Robust reconstruction of irregularly sampled geophysical time series via a sparse spectral representation

Nadia Kreimer and Mauricio D. Sacchi

Department of Physics and Institute for Geophysical Research, University of Alberta, Edmonton, Alberta, Canada
kreimer@ualberta.ca

Abstract

Reconstruction of time series with insufficient amount of samples is a common problem in geophysical signal processing. In this article, the spectral components of the signal are recovered via the minimization of an objective function that includes a sparsity constraint. We also incorporate an error re-weighting strategy to minimize the influence of outliers in the final spectral estimator of the time series. A paleoclimatic record from the Lake Baikal drilling program is used to test the algorithm.

Introduction

In signal analysis, in particular in geophysics and astronomy, we often encountered irregularly sampled time series contaminated with outliers (Ferraz-Mello, 1981; Branham, 1986). The spectral analysis of this type of series can be tackled via Fourier inversion (Sacchi et al., 1988). In this case, we seek the set of complex Fourier coefficients that can recover the irregularly sampled time series. The inclusion of weights in the misfit function permits, in addition, to control the influence of outliers. This is important, in particular, when analyzing noisy time series like those arising in paleoclimatic studies. As an example, we examine real data records of magnetic susceptibility from Lake Baikal in Siberia (Kravchinsky et al., 2007).

Theory

Classical harmonic analysis methods approximate an observed signal via a sum of complex exponentials of unknown frequencies, phases, amplitudes and number of frequencies. The latter constitutes a non-linear problem. In this paper, we propose a linear formulation where we consider a dense distribution of frequencies and we exclusively invert for the associated complex amplitudes (Sacchi et al., 1989; Bourbignon et al., 2007).

The time series is modeled via the sum of K frequencies, with $K >> N$, where N is the number of observations. Let x_k be the unknown complex spectral amplitudes and e_n the additive noise. The irregularly sampled time series can be written as follows

$$y_n = \sum_{k=-K}^{K} x_k \exp(j\, 2\, \pi \frac{k}{K} f_{\max}\, t_n) + e_n, \quad n = 1..N. \tag{1}$$

Last equation can be rewritten in matrix form as follows $y = W\,x + e$. To solve the problem we propose to use a solution that promotes *sparse* spectral estimators. For this purpose we minimize the following cost function

$$J(x) = \frac{1}{2}\|y - Wx\|^2 + \lambda \sum_{k=-K}^{K} |\,\mathrm{x}_k\,| \tag{2}$$

The first term in equation 1 is the misfit function; the second is the regularization term. We have chosen a ℓ^1 regularization to promote sparse solutions (Alliney and Ruzinsky, 1994). The trade-off parameter λ is used to control the sparsity of the solution. Following Bourbignon (2005), the minimization of equation 2 is carried out via an iterative coordinate descent (ICD) algorithm. If w_k is the $k\text{-}th$ column of matrix W, and $e_k = y - \sum_{l \neq k} w_l x_l$, the ICD algorithm can be summarized as follows

$$\mathrm{x}^{\min}{}_k = \arg\min_k \mathrm{J}(x) \Leftrightarrow \begin{cases} if\ \left|w^H{}_k\ e_k\right| \leq \lambda : x^{\min}{}_k = 0 \\ if\ \left|w^H{}_k\ e_k\right| > \lambda : \arg x^{\min}{}_k = \arg w^H{}_k\ e_k \\ and\ \left|x^{\min}{}_k\right| = \frac{1}{N}(w^H{}_k\ e_k - \lambda) \end{cases} \quad (3)$$

The iterative process stops when a predetermined misfit is reached. Sparse spectral estimates are often difficult to obtain when the time series are contaminated with outliers. Incorporating weights can alleviate this problem. In other words, we now minimize the cost function

$$J_2(x) = \frac{1}{2}\left\|P(y - Wx)\right\|^2 + \lambda \sum_{k=-K}^{K} |x_k|, \quad (4)$$

where the matrix P is diagonal with elements given by Andrew's weights (Leffler and Jay, 2008). The weights are extracted using an interactive approach. We start by computing the solution using unitary weights. Then, the residuals are used to compute a new set of weights and a new solution.

Examples

Synthetic example: Figure 1a provides a synthetic example. A monochromatic waveform contaminated with random Gaussian noise and outliers is provided. The goal is to reconstruct the series and estimate the sparse amplitude spectrum of the data using the algorithm presented in this article. The reconstruction fails when the sparsity-promoting algorithm does not consider the presence of outliers (Figures 1b and c). On the other hand, the weights were able to control the influence of outliers and the time series was properly reconstructed (Figures 1d and e).

Lake Baikal records: We test the proposed reconstruction algorithm using a real data set. The data consist of whole-core magnetic susceptibility (κ) records from the Lake Baikal drilling program in Siberia (Kravchinsky et al., 2007). Core data was converted from depth to time resulting in a paleoclimatic record of magnetic susceptibility. We have analyzed a small window of 210 samples spanning about 900Ky. The desired sampling interval was $\Delta t = 2ky$. The latter corresponds to a Nyquist frequency of 0.5 cycles/ky. The data were de-trended to filter very low frequency components prior to reconstruction and spectral estimation. First we run the algorithm using unitary weights ($P=I$). This assumes that the errors in the time series are Gaussian. The reconstruction is portrayed in Figure 3. It is evident that the reconstructed signal is adjusting outliers. The normalized sparse power spectrum of the reconstructed signal shows many non-zero spurious components. In our second test, we reconstruct the data and estimate the spectral amplitude using Andrew's weights. The results are portrayed in Figure 2. The reconstructed signal adjusts the data in a smoother way, ignoring the outliers. Logically, this makes the power spectrum less noisy. The sparse power spectrum can now identify a discrete set of frequencies that coincide with some astronomical cycles (Milankovitch periodicities). The main astronomical cycles are 640, 400, 100, 41, 23 and 19 ka (Kravchinsky et al., 2007).

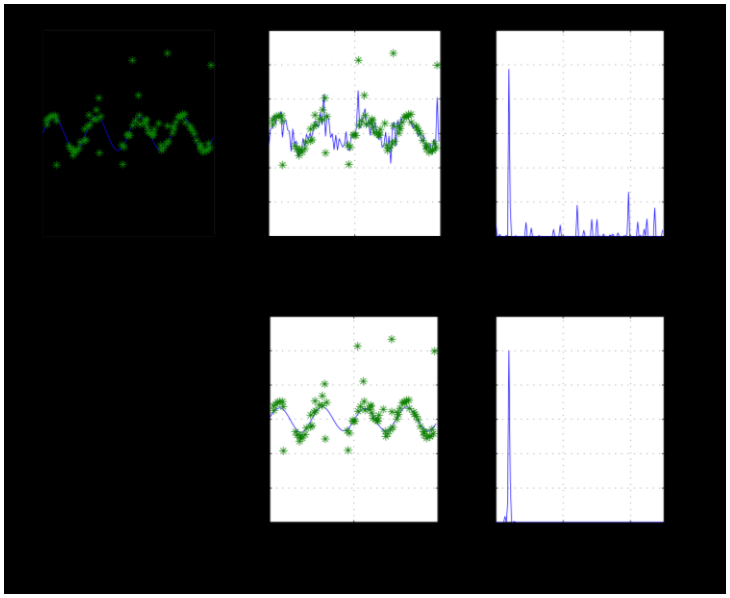

Figure 1. Synthetic example. Reconstruction of a sinusoid contaminated by outliers. a) Original signal (blue) and observations (green stars). b) Reconstruction using the sparsity promoting spectral norm. c) Sparse amplitude spectrum estimated from the data. d) Reconstruction using sparsity promoting spectral norm and weights to de-emphasize the influence of outlines. e) Sparse spectrum estimator when weights for outlier rejection are included.

Conclusions

The problem of reconstruction of irregularly sampled data is solved via the minimization of an objective function that consists of a misfit term and a ℓ^1 norm term. The last term introduces a sparsity constraint to the spectral amplitudes of the time series. An ICD method is employed for calculating the solution for this underdetermined inverse problem (Bourbignon, 2005).

The presence of outliers will deteriorate the reconstruction of the time series and produce sparse spectra dominated by artifacts. Therefore, we have proposed to include an error-reweighting scheme to minimize the influence of outliers.

References

Alliney, S., and S. A. Ruzinsky, 1994, An algorithm for the minimization of mixed L1 and L2 norms with application to Bayesian estimation: IEEE Trans. Signal Processing, 42.

Bourbignon, S., and H. Carfantan, 2008, New methods for fitting multiple sinusoids from irregularly sampled data: Statistical Methodology, 5.

Branham, R. L., 1986, Is robust estimation useful for astronomical data reduction?: Quarterly Journal RAS, 27.

Ferraz-Mello, S., 1981, Estimation of periods from unequally spaced observations: AJ, 86.

Kravchinsky, V. A., M. E. Evans, S. H. Peck, J. A., M. A. Krainov, J. W. King, and M. I. Kuzmin, 2007, A 640 kyr geomagnetic and paleoclimatic record from Lake Baikal sediments: Geophysical Journal International, 170.

Leffler, K. E., and D. A. Jay, 2008, Enhancing tidal harmonic analysis: Robust (hybrid L1 /L2) solutions: Continental Shelf Research, 29.

Sacchi, M. D., T. J. Ulrych, and C. J. Walker, 1998, Interpolation and extrapolation using a high–resolution discrete Fourier transform: IEEE Transactions on Signal Processing, 46.

Ulrych, T. J., M. D. Sacchi, and A. Woodbury, 2001, A Bayes tour of inversion: A tutorial: Geophysics, 66.

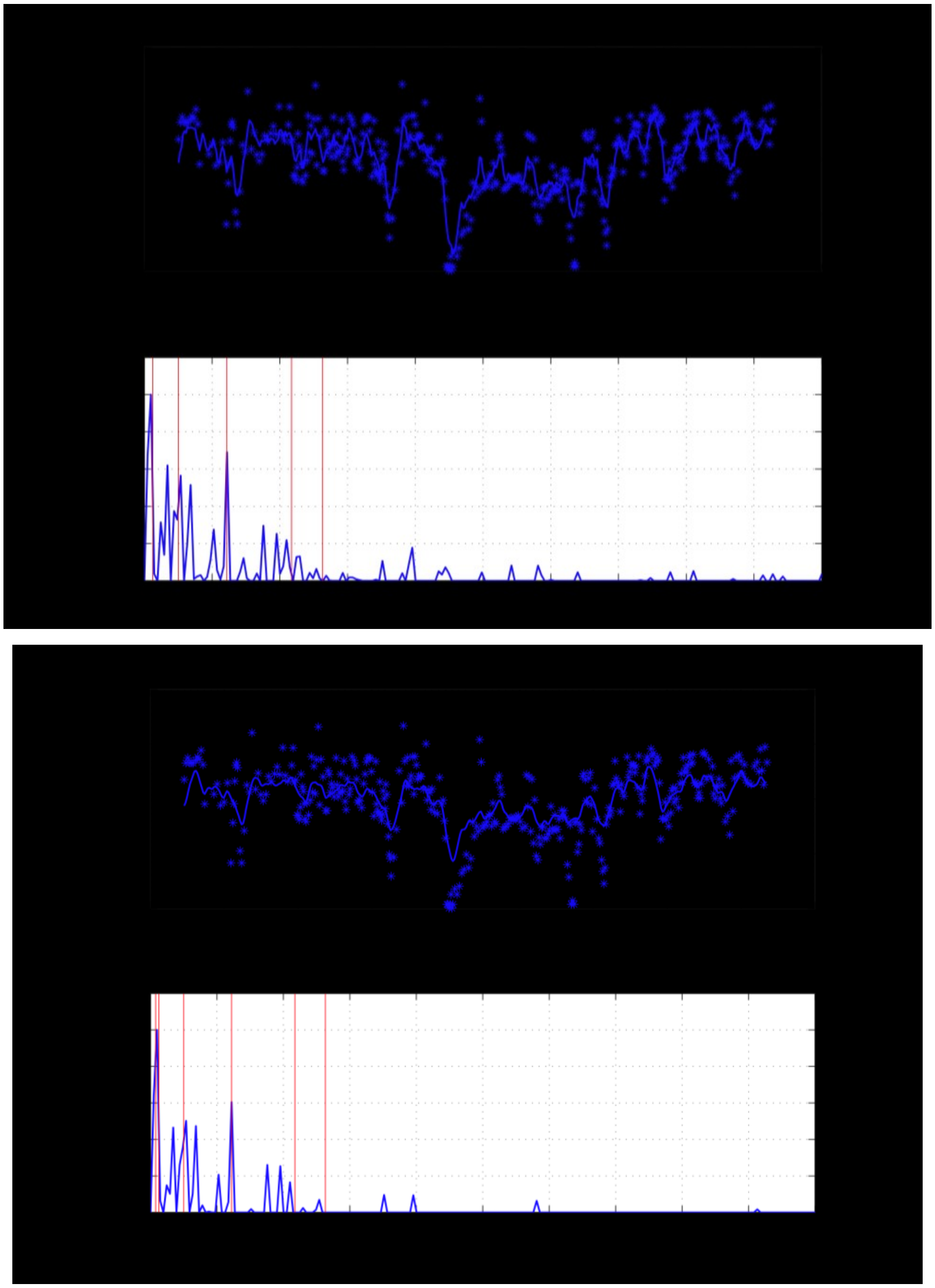

Figure 2. a) Reconstruction of a segment of the Lake Baikal magnetic susceptibility record using sparsity promoting inversion b). The associated sparse amplitude spectrum. c) Reconstruction of the time series when outliers are de-emphasized in the inversion via an error-reweighting strategy. d) Final sparse amplitude spectrum obtained with the error-reweighting strategy. Vertical red lines indicate known Milankovitch periods T=640, 400, 100, 41, 23, 19 ky.