REGULARIZED SPECTRAL ANALYSIS OF UNEVENLY SPACED DATA

Sébastien Bourguignon, Hervé Carfantan, Loïc Jahan

Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées, UMR 5572 CNRS/UPS
14 avenue Edouard Belin, 31400 TOULOUSE, France
bourgui@ast.obs-mip.fr, herve.carfantan@obs-mip.fr, loic.jahan@obs-mip.fr

ABSTRACT

High resolution spectral analysis has recently been addressed as an inverse problem, and solutions are currently proposed through the regularization framework. In this paper, we focus on regularized spectral analysis of unevenly sampled data for line spectra estimation. First, we study the structural differences of the model between regular sampling, missing data (where the sampling is regular, but with missing data) and irregular sampling cases. Then, consequences for the computation of the solution are emphasized. We propose an approximation of the irregular sampling model to compute the non quadratic regularization solution at a cost comparable to the other sampling cases. Finally, algorithmic implementation is discussed and applications to simulated data are presented.

1. INTRODUCTION

Spectral analysis is a classic signal processing problem and many parametric and non parametric methods have been proposed. Considering spectral analysis as a linear inverse problem, solutions have recently been proposed through the regularization framework [1, 2], which removed the traditional incompatibility between "high resolution" and "non parametric" methods. These works consider regularly sampled data, and only [1] mentioned the unevenly spaced data situation.

Spectral analysis of unevenly spaced noisy data is a very important topic in astrophysics and the search for line spectra is particularly interesting. In asteroseismology, for example, determining the pulsation modes of variable stars from their light curves allows to estimate some physical parameters, such as mass, temperature and chemical composition. Because of observation constraints, astronomical data are generally unevenly spaced. We deal with two kinds of unevenly spaced data. First, short time observations - from a few minutes to several days - are generally regularly sampled, but observation conditions (meteorology, object visibility…) might generate missing data. Then, long time observations - up to many years - are generally fully irregularly sampled.

In this paper we focus on the regularized spectral analysis of unevenly spaced data and study the consequences of the different sampling schemes. After stating this problem in the regularization framework as [1, 2] for line spectra estimation, we focus on some structural properties (Toeplitz, circulant…) of matrices involved in the computation of regularized solutions. This shows fundamental differences between the particular missing data case and the general irregular sampling case. We see how one can take advantage of such structural properties to compute exact or approximate solutions at a low computational cost. After discussing the algorithmic implementation, we finally show on simulations that accounting for these structures provides an accurate estimation of line spectra at a lower cost than other existing algorithms for irregularly sampled data.

2. REGULARIZATION FRAMEWORK

We consider the estimation of spectrum $\boldsymbol{x}$ from noisy data sampled at times $\{t_n\}_{n=1\dots N}$:

$$\begin{aligned} y(t_n) &= \sum_{k\in\mathcal{K}} x_k e^{j2\pi\frac{k}{P}f_{\max}t_n} + \epsilon_n \\ \Leftrightarrow \boldsymbol{y} &= \mathbf{W}\boldsymbol{x} + \boldsymbol{\epsilon} \end{aligned} \quad (1)$$

where $\boldsymbol{x}$ collects the complex spectral amplitudes $\{x_k\}_{k\in\mathcal{K}}$ associated to frequencies on the grid $\{\frac{k}{P}f_{\max}\}_{k\in\mathcal{K}}$ with $\mathcal{K} = (-P\dots P)$. To deal with an ill-posed problem – ill-conditioned or rank-deficient operator $\mathbf{W}$ if high resolution is required – a classic regularization framework consists in minimizing a penalized least squares criterion:

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2 + \lambda R(\boldsymbol{x}). \quad (2)$$

In the Bayesian framework, it corresponds to the maximum *a posteriori* (MAP) estimation of $\boldsymbol{x}$ for additive *i.i.d.* centered complex circular gaussian noise $\boldsymbol{\epsilon}$ and prior $p(\boldsymbol{x}) \propto e^{-\lambda R(\boldsymbol{x})}$, and hyperparameter $\lambda > 0$ balances between fidelity to data and confidence in prior information.

Many regularization functions have been designed to emphasize the searched line spectra characteristic. The l_1 regularization ($R(\boldsymbol{x}) = \sum_k |\boldsymbol{x}_k|$) is a standard that actually yields sparseness [3]. The non 0-differentiability of the

resulting criterion, however, requires sophisticated numerical tools, that may result very expensive for high dimension problems. In a Bayesian approach, [1] use a Cauchy prior on $\boldsymbol{x}$, producing a non convex criterion that may have local minima. In this paper we follow [2] by using the convex differentiable function $R(\boldsymbol{x}) = \sum_k \sqrt{s^2 + |x_k|^2}$. For low s it provides a very similar solution to that of l_1 regularization, while leading to a strictly convex and differentiable criterion J with no local minima. Optimization can then be performed using any descent algorithm, which require the computation of the criterion and its gradient:

$$\overrightarrow{\nabla J}(\boldsymbol{x}) = -2\mathbf{W}^\dagger(\boldsymbol{y} - \mathbf{W}\boldsymbol{x}) + \lambda\overrightarrow{\nabla R}(\boldsymbol{x})$$

Thus, products $\mathbf{W}\cdot$ and $\mathbf{W}^\dagger\mathbf{W}\cdot$ have to be performed at the lowest cost to reduce the computational cost of the solution. Next sections show how one can take advantage of the structural properties of matrix $\mathbf{W}$ *w.r.t.* the different sampling schemes.

3. OPERATOR STRUCTURE

3.1. Regular sampling

For regularly sampled data at period T_s, one can set $f_{\max}$ at the Nyquist limit $1/2T_s$ and $\mathcal{K} = \{-P+1, \ldots, P\}$, as $k = -P$ and $k = P$ give the same contribution at frequency $f_{\max}$. Relation (1) writes:

$$y(t_n) = y(nT_s) = \sum_{k=-P+1}^{P} x_k e^{j\pi\frac{kn}{P}} + \epsilon_n$$

If $2P = N$, matrix $\mathbf{W}$ is of course the Fourier kernel $\mathbf{F}_{2P} = \{\exp j\pi\frac{kn}{P}\}_{k=-P+1\ldots P}^{n=1\ldots 2P}$, relation above corresponds to the inverse Discrete Fourier Transform (DFT^{-1}) of $\boldsymbol{x}$ and $\mathbf{F}_{2P}^\dagger\mathbf{F}_{2P} = \mathbf{F}_{2P}\mathbf{F}_{2P}^\dagger = 2P\mathbf{I}_{2P}$. In the high resolution case, *i.e.* $2P \gg N$, rank-deficient operator writes $\mathbf{W} = \mathbf{Z}_{N,2P} = \{\exp j\pi\frac{kn}{P}\}_{k=-P+1\ldots P}^{n=1\ldots N}$ and matrix $\mathbf{Z}_{N,2P}^\dagger\mathbf{Z}_{N,2P} = \mathbf{F}_{2P}^\dagger\mathbf{D}_Z\mathbf{F}_{2P}$ is *circulant* (where $\mathbf{D}_Z$ is a diagonal matrix with N ones and $2P - N$ zeros).

3.2. Missing data

We now consider a regular sampling scheme with missing data: $t_n = i_n T_s$, $i_n \in \mathbb{N}$. Operator $\mathbf{W}$ writes:

$$\begin{aligned}\mathbf{M}_{N,2P} &= \{\exp j\pi\frac{ki_n}{P}\}_{k=-P+1\ldots P}^{n=1\ldots N} \\ &= \mathbf{L}_{N,2P}\mathbf{F}_{2P}\end{aligned}$$

where $\mathbf{L}_{N,2P}$ is built with lines $\{i_n\}_{n=1\ldots N}$ of identity matrix $\mathbf{I}_{2P}$. Thus, $\mathbf{M}_{N,2P}{}^\dagger\mathbf{M}_{N,2P} = \mathbf{F}_{2P}{}^\dagger\mathbf{D}_M\mathbf{F}_{2P}$ is still *circulant*, where D_M is a diagonal matrix with ones for indexes i_n and zeros elsewhere.

In both regular sampling and missing data cases, $\mathbf{W}$ writes with Fourier operators and $\mathbf{W}^\dagger\mathbf{W}$ is circulant. Thus, products $\mathbf{W}\cdot$ and $\mathbf{W}^\dagger\mathbf{W}\cdot$, then $J(\boldsymbol{x})$ and $\overrightarrow{\nabla J}(\boldsymbol{x})$, can be computed by Fast Fourier Transform (FFT) algorithms. Note also that relation $\mathbf{W}\mathbf{W}^\dagger = 2PI_N$ holds, therefore in those cases the generalized inversion of (1) leads to:

$$\hat{\boldsymbol{x}}_{GI} \triangleq \mathbf{W}^\dagger(\mathbf{W}\mathbf{W}^\dagger)^{-1}\boldsymbol{y} = \frac{1}{2P}\mathbf{W}^\dagger\boldsymbol{y}$$

which corresponds to the DFT of the zero-padded and zero-substituted data. That is, $\hat{\boldsymbol{x}}_{GI}$ is a frequency sampling of the spectrum of the signal convolved by the frequency response of the observation window. Because of the missing data, the latter has no more a sinc-like shape and may have many undesirable secondary lobes. Therefore, this solution is not acceptable. Moreover the use of windowing to lessen the lobes is not as simple as in the full sampling case.

3.3. Irregular sampling

In the general case of an irregular sampling scheme $\{t_n\}_n$, there is no theoretical maximum frequency since the spectrum is not periodic any more [4]. Thus, parameter $f_{\max}$ has to be set according to some physical prior knowledge and operator $\mathbf{W}$ writes:

$$\mathbf{W} = \mathbf{U}_{N,2P+1} = \{\exp j2\pi\frac{k}{P}f_{\max}t_n\}_{k=-P\ldots P}^{n=1\ldots N}$$

It looses the previous Fourier-like shape, which is severe for the computational cost: products $\mathbf{W}\cdot$ cannot be computed by FFT algorithms any more[1] and the computation of J requires order P^2 operations. It can be shown that:

$$\{\mathbf{U}_{N,2P+1}^\dagger\mathbf{U}_{N,2P+1}\}_{k,l} = \sum_n e^{-j2\pi\frac{(k-l)f_{\max}}{P}t_n}$$

is still Toeplitz but not circulant, and this structure can still be used to perform products $\mathbf{W}^\dagger\mathbf{W}\cdot$ at low cost.

4. MODIFIED CRITERION

To reduce the cost in the general irregular sampling case, we introduce the modified criterion (for sake of clarity, in the following we write $\mathbf{U}$ for $\mathbf{U}_{N,2P+1}$):

$$J_2(\boldsymbol{x}) = ||\mathbf{U}^\dagger\boldsymbol{y} - \mathbf{U}^\dagger\mathbf{U}\boldsymbol{x}||^2 + \lambda R(\boldsymbol{x}) \qquad (3)$$

As matrix $\mathbf{U}^\dagger\mathbf{U}$ is Toeplitz, products $\mathbf{U}^\dagger\mathbf{U}\cdot$ can be computed using $(4P+2)$-point FFTs by embedding the matrix in a twice bigger circulant one. One can easily verify also that $\overrightarrow{\nabla J_2}(\boldsymbol{x})$ can be computed at a reasonable cost. Note that the quadratic term in (3) writes:

$$Q_U(\boldsymbol{x}) = (\boldsymbol{y} - \mathbf{U}\boldsymbol{x})^\dagger(\mathbf{U}\mathbf{U}^\dagger)(\boldsymbol{y} - \mathbf{U}\boldsymbol{x})$$

with:

$$\{\mathbf{U}\mathbf{U}^\dagger\}_{k,l} = \frac{\sin(\pi(t_k - t_l)(2P+1)\frac{f_{\max}}{P})}{\sin(\pi(t_k - t_l)\frac{f_{\max}}{P})} = \gamma_P(t_k - t_l)$$

[1] Approximations of such a product using FFTs exist but are out of scope in this paper

where $\gamma_P(t)$ is the Dirichlet kernel. Thus, criteria $J_2(\boldsymbol{x})$ and $J(\boldsymbol{x})$ are not equivalent since $\mathbf{U}\mathbf{U}^\dagger$ is not proportional to identity. Actually, relation $\mathbf{U}\mathbf{U}^\dagger \propto \mathbf{I}_N$ holds if and only if the sampling is regular with eventual missing data ($t_n = i_n T_s$) and $f_{\max} = \frac{P}{2P+1}\frac{1}{T_s}$: in that case instants $\{t_k - t_l\}_{k\neq l}$ correspond to zeros of γ_P.

We can show, however, that the extradiagonal terms in $\mathbf{U}\mathbf{U}^\dagger$ almost always have negligeable values, therefore the approximation has no real impact on the solution. For example, let $\{t_n\}$ be uniformly distributed on $[O, T_{\max}]$. It is easy to show that the proportion of instants $\{t_k - t_l\}_{k\neq l}$ in the principal lobe of $\gamma_P - 1/f_{\max}$ width – is less than $\frac{1}{T_{\max} f_{\max}}$. Considering that a reasonable value for $f_{\max}$ would be *at least* $f_{\max} = \frac{1}{2}\frac{N}{T_{\max}}$ (which would be the Nyquist frequency if the N samples were evenly spaced), we show that less than $\frac{2}{N}\%$ instants $\{t_k - t_l\}_{k\neq l}$ are in the first lobe of γ_P. Hence, although minimization of $J_2(\boldsymbol{x})$ does not give exactly the same solution that minimization of $J(\boldsymbol{x})$, differences between both minimizers are expected to be not significant, which will show to be true in practice.

5. ALGORITHMIC CONSIDERATIONS

5.1. Different algorithms

Several specific algorithms have been proposed to minimize such a penalized criterion for line spectra estimation, that are mostly designed for regular sampling [1, 2]. Each iteration of such algorithms consists in solving a quadratic regularization problem, *i.e.* essentially computing a matrix inversion. Thus, these algorithms are particularly efficient when the associated matrix is *circulant*, which is not true for irregular sampling.

Here we propose to exploit the particular structure mentioned in the last section associated with descent algorithms. Recall that, for irregular sampling, the main cost is associated to the calculation of J ($\mathbf{W}$ is no more a Fourier-like matrix). The use of a conjugate gradient without line search (CGWLS) algorithm [5] is a way to avoid such a computation and is only based on successive computations of $\overrightarrow{\nabla J}$. The associated cost per iteration is then essentially that of one product $\mathbf{W}^\dagger\mathbf{W}\cdot$. However, it requires the computation of the highest eigenvalue of $\mathbf{W}^\dagger\mathbf{W}$ to state the steplength and shows slow convergence speed on simulations.

A conjugate gradient (CG) algorithm (Polak-Ribière version) can be implemented also, where the minimization *w.r.t.* each descent direction is performed by *e.g.* parabolic approximation. Each iteration requires essentially two products $\mathbf{W}\cdot$ and one product $\mathbf{W}^\dagger\mathbf{W}\cdot$ when associated to criterion J, and three products $\mathbf{W}^\dagger\mathbf{W}\cdot$ for the modified criterion J_2. Thus, this algorithm is expected to show a lower cost when associated to the minimization of criterion J_2.

5.2. Real data case

Up to now we considered a model with a *circular* complex gaussian noise: $\boldsymbol{\epsilon}$ may be complex since the hermitian symmetry of spectrum $\boldsymbol{x}$ ($x_{-k} = x_k^*$) is not guaranteed. If the data $\boldsymbol{y}$ are real, it has been shown, however, that for regular sampling the minimizer of criterion J is hermitian as J is convex [2]. This result easily extends to the minimizers of criteria J and J_2 in the general irregular sampling case.

Nevertheless, real data can be accounted for by using a model with only positive frequencies, *i.e.*:

$$\boldsymbol{y} = \mathbf{W}_+\boldsymbol{x}_+ + \mathbf{W}_+^*\boldsymbol{x}_+^* + \boldsymbol{\epsilon} = 2\Re(\mathbf{W}_+\boldsymbol{x}_+) + \boldsymbol{\epsilon} \qquad (4)$$

where $\boldsymbol{x}_+ = \{x_k\}_{k=O\ldots P}$,$\mathbf{W}_+ = \{e^{j2\pi\frac{k}{P}f_{\max}t_n}\}_{k=O\ldots P}^{n=1\ldots N}$ and $\boldsymbol{\epsilon}$ is a *real* gaussian noise. The corresponding criterion and gradient write respectively:

$$\begin{aligned} J_r(\boldsymbol{x}_+) &= ||\boldsymbol{y} - 2\Re(\mathbf{W}_+\boldsymbol{x}_+)||^2 + \lambda R(\boldsymbol{x}) \\ \overrightarrow{\nabla J}_r(\boldsymbol{x}_+) &= -4{\mathbf{W}_+}^\dagger(\boldsymbol{y} - 2\Re(\mathbf{W}_+\boldsymbol{x})) \end{aligned}$$

One can show that $\mathbf{W}_+^\dagger\mathbf{W}_+$ is still a Toeplitz matrix and $\mathbf{W}_+^\dagger\mathbf{W}_+^*$ is a Hankel matrix. Therefore, similar algorithmic conclusions hold concerning the minimization of both original and modified corresponding criteria. As model (4) deals with twice less variables than the complex data model (1), the resulting computational costs are then reduced.

6. SIMULATION RESULTS

Signal presented on figure 1 is the sum of 5 sinusoids corrupted by 10dB white gaussian noise. The $N = 250$ sampling times span $T_{\max} = 500$ days and are uniformly distributed on $[O, T_{\max}]$ with additional periodic gaps. Frequencies are set off the reconstruction grid and $f_{\max} = \frac{N}{2T_{\max}} = 0.25\ \text{day}^{-1}$. The estimation given by the non-uniform Discrete Fourier Transform is not satisfactory for peak detection, as shown on figure 1.

Minimization of criterion J is performed by CGWLS and CG algorithms, and a CG algorithm is used to compute the minimizer of modified criterion J_2. Optimal value of hyperparameter λ is selected by Hansen's L-curve criterion, which gave satisfactory results despite the lack of theoretical framework for non quadratic regularization. Both minimizers correctly localize the 5 frequencies and only slightly differ in the associated amplitudes. The minimizer of the modified criterion J_2 is presented on figure 2, in its *real data* version with $s = 10^{-3}$, $\lambda = 100$ and $P = 1000$. Figure 2(c) shows that the matrix $\mathbf{W}\mathbf{W}^\dagger$ correponding to the sampling scheme is almost diagonal, which explains the similitudes between both solutions. We note that both minimizers of J and J_2 suffer a loss in the amplitude estimation, which is inherent in the regularization process. Once the frequencies are correctly localized, however, the amplitudes of the corresponding sinusoidal model can be reestimated in terms of least squares (Fig. 2(d)).

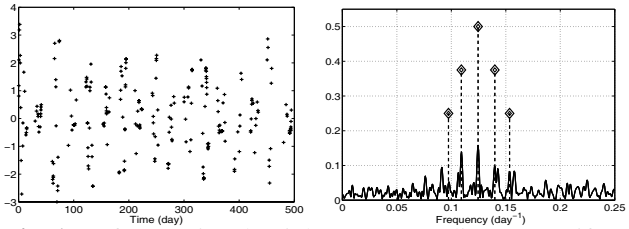
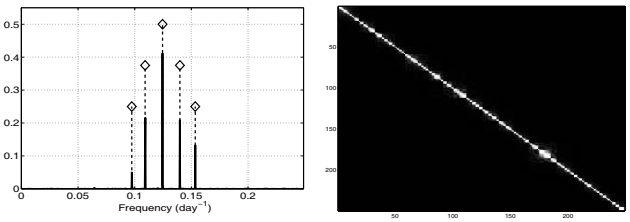
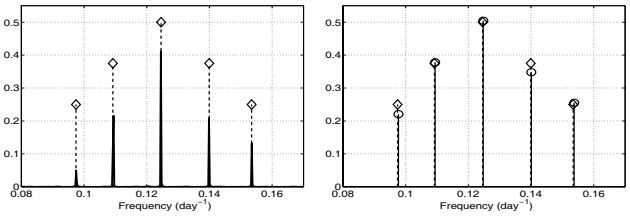


Fig. 1. Left: test signal. Right: corresponding non-uniform DFT (–) and theoretic spectrum ($\diamond$) .



(a) Estimated $|\boldsymbol{x}|$ (–) and theoretic spectrum ($\diamond$)

(c) Representation of $|\mathbf{W}\mathbf{W}^{\dagger}|$

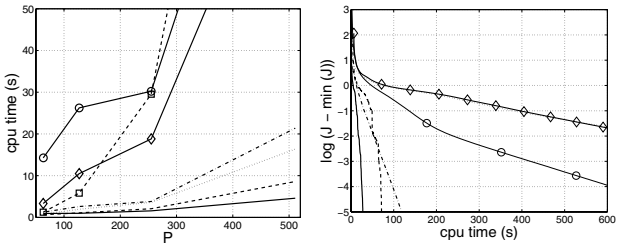


(b) Zoom on the 5 frequencies

(d) Amplitude reestimation (o) and theoretic spectrum ($\diamond$)

Fig. 2. Optimization results.

Convergence speed is compared with that of a single site update algorithm (SSUA) with a half-quadratic 1D minimization [6], which showed ability to retrieve sparse solutions, and with the IRLS algorithm presented in [1]. As they do not use the above properties, those algorithms are implemented to minimize the original criterion J. A quadratic program (QP) performing l_1 regularization is also implemented. Algorithms are tested on the previous signal, with $s = 10^{-3}$ and $\lambda = 100$, and stop at iteration m when $||\boldsymbol{x}^{(m)} - \boldsymbol{x}^{(m-1)}|| \leq \alpha$ and $J(\boldsymbol{x}^{(m)}) - J(\boldsymbol{x}^{(m-1)}) \leq \alpha$ with $\alpha = 10^{-6}$. Figure 3 shows the performances obtained by the proposed algorithms (CGWLS and CG associated to the modified criterion J_2) compared to that of SSUA, IRLS and QP. These algorithms are implemented with MATLAB, excepted QP which uses Mosek C library (`http://www.mosek.com`). The use of the modified criterion proposed in section 4 allows to reduce the computational cost of the solution, which is approximately divided by half when a model with real data as (4) is considered. The advantage of this approximation is more visible as dimension P increases. Note, however, that the selection of a lower s value dramatically increases the computational cost as the corresponding criterion tends to be not differentiable.



(a) Computation time as a function of dimension P

(b) Evolution of criteria ($P = 1000$)

Fig. 3. Algorithm performances: CG on criterion J ($\diamond$), IRLS (o), QP ($\square$), SSUA ($-\cdot$), CGWLS ($\cdots$) and CG on criterion J_2: complex data (- -) and real data (–)

7. CONCLUSION

We studied the regularized spectral analysis of unevenly sampled data. Matrices involved in the computation of the solution have shown computationnally interesting properties in the regular sampling and missing data cases, which do not generalize to the irregular sampling case. In the latter case, the criterion cannot be computed exactly using FFT, which can still be used to compute the gradient. We proposed an approximation of the initial model by introducing a modified criterion that can be computed with FFT. Therefore the regularized solution can be computed using a Conjugate Gradient algorithm at a computation cost comparable to the regular sampling and missing data cases. Such an approximation has shown on simulation results to have no impact on the solution, which provides accurate estimation of line spectra.

8. REFERENCES

[1] M.D. Sacchi, T.J. Ulrych, and C.J. Walker, Interpolation and extrapolation using a high-resolution discrete Fourier transform," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 31–38, January 1998.

[2] P. Ciuciu, J. Idier, and J.-F. Giovannelli, Regularized estimation of mixed spectra using a circular Gibbs-Markov model," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2201–2213, October 2001.

[3] J.-J. Fuchs, On sparse representations in arbitrary redundant bases," *IEEE Transactions on Information Theory*, vol. 50, pp. 1341–1344, June 2004.

[4] L. Eyer and P. Bartholdi, Variable stars: which Nyquist frequency?," *Astronomy and Astrophysics Supplement Series*, vol. 135, pp. 1–3, February 1999.

[5] Jie Sun and Jiapu Zhang, Global convergence of conjugate gradient methods without line search," *Annals of Operations Research*, vol. 103, pp. 161–173, 2001.

[6] S. Brette and J. Idier, Optimized Single Site Update Algorithms for image deblurring," in *Proc. IEEE ICIP*, Lausanne, Switzerland, Sep. 1996, pp. 65–68.