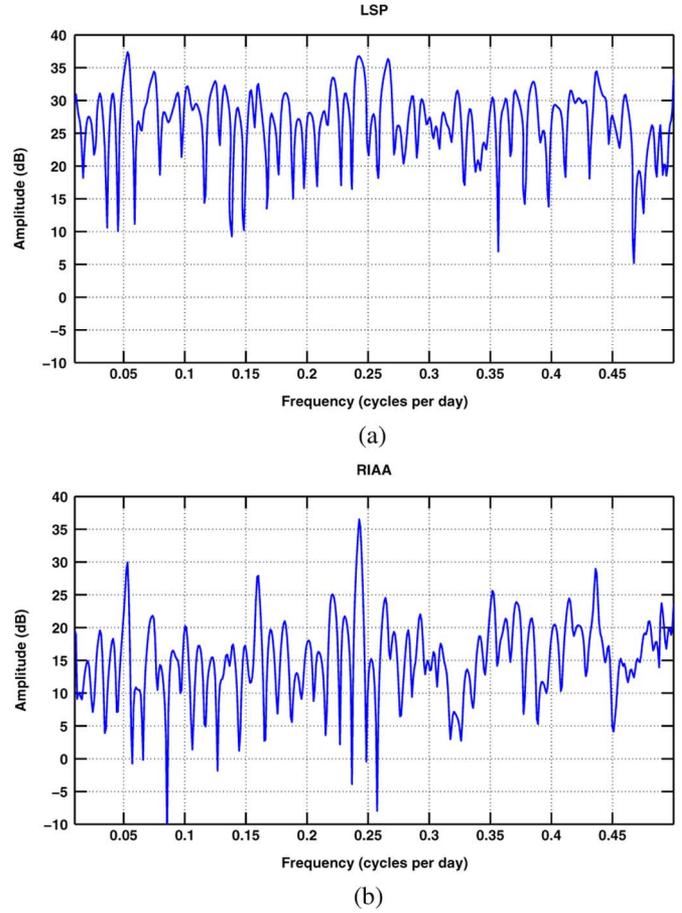


Fig. 14. Spectral estimates obtained using only the first 16 of the 38 samples in the radial velocity data set. (a) LSP and (b) RIAA.

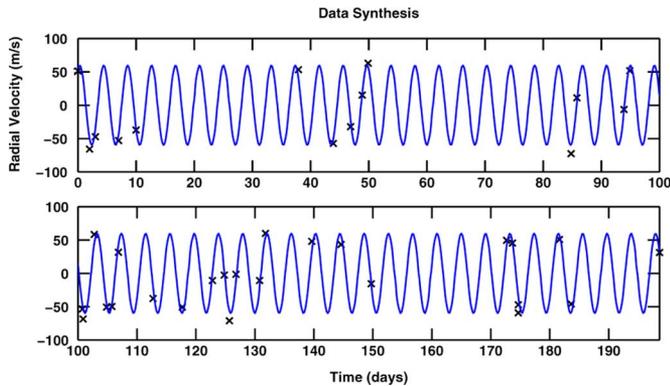


Fig. 13. Reconstructed data using the dominant sinusoidal component determined by RIAA+BIC. The “x” marks represent the original data sequence and the solid line is the synthesized waveform.

the matrix  $\mathbf{A}$  has been defined in Section V-D). For the radial velocity data, we have

$$\text{cond} \left( \sum_{p=1}^K \mathbf{A}_p \mathbf{A}_p^T \right) \approx 32646 \quad (83)$$

and therefore the corresponding sampling pattern can be considered to be pathological.

Finally, we remark on the fact that the matrix in (80) can be viewed as the real-valued counterpart of the complex-valued matrix in (71).

副本 ← 致谢  
ACKNOWLEDGMENT

We would like to thank Dr. J. Ge for providing us with the measured radial velocity data used for extrasolar planet detection, and Dr. P. Tichavsky for his critical comments on a former version of this paper.

版本 ← 评论

#### REFERENCES

- [1] M. B. Priestley, *Spectral Analysis and Time Series, Volume 1: Univariate Series.* New York: Academic, 1981.
- [2] P. Stoica and R. L. Moses, *Spectral Analysis of Signals.* Upper Saddle River, NJ: Prentice-Hall, 2005.
- [3] F. J. M. Barning, “The numerical analysis of the light-curve of 12 lacertae,” *Bull. Astronom. Inst. Netherlands*, vol. 17, no. 1, pp. 22–28, Aug. 1963.
- [4] P. Vanicek, “Approximate spectral analysis by least-squares fit,” *Astrophys. Space Sci.*, vol. 4, no. 4, pp. 387–391, Aug. 1969.
- [5] P. Vanicek, “Further development and properties of the spectral analysis by least-squares,” *Astrophys. Space Sci.*, vol. 12, no. 1, pp. 10–33, Jul. 1971.
- [6] N. R. Lomb, “Least-squares frequency analysis of unequally spaced data,” *Astrophys. Space Sci.*, vol. 39, no. 2, pp. 447–462, Feb. 1976.
- [7] S. Ferraz-Mello, “Estimation of periods from unequally spaced observations,” *Astronom. J.*, vol. 86, no. 4, pp. 619–624, Apr. 1981.
- [8] J. D. Scargle, “Studies in astronomical time series analysis. II—Statistical aspects of spectral analysis of unevenly spaced data,” *Astrophys. J.*, vol. 263, pp. 835–853, Dec. 1982.

- [9] W. H. Press and G. B. Rybicki, "Fast algorithm for spectral analysis of unevenly sampled data," *Astrophys. J.*, vol. 338, pp. 277–280, Mar. 1989.
- [10] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 560–574, Feb. 2003.
- [11] N. Nguyen and Q. Liu, "The regular Fourier matrices and nonuniform fast Fourier transforms," *SIAM J. Sci. Comput.*, vol. 21, no. 1, pp. 283–293, 2000.
- [12] L. Eyer and P. Bartholdi, "Variable stars: Which Nyquist frequency?," *Astron. Astrophys., Supp. Series*, vol. 135, pp. 1–3, Feb. 1999.
- [13] F. A. M. Frescura, C. A. Engelbrecht, and B. S. Frank, "Significance tests for periodogram peaks." [Online]. Available: <http://adsabs.harvard.edu/abs/2007arXiv0706.2225F>, E-Print arXiv:0706.2225, provided by the SAO/NASA Astrophysics Data System, Jun. 2007
- [14] P. Reegen, "SigSpec—I. Frequency- and phase-resolved significance in Fourier space," *Astron. Astrophys.*, vol. 467, pp. 1353–1371, Mar. 2007.
- [15] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1469–1484, Jun. 1996.
- [16] P. Stoica and N. Sandgren, "Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches," *Digit. Signal Process.*, vol. 16, no. 6, pp. 712–734, Nov. 2006.
- [17] T. Yardibi, M. Xue, J. Li, P. Stoica, and A. B. Baggeroer, "Iterative adaptive approach for sparse signal representation with sensing applications," *IEEE Trans. Aerosp. Electron. Syst.*, 2007, submitted for publication.
- [18] T. Yardibi, J. Li, and P. Stoica, "Nonparametric and sparse signal representations in array processing via iterative adaptive approaches," presented at the 42nd Asilomar Conf. Signals, Systems, Computers, Pacific Grove, CA, Oct. 2008.
- [19] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
- [20] E. S. Saunders, T. Naylor, and A. Allan, "Optimal placement of a limited number of observations for period searches," *Astron. Astrophys.*, vol. 455, pp. 757–763, May 2006.
- [21] P. Stoica, H. Li, and J. Li, "A new derivation of the APES filter," *IEEE Signal Process. Lett.*, vol. 6, no. 8, pp. 205–206, Aug. 1999.
- [22] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [23] D. F. Gray and K. Desikachary, "A new approach to periodogram analysis," *Astrophys. J.*, vol. 181, pp. 523–530, Apr. 1973.
- [24] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, Feb. 1996.
- [25] Y. I. Abramovich, "Compensation methods of resolution of wideband signals," *Radio Eng. Electron. Phys.*, vol. 23, pp. 54–59, Jan. 1978.
- [26] J. Ling, P. Stoica, J. Li, and Y. Abramovich, "On using cyclic algorithms for sinusoidal parameter estimation," *Electron. Lett.*, vol. 44, no. 19, pp. 1160–1161, Nov. 2008.
- [27] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 575–585, Dec. 2007.
- [28] J. Ge, J. van Eyken, S. Mahadevan, C. DeWitt, S. R. Kane, R. Cohen, A. V. Heuvel, S. W. Fleming, and P. Guo, "The first extrasolar planet discovered with a new-generation high-throughput Doppler instrument," *Astrophys. J.*, vol. 648, pp. 683–695, Sep. 2006.



**Petre Stoica** (F'94) received the D.Sc. degree in automatic control from the Polytechnic Institute of Bucharest (BPI), Bucharest, Romania, in 1979 and an honorary doctorate degree in science from Uppsala University (UU), Uppsala, Sweden, in 1993.

He is a Professor of systems modeling with the Division of Systems and Control, the Department of Information Technology at UU. Previously, he was a Professor of system identification and signal processing with the Faculty of Automatic Control and Computers at BPI. He held longer visiting positions

with Eindhoven University of Technology, Eindhoven, The Netherlands; Chalmers University of Technology, Gothenburg, Sweden (where he held a Jubilee Visiting Professorship); UU; the University of Florida, Gainesville, FL; and Stanford University, Stanford, CA. His main scientific interests are in the areas of system identification, time series analysis and prediction, statistical

signal and array processing, spectral analysis, wireless communications, and radar signal processing. He has published nine books, ten book chapters, and some 500 papers in archival journals and conference records. The most recent book he coauthored, with R. Moses, is *Spectral Analysis of Signals* (Prentice-Hall, 2005).

Dr. Stoica is on the editorial boards of six journals: the *Journal of Forecasting*; *Signal Processing*; *Circuits, Signals, and Signal Processing*; *Digital Signal Processing: A Review Journal*; *Signal Processing Magazine*; and *Multidimensional Systems and Signal Processing*. He was a Co-Guest Editor for several special issues on system identification, signal processing, spectral analysis, and radar for some of the aforementioned journals, as well as for *IEE Proceedings*. He was corecipient of the IEEE ASSP Senior Award for a paper on statistical aspects of array signal processing. He was also recipient of the Technical Achievement Award of the IEEE Signal Processing Society. In 1998, he was the recipient of a Senior Individual Grant Award of the Swedish Foundation for Strategic Research. He was also co-recipient of the 1998 EURASIP Best Paper Award for Signal Processing for a work on parameter estimation of exponential signals with time-varying amplitude, a 1999 IEEE Signal Processing Society Best Paper Award for a paper on parameter and rank estimation of reduced-rank regression, a 2000 IEEE Third Millennium Medal, and the 2000 W. R. G. Baker Prize Paper Award for a paper on maximum likelihood methods for radar. He was a member of the international program committees of many topical conferences. From 1981 to 1986, he was a Director of the International Time-Series Analysis and Forecasting Society, and he was also a member of the IFAC Technical Committee on Modeling, Identification, and Signal Processing. He is also a member of the Royal Swedish Academy of Engineering Sciences, an honorary member of the Romanian Academy, and a fellow of the Royal Statistical Society.



**Jian Li** (S'87–M'91–SM'97–F'05) received the M.Sc. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 1987 and 1991, respectively.

From April 1991 to June 1991, she was an Adjunct Assistant Professor with the Department of Electrical Engineering, The Ohio State University, Columbus. From July 1991 to June 1993, she was an Assistant Professor with the Department of Electrical Engineering, University of Kentucky, Lexington. Since August 1993, she has been with the Department of

Electrical and Computer Engineering, University of Florida, Gainesville, where she is currently a Professor. In fall 2007, she was on sabbatical leave at MIT, Cambridge, MA. Her current research interests include spectral estimation, statistical and array signal processing, and their applications.

Dr. Li is a Fellow of IET. She is a member of Sigma Xi and Phi Kappa Phi. She received the 1994 National Science Foundation Young Investigator Award and the 1996 Office of Naval Research Young Investigator Award. She was an Executive Committee Member of the 2002 International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, May 2002. She was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2005, an Associate Editor of the *IEEE Signal Processing Magazine* from 2003 to 2005, and a member of the Editorial Board of *Signal Processing*, a publication of the European Association for Signal Processing (EURASIP), from 2005 to 2007. She has been a member of the Editorial Board of *Digital Signal Processing—A Review Journal*, a publication of Elsevier, since 2006. She is presently a member of the Sensor Array and Multichannel (SAM) Technical Committee of IEEE Signal Processing Society. She is a coauthor of the papers that have received the First and Second Place Best Student Paper Awards, respectively, at the 2005 and 2007 Annual Asilomar Conferences on Signals, Systems, and Computers in Pacific Grove, CA. She is also a coauthor of the paper that has received the M. Barry Carlton Award for the best paper published in IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS in 2005.



**Hao He** (S'08) received the B.Sc. degree in electrical engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2007. He is currently working towards the Ph.D. degree with the Department of Electrical Engineering at the University of Florida, Gainesville, FL.

His research interests are in the areas of spectral estimation and radar signal processing.