Available online at www.sciencedirect.com
ScienceDirect
Statistical Methodology 5 (2008) 318–327
Statistical Methodology
Official Journal of the International Indian Statistical Association
www.elsevier.com/locate/stamet

New methods for fitting multiple sinusoids from irregularly sampled data

Sébastien Bourguignon*, Hervé Carfantan

Laboratoire d'Astrophysique de Toulouse et de Tarbes, UMR 5572 – Université Paul Sabatier - Toulouse III and CNRS, 14 avenue Edouard Belin, 31400 Toulouse, France

Received 16 April 2007; received in revised form 22 August 2007; accepted 11 October 2007

Abstract

A novel framework is proposed for the estimation of multiple sinusoids from irregularly sampled time series. This spectral analysis problem is addressed as an under-determined inverse problem, where the spectrum is discretized on an arbitrarily thin frequency grid. As we focus on line spectra estimation, the solution must be sparse, *i.e.* the amplitude of the spectrum must be zero almost everywhere. Such prior information is taken into account within the Bayesian framework. Two models are used to account for the prior sparseness of the solution, namely a Laplace prior and a Bernoulli–Gaussian prior, associated to optimization and stochastic sampling algorithms, respectively. Such approaches are efficient alternatives to usual sequential prewhitening methods, especially in case of strong sampling aliases perturbating the Fourier spectrum. Both methods should be intensively tested on real data sets by physicists.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Spectral analysis; Bayesian estimation; Sparse representations; Optimization; MCMC

1. Introduction

The search for periodicities is a very important topic in astronomical time series analysis. In particular, it is the key tool for studying stellar oscillations, since the pulsation modes of variable stars can be determined from the frequencies extracted from their light or radial velocity curves. Other fields of interest are multiple stars and exoplanet detection, which are also based on the detection of periodicities in observational data.

* Corresponding author. Tel.: +33 561 332 885; fax: +33 561 332 840.
E-mail addresses: bourgui@ast.obs-mip.fr (S. Bourguignon), Herve.Carfantan@ast.obs-mip.fr (H. Carfantan).

1572-3127/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2007.10.004

Because of observation constraints, astronomical data generally suffer from incomplete sampling. First, the day-night alternation or bad meteorological conditions may generate gaps in the time series. Second, long-time observations are generally irregularly sampled, *i.e.* the time spacing between the data points is not governed by any sampling period. Both sources of irregularities affect the observed Fourier spectrum, which is the *convolution* of the true spectrum by the *spectral window*. In particular, the spectral window may have high secondary lobes if the sampling scheme has periodic gaps. As a consequence, the Fourier spectrum of the data may show many false peaks and hide true spectral lines.

The Lomb-Scargle periodogram [1] and the date-compensated discrete Fourier transform [2] are generalizations of the classical periodogram to irregular sampling, but are only statistically valid for fitting *one* sinusoid. For multi-sine fitting, prewhitening methods such as CLEAN [3] or CLEANEST [4], widely used in astrophysics, remove iteratively the peaks in the Fourier spectrum as single frequency components. These methods, however, lack theoretical background and may fail in some cases [5].

Following recent works in the last decade [6–8], we address the problem of spectral analysis as an underdetermined inverse problem, where the spectrum is discretized on an arbitrarily thin frequency grid. As we focus on the estimation of periodicities in the data (*i.e.* of spectral lines), the solution is sparse: the discretized spectrum should be zero almost everywhere. The Bayesian framework allows to account statistically for such prior information. Two estimation strategies are proposed. The first one considers a Laplace prior, that leads to minimize a least-squares criterion penalized by an ℓ^1-norm term [9,10]. Such an approach has been adopted in a wide field of applications, including deconvolution, compressed sensing or image inpainting. For the second one, we propose to alternately model sparseness using a Bernoulli–Gaussian model [11], associated to Monte-Carlo Markov Chain (MCMC) simulation methods.

Such global approaches – all parameters are estimated jointly – represent efficient alternatives to usual sequential prewhitening methods, especially in case of strong sampling aliases perturbating the Fourier spectrum.

2. Where prewhitening methods fail

Let (t_n, y_n) define the irregularly sampled time series, where y_n is the measured amplitude at time t_n. Ideally, one wants to estimate the parameters of the multi-sine model:

$$y_n = \sum_{m=1}^{M} \alpha_m \cos(2\pi \nu_m t_n) + \beta_m \sin(2\pi \nu_m t_n). \tag{1}$$

However, this is a difficult task: model (1) is not linear in the frequency parameters, and its likelihood function was shown to be highly multi-modal [12]. Moreover, the number of frequencies, M, is unknown. Sequential prewhitening methods [3,4] try to find the best fit for model (1) by iteratively removing the peaks in the Fourier spectrum and their contribution in terms of side lobes, until some stopping condition is reached. They can be viewed as an intent to minimize, at iteration M, the least-squares misfit term:

$$J_M((\nu_m, \alpha_m, \beta_m)_{m=1..M}) = \sum_{n=1}^{N} \left(y_n - \sum_{m=1}^{M} \alpha_m \cos(2\pi \nu_m t_n) - \beta_m \sin(2\pi \nu_m t_n) \right)^2. \tag{2}$$

It is easy to build examples where sequential methods are trapped by local minima of J_M. Artificial data in Fig. 1 shows one. The sampling simulates five observing nights, so the spectral

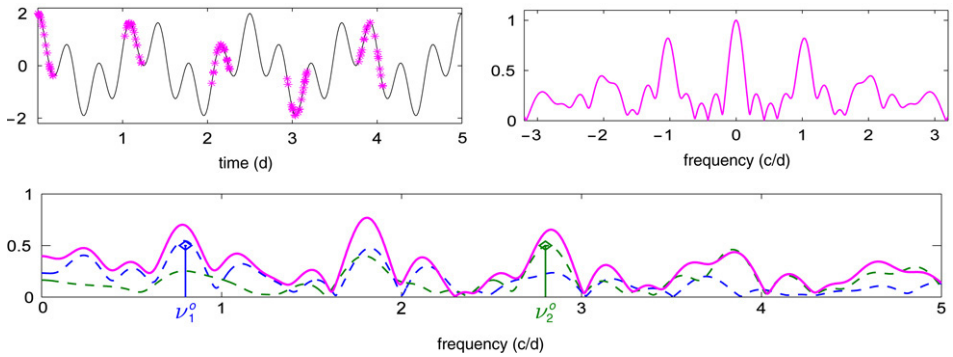


Fig. 1. An artificial example where prewhitening methods fail. Top left: available data (∗) and full signal (−). Top right: spectral window. Bottom: Fourier spectrum of the data (−) and true spectrum (⋄, ⋄). The dashed lines show the contribution of each line at ν_1^o and ν_2^o in the spectrum.

window shows high sidelobes at 1 cycle per day (c/d). The data is made of two sinusoids with critically spaced frequencies ν_1^o and ν_2^o, so that the sidelobes for each frequency superimpose at $\nu_{\text{artifact}} = (\nu_1^o + \nu_2^o)/2$, producing a false peak as the global maximum of the Fourier spectrum. Here, a prewhitening method initialized at ν_{artifact} would propagate errors through the iterations. In particular, it would lead to a local minimum of J_2 since the global minimum – zero – is obtained for $(\nu_1, \nu_2) = (\nu_1^o, \nu_2^o)$.

3. An alternative framework based on a linear model

3.1. An underdetermined linear problem

We model the data as the linear combination of an arbitrarily large number of sine waves, with equispaced frequencies $f_k = \frac{k}{K} f_{\max},\ _{k=0\ldots K}$:

$$y_n = \sum_{k=-K}^{K} x_k \exp(\mathrm{j}2\pi f_k t_n) + \epsilon_n \Leftrightarrow \boldsymbol{y} = \mathbf{W}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{3}$$

where $\boldsymbol{y} \in \mathbb{C}^N$ collects the data, $\boldsymbol{x} \in \mathbb{C}^{2K+1}$ is the discretized amplitude spectrum, $\mathbf{W}$ is the $N \times (2K+1)$ matrix $\left\{\mathrm{e}^{\mathrm{j}2\pi f_k t_n}\right\}_{n=1\ldots N,\, k=-K\ldots K}$ and $\boldsymbol{\epsilon}$ are additive perturbations such as measurement noise and model errors. Usually, the observed data are real-valued. Then, the spectrum has the hermitian symmetry $x_{-k} = x_{-k}^*$ and $\boldsymbol{\epsilon}$ is also real-valued. Note that an upper bound $f_{\max}$ of the explored frequencies has to be given; the frequency axis where oscillations are searched can also be restricted to some interval $[f_{\min}, f_{\max}]$.

The objective is to estimate the amplitude spectrum $\boldsymbol{x}$ from the noisy data $\boldsymbol{y}$. High frequency precision generally requires $2K+1 \gg N$, which leads to an underdetermined inverse problem. Among all possible solutions, a sparse one is searched, since we focus on spectral *lines*: x_k should be zero almost everywhere. Then, the non-zero values of $\boldsymbol{x}$, say $x_{k,\, k\in\mathcal{K}}$, locate the estimated frequencies at $f_{k,\, k\in\mathcal{K}}$.

It is important to understand the fundamental difference between this formulation and the classic multi-sine fitting approaches. In model (3), the frequencies are fixed and the only parameters to estimate are amplitudes. We consider a sparse vector of amplitudes, for which the whole spectrum can be estimated *jointly*. In the case of model (1), a global optimization strategy

is tricky because of the non-linearities in the frequencies ν_m, and also because the number of frequencies M is unknown: for these reasons, *sequential* methods are usually preferred, that suffer from the limitations pointed out in Section 2.

3.2. Bayesian estimation

Suppose that perturbations ϵ_n in model (3) are independent and identically distributed according to a Gaussian distribution $\mathcal{N}(0, \sigma_\epsilon^2)$. The *likelihood function* writes:

$$\mathcal{L}(\boldsymbol{y}; \boldsymbol{x}, \sigma_\epsilon^2) \propto \frac{1}{\sigma_\epsilon^N} \exp\left(-\frac{1}{2\sigma_\epsilon^2}(\boldsymbol{y} - \mathbf{W}\boldsymbol{x})^\dagger(\boldsymbol{y} - \mathbf{W}\boldsymbol{x})\right).$$

Perturbations that are not identically distributed can also be considered by adequately weighting each component of the quadratic term. This is actually the case for multisite observation campaigns, where a proper variance σ_n^2 is associated to every measurement y_n.

Bayesian estimation relies on a probabilistic description of the prior information. Let $p(\boldsymbol{x}|\boldsymbol{\theta})$ be the prior distribution of the spectrum, with *hyperparameters* $\boldsymbol{\theta}$, the *posterior distribution* writes:

$$p(\boldsymbol{x}|\boldsymbol{y}, \sigma_\epsilon, \boldsymbol{\theta}) \propto \mathcal{L}(\boldsymbol{y}; \boldsymbol{x}, \sigma_\epsilon)\, p(\boldsymbol{x}|\boldsymbol{\theta}). \tag{4}$$

The goal is to model sparseness through an adequate prior distribution. Without more information, there is no reason to consider any correlation between the spectral components, so it is natural to consider that variables x_k are independent and identically distributed (*i.i.d.*): $p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_k p(x_k|\boldsymbol{\theta})$, where $p(x_k|\boldsymbol{\theta})$ should be at the same time high-valued in zero and sufficiently heavy-tailed (compared to the Gaussian distribution) to model sparseness. In this paper, we propose to use two kinds of priors:

(i) the *Laplace prior distribution* has become the most usual choice, since the corresponding *Maximum A Posteriori* (MAP) estimate is generally sparse [10];
(ii) the *Bernoulli–Gaussian prior distribution* [11] is a composite probabilistic description which explicitly expresses sparseness, although it received much less attention.

4. First approach: Laplace prior/ℓ^1-norm penalisation

Let us consider a Laplace distribution on the amplitudes: $p(x_k|\sigma_x) \propto \frac{1}{\sigma_x} \exp\left(-\frac{1}{\sigma_x}|x_k|\right)$.

4.1. Maximum A Posteriori estimation: A powerful optimization strategy

The MAP estimate $\widehat{\boldsymbol{x}}$ equivalently minimizes the penalized least-squares criterion:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \tag{5}$$

for $\lambda = \sigma_\epsilon^2/\sigma_x$, where $\|\cdot\|_2$ is the Euclidean norm and $\|\cdot\|_1$ is the ℓ^1-norm: $\|\boldsymbol{x}\|_1 = \sum_k |x_k|$. The ℓ^1-norm penalization has been widely studied in the past ten years and is known as a *Basis Pursuit De-Noising* (BPDN) technique [9]. In application to sparse spectral analysis, other penalization functions were proposed. A Cauchy prior was used in [7], which is efficient for sparseness, but gives a non-convex criterion: optimization may be trapped by local minima. In [8], a strictly convex and differentiable approximation of the ℓ^1-norm was proposed, associated

with fast FFT-based algorithms. However, it only yields approximately sparse solutions, and such algorithms are not efficient with irregularly sampled data. Let us remark that the ℓ^1-norm penalization is the only ℓ^p-norm penalization that achieves sparseness ($p \leq 1$) and yields a convex criterion ($p \geq 1$).

The maximization of (5) is a convex optimization problem (although the criterion is not strictly convex if $2K + 1 > N$), for which convergent optimization procedures exist. Note that variables are complex-valued, which yields more accurate estimates than real variables [5], but many optimization algorithms used for BPDN, based on linear programming, are not applicable. A classical solution in this case formulates (5) as a second order cone program (SOCP), with specific algorithms. In our case, we computed $\widehat{\boldsymbol{x}}$ by iterating coordinatewise – scalar – minimizations, which can be performed at low cost [13]: see [5] for details. In practice, it gave more satisfactory results than SOCP in terms of computational time.

Note that amplitudes are systematically underestimated: indeed, the minimizer of (5) does not fit exactly the data since amplitudes are penalized. However, this bias can be corrected posteriorly by least-squares, once the frequencies are correctly located: model (1) with *known* frequencies is linear and largely over-determined. Optimizing the location of the frequencies off the reconstruction grid would also be possible, but it is out of the scope of this paper.

4.2. Interpretation of the regularization parameter

Despite its Bayesian interpretation ($\lambda = \sigma_\epsilon^2/\sigma_x$), λ is unknown in practice, and must be fixed to obtain a satisfactory solution. This is an open problem of regularization theory. However, in our case, parameter λ has a physical interpretation. Indeed, $\widehat{\boldsymbol{x}}$ satisfies [14]:

$$\begin{cases} \forall k/\widehat{x}_k = 0 : |r_k| \leq \lambda, \\ \forall k/\widehat{x}_k \neq 0 : r_k + \lambda \dfrac{\widehat{x}_k}{|\widehat{x}_k|} = 0, \end{cases}$$

where $\boldsymbol{r} = \mathbf{W}^\dagger(\mathbf{W}\widehat{\boldsymbol{x}} - \boldsymbol{y})$ is merely the discretized Fourier spectrum of the residual $\mathbf{W}\widehat{\boldsymbol{x}} - \boldsymbol{y}$. Thus, parameter λ can be interpreted as the maximum peak amplitude allowed in the residual spectrum. In practice, one can set λ to some percentage of the maximum of the Fourier spectrum of the data, *e.g.* 10%: $\lambda = \max |\mathbf{W}^\dagger \boldsymbol{y}|/10$.

5. Second approach: Bernoulli–Gaussian prior

To simplify the description, let us switch from model (3) to its equivalent formulation with real variables:

$$y_n = a_0 + \sum_{k=1}^{K} a_k \cos(2\pi f_k t_n) + b_k \sin(2\pi f_k t_n) + \epsilon_n \Leftrightarrow \boldsymbol{y} = \mathbf{H}\boldsymbol{s} + \boldsymbol{\epsilon}, \tag{6}$$

where $\boldsymbol{s}_k = [a_k, b_k]$ (with $s_0 = a_0$), $\boldsymbol{s} = [\boldsymbol{s}_0, \ldots, \boldsymbol{s}_K]$ are the $2K + 1$ real-valued unknown amplitudes, $\mathbf{H}_{n,0} = 1, \mathbf{H}_{n,2k-1} = \cos(2\pi f_k t_n)$ and $\mathbf{H}_{n,2k} = \sin(2\pi f_k t_n)$ for $n = 1 \ldots N$, $k = 1 \ldots K$.

5.1. Probabilistic description

The Bernoulli–Gaussian (B–G) model was introduced in application to seismic deconvolution [11], and first used for spectral analysis in [6]. It can be decomposed in:

– a sequence of binary variables $\boldsymbol{q} = \{q_0, \ldots, q_K\}$, that indicate the existence of a spectral line at each point of the frequency grid. The q_k are supposed independent and identically distributed according to a Bernoulli distribution with parameter γ:

$$\Pr(q_k = 1|\gamma) = \gamma$$

is the prior probability of existence of a spectral line at frequency f_k. Then, the value of γ controls the sparsity in the solution; *e.g.* for $\gamma = 0.1$, 10% of the amplitudes in average are non-zero.

– conditionally to the location by the Bernoulli variables, amplitude variables $(a_k|\, q_k = 1)$ and $(b_k|\, q_k = 1)$ for the corresponding spectral line at frequency f_k, are supposed independent and centered Gaussian with variance σ^2, so that:

$$p(\boldsymbol{s}_0|\boldsymbol{q}_0 = 1) = g_1(\boldsymbol{s}_0, \sigma^2), \qquad p(\boldsymbol{s}_k|\boldsymbol{q}_k = 1) = g_2(\boldsymbol{s}_k, \sigma^2\mathbf{I}_2), \quad k = 1 \ldots K,$$

where $g_m(\boldsymbol{s}, \boldsymbol{\Sigma})$ stands for the m-variate centered Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}$. Variables $(a_k|q_k = 0)$ and $(b_k|q_k = 0)$ are zero by definition of q_k, *i.e.* $(\boldsymbol{s}_k|q_k = 0)$ follows a bi-dimensional Dirac distribution: $p(\boldsymbol{s}_k|\boldsymbol{q}_k = 0) = \delta_2(\boldsymbol{s}_k)$ and $p(\boldsymbol{s}_0|\boldsymbol{q}_0 = 0) = \delta_1(\boldsymbol{s}_0)$.

In the following, we will denote $\boldsymbol{\theta}$ the set of hyperparameters: $\boldsymbol{\theta} = (\gamma, \sigma^2, \sigma_\epsilon^2)$.

Contrarily to the Laplace case in Section 4, the maximization of the posterior distribution $p(\boldsymbol{q}, \boldsymbol{s}|\boldsymbol{y}, \boldsymbol{\theta})$ is a difficult problem for the B–G prior. Indeed, it is a *combinatorial* problem, with 2^{K+1} configurations for $\boldsymbol{q}$: exploration algorithms are necessarily suboptimal. In particular, empirical results [6,14] showed that classical exploration methods may fall in local modes, characterized by frequency splitting. An alternative to optimization consists in using stochastic sampling methods such as MCMC. The idea, formerly proposed in [15] for the B–G model in seismic deconvolution, is precisely to explore as completely as possible the distribution $p(\boldsymbol{q}, \boldsymbol{s}|\boldsymbol{y}, \boldsymbol{\theta})$, and compute a more subtle estimator than MAP.

5.2. Improving the frequency precision

As the use of MCMC requires a high computational load, it is unreasonable to consider a similar size for the frequency grid to that used for the Laplace prior. This considerably limits the frequency precision. Dublanchet [6] proposed to add frequency shifts in the model, say $\{df_k\}_{k=1\ldots K}$ in $\Delta f = [0, \frac{f_{\max}}{K}[$, that take into account frequency deviations from the fixed grid locations. That is, model (6) is changed for the *extended* model:

$$\begin{aligned} y_n = \; & a_0 + \sum_{k=1}^{K} a_k \cos(2\pi(f_k - df_k)t_n) + b_k \sin(2\pi(f_k - df_k)t_n) + \epsilon_n \\ \Leftrightarrow \; & \boldsymbol{y} = \mathbf{H}_{\mathbf{df}}\boldsymbol{s} + \boldsymbol{\epsilon}, \end{aligned} \tag{7}$$

where $\mathbf{H}_{\mathbf{df}}$ is the $N \times 2K + 1$ matrix with terms $\mathbf{H}_{\mathbf{df}_{n,0}} = 1$, $\mathbf{H}_{\mathbf{df}_{n,2k-1}} = \cos(2\pi(f_k - df_k)t_n)$ and $\mathbf{H}_{\mathbf{df}_{n,2k}} = \sin(2\pi(f_k - df_k)t_n)$ for $n = 1 \ldots N$, $k = 1 \ldots K$. Note that $(q_k = 1)$, $k = 1 \ldots K$ now indicates a spectral line in the frequency interval $]f_{k-1}, f_k]$. In the following, df_k are supposed random with a uniform distribution on Δf.

5.3. MCMC unsupervised procedure

The goal of Bayesian estimation is to estimate all parameters $(\boldsymbol{q}, \boldsymbol{s}, \mathbf{df})$ from their joint posterior distribution $p(\boldsymbol{q}, \boldsymbol{s}, \mathbf{df}|\boldsymbol{y}, \boldsymbol{\theta})$. MCMC methods allow to perform such estimation from

samples of unknown parameters that are distributed according to the distribution of interest. Fully *unsupervised* estimation can be considered by sampling $\boldsymbol{q}, \boldsymbol{s}, \mathbf{df}$ and $\boldsymbol{\theta}$ from the *joint* distribution: $p(\boldsymbol{q}, \boldsymbol{s}, \mathbf{df}, \boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{q}, \boldsymbol{s}, \mathbf{df}|\boldsymbol{y}, \boldsymbol{\theta})p(\boldsymbol{\theta})$. With appropriate prior distributions $p(\boldsymbol{\theta})$ on the hyperparameters, the global computational cost of the procedure is unchanged. We consider here a uniform prior on γ and Jeffreys' uninformative priors [16] on the variances: $p(\sigma^2) \propto \sigma^{-2}$, $p(\sigma_\epsilon^2) \propto \sigma_\epsilon^{-2}$. Such choices guarantee the integrability of $p(\boldsymbol{q}, \boldsymbol{s}, \mathbf{df}|\boldsymbol{y}, \boldsymbol{\theta})p(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, which is necessary to sample correctly the joint distribution [16]. Then, no parameter needs to be tuned in the procedure.

5.4. Estimation strategy

To obtain samples $(\boldsymbol{q}^{(u)}, \boldsymbol{s}^{(u)}, \mathbf{df}^{(u)}, \boldsymbol{\theta}^{(u)}) \sim p(\boldsymbol{q}, \boldsymbol{s}, \mathbf{df}, \boldsymbol{\theta}|\boldsymbol{y})$, $u = 1 \dots U$, we use a Gibbs sampler, with additional Metropolis–Hastings steps to sample the frequency shifts (see [17] for details). Estimation is then performed in the following way.

(i) First, estimate the posterior mean of the Bernoulli sequence:

$$\widehat{\boldsymbol{q}}_{\mathrm{PM}} = \frac{1}{U}\sum_u \boldsymbol{q}^{(u)} \simeq \mathbb{E}[\boldsymbol{q}|\boldsymbol{y}].$$

$\widehat{\boldsymbol{q}}_{\mathrm{PM}}$ takes values between 0 and 1, and its kth element is an estimation of the probability of detection of a spectral line in the interval $]f_{k-1}, f_k]$.

(ii) Threshold $\widehat{\boldsymbol{q}}_{\mathrm{PM}}$ by keeping only the locations corresponding to a significant *confidence level*, *e.g.* $\alpha = 95\%$: build $\boldsymbol{q}_\alpha$ with ones where $\widehat{\boldsymbol{q}}_{\mathrm{PM}} > \alpha$ and zeros elsewhere.

(iii) Estimate the posterior mean $\widehat{\mathbf{df}}_{\mathrm{PM}}$ of the frequency shifts and their variances by computing means and variances of samples $\mathbf{df}^{(u)}$ for the configurations $\boldsymbol{q}^{(u)} = \boldsymbol{q}_\alpha$.

(iv) Estimate similarly the posterior mean of the hyperparameters, $\widehat{\boldsymbol{\theta}}_{\mathrm{PM}}$. Finally, estimate the amplitude parameters: $(\boldsymbol{s} \mid \boldsymbol{q} = \boldsymbol{q}_\alpha, \mathbf{df} = \widehat{\mathbf{df}}_{\mathrm{PM}}, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_{\mathrm{PM}})$ is Gaussian, so its mean and covariance have analytical expressions.

6. Simulation results

We first show that the proposed methods are not as sensitive toward sampling artifacts as prewhitening methods are. Then, we illustrate their behaviour on more realistic data.

6.1. Robustness toward sampling artifacts

Both methods were applied to the critical example introduced in Section 2. Results are shown in Fig. 2, that reveal little sensitivity of the methods toward sampling artifacts: the two spectral lines are correctly located in frequency, while prewhitening methods are trapped by the aliased maximum in the Fourier spectrum.[1]

6.2. Results on realistic artificial data

The signal in Fig. 3 is the sum of five sinusoids ranging from 3 to 4.8 c/d, with 10 dB additive Gaussian noise. The $N = 250$ data points cover 16 days and the data are irregularly sampled, with additional daily gaps, so that the spectral window shows high sidelobes at 1 c/d. The Fourier spectrum of this *observed* data is thus unreadable.

[1] Prewhitening results were obtained using the *Period04* package: http://www.univie.ac.at/tops/Period04.

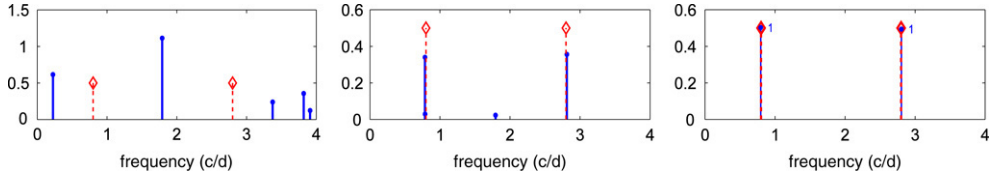


Fig. 2. Estimation results on the critical data from Fig. 1. Left: results of a prewhitening strategy. Centre: spectrum $|\widehat{\boldsymbol{x}}|$ obtained by the ℓ^1-penalization approach. Right: spectrum obtained by the Bernoulli–Gaussian approach, with their estimated probabilities $\widehat{\boldsymbol{q}}_{\mathrm{PM}}$. The diamonds indicate the true spectral lines.

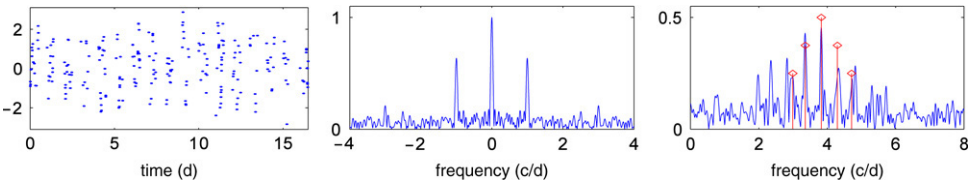


Fig. 3. Artificial data. Available time series (left), spectral window (centre) and Fourier spectrum (right). The diamonds indicate the true spectral lines.

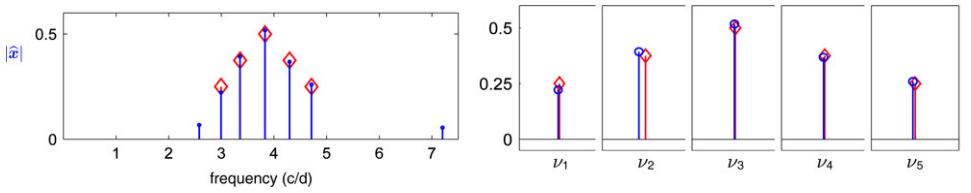


Fig. 4. Prewhitening results on the artificial data from Fig. 3. Left: whole frequency range. Right: zoom on the estimated spectrum around each true frequency; the width of every window is 0.03 c/d.

Note that a well-monitored prewhitening procedure leads to quite acceptable results, with low-valued artifacts, however, as shown in Fig. 4.

Fig. 5 shows the results obtained with the ℓ^1-penalization approach. The number of positive frequencies was set to $K_{\ell^1} = 1000$ and parameter λ was tuned heuristically to 20% of the maximum of the Fourier spectrum of the data (see Section 4.2). The estimated spectrum is sparse, with non-zero values only at the closest approximation of the true frequencies on the grid. Fig. 5 right shows that posterior amplitude re-estimation also gives accurate amplitudes. Note that the frequency precision may also be improved once the lines have been correctly detected, but this is out of the scope of this paper.

The results of the Bernoulli–Gaussian estimation are shown in Fig. 6, with $K_{\mathrm{BG}} = 500$ frequencies. The probabilities of detection are estimated to 1 for the five correct frequency locations, and very low elsewhere. The introduction of continuous-valued frequency variables allows to considerably improve the frequency precision. Note that the minimizer of criterion (5) for ℓ^1-penalization was computed in a few seconds, whereas several minutes were necessary to draw a sufficient number of samples of the Bernoulli–Gaussian posterior distribution, but the latter method yields much more information such as detection probabilities and uncertainties on the parameters.

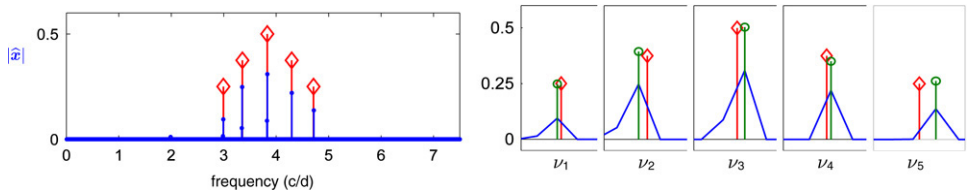


Fig. 5. Results for the ℓ^1-penalization approach on the artificial data from Fig. 3. Left: whole frequency range. Right: zoom around each true frequency; the width of every window is $4 f_{\max}/K_{\ell 1} = 0.03$ c/d. The full line indicates estimator $|\widehat{x}|$ and the circles correspond to the posterior amplitude re-estimation. The diamonds indicate the true spectral lines.

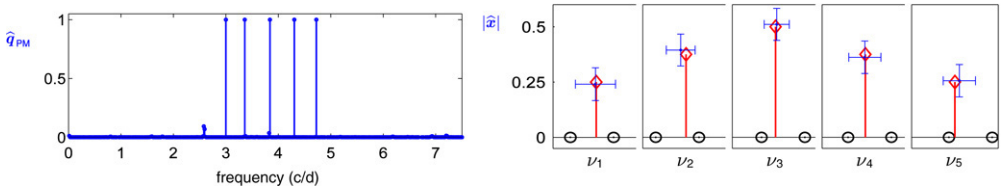


Fig. 6. Results for the Bernoulli–Gaussian approach on the artificial data from Fig. 3. Left: estimated probabilities $\widehat{\boldsymbol{q}}_{\mathrm{PM}}$ for the whole frequency range. Right: zoom on the estimated spectrum around each true frequency; the width of every window is $2 f_{\max}/K_{\mathrm{BG}} = 0.03$ c/d, and the circles locate the frequency grid. Error bars correspond to $\pm 3\widehat{\sigma}$, where $\widehat{\sigma}$ is the estimated standard deviation for each parameter.

7. Conclusion

We proposed two approaches to the problem of estimating multiple sinusoids from irregularly sampled data. The use of a model with a high number of frequencies on a grid, combined with adequate sparse modelling, allows to define robust estimators, with accurate frequency location. In particular, compared to classical sequential prewhitening methods, more robustness is achieved toward the sampling artifacts caused by periodic gaps in the data.

The first method is based on ℓ^1-penalization. An efficient optimization algorithm can be used, so that the frequency axis can be discretized with a very high precision. The resulting estimator is obtained at a low computational cost and its accuracy depends on the correct tuning of the regularization parameter. Practical rules generally perform well for this. On the contrary, the MCMC procedure for the Bernoulli–Gaussian approach is fully unsupervized and provides additional probabilities of detection and estimation of uncertainties. Obtaining such information has a price, however: it requires a much higher computational load and needs critical points to be tackled that are still open, such as automatic convergence monitoring of the MCMC algorithm.

Many perspectives are expected for both methods. First, simulations are needed to assess estimators properties such as resolution limit, sensitivity to noise and to model errors. Then, model generalizations open challenging questions, such as the estimation of non-stationarities (evolution of the frequency with the time, temporary excitation of the oscillation modes) or the identification of periodic signals. Researchers are invited to test both methods intensively and send feed-back, comments and suggestions to the authors. An implementation of the first method is available at http://www.ast.obs-mip.fr/SparSpec. A MATLAB code for the second one is available on request.

References

[1] J.D. Scargle, Studies in astronomical time series analysis. II—Statistical aspects of spectral analysis of unevenly spaced data, The Astrophysical Journal 263 (1982) 835–853.
[2] S. Ferraz-Mello, Estimation of periods from unequally spaced observations, The Astronomical Journal 86 (1981) 619–624.
[3] D.H. Roberts, J. Lehar, J.W. Dreher, Time series analysis with CLEAN. I. Derivation of a spectrum, The Astronomical Journal 93 (4) (1987) 968–989.
[4] G. Foster, The Cleanest Fourier spectrum, The Astronomical Journal 109 (4) (1995) 1889–1902.
[5] S. Bourguignon, H. Carfantan, T. Böhm, SparSpec: A new method for fitting multiple sinusoids with irregularly sampled data, Astronomy & Astrophysics 462 (2007) 379–387.
[6] F. Dublanchet, Contribution de la méthodologie bayésienne à l'analyse spectrale de raies pures et à la goniométrie haute résolution, Ph.D. Thesis, Université de Paris-Sud, October 1996.
[7] M.D. Sacchi, T.J. Ulrych, C.J. Walker, Interpolation and extrapolation using a high-resolution discrete Fourier transform, IEEE Transactions on Signal Processing 46 (1) (1998) 31–38.
[8] P. Ciuciu, J. Idier, J.-F. Giovannelli, Regularized estimation of mixed spectra using a circular Gibbs–Markov model, IEEE Transactions on Signal Processing 49 (10) (2001) 2201–2213.
[9] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1) (1999) 33–61.
[10] J.J. Fuchs, Recovery of exact sparse representations in the presence of bounded noise, IEEE Transactions on Information Theory 51 (10) (2005) 3601–3608.
[11] J. Kormylo, J.M. Mendel, Maximum-likelihood detection and estimation of Bernoulli–Gaussian processes, IEEE Transactions on Information Theory 28 (1982) 482–488.
[12] P. Stoica, R.L. Moses, B. Friedlander, T. Söderström, Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements, IEEE Transactions on Acoustics, Speech and Signal Processing 37 (3) (1989) 378–392.
[13] S. Sardy, A. Bruce, P. Tseng, Block coordinate relaxation methods for nonparametric wavelet denoising, Journal of Computational and Graphical Statistics 9 (2000) 361–379.
[14] S. Bourguignon, Analyse spectrale à haute résolution de signaux irrégulièrement échantillonnés : Application à l'Astrophysique, Ph.D. Thesis, Université Paul Sabatier – Toulouse III, France, December 2005.
[15] Q. Cheng, R. Chen, T.-H. Li, Simultaneous wavelet estimation and deconvolution of reflection seismic signals, IEEE Transactions on GeoScience and Remote Sensing 34 (2) (1996) 377–384.
[16] C. Robert, Méthodes de Monte-Carlo par Chaînes de Markov, Economica, Paris, 1996.
[17] S. Bourguignon, H. Carfantan, Spectral analysis of irregularly sampled data using a Bernoulli–Gaussian model with free frequencies, in: IEEE ICASSP, Toulouse, France, May 2006, pp. 516–519.