PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC, 112:74–90, 2000 January
© 2000. The Astronomical Society of the Pacific. All rights reserved. Printed in U.S.A.

On the Reconstruction of Irregularly Sampled Time Series

ROBERTO VIO

Chip Computers Consulting s.r.l., Viale Don L. Sturzo 82, S.Liberale di Marcon, 30020 Venice, Italy; and ESA *IUE* Observatory, Apartado 50727, 28080 Madrid, Spain; robertovio@tin.it

THOMAS STROHMER

Department of Mathematics, University of California, Davis, CA 95616; strohmer@math.ucdavis.edu

AND

WILLEM WAMSTEKER

ESA *IUE* Observatory, Apartado 50727, 28080 Madrid, Spain; ww@vilspa.esa.es

Received 1999 June 14; accepted 1999 October 18

ABSTRACT. We consider the question of numerical treatment of irregularly sampled time series. This problem is quite common in astronomy because of factors such as the day-night alternation, weather conditions, nonobservability of the objects under study, etc. For this reason an extensive literature is available on this subject. Most of the proposed techniques, however, are based on heuristic arguments, and their usefulness is essentially in the estimation of power spectra and/or autocovariance functions. Here we propose an approach, based on the reasonable assumption that many signals of astronomical interest are the realization of band-limited processes, which can be used to fill gaps in experimental time series. By using this approach we propose several reconstruction algorithms that, because of their regularization properties, yield reliable signal reconstructions even in case of noisy data and large gaps. A detailed description of these algorithms is provided, their theoretical implications are considered, and their practical performances are tested via numerical experiments. MATLAB software implementing the methods described in this work is obtainable by request from the authors.

1. INTRODUCTION

In the past many numerical/statistical techniques have been developed for the analysis of unevenly sampled time series. Rather roughly, these techniques can be split into two (overlapping) groups according to the task for which they have been designed: (1) test for the presence of one or more periodic components and (2) reconstruction on a regular grid.

Although in principle a time series resampled on a regular grid could be tested for periodicity by using classical tools (e.g., Fourier methods), this approach is not efficient. In fact, testing for periodicity usually requires only the estimation of a statistical index. That is certainly an easier task than a complete reconstruction of a signal.

To the first group belong such techniques as folding methods (see Heck, Manfroid, & Mersch 1985 and references therein), periodogram techniques (Scargle 1982; Press et al. 1992), CLEAN deconvolution (Roberts, Lehar, & Dreher 1987), wavelet approaches (e.g., Foster 1996). In general, the techniques belonging to the second group are useful when the analysis of time series is only an intermediate step of a more general analysis as, for example, in echo-mapping or time-delay problems. In this paper we will consider only these kinds of methods.

Recently, several papers have been presented on such subjects as structure detection (especially in the context of X-ray and γ-ray signal analysis; see Scargle 1998) and Bayesian modeling (Rayner & Fitzgerald 1998). These approaches, however, are very specific to particular problems, and therefore we will not consider them further.

2. RECONSTRUCTION TECHNIQUES

In the current literature only a few reliable methods are available for the reconstruction of an uneven time series (see Adorf 1995[1] for a short review). The reason is that, without a priori information (a typical situation in astronomy), there

[1] Also available via http://www.stsci.edu/stsci/meetings/adassIV/adorfhm2.html.

are not many possibilities for obtaining reliable results. For example, a possible solution is to recover missing data through interpolation with polynomials or splines. Although this practice is very popular, it is difficult to verify the quality of the approximation. That is because very often there is no reason for splines and/or polynomials to have any relationship with the process underlying the observed time series. The selection of a priori information, however, is not an easy task. If the choice is correct, then methods may be expected to perform satisfactorily, otherwise misleading results are possible. For this reason, in many situations it is possible to make only very general assumptions.

Among the available methods, we describe the most popular ones:

1. *Autoregressive maximum entropy interpolation.*—Fahlman & Ulrych (1982) designed these techniques for filling the gaps in time series with regular sampling but missing data segments. Their idea is to start with estimating an initial autoregressive (AR) model of order L less than the length of the shortest known data segment:

$$x[t]_{t \in \mathscr{L}} = \sum_{k=1}^{L} a[k]x[t-k] + e[t] ,$$

where $\mathscr{L}$ is the set of indices corresponding to the observed data, $\{x[t]\}_{t \in \mathscr{L}}$ is the observed time series, $\{a[k]\}_{k=1}^{L}$ are the constant coefficients of the model, and $\{e[t]\}_{t \in \mathscr{L}}$ is the realization of a discrete white-noise process. A maximum Burg-Entropy algorithm is used to predict the unknown data values $\{x[t]\}_{t \notin \mathscr{L}}$. A new AR model is estimated from the entire data series with known and predicted values, and iteration proceeds until convergence. The rationale behind such an approach is that, hopefully, an AR model with a sufficiently high order is able to capture the dynamical properties of the process that generates the observed time series. One of the most serious drawbacks is that usually a rather high value of L (even of order 100–200) is required, and therefore the time series must be rather long. Another limitation is that this method is strictly based on the hypothesis of linear and stationary processes. For these reasons, the results are satisfactory only in the case of time series with simple time evolutions. For example, in the analysis of more complex signals, Brown & Christensen-Dalsgaard (1990) have been forced to improve the method via inclusion of a bandpass filter preprocessing step. Their approach, however, can be used only in case of particular signals characterized by power spectra with narrow peaks more or less evenly distributed in frequency.

2. *Linear least-squares method.*—Rybicki & Press (1992) suggest a method, based on a linear model similar to the previous one, that holds for general sampling patterns. The observed time series are resampled on a regular time grid through a least-squares estimator that requires the knowledge of both the autocovariance matrix of the signal and the covariance matrix of the measurement errors. For this reason, usually this method does not require the fit of very large models and therefore can be applied to reasonably short time series. The main drawback lies in the fact that in many situations the autocovariance matrix of the process is not known a priori and must be estimated through the time series itself. This, again, requires the condition of stationarity of the process. Moreover, in the presence of uneven sampling, the estimation of an autocovariance function is a problem whose difficulty is comparable to the reconstruction of the time series itself.

3. *Scargle's DFT interpolation.*—Scargle (1989) suggests a reconstruction technique based on the inversion of his definition of the discrete Fourier transform (DFT) for uneven time series (Scargle 1982). The main limitation of such an approach is that in reality this DFT has been designed only for retaining, in the case of pure noise, the simple statistical behavior of the evenly spaced case. However, the effects of irregular sampling are not removed. That means that the inversion of the DFT will provide a time series which is the result of a mix of the characteristics of the process and those of the sampling pattern.

At first sight, the condition of linearity required by the first two methods could seem to be a reasonable assumption. In reality, as shown by Vio et al. (1992), this condition is not satisfied by many astronomical time series. Also, since in general stationarity is not a characteristic for nonlinear processes, this condition could be violated by many real signals.

An additional comment regards the fact that with these methods it is difficult to check whether the reconstructed signals somehow resemble the true ones (that is not claimed by the authors themselves). One can only assert that the reconstructions share some statistical properties (e.g., mean value, covariance structure) with the process under study. In fact, these methods are useful principally in the estimation of power spectra or other statistical quantities.

3. A GENERAL APPROACH

In order to develop a more general approach, we start with the observation that signals $x(t)$, characteristic of many physical systems with unpredictable time evolution, can be described in terms of the following stochastic model (Vio et al. 1992):

$$\dot{x}(t) = f[x(t), \boldsymbol{u}(t), \boldsymbol{w}(t), t] .$$

Here $f(.)$ represents the dynamics of the system under study, $\boldsymbol{u}(t)$ is a vector containing the time evolution of processes

that influence $f(.)$, whereas $\boldsymbol{w}(t)$ is a vector containing the realization of continuous white-noise processes that disturb the time evolution of $x(t)$. Often (although not always) the function $f(.)$ acts as a time-dependent, low-pass filter. The consequence is that the energy of the power spectrum of $x(t)$ is concentrated at the lowest frequencies whereas it is zero (or approximately zero) at the highest ones. For instance, this is a common characteristic of the light curves of many astrophysical sources. That means that signals typical of these objects can be considered as the realization of band-limited processes. In the following sections it will be shown that with this simple a priori information it is possible to derive efficient and robust reconstruction methods.

3.1. Reconstruction of Discrete, Band-limited Signals: Noise-free Data

Let us start by considering the simplest case where a band-limited signal $x(t)$ has been sampled on a regular time grid but some data are missing. This situation is typical of observations with a relatively good control of the experimental conditions but suffering from temporary failures (e.g., satellite observations). Here the problem consists of obtaining the entire time series $\{x[n]\}_{n=0}^{N-1} \equiv \boldsymbol{x}$ (square brackets mean a discrete sampling variable) when a subset $\{x^o[l]\}_{l \in \mathscr{L}} \equiv \boldsymbol{x}^o$ (^{o} means observed quantity) with N_0 ($<N$) elements is available. Here $\mathscr{L} \subset \{0, 1, \ldots, N-1\}$. Even though $x(t)$ is a continuous signal, this problem corresponds to the reconstruction of a discrete time series. In the case of noise-free data the available information can be modeled as

$$x^o[l] = \begin{cases} x[l] & \text{if } l \in \mathscr{L}, \\ \text{undefined} & \text{if } l \notin \mathscr{L}, \end{cases} \tag{1}$$

where $\boldsymbol{x}$ satisfies the spectral condition

$$\hat{x}[k] = 0; \quad k \in \mathscr{K} . \tag{2}$$

Here k is an integer parameter with values in the range $[-N/2 + 1, N/2]$ if N is even and $[-N/2, N/2]$ otherwise; $\mathscr{K} = \{k: M < |k| \leq N/2\}$, where M is an integer index representing the bandwidth of the signal; and

$$\hat{x}[k] = \sum_{l=0}^{N-1} x[l] e^{-i2\pi kl/N} \tag{3}$$

is the DFT of $\boldsymbol{x}$. By inserting equation (1) into equation (3), it can be shown, after trivial algebra, that the missing data can be recovered via

$$\sum_{n \notin \mathscr{L}} x[n] e^{-i2\pi kn/N} = -\sum_{l \in \mathscr{L}} x^o[l] e^{-i2\pi kl/N}; \quad k \in \mathscr{K} . \tag{4}$$

In matrix notation this equation can be written as

$$\mathbf{A}\boldsymbol{y} = \boldsymbol{b} , \tag{5}$$

where $\mathbf{A}$ is an $(N - 2M - 1) \times (N - N_0)$ matrix,

$$\mathbf{A} = (e^{-i2\pi kn/N})_{k \in \mathscr{K}, n \notin \mathscr{L}} ,$$

$\boldsymbol{y} \equiv \{x[n]\}_{n \notin \mathscr{L}}$ is the vector containing the missing data to reconstruct, and $\boldsymbol{b}$ is the DFT of $\boldsymbol{x}^o$ at the frequencies $\{k\} \in \mathscr{K}$. This last term can be efficiently calculated via fast Fourier transform by filling the gaps of the observed time series with zeros.

This model tells us an interesting point about the possibility of recovering a band-limited time series. In fact, in the present case, model (5) is a linear system of $(N - 2M - 1)$ equations in $(N - N_0)$ unknowns. At least in principle, that means that $\boldsymbol{x}$ can be recovered if the number of missing data is smaller than $(N - 2M - 1)$, and this independently of the sampling pattern. In other words, there is no Nyquist limit (see Fig. 1)! One could be surprised that the discrete reconstruction does not present a critical sampling rate similar to the Nyquist limit typical of continuous signals. We have to remind ourselves, however, that here we have not attempted the ambitious task of reconstructing a continuous signal starting from discrete observations. The two problems are deeply different. In fact, according to the well-known Shannon's sampling theorem, if the sampling frequency is larger than the Nyquist limit, a continuous signal $x(t)$ can be

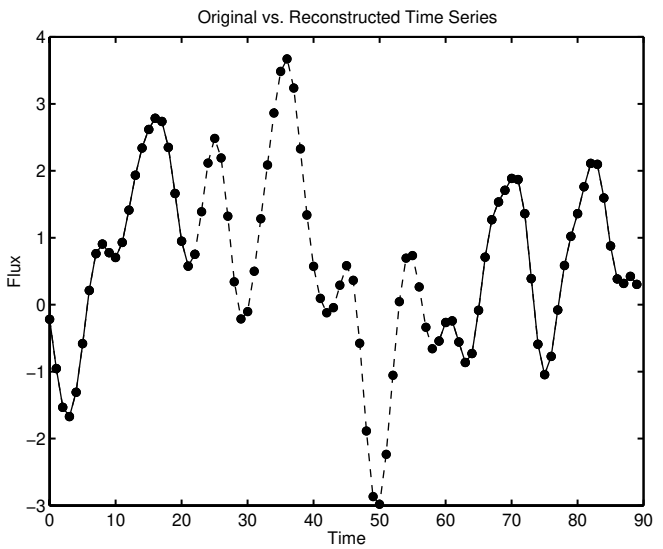

FIG. 1.—Reconstruction, via eq. (4), of a discrete signal constituted by 90 points but with only 53 points effectively available. The bandwidth is 20 time steps and the Nyquist time interval (i.e., the largest gap that can be reconstructed according to Shannon's sampling theorem), equal to $N/(2M + 1)$, is about 2.2 time steps. Solid line + filled circles = original signal, dashed line = reconstructed time series.

reconstructed from its sampled version $\{x[t]\}$ via the convolution

$$x(t) = (x * \text{sinc})[t] ,$$

where

$$\text{sinc}\,(t) = \frac{\sin\,(\pi t/T)}{\pi t/T} ,$$

and T is the sampling time step. It is trivial to show that this convolution can be written in form (5). But now, contrary to the problem we are considering in this work, $\boldsymbol{y}$ is known and $\boldsymbol{b}$ must be calculated. In other words, Shannon's reconstruction requires a convolution, whereas through model (4) we carry out a sort of "deconvolution" (more properly, an inversion operation). A detailed analysis of the relation of the finite-dimensional and the continuous reconstruction problems has been carried out by Gröchenig (1999).

Another interesting point regarding model (4) is that no assumption is made about the stationarity and/or linearity of the processes.

The most serious problem in the solution of equation (5) is that in general the matrix $\mathbf{A}$ is ill-conditioned. This means that we need numerical techniques that are able to handle this difficulty. In many situations that does not represent an insuperable limitation since various advanced algorithms are available in many software packages. The only concern regards the case of very long time series since the roundoff errors and the storage of $\mathbf{A}$ can create some problems. In such a situation an iterative approach such as the conjugate gradient (CG) method could be better suited.

3.1.1. Conjugate Gradient Method

The idea behind this technique arises from the observation that the solution of the linear system of equations $\mathbf{T}\boldsymbol{z} = \boldsymbol{u}$—where $\mathbf{T}$ is a $N \times N$ square, symmetric, and positive definite matrix—is equivalent to the minimization of the following quadratic form,

$$F(\boldsymbol{z}) = \tfrac{1}{2}\boldsymbol{z}'\mathbf{T}\boldsymbol{z} - \boldsymbol{z}'\boldsymbol{u} ,$$

since at the minimum it is

$$\nabla F(\boldsymbol{z}) = \mathbf{T}\boldsymbol{z} - \boldsymbol{u} = 0 .$$

The name of the method stems from the fact that minimum is sought along a certain set of direction vectors $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N$ with the property that

$$\boldsymbol{s}_i'\mathbf{T}\boldsymbol{s}_j = 0 \quad \forall i \neq j .$$

These vectors are called *conjugate directions*. CG is so interesting because in principle it is very fast. In fact, it can be shown that by defining a succession of search directions $\boldsymbol{s}_i$, the minimum of $F(\boldsymbol{z})$ is reached in a number of exact line searches at most equal to N (Fletcher 1993). In practice, operations are carried out in finite-precision arithmetic, hence termination after N iterations is not guaranteed. On the other hand, the rate of convergence of CG essentially depends on the clustering of the singular values of the matrix $\mathbf{A}$, and therefore often much fewer than N iterations are required.

In the present context described by equation (4), the matrix $\mathbf{A}$ is not square. This problem can be solved by setting $\mathbf{T} = \mathbf{A}'\mathbf{A}$, $\boldsymbol{z} = \boldsymbol{y}$, and $\boldsymbol{u} = \mathbf{A}'\boldsymbol{b}$. Such an approach is reasonable when the number of missing data is small and the spectral bandwidth is large. However, in more general situations it is not efficient. An alternative solution is the design of a model in the Fourier space. In fact, the DFT of $\boldsymbol{x}^o$ is given by

$$\hat{x}^o[l] = \sum_{j \in \mathscr{L}} x^o[j] e^{-i2\pi lj/N} , \tag{6a}$$

whereas the relationship between the observed data and the true DFT is given by

$$x^o[j]_{j \in \mathscr{L}} = \frac{1}{N} \sum_{k=-M}^{M} \hat{x}[k] e^{i2\pi kj/N} . \tag{6b}$$

By inserting equation (6b) in equation (6a), we obtain

$$\hat{x}^o[l] = \frac{1}{N} \sum_{k=-M}^{M} \hat{x}[k] \sum_{j \in \mathscr{L}} e^{i2\pi(k-l)j/N} . \tag{6c}$$

It is trivial to show that also this equation can be written in form (5), with $\mathbf{A}$ being an $N \times (2M + 1)$ matrix, and solution $\boldsymbol{y}$ providing the DFT, say, $\hat{\boldsymbol{x}}$, of $\boldsymbol{x}$. However, if in equation (6c) we limit index l to values in the interval $[M, -M]$, then $\mathbf{A}$ becomes square with sizes $(2M + 1) \times (2M + 1)$. It is not difficult to show that if equation (6b) is written in the form $\mathbf{V}\hat{\boldsymbol{x}} = \boldsymbol{x}^o$, then equation (6c) with $-M \leq l \leq M$ corresponds to the solution of the system of *normal equations* $\mathbf{A}\hat{\boldsymbol{x}} = \boldsymbol{b}$ with $\mathbf{A} = \mathbf{V}'\mathbf{V}$ and $\boldsymbol{b} = \mathbf{V}'\boldsymbol{x}^o$. The interesting point is that in this way $\mathbf{A}$ becomes not only a square matrix but also symmetric, positive definite, and of Toeplitz type. That permits the realization of a very fast implementation of the CG algorithm that moreover requires only the storage of the first row of this matrix. In Appendix A details of such an implementation are provided (see also Feichtinger, Gröchening, & Strohmer 1995). The only price to pay for working with the *normal equations* is that usually $\mathbf{A}$ becomes more ill-conditioned since the values of its smallest singular values decrease. In many practical applications, however, such singular values are filtered out by the pro-

cedures for noise removal (see below). Therefore, this approach can be safely used.

3.2. Reconstruction of Discrete, Band-limited Signals: Noisy Data

In case of data contaminated by noise, equation (1) must be replaced by

$$x^o[l] = \begin{cases} x[l] + \delta[l] & \text{if } l \in \mathscr{L}, \\ \text{undefined} & \text{if } l \notin \mathscr{L}, \end{cases}$$

where $\delta[l]$ is the measurement error corresponding to the lth observation. The consequence is that the solution of equation (5) will not provide $\boldsymbol{y}$, but only its least-squares estimate $\tilde{\boldsymbol{y}}$. Moreover, in case of the model based on equation (4), the matrix $\mathbf{A}$ must be adjusted to take into account that now the time series must be reconstructed also in correspondence with the observed data. That can be accomplished by forming the $(N_0 + N - 2M - 1) \times N$ matrix

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{e}_{l \in \mathscr{L}} \\ (e^{-i2\pi kn/N})_{k \in \mathscr{K}, n} \end{pmatrix}, \tag{7a}$$

where $n = 0, 1, 2, \ldots, \mathrm{N}-1$ and $\boldsymbol{e}_l$ is an N-dimensional vector whose elements are all equal to zero except the one at position l, which has unitary value. At the same time $\boldsymbol{b}$ must be modified to an $(N_0 + N - 2M - 1)$ vector whose first N_0 elements contain $\boldsymbol{x}^o$ and the others are equal to zero:

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{x}^o \\ 0 \end{pmatrix}. \tag{7b}$$

In the presence of errors the ill-conditionedness of the matrix $\mathbf{A}$ becomes critical since the errors of $\boldsymbol{b}$ will appear in the least-squares solution,

$$\tilde{\boldsymbol{y}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{b},$$

enormously amplified by the $(\mathbf{A}'\mathbf{A})^{-1}$ term. The consequence is that $\| \tilde{\boldsymbol{y}} \|$ will be disastrously large (note: "$\| \boldsymbol{a} \|$" means Euclidean norm of the vector $\boldsymbol{a}$ which, in the case of a column vector, is equal to $\sqrt{\boldsymbol{a}'\boldsymbol{a}}$). In the literature many techniques have been developed for overcoming this problem. The reason for such a large number of methods is that, in general, the stabilization of the solution implies that $E[\tilde{\boldsymbol{y}}] \neq \boldsymbol{y}$ ($E[.]$ means expected quantity). These methods try to obtain a stable solution with the introduction of the smallest possible bias. More or less explicitly, that requires the adoption of some a priori hypothesis for the problem under study. A method can be expected to work only if such a hypothesis is correct. Accordingly, if a particular context is not considered, no method can be claimed superior to the others.

Here we present two general methods. The first is based on a classical approach that, however, at least to the best of our knowledge, has never been used before for the problem we are treating. The second is an algorithm not yet presented in the astrophysical literature.

3.2.1. Regularization Method (RM)

The first approach we consider is the *regularization* method (RM). This is a classical technique for which extensive literature is available (for a comprehensive presentation, see Hansen 1998; whereas for a short introduction to the subject, see Press et al. 1992 or Vio, Horne, & Wamsteker 1994). For this reason we give only a brief description, and for details we refer to the cited literature.

Since the primary difficulty with discrete ill-conditioned problems is the huge norm of the solutions, one possibility for improving results is requiring that $\| \boldsymbol{y} \|$ has to be small. This can be achieved by defining a *regularized* solution $\tilde{\boldsymbol{y}}_\lambda$ as the minimizer of the following weighted combination of the residual norm and the side constraint (Tikhonov et al. 1990):

$$\tilde{\boldsymbol{y}}_\lambda = \operatorname{argmin} \left(\| \mathbf{A}\boldsymbol{y} - \boldsymbol{b} \|^2 + \lambda^2 \| \mathbf{L}\boldsymbol{y} \|^2 \right). \tag{8}$$

Here λ is a scalar, called the *regularization parameter*, that controls the weight given to minimization of the side constraint relative to minimization of the residual norm, whereas $\mathbf{L}$ is a matrix chosen such that $\| \mathbf{L}\boldsymbol{y} \|$ is small. Increasing λ pulls the solution away from minimizing the residual norm in favor of minimizing the side constraint. In case of the model based on equations (7a) and (7b), a typical choice for $\mathbf{L}$ is the discrete approximation of a derivative operator of given order d. This choice effectively provides stable solutions. For example, if $d = 1$ we have

$$\mathbf{L} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ \vdots & & & & \vdots & & & & \vdots \\ 0 & \ldots & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & \ldots & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

Here it is assumed that the first derivative of the process is not too far from zero (in other words, a constant quantity) and therefore a very smooth signal. The larger the order of d, the rougher the solution will be. For higher order approximations of a derivative operator, see Press et al. (1992). In principle, RM could be used also with the model based on equations (6a)–(6c). In this case, however, choosing $\mathbf{L}$ to be the discrete approximation of a derivative operator usually does not work, and unfortunately in general a sensible alternative is not available.

Once we have fixed $\mathbf{L}$ (see discussion below), the question remains which value to give to the regularization parameter

λ. Also for this difficult problem extensive literature is available. Here we mention only the generalized cross-validation technique (GCV), which in our numerical experiments has proved to be reliable. This method is based on the philosophy that if an arbitrary element b_i of the right-hand side $\boldsymbol{b}$ is left out, then the corresponding regularized solution should predict this observation well. This leads to choosing λ as

$$\lambda = \operatorname{argmin}\left(\frac{\| \mathbf{A}\tilde{\boldsymbol{y}}_\lambda - \boldsymbol{b} \|^2}{[\operatorname{trace}(\mathbf{I} - \mathbf{A}\mathbf{A}^I)]^2}\right), \tag{9}$$

where $\mathbf{A}^I = (\mathbf{A}'\mathbf{A} + \lambda^2\mathbf{L})^{-1}\mathbf{A}'$ is the matrix which produces the regularized solution $\tilde{\boldsymbol{y}}_\lambda$ when multiplied with $\boldsymbol{b}$, i.e., $\tilde{\boldsymbol{y}}_\lambda = \mathbf{A}^I\boldsymbol{b}$. A more detailed description of GCV and an efficient numerical solution of equations (8) and (9) is provided in Hansen (1998).

3.2.2. *Multilevel Algorithm (ML)*

In the previous sections the bandwidth M of the signal is a fixed parameter since it is assumed to be a priori known. However, this assumption may not be justified in some situations. In this section we describe a so-called *multilevel algorithm* (ML) that iteratively adapts to the optimal bandwidth during the reconstruction procedure. ML is also useful in the situation when the bandwidth is known a priori. The reason is that in this approach we consider the bandwidth M as free parameter and use it as a regularization parameter. In fact, it is intuitable that if M is too large, the reconstruction may be strongly oscillating due to overfitting of the noisy data by the high-frequency components. On the other hand, if M is too small, then the reconstruction may be very smooth, but poorly fitting the given data. The idea is not original, being already hinted by Swan (1982). This author, however, did not provide any criterion for the choice of the "optimal" value of M. The aim of this section is to provide such a criterion.

When we solve a linear system $\mathbf{T}\boldsymbol{z} = \boldsymbol{u}$ for noisy data $\boldsymbol{u}$ using CG, the following happens: the iterates $\boldsymbol{z}_k$ may diverge for $k \to \infty$, but in the beginning of the iteration the error propagation will be limited (see Hanke 1995). This means that the quality of the solution depends on how many iterative steps can be performed until the iterates begin to diverge. The idea is now to stop the iteration at about the point where divergence sets in. In other words, the iteration count is the regularization parameter which remains to be controlled by an appropriate stopping rule. In the context of the reconstruction of irregularly sampled time series, this means that CG fits first the low-frequency components of the signal and in the later iterations the higher frequency components. At some point, however, CG tends to fit more and more the noise in the data and iterations must be stopped.

Another consideration regards the fact that if the data are distorted by noise, it does not make sense to ask for a solution $\tilde{\boldsymbol{x}}$ that satisfies $\sum_{l\in\mathscr{L}} |x^o[l] - \tilde{x}[l]|^2 = 0$. All we should ask for is a solution $\tilde{\boldsymbol{x}}$ that satisfies

$$\left(\sum_{l\in\mathscr{L}} |x^o[l] - \tilde{x}[l]|^2\right)^{1/2} \le \delta\,, \tag{10}$$

where $x^o[l] = x[l] + \delta[l]$ and $(\sum_{l\in\mathscr{L}} |\delta[l]|^2)^{1/2} \le \delta$.

The method we propose is based on the following idea. Choose the initial bandwidth or *level* $M_0 = 0$ and compute the corresponding approximation $\tilde{\boldsymbol{x}}_0$ iteratively by using the CG algorithm and then calculate the reconstruction error $\sum_{l\in\mathscr{L}} |x^o[l] - \tilde{x}_0[l]|^2$. If the error satisfies condition (10), we stop and take $\tilde{\boldsymbol{x}}_0$ as our solution. If condition (10) is not satisfied, we continue by increasing the bandwidth to $M_1 = 1$. Now we use the approximation $\tilde{\boldsymbol{x}}_0$ as starting value for the CG method at the next higher level M_1 and compute an approximation $\tilde{\boldsymbol{x}}_1$. We again check whether our approximation satisfies condition (10) and either terminate the algorithm or proceed to the next higher level $M_2 = 2$. We proceed until we reach a bandwidth (*level*) M_* for which the error of the reconstruction $\tilde{\boldsymbol{x}}_{M_*}$ satisfies for the first time condition (10). In this procedure the only open question regards the stopping criterion for the CG iterations at each level M_k. In fact, we have pointed out that choosing an appropriate stopping criterion is crucial when using the CG method, in connection with noisy data, to ensure convergence of the iterates $\tilde{\boldsymbol{x}}_{M_k}^{(n)}$ for $n \to \infty$ at a given level M_k. If the level M_k is too small, then the iterates $\tilde{\boldsymbol{x}}_{M_k}^{(n)}$ will never satisfy condition (10), and further iterations at this level will not lead to an improvement. Instead, it is better to stop the iterations and proceed to the next level M_{k+1} using the last approximation $\tilde{\boldsymbol{x}}_{M_k}^{(n)}$ from level M_k as starting value, as described above. Therefore, the crucial question is how to choose the stopping criterion that tells us whether we should switch to the next level M_{k+1}. It has been shown by Scherzer & Strohmer (1998) that the appropriate stopping criterion is given when the iterates $\tilde{\boldsymbol{x}}_{M_k}^{(n)}$ at level M_k satisfy for the first time

$$\sum_{l\in\mathscr{L}} |\tilde{x}_{M_k}^{(n)}[l] - x^o[l]|^2 \le 2(\delta + \| P_k\boldsymbol{x} - \boldsymbol{x} \|)\,\| \boldsymbol{x}^o \|\,,$$

where $P_k\boldsymbol{x}$ is the projection of the original signal $\boldsymbol{x}$ on the space of band-limited signals with bandwidth M_k:

$$P_k x[l] = \frac{1}{N}\sum_{s=-M_k}^{M_k} \hat{x}[s] e^{i2\pi sl/N}\,.$$

This stopping criterion, however, is based on a "worst case estimate" approach. Numerical experiments and the heuristic arguments of Hanke (1995) indicate that better results

can be obtained with the criterion

$$\sum_{l \in \mathscr{L}} |\tilde{x}^{(n)}_{M_k}[l] - x^o[l]|^2 \leq 2(\delta + \| P_k \boldsymbol{x} - \boldsymbol{x} \| \, \| \boldsymbol{x}^o \|) .$$

Note that in both criteria $\| P_k \boldsymbol{x} - \boldsymbol{x} \|$ cannot be computed exactly since solution $\boldsymbol{x}$ is not known. In Appendix B, besides a pseudocode for the implementation of ML, we present a simple but robust procedure that allows us to estimate $\| P_k \boldsymbol{x} - \boldsymbol{x} \|$ at level M_k recursively using the information of the samples $\boldsymbol{x}^o$ and the approximation $\tilde{\boldsymbol{x}}^{(n)}_{M_{k-1}}$ at level M_{k-1}.

3.3. Reconstruction of Continuous, Band-limited Signals on a Regular Time Grid: Noise-free Data

Often the result of an observational campaign consists of a time series $\{x^o[t_j]\} \equiv \boldsymbol{x}^o$ obtained by sampling a continuous signal $x(t)$ on an set of N_t unevenly spaced time instants $\{t_j\}_{j=1}^{N_t}$. In situations like these, the problem consists of computing a reconstruction $\{x[t]\} \equiv \boldsymbol{x}$ of $x(t)$ on a regular grid of N time instants $\{t\}$. Again, we are dealing with a discrete problem, but here the signal to reconstruct cannot be considered discrete since in general the sampling pattern has no density limitations and $\{t_j\}_{j=1}^{N_t} \not\subset \{t\}$.

For ease of formalism, and without loss of generality, we can assume that $t = 0, 1, \ldots, N-1$ and that $t_1 = 0$ and $t_{N_t} = N - 1$. Under the hypothesis that $x(t)$ is a band-limited process and assuming for the moment that we have noise-free data, the available information is in the form of N_t observations

$$x^o[t_j] = x(t_j); \qquad j = 1, \ldots, N_t \tag{11a}$$

satisfying the spectral condition

$$\hat{x}[k] = 0; \quad k \in \mathscr{K} . \tag{11b}$$

Here k is an integer parameter with values in the range $[-N/2 + 1, N/2]$ if N is even and $[-N/2, N/2]$ otherwise; $\mathscr{K} = \{k: M < |k| \leq N/2\}$, where M is an integer index representing the bandwidth of the signal; and

$$\hat{x}[k] = \sum_{t=0}^{N-1} x[t] e^{-i2\pi kt/N} \tag{11c}$$

is the DFT of $\boldsymbol{x}$. Note that the spectral bandwidth M regards the DFT of $\boldsymbol{x}$, not that of $\boldsymbol{x}^o$.

It is possible to reconstruct the time series $\boldsymbol{x}$ by coupling the observed data $\boldsymbol{x}^o$ and the spectral information $\{\hat{x}[k]\} \equiv \hat{\boldsymbol{x}}$ in some different ways:

Model 1.—

$$x^o[t_j] = \frac{1}{N} \sum_{k=-M}^{M} \hat{x}[k] e^{i2\pi kt_j/N} . \tag{12}$$

With this model we estimate $\hat{\boldsymbol{x}}$ by imposing that its inverse DFT, calculated at the time instants $\{t_j\}_{j=1}^{N_t}$, provides the observed data.

Model 2.—By inserting equation (11c) into equation (12), it is possible to obtain

$$\begin{aligned} x^o[t_j] &= \frac{1}{N} \sum_{t=0}^{N-1} x[t] \sum_{k=-M}^{M} e^{i2\pi k(t_j - t)/N} \\ &= \sum_{t=0}^{N-1} x[t] \frac{\sin [\pi(2M+1)(t_j - t)/N]}{N \sin [\pi(t_j - t)/N]} . \end{aligned} \tag{13}$$

This means that, according to this model, the observed time series $\boldsymbol{x}^o$ is the result of convolving $\boldsymbol{x}$ with the Dirichlet kernel. Notice that, conversely to model 1, here the spectral information is not explicitly considered.

Model 3.—Although sequence $\boldsymbol{x}^o$ does not constitute an even time series, it is possible to calculate the corresponding DFT $\{\hat{x}^o[l]\} \equiv \hat{\boldsymbol{x}}^o$ on the same regular grid of $\hat{\boldsymbol{x}}$:

$$\hat{x}^o[l] = \sum_{j=1}^{N_t} x^o[t_j] e^{-i2\pi lt_j/N} . \tag{14a}$$

By inserting equation (12) into equation (14a), we obtain

$$\hat{x}^o[l] = \frac{1}{N} \sum_{k=-M}^{M} \hat{x}[k] \sum_{j=1}^{N_t} e^{i2\pi(k-l)t_j/N} . \tag{14b}$$

This model is similar to the previous one, but now the observed data are not explicitly considered.

In principle, all the models presented before are equivalent. In effect, in matrix notation the three models can be again written in the form (5). Here, however, $\mathbf{A}$ contains the exponential terms that appear in equations (12), (13), and (14b) whereas, according to the particular case, $\boldsymbol{y}$ contains the reconstructed time series $\boldsymbol{x}$ or the corresponding $\hat{\boldsymbol{x}}$, while $\boldsymbol{b}$ contains the observed data $\boldsymbol{x}^o$ or the corresponding $\hat{\boldsymbol{x}}^o$. In practice, these models have different properties and allow us to examine the problem from different perspectives. For example, models 1 and 3 tell us an interesting point about the possibility of recovering an uneven band-limited time series. In fact, equations (12) and (14b) correspond to two linear systems with N_t equations in $2M + 1$ unknowns. This means that, at least in principle, time series $\boldsymbol{x}$ can be recovered if N_t is as large as $2M + 1$, and this independently of the sampling pattern. We already found a similar property in the case of regularly sampled time series with missing data. Here, however, things are much less favorable: if we consider two close time instants t and $t + dt$, it is trivial to show that

$$e^{i2\pi k(t+dt)} \simeq e^{i2\pi kt}(1 + 2\pi k \, dt) .$$

In other words, with k fixed, the exponential at time $t + dt$ differs from that at time t only by a constant factor that does not depend on t. That implies that if time instants $\{t_j\}_{j=1}^{N_t}$ are too close to each other, **A** can be so ill-conditioned as to become numerically singular. This unpleasant behavior can be significantly reduced by introducing weights that compensate for these clustering effects; see Appendix A. Similar situations turn up in the presence of very irregular sampling patterns, for example, when the observations are clustered in densely populated groups followed by large gaps. In such an occurrence there is no way to improve the condition number of A; even preconditioning does not give significant improvements: this means unsatisfactory if not disastrous results. Sometimes, even in the absence of measurements errors a regularization approach is required to obtain reasonable results.

Model 2 tells us that if we want to recover $\boldsymbol{x}$ without resorting to the Fourier space, we have to face a deconvolution problem. In fact, equation (13) is the discrete analog of the well-known Shannon's sampling theorem for continuous signals. In this particular case, if $N_t < N$, the matrix **A** has more columns than rows and the problem is not only ill-conditioned but singular. As is well known, this requires great care in the numerical computation of the solution.

Also from a computational point of view these models have different properties. For example, model 3 is by far the most interesting since, by adopting the same arguments as used for the discrete signals (see eq. [6c]), **A** can be reduced to a square matrix with sizes $(2M + 1) \times (2M + 1)$ by limiting the spectral index l in equation (14b) to have values in the interval $-M \leq l \leq M$. Apart the smallest dimensions, the great benefit is that the matrix **A** becomes Hermitian and Toeplitz. That permits a very efficient implementation of the CG algorithm with respect to both the storage requirement and the number of operations (see Appendix A).

3.4. Reconstruction of Continuous, Band-limited Signals on a Regular Time Grid: Noisy Data

The three models presented above are rather obvious, and already in 1982 they were suggested as tools for signal recovering (Kuhn 1982; Swan 1982). Their use, however, has been rather limited if not null. Again, the reason lies in the severe ill-conditionedness of the matrix **A** that, in case of data contaminated by errors, makes the direct solution of equation (5) very unstable. In order to partially overcome this problem, it is possible to use the same machinery introduced in the reconstruction of the discrete time series.

4. NUMERICAL EXPERIMENTS

From the discussion in the previous sections, it is intuitable that in a given experimental context performances such as those shown in Figure 1 cannot be expected. Unfortunately, the actual capabilities of the reconstruction techniques are very difficult to be fixed since they critically depend on factors as the time sampling pattern, the characteristic of the signal, and the noise level. The only possibility is to resort to numerical experiments. However, before we proceed, a few remarks have to be made, since both RM and ML require the a priori choice of some parameters.

4.1. Practicalities

In the case of ML, the only free parameter is δ, which can be estimated via the standard deviation of the measurement errors. This estimate usually is sufficiently good. In any case, if the sampling is not characterized by extremely large gaps, from our numerical experiments it appears that the results provided by ML do not depend critically on the value of this parameter.

In the case of RM, one has to fix the bandwidth M of the signal and to decide the form of the operator **L**. When there are not too many large gaps, a reliable estimate of the parameter M can be obtained by filling them via linear interpolation to obtain a regularly spaced time series. An estimate of the bandwidth can now be obtained by computing the DFT of this time series and taking those frequency components whose absolute value is larger than the noise level. In any case, also for this parameter, our numerical simulations indicate that its exact value is not so crucial for the final result. The situation is different for the choice of **L**. In our experiments we have chosen this matrix to be the discrete derivative operator of order d. With this choice, and adopting model 2 in the case of reconstruction of continuous signals (the only one that does not consider explicitly the spectral information), it is implicitly assumed that the process can be approximated by means of a differential equation whose dth derivative is equal to zero. In practice, larger and larger d and more and more degrees of freedom are allowed for the reconstruction. Here the problem is that the "optimal" value of d is not known a priori. Again, according to our simulations it appears that, if the sampling is not characterized by large gaps, results do not depend too much on the exact value of d. In fact, the standard deviation, say, σ_x, of the reconstruction residuals $\{x^o[j] - \tilde{x}[j]\}_{j \in \mathcal{L}}$ (for the discrete case), or $\{x^o[t_j] - \tilde{x}^o[t_j]\}_{j=1}^{N_t}$ (for the continuous case), is a monotonic and decreasing function of d that eventually settles on a plateau at a level given by the standard deviation of the noise. The reason for such a behavior is that when parameter d is too small, then the capability of models to fit the data is limited and σ_x tends to be large. On the other hand, if d is large enough, then the expected σ_x is equal to the noise level.

In case of time series characterized by large gaps, the situation can be very different. In particular, the monotonicity of the relationship between σ_x and d frequently is

lost. According to our numerical simulation, in situations like this one the criterion of adopting the order d corresponding to the smallest σ_x usually works. The reason is that when σ_x grows for increasing values of d, it means that the larger number of degrees of freedom allowed for the reconstruction is not used correctly by the model (that is typical of all ill-conditioned problems). Unfortunately, this criterion can fail when σ_x is an increasing function of d. Situations like this one are typical of bad sampling patterns and can be an indication for possible troubles. For example, it is possible that, in correspondence to certain values of d, σ_x is much smaller than the noise level. When that happens, the choice of d for which σ_x is the nearest possible to the noise level is still able to provide reasonable results. The reason is that a "good" solution must be characterized by a σ_x that is close to the noise level. A limitation of this approach is that it requires the knowledge of the noise level.

4.2. Reconstruction Problems

Figure 2 shows the RM and ML reconstructions of a short time series with 90 points, bandwidth equal to 20 time steps, Nyquist interval (i.e., the largest gap that can be reconstructed according to Shannon's sampling theorem) equal to about 2.2 time steps and contaminated by a discrete, Gaussian, white-noise process with standard deviation of 0.2. For comparison, the same figure also shows the reconstruction obtained with the classic cubic-spline interpolation (CS), which does not take into account the spectral bandwidth of the time series. The signal is sampled on a regular time grid with about one-third of the data (27) removed in such a way that most of the "peaks" are cut. This constitutes really a pathological sampling pattern simulating an extreme situation. In fact, the original time series is characterized by a standard deviation of about 1.4, but the standard deviation for the observed data is about 0.7. This means that data are able to explain only 25% of the power per unit of time step of the underlying discrete process. Figure 3 shows the same situation when half of the data are randomly removed. Also this case simulates a difficult situation since the mean time interval between two adjacent data is very close to the Nyquist limit of the signal. The choice of these examples is based on the fact that the sampling patterns of the astronomical time series are often a combination of these situations. We use only 90 points in order to test the performances of algorithms in the case of a very limited data set. In order to test the stability of the algorithms, in Figures 4 and 5 a total of 10 reconstructions are presented regarding the same time series shown in the upper panels of Figures 2 and 3 but with different realizations of the white noise contaminating the data (since the CS reconstructions are very similar to each other, they are not shown).

From these figures it appears that RM and ML are able to provide good results even in these unfavorable situations. Moreover, their performances are superior to those provided by the classical approach. According to the results of the numerical experiments we have carried out with different noise levels, this can be considered as a rather general result. That is especially true in the case of sampling patterns that

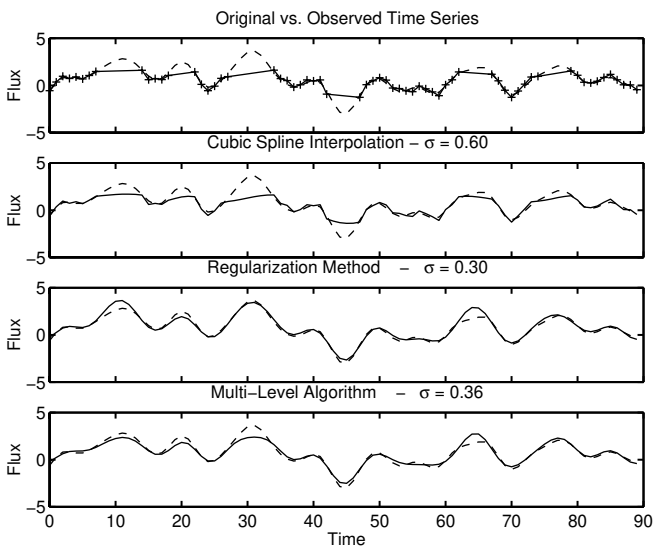

FIG. 2.—Reconstructions, via CS, RM, and ML methods of a noisy, discrete signal constituted by 90 points but with only 63 points effectively available. The bandwidth is 20 time steps, the Nyquist time interval is about 2.2 time steps, and the noise standard deviation is equal to 0.2. The sampling has been chosen in such a way as to remove most of the "peaks" of the signal. σ means standard deviation of the residuals between the true and the reconstructed time series. Dashed line = original signal, solid line = reconstructed signal as indicated in each panel title, plus signs = observed data.

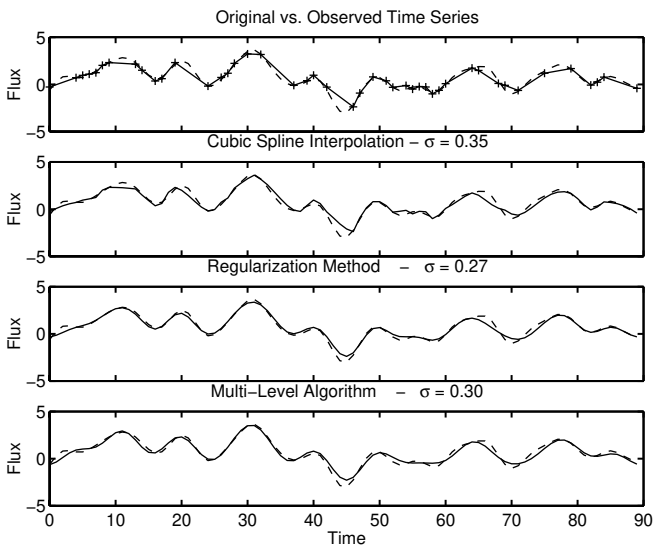

FIG. 3.—As in Fig. 2 with the difference that here there are 45 available points and the sampling is discrete and randomly uniform.

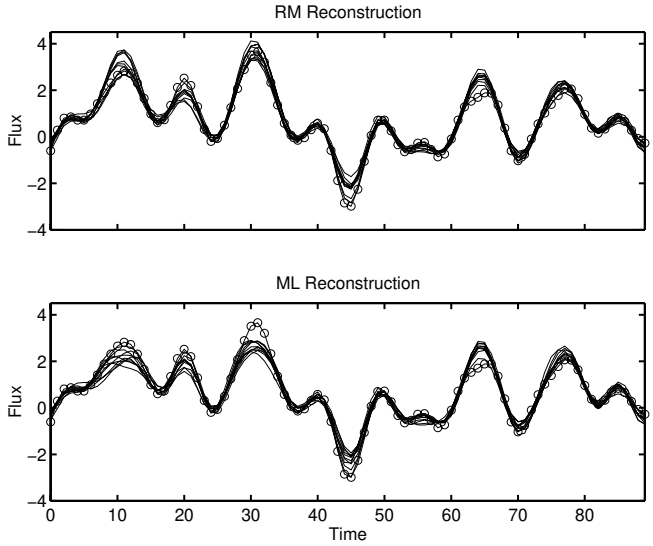

FIG. 4.—Reconstructions of 10 time series obtained by using the same signal shown in the upper panel of Fig. 2 but with a different realization of the noise process. Solid line = reconstructed signal as indicated in each panel title, solid line + open circles = original signal.

present some large gaps. In fact, methods such as CS are not able to reconstruct the missing peaks: they do not correspond to any underlying "physical" model that could be included as additional information in order to improve the reconstruction. Of course, this conclusion does not mean that splines cannot be used in reconstruction problems, but only that in order to provide reliable results they have to incorporate the information regarding the spectral bandwidth of the time series. For example, that could be accomplished by an opportune choice of the number and position of the knots. Since in this way the spectral bandwidth of the reconstruction is controlled by these two quantities, now the problem becomes how to determine them in an "optimal" manner. At this point, however, it is not clear why a spline approach would be adopted, which must be somehow readapted for the present task, instead of RM or ML, which have be expressly designed for it.

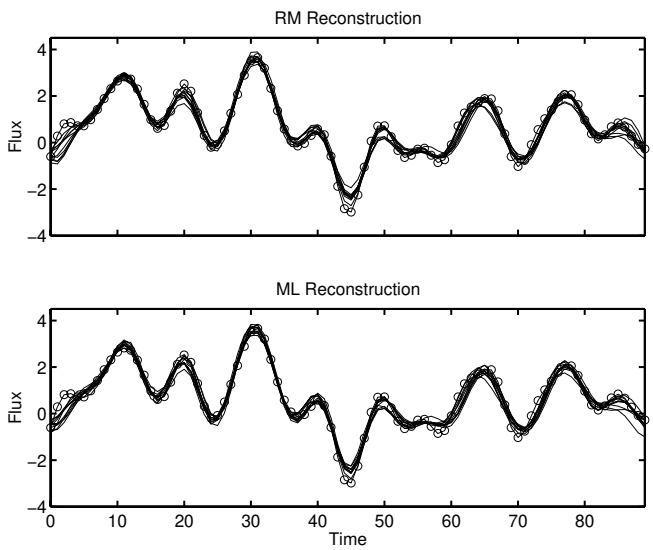

FIG. 5.—Reconstructions of 10 time series obtained by using the same signal shown in the upper panel of Fig. 3 but with a different realization of the noise process. Solid line = reconstructed signal as indicated in each panel title, solid line + open circles = original signal.

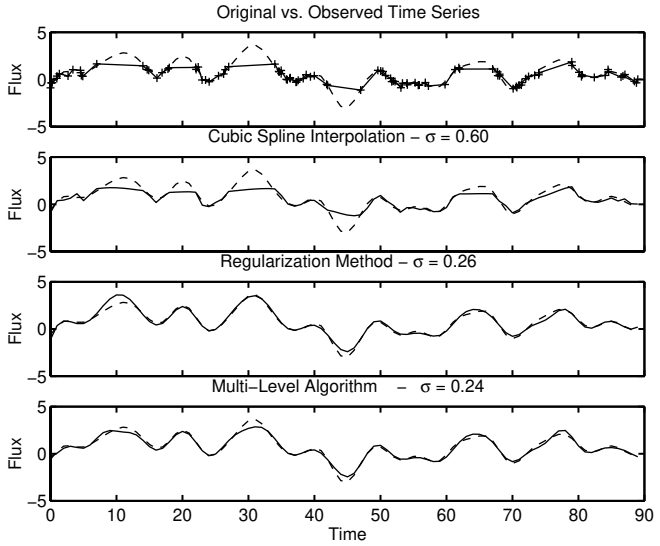

FIG. 6.—Reconstructions, via CS, RM, and ML methods on an even grid of 90 time instants, of a signal sampled on 100 time instants uniformly distributed in the regions external to the gaps shown in Fig. 2. The signal bandwidth is 20 time steps (with respect to the resampling time grid), the Nyquist time interval is about 2.2 time steps, and the noise standard deviation is equal to 0.2. σ means standard deviation of the residuals between the true and the reconstructed time series. Dashed line = original signal, solid line = reconstructed signal as indicated in each panel title, plus signs = observed data.

As expected, performances of the three methods tend to level out in the presence of a high noise level. That is because the stabilization requires a strong smoothing action with the consequent degradation of the final result. The value of this level depends critically on the characteristics of both the signal and the sampling geometry.

In an extensive set of numerical experiments, not presented here, we have tried to determine the maximum length of gaps that RM and ML are able to fill up correctly. The results have not been impressive: typically the largest gaps are similar to those shown in Figure 2, corresponding to 2–3 times the value of the Nyquist interval, where the signal shows only a "fluctuation." The problem is that, without any additional information, there is no way to discriminate among the solutions that in a given gap show different behaviors (e.g., a different number of "peaks"): these solutions are all compatible with the same spectral bandwidth and the same set of data! A point to be noted is that usually such solutions are characterized by different power spectra.

Therefore, a way to improve the results could be the imposition of some constraints regarding power spectra. Unfortunately, in general this piece of information is not available for astronomical signals, and therefore in the present paper we will not consider this possibility any further. We stress that, lacking any additional information, the inability to correctly fill up large gaps is a limit of any method, not only of those presented here.

Figures 6 and 7 and Figures 8 and 9 correspond, respectively, to Figures 2 and 3 and Figures 4 and 5 in the case of a continuous, randomly uniform sampling. For Figures 6 and 8, however, the gaps visible in Figure 2 are not permitted to be occupied by any sampling point. Also in this experiment results provided by RM and ML are satisfactory and superior to those provided by CS. Here, however, we have

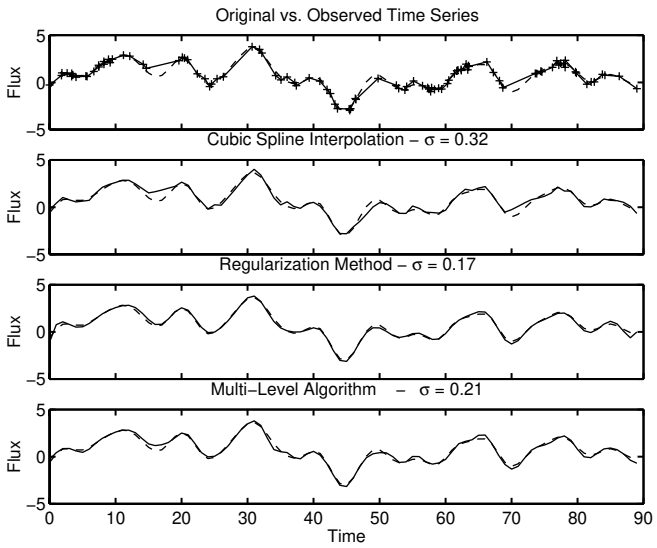

FIG. 7.—As in Fig. 6 with the difference that here the sampling is continuous, randomly uniform in the interval 0–89.

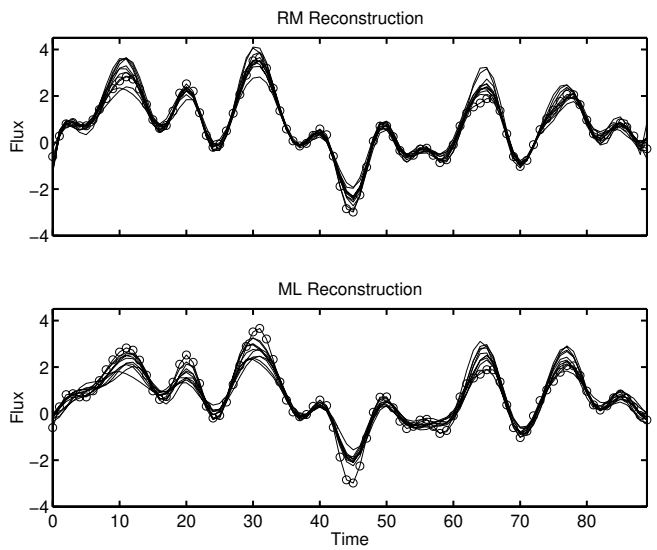

FIG. 8.—Reconstructions of 10 time series obtained by using the same signal shown in the upper panel of Fig. 6 but with a different realization of the noise process. Solid line = reconstructed signal as indicated in each panel title, solid line + open circles = original signal.

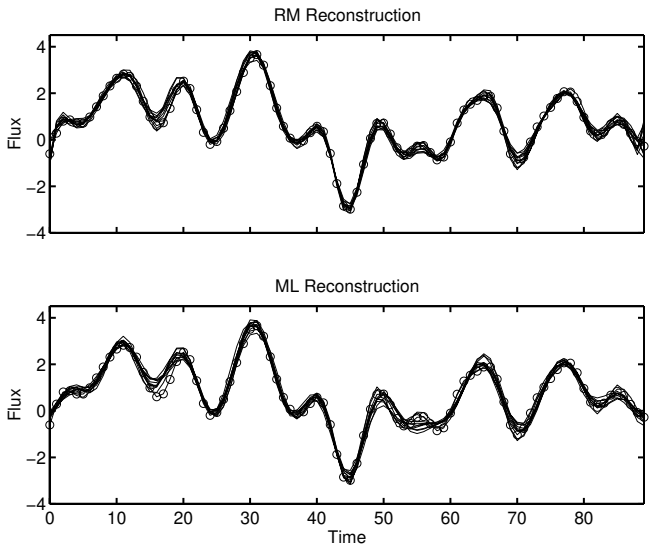

FIG. 9.—Reconstructions of 10 time series obtained by using the same signal shown in the upper panel of Fig. 7 but with a different realization of the noise process. Solid line = reconstructed signal as indicated in each panel title, solid line + open circles = original signal.

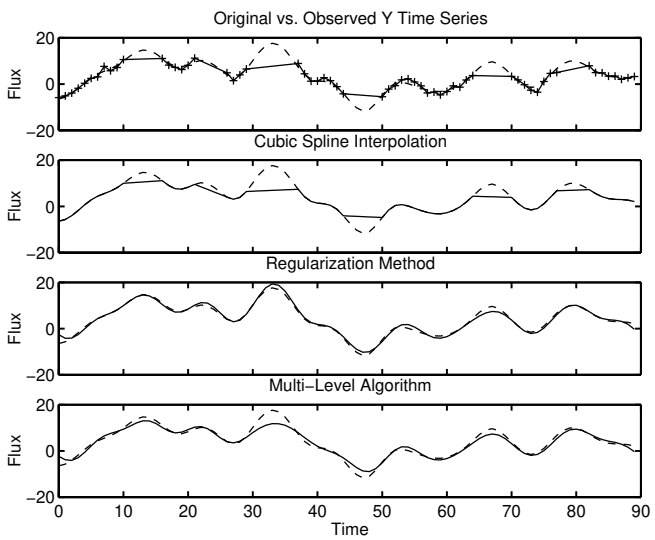

FIG. 10.—Reconstructions, via CS, RM, and ML methods of time series y that has been obtained from time series x, shown in the upper panel of Fig. 2, via $y[t] = \sum_{k=0}^{4} x[t-k]$ (see text). This time series is composed of 90 points, but only 60 points are effectively available. The bandwidth is 20 time steps, the Nyquist time interval is about 2.2 time steps, and the noise standard deviation is equal to 1.0. The sampling has been chosen in such a way to remove most of the "peaks" of the signal. σ means standard deviation of the residuals between the true and the reconstructed time series. Dashed line = original signal, solid line = reconstructed signal as indicated in each panel title, plus signs = observed data.

been forced to use a larger number of data (100) than before, for the following two reasons:

1. Model 2 is characterized by a matrix **A** that usually is more ill-conditioned than in the case of discrete signals. This means that we have to expect worse results than those obtained for the discrete algorithm even in the case where the same set of data is analyzed.

2. A continuous, randomly uniform sampling can yield sampling patterns with very close time instants. Since, as seen before, the columns of the matrix **A** corresponding to two very close observations are almost linearly dependent, this point indicates that the number of "useful" data is smaller than the number of the data effectively used in the reconstruction.

Especially this last point is very interesting since it indicates that the practice in various astronomical works of collecting the largest possible amount of data without any kind of attention to the final sampling pattern can be simply useless.

4.3. Echo-mapping Problems

In order to test the quality of results that can be obtained when the reconstruction of an uneven time series is only the intermediate step of a more general analysis, we have simulated an echo-mapping problem. As is well known, this kind of problem arises when the radiation y, emitted by a source at a given wavelength, is due to the reprocessing of radiation x emitted in a different part of the electromagnetic spectrum. In various situations such a reprocessing can be described through the following simple model:

$$y(t) = \int_{\tau=0}^{\Delta} h(\tau)x(t-\tau)d\tau \ ,$$

where Δ corresponds to the reprocessing timescale. What it is of interest here is $h(.)$ since this function can provide useful information about the physical system under study. Usually quantities $x(t)$ and $y(t)$ are available only at a set of discrete time instants, and therefore the previous model must be discretized as

$$y[l] = \sum_{k=0}^{N} x[l]h[l-k] \ .$$

It is easy to see that this problem can be readily solved only in the case of signals sampled on regular time grids. Unfortunately, often this condition is not satisfied by the experimental time series, and therefore their reconstruction is required. The main problem in this operation is that the estimation of $h[.]$ is an ill-conditioned problem. This means that the quality of the reconstruction must be rather good, otherwise unstable results have to be expected. Figures 2, 10, and 11 show an example of results obtainable with RM, ML, and CS when time series x and y are sampled on a regular grid (that is only for ease of simulation of the two signals), about one-third of the data are removed and $y[l] = \sum_{k=0}^{4} x[l-k]$ (i.e., $h[k] = 1$ for $k = 0, \ldots, 4$). The noise level is equal to 0.2 for signal x and 1.0 for signal y. The two time series have been separately reconstructed, and then function $h[.]$ has been estimated via a Tikhonov inversion approach with the additional constraint of positivity of the solution (for more details, see Vio et al. 1994). Also in this experiment the sampling pattern is pathological since we are simulating again an extreme situation. As before, results provided by RM and ML are by far better than those obtained with CS.

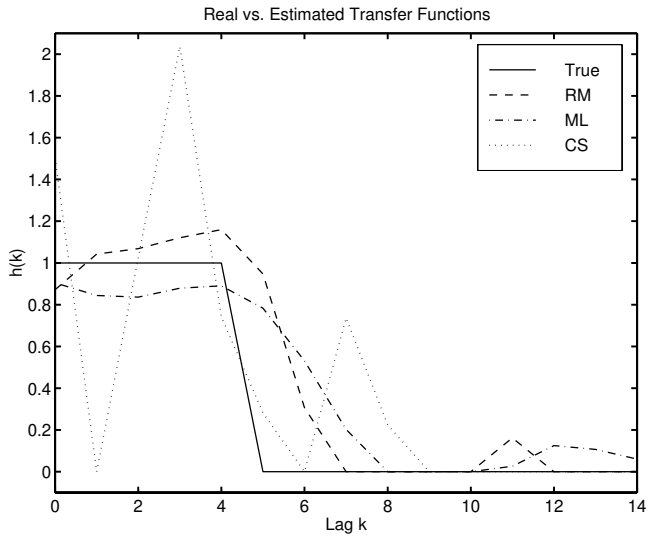

FIG. 11.—Transfer functions obtained via echo-mapping method by using the reconstructed time series of Figs. 2 and 10. RM = regularization method, ML = multilevel algorithm, CS = cubic-spline interpolation.

4.4. Some Considerations

From the results presented above it could appear that RM and ML are two equivalent methods that can be indifferently interchanged. In reality these two approaches have different properties. For example, our numerical experiments have shown that RM has a greater ability than ML in filling large gaps. That is undoubtedly an important and very useful characteristic. At the same time, however, RM appears less stable than ML, and that can create some troubles in situations of large noise level. Another question is the computational cost. In fact, ML requires only $K \times N \times \log N$ operations, where K is the number of iterations for convergence (typically some tens for time series with some hundreds points) and N is the dimension of the square matrix **A**. RM is much more time consuming, requiring $\propto \alpha MN + \beta N^3$ operations (α and β are constants depending on the particular numerical method used) in the case where **A** is rectangular with dimensions $M \times N$. That

limits the use of RM in the case of very long time series. From these considerations it is clear that the choice between the two methods depends on the experimental context in which one is operating. In any case, since any inversion problem must be treated with much care, it is advisable to use RM and ML together: very different results will indicate that something has gone wrong in the analysis.

5. SUMMARY AND CONCLUSIONS

In this paper we have considered the problem of the reconstruction of irregularly sampled time series. From the presented considerations it appears that the assumption of band-limited processes (namely, processes whose Fourier transform is equal to zero above a threshold frequency) is a reasonable and yet powerful assumption that permits us to face this problem from well-defined theoretical grounds. In spite of that, we have seen that from the numerical point of view some drawbacks still remain since the resulting numerical problems are ill-conditioned. We have presented some algorithms that are able to deal with this situation and have tested their performances and limits via numerical experiments. The results we have obtained indicate that, although these algorithms cannot be considered as the solution of all the problems, they are effectively able to recover even extreme situations. An additional result of this work is that, from both theoretical arguments and numerical simulation, the practice of collecting data without any care for the sampling appears simply useless, if not risky, for the reliability of results.

APPENDIX A

CONJUGATE GRADIENT ALGORITHM (CG)

We start with the consideration that if $\mathbf{V}$ is a Hermitian, positive definite $N \times N$ matrix and $\mathbf{V}\boldsymbol{z} = \boldsymbol{u}$, then the solution of this system of linear equations is equivalent to the minimization of the following quadratic form:

$$F(\boldsymbol{z}) = \tfrac{1}{2}\boldsymbol{z}'\mathbf{V}\boldsymbol{z} - \boldsymbol{z}'\boldsymbol{u} .$$

That is because this function is minimized when

$$\nabla F(\boldsymbol{z}) = \mathbf{V}\boldsymbol{z} - \boldsymbol{u} = 0 .$$

A set of vectors $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N$ is said to be a conjugate set if

$$\boldsymbol{s}_i'\mathbf{V}\boldsymbol{s}_j = 0 \quad \forall i \neq j .$$

These vectors are called *conjugate directions*. They are so interesting because it can be shown that, by defining a succession of search directions $\boldsymbol{s}_i$, the minimum of $F(\boldsymbol{z})$ is reached in at most N exact line searches (Fletcher 1993). Again, it can be shown that the following algorithm, indicated by $\boldsymbol{z} = \mathrm{CG}(\mathbf{V}, \boldsymbol{z}_0, \boldsymbol{u}, \Psi)$, makes use of these "special" directions:

$$\begin{aligned}
\boldsymbol{r}_0 &= \boldsymbol{u} \\
\boldsymbol{p}_0 &= \boldsymbol{r}_0 \\
n &= 0 \\
&\quad \text{do while } \Psi \text{ is TRUE} \\
n &= n + 1 \\
\alpha_n &= \frac{\boldsymbol{r}_{n-1}'\boldsymbol{r}_{n-1}}{\boldsymbol{p}_{n-1}'\mathbf{V}\boldsymbol{p}_{n-1}} \\
\boldsymbol{z}_n &= \boldsymbol{z}_{n-1} + \alpha_n\boldsymbol{p}_{n-1} \\
\boldsymbol{r}_n &= \boldsymbol{r}_{n-1} - \alpha_n\mathbf{V}\boldsymbol{p}_{n-1} \\
\beta_n &= \frac{\boldsymbol{r}_n'\boldsymbol{r}_n}{\boldsymbol{r}_{n-1}'\boldsymbol{r}_{n-1}} \\
\boldsymbol{p}_n &= \boldsymbol{r}_n + \beta_n\boldsymbol{p}_{n-1} \\
&\quad \text{end do} \\
\boldsymbol{z} &= \boldsymbol{z}_n
\end{aligned}$$

where $\boldsymbol{z}_0$ is the starting guess for the solution (often $\boldsymbol{z}_0 = 0$) and Ψ represents the stopping criterion. A typical choice is

$\Psi = \| \mathbf{V}z - \boldsymbol{u} \| < \text{tol} \times \| \boldsymbol{u} \|$, where "tol" is a fixed parameter with values of order 10^{-7} in single-precision or 10^{-15} in double-precision arithmetic. Alternative choices are possible (see Appendix B).

It is easy to check that all the steps of this algorithm require operations between vectors and/or a scalar. Only for the calculation of scalar α_n does one needs to carry out a multiplication between a matrix and a vector (the result of such multiplication can be used also for the calculation of vector $\boldsymbol{r}_n$). This is the most expensive part of the procedure.

An interesting characteristic of CG is that it can be easily adapted to exploit particular structures of the matrix **V**. That can be very useful in the problem of the reconstruction of uneven time series. In fact, according to model 3 based on equation (14b), the matrix that corresponds to **V** is Hermitian and Toeplitz. Recall that an $N \times N$ matrix of this type has the form

$$\mathbf{V} = \begin{pmatrix} v_0 & v_1 & v_2 & \dots & v_{N-2} & v_{N-1} \\ \bar{v}_1 & v_0 & v_1 & \dots & v_{N-3} & v_{N-2} \\ \dots & & & \dots & & \\ \bar{v}_{N-2} & \bar{v}_{N-3} & \bar{v}_{N-4} & \dots & v_0 & v_1 \\ \bar{v}_{N-1} & \bar{v}_{N-2} & \bar{v}_{N-3} & \dots & \bar{v}_1 & v_0 \end{pmatrix} .$$

Hence the whole information of **V** is contained in one row. This fact can be exploit to speed up the procedure. Indeed, the multiplication of a vector $\boldsymbol{p}$ by the Toeplitz matrix **V** can be carried out via fast Fourier transform. In order to do so, however, we have to embed **V** into a *circulant matrix*.

We remind that a circulant matrix has the following form:

$$\mathbf{C} = \begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_{N-2} & c_{N-1} \\ c_{N-1} & c_0 & c_1 & \dots & c_{N-3} & c_{N-2} \\ \dots & & & \dots & & \\ c_2 & c_3 & c_4 & \dots & c_0 & c_1 \\ c_1 & c_2 & c_3 & \dots & c_{N-1} & c_0 \end{pmatrix} .$$

For the problem at hand the important point is that the matrix-vector multiplication $\mathbf{C}\boldsymbol{g} = \boldsymbol{h}$ is identical to a discrete convolution $\boldsymbol{c} * \boldsymbol{g} = \boldsymbol{h}$, where $\boldsymbol{c}$ is the first column of **C**. As is well known, in the Fourier domain this convolution becomes $\hat{\boldsymbol{c}}\hat{\boldsymbol{g}} = \hat{\boldsymbol{h}}$ and $\boldsymbol{h}$ can be recovered via an inverse Fourier transform. In numerical implementations these steps can be sped up by the use of fast Fourier transforms.

To compute $\boldsymbol{q} = \mathbf{V}\boldsymbol{p}$ via fast Fourier transform we have to embed **V** into a $(2N-1) \times (2N-1)$ circulant matrix $\mathbf{C}_V$ whose first row is

$$\boldsymbol{c}_V = (v_0, v_1, \dots, v_{N-2}, v_{N-1}, \bar{v}_{N-1}, \bar{v}_{N-2}, \dots, \bar{v}_2, \bar{v}_1) ,$$

and augment the vector $\boldsymbol{p}$ to a vector of length $2N-1$ as follows:

$$\boldsymbol{p}_C = (p_0, p_1, \dots, p_{N-1}, 0, \dots, 0) .$$

It is easy to see that, after having computed $\boldsymbol{q}_C = \mathbf{C}_V \boldsymbol{p}_C$ we obtain $\boldsymbol{q}$ by taking only the first N entries of $\boldsymbol{q}_C$.

An additional improvement comes from the observation that the DFT algorithms are very efficient when the length of vectors is equal to 2^k, where k is a positive integer. Therefore, it is convenient to augment vector $\boldsymbol{c}_V$ to a length of 2^k via zero padding:

$$\boldsymbol{c}_V = (v_0, v_1, \dots, v_{N-1}, 0, \dots, 0, \bar{v}_{N-1}, \bar{v}_{N-2}, \dots, \bar{v}_2, \bar{v}_1) .$$

A similar operation must be also carried out for $\boldsymbol{p}_C$.

In the problem of the reconstruction of continuous signals, the arguments presented above can be directly applied by making the substitutions $\mathbf{V} = \mathbf{A}$ (with element provided by eq. [14b]) and $\boldsymbol{p} = \boldsymbol{p}_{n-1}$. Here, however, there is an additional drawback: in the case of very irregular sampling patterns, or large gaps, the CG algorithm may fail because the reconstruction can be dominated by the most frequently sampled part of the signal. Theoretical arguments and numerical experiments (Strohmer 1993)[2] have shown that in order to reduce the importance of this problem, a possible approach is to weight the jth

[2] Available via http://tyche.mat.univie.ac.at/papers/1993/str1193.html.

element of vector $\boldsymbol{x}^o$ by the factors

$$w_j = \begin{cases} \dfrac{1+t_2}{2} & j = 1, \\ \dfrac{t_{j+1}-t_{j-1}}{2} & j \neq 1,\ N_t, \\ \dfrac{N-t_{j-1}}{2} & j = N_t \end{cases}$$

(note: remember that according to the convention adopted in this work $t_1 = 0 < t_2 < \cdots < t_{N_t} = N - 1$). In our algorithm, this weighting can be carried out by replacing equation (14a) with

$$\hat{x}^o[l] = \sum_{j=1}^{N_t} w_j x^o[t_j] e^{-i2\pi l t_j/N} .$$

This operation transforms $\mathbf{A}$ into a new matrix $\mathbf{A}_w$, but the procedure remains the same.

In conclusion the fast algorithm can be arranged in the five steps:

1. Compute the weight $\{w_j\}_{j=1}^{N_t}$.
2. Calculate the first row $\boldsymbol{a}_w$ of $\mathbf{A}_w$.
3. Calculate $\boldsymbol{p}_0 = \hat{\boldsymbol{x}}^o$.
4. Apply CG to $\mathbf{A}_w \hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^o$ with the matrix-vector multiplication necessary to the evaluation of the coefficient α_n carried out via the discrete fast Fourier transform method described above.
5. Transfer the result via inverse Fourier transform back to the signal space.

The calculations in steps 2 and 3 can be carried out quickly by making use of the algorithm of Beylkin (1995). In the case of regularly sampled time series with missing data, the arguments are exactly the same, with the elements of $\mathbf{A}$ given by equation (6c). Here calculations can be made even faster by calculating the elements of $\mathbf{A}$ and $\hat{\boldsymbol{x}}^o$ through fast Fourier transform via zero padding of the elements of the time series that are not available.

APPENDIX B

EFFICIENT IMPLEMENTATION OF THE MULTILEVEL ALGORITHM (ML)

In this appendix we present an efficient implementation of the multilevel algorithm, based on equation (14b), for the reconstruction of continuous signals (the extension to the discrete case is trivial). As shown in Appendix A, in this context matrix $\mathbf{A}$ is Hermitian and Toeplitz, and that permits a very fast approach via the CG algorithm. Starting from this consideration, a procedure for an efficient implementation of ML is the following (see main text):

$$\begin{aligned} k &= 0 \\ M_k &= 0 \\ \hat{\boldsymbol{x}}_{M_k} &= 0 \\ \tilde{\boldsymbol{x}}^o_{M_k} &= 0 \\ &\quad \text{do while } \| \tilde{\boldsymbol{x}}^o_{M_k} - \boldsymbol{x}^o \| > \delta \\ k &= k + 1 \\ M_k &= M_{k-1} + 1 \\ \hat{\boldsymbol{x}}_{M_k} &= \mathrm{CG}(\mathbf{A}_{M_k}, \hat{\boldsymbol{x}}_{M_{k-1}}, \hat{\boldsymbol{x}}^o, \Psi) \\ \tilde{\boldsymbol{x}}_{M_k} &= \mathrm{IDFT}(\hat{\boldsymbol{x}}_{M_k}) \\ M_* &= M_k \\ &\quad \text{end do} \\ \tilde{\boldsymbol{x}} &= \tilde{\boldsymbol{x}}_{M_*} \end{aligned}$$

where N_t is the number of available data, IDFT(.) means *inverse discrete Fourier transform*, CG(.) means *conjugate gradient solution* (see Appendix A), $\mathbf{A}_{M_k}$ means that the size of this matrix is $(2M_k + 1) \times (2M_k + 1)$,

$$\Psi = \| \tilde{\boldsymbol{x}}^{o(n)}_{M_k} - \boldsymbol{x}^o \| > 2(\delta + \| P_k \boldsymbol{x} - \boldsymbol{x} \| \, \| \boldsymbol{x}^o \|) \tag{B1}$$

represents the level-dependent stopping criterion for CG according to Hanke (1995), the expression $\| \tilde{\boldsymbol{x}}^o_{M_k} - \boldsymbol{x}^o \| \geq \delta$ in the *while loop* is the level-independent stopping criterion and

$$P_k x[l] = \frac{1}{N} \sum_{s=-M_k}^{M_k} \hat{x}[s] e^{i2\pi sl/N}$$

denotes the projection of $\boldsymbol{x}$ in the space of the band-limited signals with bandwidth $\leq M_k$. In the criterion (B1) the term $\tilde{\boldsymbol{x}}^{o(n)}_{M_k}$ is the solution, resampled on the same time grid of the available time series, corresponding to the nth iteration of CG(.) at level M_k:

$$\tilde{\boldsymbol{x}}^{o(n)}_{M_k} = \text{DSINC}\,(M_k, \tilde{\boldsymbol{x}}^{(n)}_{M_k}) \,,$$

where DSINC (.) means the discrete convolution given by equation (13). This step is necessary since usually the reconstruction time grid has no superimposition with the sampling time pattern of the signal.

At this point, the only still open question regards the stopping criterion (B1) requiring the a priori knowledge of solution $\boldsymbol{x}$ that we are trying to calculate. This means that if we want to use this criterion we have to find out a method for estimating it. The aim of this appendix is to provide such a method.

Scherzer & Strohmer (1998) have shown that as long as $\mathbf{A}$ is not very ill-conditioned, $\boldsymbol{x} \in P_M \boldsymbol{x}$, and given noisy samples $x^o[t_j] = x[t_j] + \delta[t_j]$ (where as usual $(\sum_{j=1}^{N_t} |\, \delta[t_j] \,|^2)^{1/2} \leq \delta$), the following estimate holds:

$$\left| \left(\sum_{j=1}^{N_t} |\, x^o[t_j] \,|^2 w_j \right)^{1/2} - \| \boldsymbol{x} \| \right| \approx \delta \,, \tag{B2}$$

where $\{w_j \equiv (t_{j+1} - t_{j-1})/2\}$ are the weights introduced in Appendix A.

For a recursive estimation of $\| P_k \boldsymbol{x} - \boldsymbol{x} \|$ let us start at level $M_k = 1$ where $\| \boldsymbol{x} - P_1 \boldsymbol{x} \|$ must be estimated. According to criterion (B2), we have

$$\| \boldsymbol{x} - P_1 \boldsymbol{x} \|^2 \leq \| \boldsymbol{x} - P_0 \boldsymbol{x} \|^2 = \| \boldsymbol{x} \|^2 \approx \sum_{j=1}^{N_t} |\, x^o[t_j] \,|^2 w_j \,.$$

Using this estimate we run the CG algorithm until the stopping criterion (B1) is satisfied, at iteration n, say. In this way an approximation $\tilde{\boldsymbol{x}}^{(n)}_1$ is obtained and we can switch to the next level $M = 2$. To apply the multilevel algorithm at this level we first have to estimate $\| \boldsymbol{x} - P_2 \boldsymbol{x} \|^2$. Similar to above we have that

$$\| \boldsymbol{x} - P_2 \boldsymbol{x} \|^2 = \| \boldsymbol{x} \|^2 - \| P_2 \boldsymbol{x} \|^2$$

Here, as before, $\| \boldsymbol{x} \|^2 \approx \sum_{j=1}^{N_t} |\, x^o[t_j] \,|^2 w_j$, and we can use the approximation $\| P_2 \boldsymbol{x} \|^2 \approx \| \tilde{\boldsymbol{x}}^{(n)}_1 \|^2$ to get $\| \boldsymbol{x} - P_2 \boldsymbol{x} \|^2 \approx \sum_{j=1}^{N_t} |\, x^o[t_j] \,|^2 w_j - \| \tilde{\boldsymbol{x}}^{(n)}_1 \|^2$.

Consequently, for each level M_{k+1}, we can estimate

$$\| \boldsymbol{x} - P_{k+1} \boldsymbol{x} \|^2 \approx \sum_{j=1}^{N_t} |\, x^o[t_j] \,|^2 w_j - \| \tilde{\boldsymbol{x}}^{(n)}_{M_k} \|^2 \,.$$

From the viewpoint of computational efficiency it is worth mentioning that there is a strong relationship between the linear systems $\mathbf{A}\boldsymbol{y} = \boldsymbol{b}$ for different levels M_k. Let $\mathbf{A}_{M^*}$ and $\mathbf{A}_{M^*+1}$ denote the Toeplitz matrices at level M^* and level $M^* + 1$, respectively. We adopt a similar notation for the associated right-hand sides $\boldsymbol{b}_{M^*}$ and $\boldsymbol{b}_{M^*+1}$. Then it is easy to see that the

$(2M^* + 1) \times (2M^* + 1)$ matrix $\mathbf{A}_{M^*}$ is embedded in the $(2M^* + 3) \times (2M^* + 3)$ matrix $\mathbf{A}_{M^*+1}$ in the following way:

$$\mathbf{A}_{M^*+1} = \begin{pmatrix} a_0 & \cdots & \bar{a}_{2(M^*+1)} \\ \vdots & \mathbf{A}_{M^*} & \vdots \\ a_{2(M^*+1)} & \cdots & a_0 \end{pmatrix}.$$

A similar relation holds for $\boldsymbol{b}_{M^*}$ and $\boldsymbol{b}_{M^*+1}$ (but of course this is not true for the solutions $\boldsymbol{y}_{M^*}$ and $\boldsymbol{y}_{M^*+1}$ of these systems). Thus when we switch from one level to the next level, we have to compute only a few new entries to establish the new linear system of equations.

Thomas Strohmer acknowledges partial support by NSF grant 9973373.

REFERENCES

Adorf, H. M. 1995, in ASP Conf. Ser. 77, Astronomical Data Analysis Software and System IV, ed. R. A. Shaw, H. E. Payne, & J. J. Hayes (San Francisco: ASP), 460
Beylkin, G. 1995, Appl. Comput. Harmonic Anal., 2, 363
Brown, T. M., & Christensen-Dalsgaard, J. 1990, ApJ, 349, 667
Fahlman, G. G., & Ulrych, T. J. 1982, MNRAS, 199, 53
Feichtinger, H. G., Gröchening, K., & Strohmer, T. 1995, Numerische Mathematik, 69, 423
Fletcher, R. 1993, Practical Methods of Optimization (New York: Wiley)
Foster, G. 1996, AJ, 112, 1709
Gröchenig, K. 1999, Math. Comput., 68, 749
Hanke, M. 1995, Conjugate Gradient Type Methods for Ill-posed Problems (New York: Longman Scientific & Technical)
Hansen, P. C. 1998, Rank-Deficient and Discrete Ill-posed Problems (Philadelphia: SIAM)
Heck, A., Manfroid, J., & Mersch, G. 1985, A&A, 59, 63
Kuhn, J. R. 1982, AJ, 87, 196
Press, H. W., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes (Cambridge: Cambridge Univ. Press)
Rayner, P. J. W., & Fitzgerald, W. J. 1998, in Signal Analysis and Prediction 1: Proc. First European Conference on Signal Analysis and Prediction, ed. A. Prochazka, J. Uhlir, P. J. W. Rayner, & N. G. Kingsbury (Boston: Birkhäuser), 65
Roberts, D. H., Lehar, J., & Dreher, J. W. 1987, AJ, 93, 968
Rybicki, G. B., & Press, W. H. 1992, ApJ, 398, 169
Scargle, J. D. 1982, ApJ, 263, 835
———. 1989, ApJ, 343, 874
———. 1998, ApJ, 504, 405
Scherzer, O., & Strohmer, T. 1998, Numer. Funct. Anal. Opt., 19, 353
Strohmer, T. 1993, Ph.D. thesis, Univ. Wien
Swan, P. R. 1982, AJ, 87, 1608
Tikhonov, A. N., Goncharsky, A. V., Stepanov, V. V., & Yagola, A. G. 1990, Numerical Methods for the Solution of Ill-posed Problems (London: Kluwer)
Vio, R., Cristiani, C., Lessi, O., & Provenzale, A. 1992, ApJ, 391, 518
Vio, R., Horne, K., & Wamsteker, W. 1994, PASP, 106, 1091