SPECTRAL ESTIMATION OF UNEQUALLY-SPACED DATA VIA RADIAL BASIS FUNCTION INTERPOLATION

K. Nedir, J.-M. Vesin, and L. Eyer*
Signal Processing Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
* Observatoire de Genève
Genève, Switzerland

RÉSUMÉ

Un problème couramment rencontré dans de nombreux domaines (économie, médecine, astronomie, ...) est celui de l'estimation spectrale de données échantillonnées irrégulièrement. Nous proposons dans cet article une méthode basée sur une interpolation des données par des fonctions radiales, suivie d'un ré-échantillonnage régulier et d'une estimation spectrale classique. Nous montrons qu'il est possible, sous l'hypothèse d'une connaissance au moins partielle de la bande de fréquence utile du signal étudié, de définir un critère permettant d'optimiser l'interpolation effectuée et donc l'estimation spectrale. Nous présentons des résultats de tests réalisés sur des signaux simulés et expérimentaux.

ABSTRACT

A frequently encountered problem in various domains (economics, medicine, astronomy, ...) is the one of spectral estimation of irregularly-sampled data. In this paper we propose a method based on a radial basis function interpolation of the data followed by a regular re-sampling and a classical power density spectrum estimation. We show that it is possible, under the assumption of prior partial knowledge about the frequency range of the signal under study, to define a criterion allowing to optimize the interpolation and thus the spectral estimation. We present results of tests performed both on real and experimental signals.

I. INTRODUCTION

In many applications, one is confronted with the problem of dealing with unequally-spaced data series. There are several causes to irregular sampling: some data points may be missing in otherwise equally spaced data (e.g. economic data available daily, except on Saturdays and Sundays); the data points may inherently be obtained at random locations in time or space; data obtained manually (medical data such as blood pressure or body temperature) or data obtained over a long period of time (astronomic and environmental measurements such as the light variations of a pulsating star, or the stratospheric ozone thickness) are very likely to suffer from unequal spacing, with random sample spacings and possibly extended gaps.

Different methods have been elaborated to analyze such irregularly-spaced data. Stellingwerf [1] carried out a period determination using phase dispersion minimization, which seems well suited to short data series of non-sinusoidal periodic signals. Shumway [2] studied maximum likelihood estimators for parameters in missing data problems. Deeming [3] stated that the discrete Fourier transform of the irregularly-spaced samples is the convolution of the Fourier transform of the unknown underlying continuous signal from which only unequally-spaced samples are known, with a spectral window depending only on the times of observation.

In this paper, we propose a spectral analysis scheme via an interpolation of the data. The approach consists first in interpolating the available irregularly-spaced data samples, then in performing digital spectral analysis on the regularly re-sampled interpolated signal. Some special feature of the spectral analysis is used as a criterion to select the free parameter of the interpolation process. Specifically, it is assumed that a frequency band can be provided, out of which the analyzed signal is very unlikely to have significant spectral components. The interpolation scheme chosen is a Radial Basis Function (RBF) one [4, 5] with a free parameter designated by s. The value of s is taken so as to maximize the power spectrum of the interpolated signal in the specific frequency band selected. Note that, although spectral estimation was the main motivation for this study, a direct byproduct is data interpolation itself, a problem of current interest in many branches of time series analysis [8].

Section II first reviews briefly the features of RBF interpolation in the special case of Gaussian radial functions. The interpolation optimization via the spectral band power criterion is then introduced. Finally, some results of the method are illustrated in Section III on simulated signals and in section IV on real astronomical data.

II. OUTLINE OF THE METHOD

A. Radial Basis Function Interpolation

The Radial Basis Function (RBF) interpolation provides an approximation of a function given a sequence of samples of this function at irregularly-spaced time locations. We shall concentrate here on the real-valued, univariate case. Generalization and other relevant issues may be found in Powell [4].

Let us denote by $\{f(t_i) : i = 1,2,\ldots,N\}$ the available data samples, and by $\{t_i : i = 1,2,\ldots N\}$ the strictly increasing set of sampling times. The interpolated continuous function v is then given by

$$v(t) = \sum_{i=1}^{N} c_i h(|t - t_i|) \quad (1)$$

where the $\{c_i : i = 1,2,\ldots,N\}$ are the *interpolation coefficients* , and where h is a function from $\mathbf{R}^+$ to $\mathbf{R}$. If we define $r_i = |t-t_i|$, then each function $h(r_i)$, called *radial basis function* , depends only on the distance r_i from the reference sample i . In this paper, we shall focus on the *Gaussian case*

$$h(x) = \exp(-x^2/s^2) \quad (2)$$

with standard deviation (sometimes called *bandwidth*) s, though other relevant choices for h exist. The N *interpolation conditions*

$$v(t_i) = f(t_i) \qquad i = 1,2,\ldots,N \quad (3)$$

must be satisfied. With eq. (1), this leads to

$$v(t_k) = \sum_{i=1}^{N} c_i h(|t_k - t_i|) = f(t_k) \quad (4)$$
$$k = 1,2,\ldots,N.$$

This is a square linear system in the c_i coefficients, namely $Ac = v$, in which the elements of matrix A are given by

$$A_{ik} = h(|t_k - t_i|) \quad i,k = 1,2,\ldots,N \quad (5)$$

For many choices of RBFs, the *interpolation matrix* A defined by eq. (5) is guaranteed to be non-singular under very mild or no restrictions on the locations t_i of the interpolation points. Micchelli [6] has shown that the interpolation matrix A is always non-singular for Gaussian RBF if the t_i are all distinct. This important result allows the interpolation coefficients c_i to be uniquely defined.

B. Spectral Criterion for the RBF Interpolation

It should be once again emphasized that the only free parameter of the Gaussian RBF interpolation is the standard deviation s of expression (2), which states that the interpolated signal is a linear combination of N Gaussian functions placed at each of the unequally-spaced times t_i .

The original feature of this paper lies in the way the parameter s of the Gaussian RBF is selected. It is assumed that the frequency components of interest are known a priori to lie in the band $[f_L, f_H]$. The interpolated function $v(t)$ given by eq. (1) is then regularly sampled :

$$v[n] = v(nT_s) \qquad t_1 \le nT_s \le t_N \quad (6)$$

and a power spectral density (PSD) estimate is computed. Either non-parametric or parametric (AR model based) estimators [7] can be used. The effect of the Gaussian parameter s on the power spectrum can be described as follows: (i) For too small a value of s, the signal $v[n]$ is most unsmooth, with non-zero samples only in the close vicinity of the original unequally-spaced points: the resulting PSD is quite flat. (ii) For too large an s, the signal $v[n]$ is oversmoothed, and its PSD is concentrated in the lower frequencies. Hence, the spectral criterion chosen is the power in the band $[f_L, f_H]$ normalized by the total estimated spectrum power. We express this relative power criterion as:

$$p_{[f_L, f_H]} = \int_{f_L}^{f_H} \hat{P}(f)df \Bigg/ \int_{0}^{f_s/2} \hat{P}(f)df \quad (7)$$

where $\hat{P}(f)$ is the PSD estimate and f_S is the regular sampling frequency. The value of s that maximizes eq. (7) is selected as the parameter value for the Gaussian RBF interpolation of the unequally-spaced data, and this yields the corresponding PSD estimate.

It should also be mentioned that the interpolation matrix tends towards singularity as s becomes larger. In such a case, the interpolation is no longer achievable in finite precision arithmetic. This is an upper bound to the range of investigation for s.

C. Use of multiple bandwidth RBF interpolation

It is well known that better interpolation results are obtained if one uses a different RBF at each data point, i.e., eq. (1) is changed to:

$$v(t) = \sum_{i=1}^{N} c_i h_i(|t - t_i|) \quad (8)$$

with:

$$h_i(x) = \exp(-x^2/s_i^2) \quad (9)$$

It is clearly impossible in practice to try to optimize the interpolation with respect to all the parameters $\{s_i\}$. We developed instead a heuristic formulation in which the parameter s_i corresponding to a particular t_i of the set of sampling times is proportional to the average difference between t_i and the sampling times which precede and follow it, that is:

$$s_i = d\,\frac{(t_i - t_{i-1}) + (t_{i+1} - t_i)}{2} \quad (10)$$

This choice can be justified as follows: if the sampling time t_i is close to t_{i-1} and t_{i+1}, then the influence of the corresponding sample should not extend too far, and accordingly s_i must be small. On the other hand, if t_i is isolated, then the influence of the sample must increase, and s_i must be large. An advantage of this scheme is that, like in the above single RBF one, optimization has to be performed on one parameter only, i.e. the scaling factor d in (10). Experiments performed show that this modified method often yields better results than the basic one.

III. RESULTS ON SIMULATIONS

A signal consisting of two sinusoids ($f_1 = 0.10$, $f_2 = 0.13$ Hz) embedded in white noise (SNR=10dB) has been Poisson-sampled (mean of Poisson process = 4 sec.), providing 40 unequally-spaced data samples. A resampling at $f_s = 0.8$ Hz has been performed. Fig.1 shows the optimization curve $p_{[f_L, f_H]}$ with respect to the parameter s , with $f_L = 0.05$ Hz and $f_H = 0.20$ Hz. Figures 2-4 show three AR(20) spectral estimates (in dB vs Hz): Fig.2 corresponds to the case where $s = 0.10$ has been used to interpolate the data, Fig.3 corresponds to the case $s = 4.0$, whereas Fig.4 represents the spectrum obtained in the case $s = 1.8$, which is the optimal value given by figure.1. This last spectrum exhibits two strong peaks corresponding to the two sinusoids of the signal, while the effects of respectively under- and oversmoothing are clearly visible on figures 2 and 3.

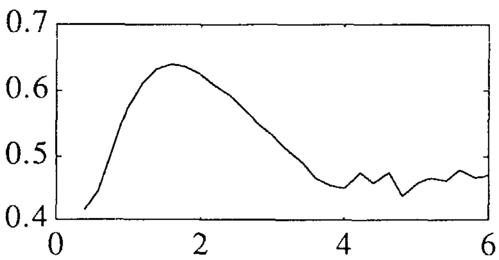

Fig.1 : Relative power in freq. band [0.05, 0.20] vs s

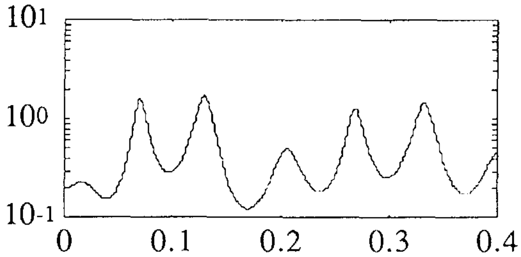

Fig.2 : AR PSD estimate for $s = 0.1$

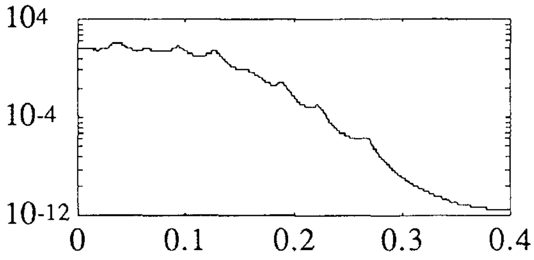

Fig.3 : AR PSD estimate for $s = 4.0$

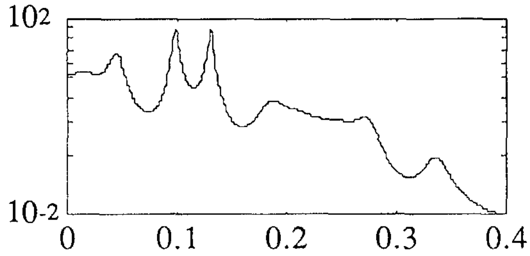

Fig.4 : AR PSD estimate for $s = 1.6$

APPLICATION TO AN EXPERIMENTAL SIGNAL

The multiple RBF method of section II *C* has been applied to a photometric data series obtained from a variable star. This series is composed of 171 samples covering a time period of 555 days, which corresponds to an average of 0.308 samples/day. Standard astronomical tables indicate that the light variation of this star exhibits three spectral components corresponding to frequencies of $f_1 = 0.58\ 10^{-2}$, $f_2 = 0.91\ 10^{-2}$, and $f_3 = 1.37\ 10^{-2}$ [day^{-1}]. The method used to estimate these frequencies has not been cited, and has probably not been employed on this particular photometric series.
The frequency interval was chosen to be $[f_L, f_H] = [0.05, 0.20]$ day^{-1}. Note that in the case of stellar photometric data additional information on the star under study (size, age, ...) usually provides an acceptable frequency interval. Figure 5 shows the evolution of the relative power criterion with respect to the scaling parameter d in (10) or twenty values between 0.1 and 2. A resampling frequency $f_s = 0.05$ was selected and an AR(10) power spectral density estimation was used.

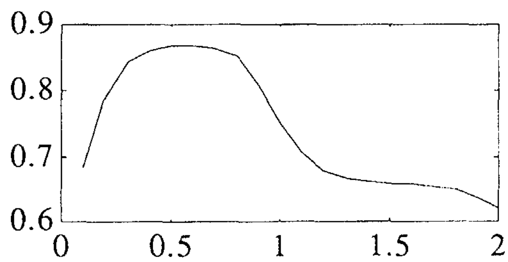

Fig.5 : Relative power in freq. band [0.05, 0.20] day^{-1} vs d

The optimum is reached for $d = 0.6$. This parameter value yields the power spectral density estimate displayed on figure 6.

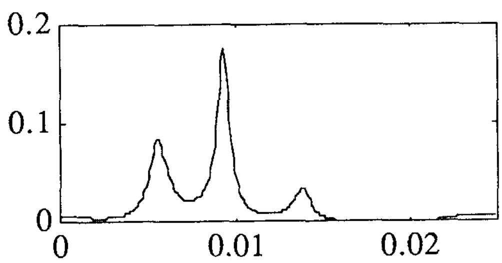

Fig.6: optimal PSD estimate (horiz. [day^{-1}]).

The three peaks are clearly visible. Their locations, namely 0.56 10^{-2}, 0.93 10^{-2}, and 1.38 10^{-2} day^{-1}, compare very favorably with the announced values. On the other hand, Deeming's method [3] performs rather poorly on this series, as illustrated by figure 7. The third peak is barely visible, and a spurious low frequency one is present.

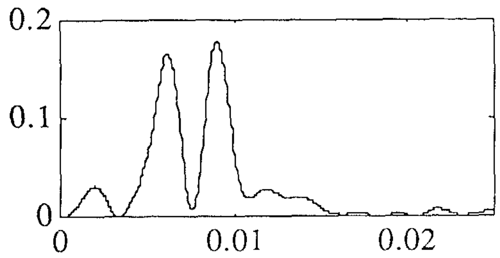

Fig.7: PSD estimate using Deeming's method

CONCLUSION

In this paper we have presented a spectrum based RBF interpolation scheme providing a technique for power density spectrum estimation of irregularly-sampled time series. This scheme has been shown to perform well both on simulated and experimental signals. Further work aims at using other interpolation techniques and spectral criteria, such as flatness of the PSD outside the band of interest.

REFERENCES

[1] R.T. Stellingwerf, "Period determination using phase dispersion minimization," *Astro. J.*, **224**, pp.953-960, Sept. 1978.

[2] R.H. Shumway, "Some applications of the EM algorithm to analyzing incomplete time series Data," in *Time Series Analysis of Irregularly Observed Data*, E. Parzen ed., Springer Verlag 1984.

[3] T.J. Deeming, "Fourier analysis with unequally-Spaced data," *Astro. and Space Sc.*, **36**, pp.137-158, 1975.

[4] M.D. Powell, *The Theory of Radial Basis Function Approximation in 1990* , in *Advances in Numerical Analysis* , W. Light ed., Oxford Science Publications 1992.

[5] T. Poggio, and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481-1497.

[6] C.A. Micchelli, "Interpolation of scattered data: Distance matrices and conditionally positive definite functions," *Constr. Approx.*, **2**, pp.11-22, 1986.

[7] S.L. Marple, *Digital Spectral Analysis with Applications*, Prentice-Hall 1987.

[8] P. Hall and I. Johnstone, "Empirical functionals and efficient smoothing parameter selection," *J.R. Statist. Soc. B*, vol. 54, no. 2, 1992, pp. 475-530.