Signal Processing 90 (2010) 2133–2140

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A robust periodogram for high-resolution spectral analysis

Ta-Hsin Li

Department of Mathematical Sciences, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598-0218, USA

ARTICLE INFO

Article history:
Received 16 October 2009
Received in revised form
15 January 2010
Accepted 16 January 2010
Available online 1 February 2010

Keywords:
Frequency estimation
Harmonic
Heavy tail
Least absolute deviations
Non-Gaussian
Nonlinear
Outlier
Periodogram
Regression
Robust
Sinusoid
Spectrum

ABSTRACT

Periodogram is an important tool for analyzing time series of mixed spectra that can be decomposed as sinusoids plus noise. While effective in many situations, the ordinary periodogram has two major shortcomings: it cannot resolve sinusoids whose frequencies are separated by less than 1 cycle per unit time; it does not possess sufficient robustness against heavy-tailed noise such as outliers. An alternative periodogram is introduced in this article with the aim of improving the frequency resolution as well as the robustness of the ordinary periodogram. The new periodogram, called bivariate ℓ_1−periodogram, is derived from the maximum likelihood method of multiple frequency estimation under the assumption of Laplace white noise. The desired high-resolution and robustness property of the bivariate ℓ_1−periodogram is confirmed by simulation studies. Superior statistical efficiency over alternative methods is also demonstrated.

1. Introduction

A time series of mixed spectrum is a random sequence that can be decomposed as sinusoids plus noise. The Fourier-transform-based periodogram is a widely used tool for analyzing such time series. Important applications of periodogram include detection of hidden periodicities and estimation of unknown sinusoidal parameters (amplitude and frequency) [1]. Periodogram analysis often yields satisfactory results. For example, it is well known that very accurate frequency estimates of the sinusoidal components can be obtained from the local maxima of a periodogram [2]. However, in order to obtain such results, two critical preconditions have to be satisfied. One condition is that the frequencies of the sinusoids must be well separated. Indeed, it is well known that the periodogram cannot resolve sinusoidal frequencies that differ by less than 1 cycle per unit time. Another condition is that the noise must be free of outliers because the periodogram lacks the necessary robustness against outlier contamination [3].

To overcome these obstacles, we introduce a new periodogram, called bivariate ℓ_1−periodogram, that has demonstrable advantages over the traditional periodogram in resolving closely spaced frequencies with desired robustness. The bivariate ℓ_1−periodogram is derived from the maximum likelihood method for estimating the frequencies of two sinusoids in Laplace white noise. It is the counterpart of a similar periodogram, called bivariate ℓ_2−periodogram, derived under the assumption of Gaussian white noise as a natural extension of the traditional univariate periodogram. The bivariate ℓ_2−periodogram has higher resolution than the univariate periodogram but still lacks robustness due to the inherit Gaussian assumption. The bivariate ℓ_1−periodogram provides both high resolution and robustness, thanks to its root in the least-absolute-deviations (LAD) criterion

E-mail address: thl@us.ibm.com

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.01.012

instead of the least-squares (LS) criterion. Related works on the LAD method for spectral analysis can be found in [3–6]. Some computational issues of the LAD method are discussed in [7–9], for example. Note that the bivariate periodograms discussed in this paper should not be confused with the periodogram–coperiodogram matrix of vector-valued time series or the biperiodogram based on the third-order cumulants of scalar time series.

The rest of the paper is organized as follows. Section 2 discusses the Gaussian maximum likelihood method of frequency estimation and the corresponding bivariate ℓ_2-periodogram. Section 3 introduces the bivariate ℓ_1-periodogram based on the Laplace maximum likelihood approach and demonstrates its robustness and high resolution property for spectral analysis. Concluding remarks are given in Section 4.

2. Bivariate ℓ_2-periodogram

Consider the problem of estimating the frequencies of two complex sinusoids in Gaussian white noise from a data record $\mathbf{y} := [y_1, \ldots, y_n]^T$ of length n that satisfies

$$\mathbf{y} = \mathbf{F}(\boldsymbol{\omega}_0)\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}. \tag{1}$$

In this expression, $\boldsymbol{\beta}_0 \in \mathbb{C}^2$ and $\boldsymbol{\omega}_0 \in \Omega := (0, 2\pi) \times (0, 2\pi) \backslash \{(\omega_1, \omega_2) : \omega_1 = \omega_2\}$ are unknown constants, $\mathbf{F}(\boldsymbol{\omega}) := [\mathbf{f}(\omega_1), \mathbf{f}(\omega_2)]$ is an n-by-2 matrix with $\boldsymbol{\omega} := (\omega_1, \omega_2)$ and $\mathbf{f}(\omega) := [\exp(i\omega), \ldots, \exp(in\omega)]^T \in \mathbb{C}^n$, and $\boldsymbol{\varepsilon} \in \mathbb{C}^n$ is the noise vector comprising i.i.d. complex Gaussian random variables with mean zero and unknown variance $\sigma^2 > 0$. Under the Gaussian assumption, the maximum likelihood estimator of the sinusoidal parameter $(\boldsymbol{\beta}_0, \boldsymbol{\omega}_0)$ is given by

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\omega}}) := \arg \min_{\boldsymbol{\beta} \in \mathbb{C}^2, \boldsymbol{\omega} \in \Omega} \|\mathbf{y} - \mathbf{F}(\boldsymbol{\omega})\boldsymbol{\beta}\|^2, \tag{2}$$

where $\|\cdot\|$ denotes the ℓ_2-norm of complex vectors, i.e., the square root of the sum of squared real and imaginary parts of all components.

The joint optimization problem in (2) can be solved by first optimizing $\boldsymbol{\beta}$ for fixed $\boldsymbol{\omega}$. In doing so, the Gaussian maximum likelihood (GML) frequency estimator $\hat{\boldsymbol{\omega}}$ in (2) can be expressed as

$$\hat{\boldsymbol{\omega}} = \arg \max_{\boldsymbol{\omega} \in \Omega} \{\|\mathbf{y}\|^2 - \|\mathbf{y} - \mathbf{F}(\boldsymbol{\omega})\hat{\boldsymbol{\beta}}(\boldsymbol{\omega})\|^2\}, \tag{3}$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{\omega}) := \{\mathbf{F}^H(\boldsymbol{\omega})\mathbf{F}(\boldsymbol{\omega})\}^{-1}\mathbf{F}^H(\boldsymbol{\omega})\mathbf{y}$ is the least-squares (LS) solution that minimizes $\|\mathbf{y} - \mathbf{F}(\boldsymbol{\omega})\boldsymbol{\beta}\|^2$ with respect to $\boldsymbol{\beta} \in \mathbb{C}^2$ for fixed $\boldsymbol{\omega} \in \Omega$. It is evident that in Eq. (3) the nonnegative bivariate function

$$G_n(\boldsymbol{\omega}) := \|\mathbf{y}\|^2 - \|\mathbf{y} - \mathbf{F}(\boldsymbol{\omega})\hat{\boldsymbol{\beta}}(\boldsymbol{\omega})\|^2 \tag{4}$$

plays the same role as the traditional univariate periodogram whose global maximizer coincides with the GML frequency estimator for a single complex sinusoid in Gaussian white noise [10].

Numerical and analytical studies show [11,12] that the GML estimator $\hat{\boldsymbol{\omega}}$ is capable of producing very accurate frequency estimates even if the frequency separation is less than the resolution limit of the traditional periodogram, which is equal to $2\pi/n$ (or 1 cycle per unit time). Therefore, a graphical plot of $G_n(\boldsymbol{\omega})$ as a function of ω_1 and ω_2 should reveal a prominent peak near the true signal frequency $\boldsymbol{\omega}_0$ but away from the diagonal line $\omega_1 = \omega_2$. This gives the idea of regarding $G_n(\boldsymbol{\omega})$ in (4) as a bivariate periodogram and using it graphically to uncover closely spaced hidden frequencies that cannot be resolved by means of the traditional periodogram. We call $G_n(\boldsymbol{\omega})$ the bivariate ℓ_2-periodogram. Because of symmetry, it suffices to consider the bivariate ℓ_2-periodogram above (or below) the diagonal line. The high-resolution property of the bivariate ℓ_2-periodogram is demonstrated by the simulation example shown in Fig. 1.

The time series shown in Fig. 1 is of length n=50 and comprises three complex sinusoids in simulated Gaussian white noise:

$$y_t = \exp(it\omega_{01}) + \exp(it\omega_{02}) + 1.3\exp(it\omega_{03}) + \varepsilon_t \quad (t = 1, \ldots, n).$$

The first two frequencies $(\omega_{01} := 2\pi \times 0.12$ and $\omega_{02} := 2\pi \times 0.13)$ are separated only by one-half of the resolution limit of the traditional periodogram, i.e., by $2\pi \times 0.5/n$, whereas the third frequency $(\omega_{03} := 2\pi \times 0.25)$ is well separated from the others. As shown in Fig. 1, the univariate periodogram is unable to resolve ω_{01} and ω_{02}, as it produces only two, rather than three, spectral peaks. The bivariate ℓ_2-periodogram, on the other hand, successfully reveals the closely spaced frequencies by portraying them as a well-defined local maximum near the signal frequency $(2\pi \times 0.12, 2\pi \times 0.13)$ and away from the diagonal line. Notice the absence of such local maxima near the diagonal line around frequency $2\pi \times 0.25$ where only a single sinusoid is present. The larger maximum near $(2\pi \times 0.125, 2\pi \times 0.25)$ corresponds to the two spectral peaks seen in the univariate periodogram.

Frequency estimates can be obtained by finding the local maxima of the bivariate ℓ_2-periodogram numerically. For demonstration, we employ a general-purpose optimization routine called `optim` in the open-source software R. It is an implementation of the iterative Nelder–Mead algorithm [13] for unconstrained optimization without the differentiability requirement. To avoid the complication of initialization techniques, we simply choose the true signal frequencies as initial value. For the bivariate ℓ_2-periodogram shown in Fig. 1, the optimization routine with initial value $(2\pi \times 0.12, 2\pi \times 0.13)$ produces a frequency estimate $(2\pi \times 0.1175, 2\pi \times 0.1275)$ with average absolute error $2\pi \times 0.0025$.

3. Bivariate ℓ_1-periodogram

Although effective as a high-resolution spectral analyzer, the bivariate ℓ_2-periodogram is not robust in the presence of outlier contamination, as demonstrated by Fig. 2. In this example, the noise is the same as that in Fig. 1 except for a large outlier contained in its real part at t=12. The outlier dramatically alters the appearance of the bivariate ℓ_2-periodogram. The local maximum in Fig. 1 that corresponds to the closely spaced frequencies disappears completely in Fig. 2. Even the spectral peak near $(2\pi \times 0.125, 2\pi \times 0.25)$ becomes much less prominent, as does the first spectral peak in the univariate periodogram. The frequency estimate