AN ITERATIVE NON-PARAMETRIC APPROACH TO THE ESTIMATION OF POLYSPECTRA

Naveed R. Butt and Andreas Jakobsson

Center For Mathematical Sciences, Lund University
Box 118, SE 22100, Lund, Sweden
Phone: + (46) 46 22 285 50, fax: + (46) 46 22 246 23
Email: naveed@maths.lth.se; andreas.jakobsson@matstat.lu.se

ABSTRACT

A number of powerful tools for analyzing linear and nonlinear data sets are based on various spectral measures. In particular, the bispectrum is commonly used for testing Gaussianity and linearity. Due to their inherent robustness to model assumptions, non-parametric estimators of the polyspectra are of particular importance. Unfortunately, the most commonly used non-parametric estimator, the windowed-periodogram, suffers from large sidelobes and fails to provide high-resolution estimates. In this paper, we develop a non-parametric estimator that utilizes the recently introduced iterative adaptive approach (IAA) to provide high-resolution estimates of the polyspectra for nonlinear data. Using the IAA method, we first obtain estimates of the spectral amplitudes and the covariance matrix iteratively, and then use the spectral amplitudes to form accurate estimates of the polyspectra. The developed estimator can be extended to the application of unevenly sampled data, and can also be used in the statistically efficient estimation of coherence polyspectra. The effectiveness of the proposed estimator is demonstrated with both real and simulated data.

1. INTRODUCTION

Spectral analysis is an important data analysis tool that finds applications in a wide variety of fields, including speech processing, telecommunications, radar and sonar systems, biomedical and seismic signal processing, and economics. Two particulary useful spectral measures are the power spectrum and the bispectrum [1], [2]. While the power spectrum shows the contribution of different frequencies to the formation of a signal, the bispectrum indicates any possible couplings between these frequencies. The power spectrum is generally used for analyzing linear and stationary processes, while the bispectrum finds usage in the analysis of nonlinear and/or nonstationary processes (see, e.g., [2], [3] and [4]). Some of the more important uses of the bispectrum are in testing measured sequences for Gaussianity and linearity. For these, and several other applications, it is desired to get consistent high-resolutions estimates of the polyspectra from available data. The most commonly used estimator for polyspectra, the periodogram and its windowed versions, suffer from either low resolution or high leakage, or both, and to achieve sufficient spectral resolution, the duration of the observation window has to be long. Furthermore, these methods only allow for uniformly sampled data. However, in a wide range of applications the measured data could be unevenly sampled, or might suffer from lost samples (see, e.g., [5–10]). In this regard, a new weighted least squares (WLS) based non-parametric approach, the so-called iterative adaptive approach (IAA), has recently been shown to provide high-resolution estimates of the power spectrum [11] and the magnitude squared coherence (MSC) [12] for both uniformly sampled and non-uniformly sampled data. The IAA-based estimators first make a frequency-domain reformulation of the given data, and then use it in a WLS fitting criterion to iteratively obtain estimates of the spectral amplitudes and the covariance matrix. In this work, we develop and discuss IAA-based estimators for polyspectra. Using the IAA method, we first obtain estimates of the spectral amplitudes and the covariance matrix iteratively, and then use the spectral amplitudes to form high-resolution estimates of the polyspectra. Without loss of generality, the development is carried out for the bispectrum, but the technique can be easily extended analogously to higher-order spectra. Furthermore, the work can also be extended along the lines of [12] to obtain statistically efficient estimates of the coherence polyspectra from unevenly sampled data.

This paper is organized as follows; in the next section, we provide a brief review of cumulants and polyspectra, and their estimation. In Section 3, we develop IAA-based estimators for the bispectrum. The performance of the proposed estimators is evaluated using real and simulated data in Section 4.

Notation: $(\cdot)^T$ and $(\cdot)^*$ are used to represent the transpose and the complex conjugate transpose, respectively. Vectors are denoted with bold letters, $\mathbf{x}$, while scalars are in light-face, x.

2. PRELIMINARIES

In this section, we briefly review the concepts of cumulants and polyspectra followed by a discussion on how these are commonly estimated. The second and third order cumulants of a zero mean[1] random process x are defined as

$$C_{xx}(k) = E\{x^*(n)x(n+k)\} \quad (1)$$
$$C_{xxx}(k,l) = E\{x^*(n)x(n+k)x(n+l)\}, \quad (2)$$

where $E\{\cdot\}$ is the expectation operator. The second-order cumulant, C_{xx}, is also often referred to as the autocovariance sequence. The power spectrum is defined as the Fourier transform of the second-order cumulant and represents the frequency content of the series, i.e.,

$$S_{xx}(\omega_1) = \sum_{k=-\infty}^{\infty} C_{xx}(k)e^{-j\omega_1 k} \quad (3)$$
$$= \mathcal{X}^*(\omega_1)\mathcal{X}(\omega_1), \quad (4)$$

where $\mathcal{X}$ represents the Fourier transform of x. Similarly, the bispectrum is defined as the Fourier transform of the

This work was supported in part by the Swedish Research Council and Carl Trygger's foundation, Sweden.

[1] A given process can always be made zero-mean by subtracting the mean from it.

 ISSN 2076-1465

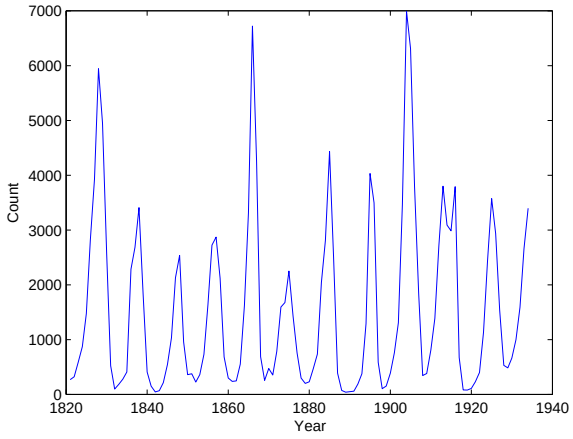

Figure 1: Canadian lynx data.

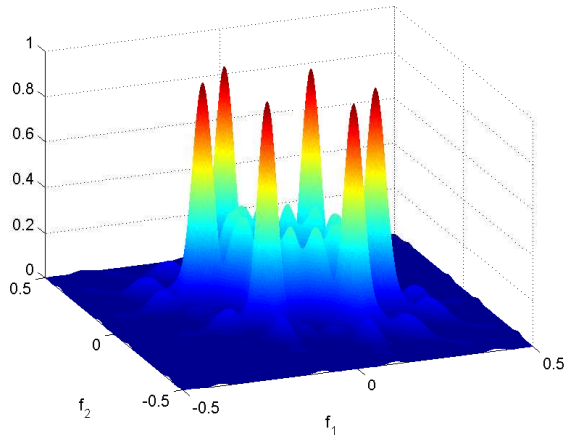

Figure 2: Lynx data: windowed-periodogram based estimate of the bispectrum.

third-order cumulant,

$$S_{xxx}(\omega_1,\omega_2) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} C_{xxx}(k,l)e^{-j\omega_1 k}e^{-j\omega_2 l} \quad (5)$$

$$= \mathcal{X}^*(\omega_1+\omega_2)\mathcal{X}(\omega_1)\mathcal{X}(\omega_2). \quad (6)$$

The bispectrum is a function of two frequencies, and as mentioned in the introduction, it is commonly used in testing the linearity and Gaussianity of a given sequence (see, e.g., [2], [3] and [4]). We note that unlike the power spectrum, the bispectrum is complex-valued.

Typically, one has to estimate the cumulants and polyspectra from a limited set of samples. For instance, consider a vector $\mathbf{x}$ containing N samples of a time-series, i.e.,

$$\mathbf{x} = \begin{bmatrix} x(1) \\ \vdots \\ x(N) \end{bmatrix}. \quad (7)$$

The estimates of the second and third-order cumulants of $\mathbf{x}$ may then be obtained through averaging as,

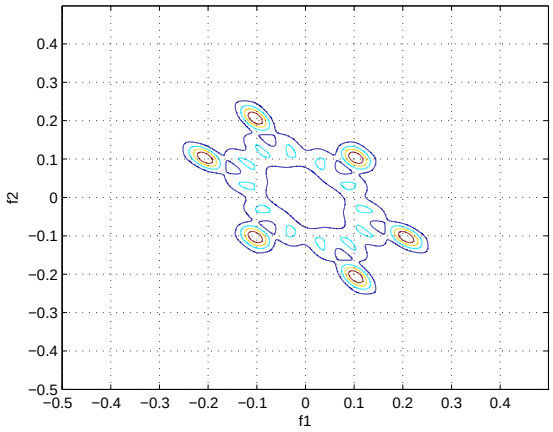

Figure 3: Lynx data: Contour plot of the windowed-periodogram based estimate of the bispectrum.

$$\hat{C}_{xx}(k) = \frac{1}{N}\sum_{n=1}^{N-\gamma_k} x^*(n)x(n+k) \quad (8)$$

$$\hat{C}_{xxx}(k,l) = \frac{1}{N}\sum_{n=1}^{N-\gamma_{kl}} x^*(n)x(n+k)x(n+l), \quad (9)$$

where

$$\gamma_k = \max(0,k) \quad (10)$$

$$\gamma_{kl} = \max(0,k,l) \quad (11)$$

$$k,l > 0. \quad (12)$$

Then, using $\hat{C}_{xx}(k)$ and $\hat{C}_{xxx}(k,l)$ as obtained from (8) and (9), estimates of the power spectrum and bispectrum may be obtained through the windowed-periodogram method as shown in equations (13)-(14) on the next page, where the window sequences, $v(k)$ and $w(k,l)$, are included to make the estimators asymptotically unbiased and consistent (i.e., $\lim_{N\to\infty}\hat{S} = S$). Further details on the selection of $v(k)$ and $w(k,l)$ may be found in [4]. The simplicity of the periodogram approach makes it a very useful estimation tool for general applications. However, as is well-known, the periodogram method does not generally provide high resolution estimates [1]. We refer the reader to [11] and [12] where some of these issues for the power spectrum have been addressed.

3. IAA-BASED ESTIMATION OF POLYSPECTRA

Given the data vector $\mathbf{x}$, a grid-dependent frequency-domain representation of $\mathbf{x}$ may be obtained along the lines of [13], by selecting I grid points in the frequency domain with corresponding frequencies $\{\omega_i\}_{i=1}^I$, as

$$x(n) = \sum_{i=1}^{I} \alpha_{\omega_i} e^{j\omega_i n}, \qquad n = 1,\ldots,N, \quad (15)$$

where α_{ω_i} is the (unknown) complex-valued spectral amplitude of the ith frequency, including any corrupting noise elements. It is worth noting that no signal model has been

$$\hat{S}_{xx}(\omega_1) = \sum_{k=-N-1}^{N-1} v(k)\hat{C}_{xx}(k)e^{-j\omega_1 k} \quad (13)$$

$$\hat{S}_{xxx}(\omega_1,\omega_2) = \sum_{k=-N-1}^{N-1}\sum_{l=-N-1}^{N-1} w(k,l)\hat{C}_{xxx}(k,l)e^{-j\omega_1 k}e^{-j\omega_2 l} \quad (14)$$

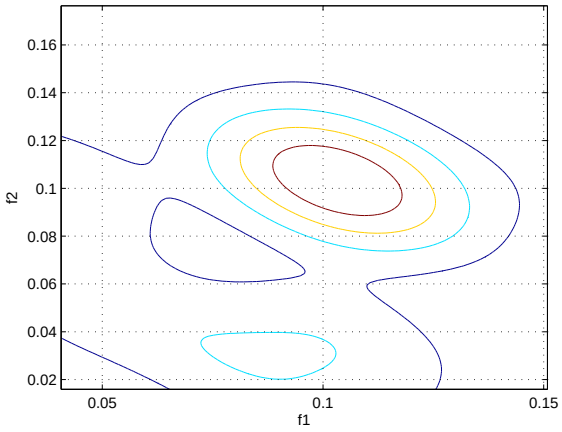

Figure 4: Zoom-in of Figure 3.

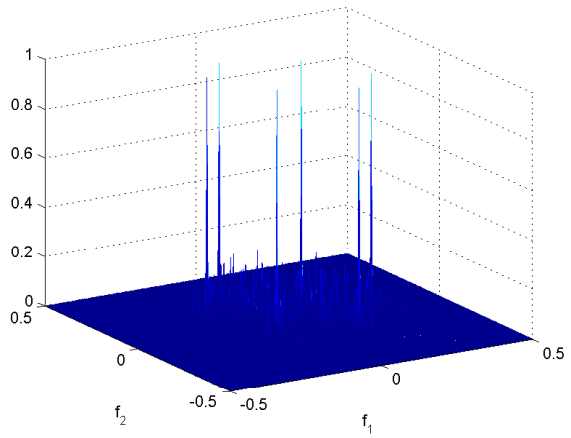

Figure 5: Lynx data: IAA-based estimate of the bispectrum.

assumed; rather the signal is made from the contribution corresponding to each of the frequency grid points. There is also no corrupting noise term as is typical in model-based methods describing the data as a signal and a noise part. The contribution of any noise component is instead implicitly described via its contribution to α_{ω_i}. Further, defining

$$\mathbf{a}(\omega_i) = \begin{bmatrix} e^{j\omega_i} & \cdots & e^{j\omega_i N} \end{bmatrix}^T \quad (16)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}(\omega(1)) & \cdots & \mathbf{a}(\omega(I)) \end{bmatrix}^T \quad (17)$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{\omega_1} & \cdots & \alpha_{\omega_I} \end{bmatrix}^T, \quad (18)$$

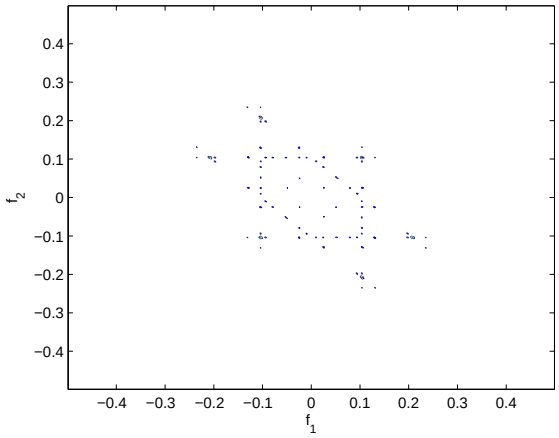

Figure 6: Lynx data: contour plot of the IAA-based estimate of the bispectrum.

where $\mathbf{a}(\omega_i)$ represents the Fourier vector corresponding to the sampling times of $\mathbf{x}$, at the frequency ω_i, leads to the vector representation

$$\mathbf{x} = \mathbf{A}\boldsymbol{\alpha}. \quad (19)$$

The covariance matrix, $\mathbf{R}$, may thus be written as

$$\mathbf{R} = \sum_{i=1}^{I} |\alpha_{\omega_i}|^2 \mathbf{a}(\omega_i)\mathbf{a}^*(\omega_i). \quad (20)$$

For each frequency, ω_i, the interference and noise covariance matrix, $\mathbf{Q}(\omega_i)$, defined as the contribution from all grid points other than ω_i, may be formed as

$$\mathbf{Q}(\omega_i) = \sum_{p\neq i,\ p=1}^{I} |\alpha_{\omega_i}|^2 \mathbf{a}(\omega_p)\mathbf{a}^*(\omega_p) \quad (21)$$

An estimate of the spectrum amplitude at frequency ω_i, say $\hat{\alpha}_{\omega_i}$, may now be formed using the weighted least squares problem

$$\hat{\alpha}_{\omega_i} = \arg\min_{\alpha_{\omega_i}} \left\| \mathbf{x} - \alpha_{\omega_i}\mathbf{a}(\omega_i) \right\|^2_{\mathbf{Q}(\omega_i)}, \quad (22)$$

where $\|\cdot\|_{\mathbf{Q}(\omega_i)}$ represents the weighted 2-norm. As shown in [11], the solution to this problem may be obtained as

$$\hat{\alpha}_{\omega_i} = \frac{\mathbf{a}^*(\omega_i)\hat{\mathbf{R}}^{-1}\mathbf{x}}{\mathbf{a}^*(\omega_i)\hat{\mathbf{R}}^{-1}\mathbf{a}(\omega_i)} \quad (23)$$

$$\hat{\mathbf{R}} = \sum_{i=1}^{I} |\hat{\alpha}_{\omega_i}|^2 \mathbf{a}(\omega_i)\mathbf{a}^*(\omega_i), \quad (24)$$

$$\{\omega_i\}_{i=1}^{I} \quad = \quad \{\omega_1(p_1)\}_{p_1=1}^{P} \bigcup \{\omega_2(p_2)\}_{p_2=1}^{P} \bigcup \{\omega_1(p_1)+\omega_2(p_2)\}_{p_1,p_2=1}^{P} \tag{25}$$

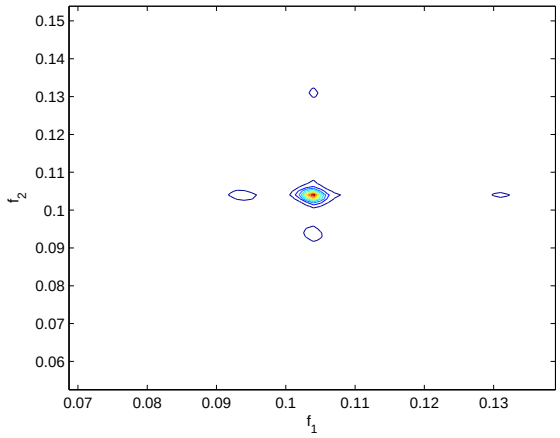

Figure 7: Zoom-in of Figure 6.

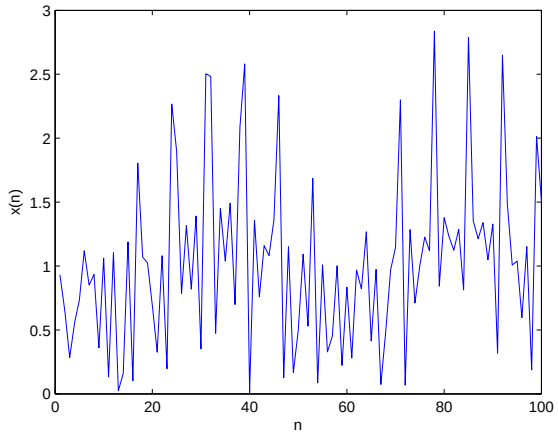

Figure 8: Coupled sinusoids data.

where (23) and (24) are solved iteratively by initializing $\hat{\mathbf{R}}$ to an identity matrix. Assuming that the desired $P \times P$ grid for the bispectrum is formed from $\{\omega_1(p_1)\}_{p_1=1}^{P}$ and $\{\omega_2(p_2)\}_{p_2=1}^{P}$, we may form the frequency grid for IAA, $\{\omega_i\}_{i=1}^{I}$ according to (25) on the top of the page, where $\bigcup$ represents union of sets. The IAA-based estimate of the power spectral density and the bispectrum can thus be formed as

$$\hat{S}_{xx}^{\alpha}(\omega_1) \quad = \quad \hat{\alpha}_{\omega_1}^{*}\hat{\alpha}_{\omega_1} \tag{26}$$

$$\hat{S}_{xxx}^{\alpha}(\omega_1+\omega_2) \quad = \quad \hat{\alpha}_{\omega_1+\omega_2}^{*}\hat{\alpha}_{\omega_1}\hat{\alpha}_{\omega_2} \tag{27}$$

where the superscript α has been added to show that these are IAA-based estimates. We note that IAA-based estimates of higher polyspectra may be obtained analogously, and extension to the non-uniformly sampled data sets can be formed reminiscent to the approach in [12].

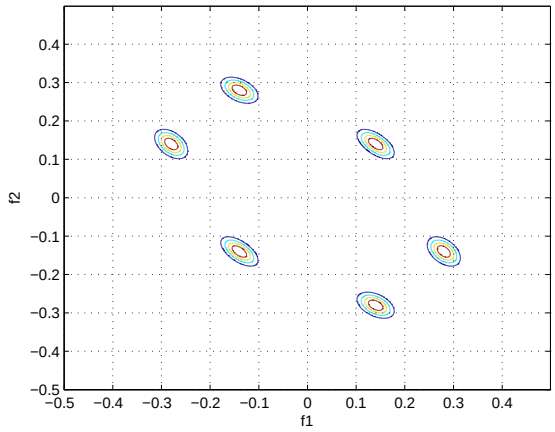

Figure 9: Coupled sinusoids: contour plot of the windowed-periodogram estimate of bispectrum.

4. EXAMPLES

In this section, we demonstrate the effectiveness of the proposed IAA-based estimators through application to real and simulated data. In each case, the suggested approach is compared to the popular windowed-periodogram method[2].

4.1 Canadian Lynx Data

The first time-series studied here is the Canadian lynx data that consists of the annual number of Canadian lynx trapped in the Mackenzie River district of Northwest Canada for the years 1821-1934 (see [14] for further details). The trapped-lynx count is plotted versus time in Figure 1. Figures 2-4 show the windowed-periodogram estimate of the bispectrum using the Rao-Gabr window. The bispectrum estimate shows a rather wide peak at around (0.1,0.1) (and other symmetric locations). This indicates the possibility of quadratic frequency coupling. The IAA-based estimate of the bispectrum is shown in Figures 5-7. As is clear from these figures, the IAA-based estimate results in much sharper peaks. The basic peak is at (0.104, 0.104), which gives a period of 9.6154 years, and its coupled harmonic at 4.8077 years.

4.2 Coupled Sinusoids

To show the importance of obtaining high resolution bispectrum estimates, we also consider simulated data representing closely-spaced coupled sinusoids. We generate 100 samples of a signal, x_n, having three sinusoidal components

$$x_n \quad = \quad \sum_{r=1}^{3} \alpha_r \cos(2\pi f_r n + \phi_r) + w_n, \tag{28}$$

where α_r and f_r denote the amplitude and frequency of the rth sinusoid, respectively; ϕ_1 and ϕ_2 are independent uniformly distributed random variables between 0 and 2π; and

[2]A particularly useful implementation is available in the Matlab toolbox 'HOSA', http://www.mathworks.com/matlabcentral/fileexchange/3013.

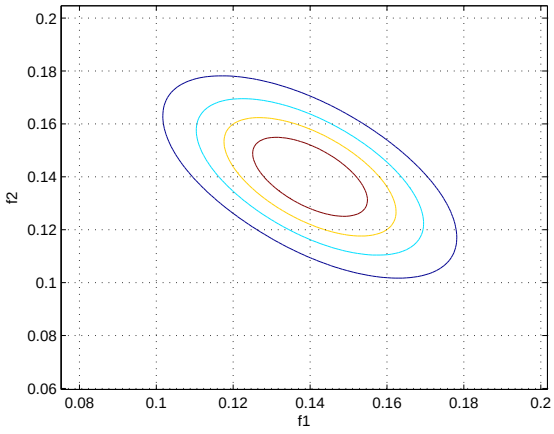

Figure 10: Zoom-in of Figure 9.

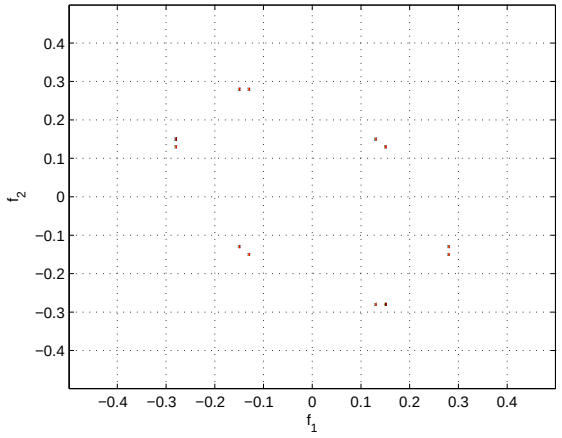

Figure 11: Coupled sinusoids: contour plot of the IAA-based estimate of the bispectrum.

the noise term, w_n, represents zero-mean Gaussian random noise. In the simulations, we set $\alpha_r = 1, \forall r$, $f_1 = 0.13$ Hz, $f_2 = 0.15$ Hz, $f_3 = f_1 + f_2$, and $\phi_3 = \phi_1 + \phi_2$. The simulated data is shown in Figure 8. The windowed-periodogram based estimate of the bispectrum, shown in Figures 9 and 10, fails to resolve the closely-spaced peaks. Figures 11-12 show that the corresponding IAA-based estimator provides accurate estimate of the bispectrum, showing the peaks to be at the correct frequency locations.

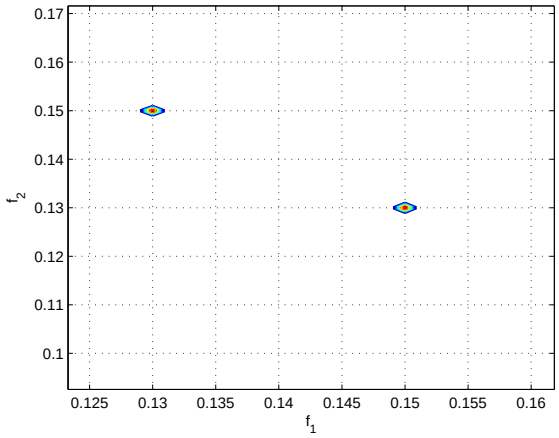

Figure 12: Zoom-in of Figure 11.

REFERENCES

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, N.J.: Prentice Hall, 2005.

[2] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. The Edinburgh Building, Cambridge CB2 2RU, UK.: Cambridge University Press, 2004.

[3] T. S. Rao, "Analysis of nonlinear time series (and chaos) by bispectral methods," *Santa Fe Institute Studies in the Sciences of Complexity Proceedings*, vol. 12, pp. 199–222, 1992.

[4] H. Madsen and J. Holst, *Modelling Non-linear and Non-stationary Time Series*. IMM, 2000.

[5] A. Afifi and R. Elashoff, "Missing observations in multivariate statistics I: Review of the literature," *J. Amer. Statist. Assoc.*, pp. 595–604, 1966.

[6] T. Orchard and M. A. Woodbury, "A missing information principle: theory and applications," *Prox. Sixth Berkeley Symp. on Math. Statist. and Prob.*, pp. 697–715, 1972.

[7] J. D. Scargle, "Studies in astronomical time series analysis. ii. statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, pp. 835–853, 1982.

[8] M. Schulz and K. Stattegger, "Spectrum: Spectral analysis of unevenly spaced paleoclimatic time series," *Computers and Geosciences*, vol. 23, no. 9, pp. 929–945, 1997.

[9] J. L. Schafer and J. W. Graham, "Missing data - our view of the state of the art." *Psychological Methods*, pp. 147–177, 2002.

[10] Y. Wang, J. Li, and P. Stoica, *Spectral Analysis of Signals, The Missing Data Case*. USA: Morgan & Claypool Publishers, 2005.

[11] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Transactions on Aerospace and Electronic Systems. In Press.*, 2009.

[12] N. R. Butt and A. Jakobsson, "Coherence spectrum estimation from non-uniformly sampled sequences," *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 339–342, April 2010.

[13] P. Stoica, J. Li, and J. Ling, "Missing data recovery via a nonparametric iterative adaptive approach," *Signal Processing Letters, IEEE*, vol. 16, no. 4, pp. 241–244, April 2009.

[14] M. Bulmer, "A statistical analysis of the 10-year cycle in Canada," *J. Anim. Ecol.*, vol. 43, pp. 701–715, 1974.