Fast Successive Spectral Estimation of Irregularly Sampled Data

Peter A. Parker
Los Alamos National Laboratory
P. O. Box 1663, MS D466
Los Alamos, NM 87545
peter.parker@lanl.gov

***Abstract*—Techniques for estimation of a dense spectrum from irregularly sampled data are typically either very processing intensive or suffer from large amounts of bias in the estimate due to signal leakage. This paper proposes an estimation algorithm that has similarities to the successive interference cancellation algorithms from the communications literature. The algorithm successively estimates larger amplitude frequency components first and then subtracts out those estimates before continuing on to lower amplitudes. The algorithm is able to maintain the low complexity of an FFT-based algorithm while overcoming the poor bias performance typically associated with those algorithms.**

I. Introduction

Spectral estimation techniques for the case when data are sampled irregularly exist mainly for the case of sparse spectra (see [1] for a review paper of techniques). That is, when the spectrum only has a few discrete sources such as single frequency tones. For the case of a dense spectrum, there are far fewer options [2], [3], [4], [5] as well as using standard spectral estimation techniques after interpolation. These options can be classified into $\mathcal{O}\left(N \log N\right)$ type algorithms like interpolation of the data followed by standard periodogram processing, $\mathcal{O}(N^2)$ algorithms that use the Discrete Fourrier Transform directly on the irregularly spaced samples, and $\mathcal{O}(N^3)$ algorithms that rely on Maximum Likelihood or Least Squares principals. This paper proposes an ad hoc $\mathcal{O}(N \log N)$ algorithm that has performance comparable to the much more complex algorithms.

The case of a dense spectrum estimation with irregular sampling occurs in fields dealing with estimating parameters of a turbulent flow [6], [7], [8]. These may be either one or two dimensional data sets, so it is important that the algorithm be extensible to multiple dimensions. For these turbulent flow data problems, the spectrum is usually described by a power law or some other monotonically decreasing spectrum. The algorithm presented takes advantage of this fact, although it may be extended to dealing with other smooth spectral functions.

II. Fast Spectral Estimation

A. Algorithm Description

The proposed algorithm overcomes the signal leakage problem of other fast algorithms by successively estimating the largest amplitude frequency component first and then subtracting out this component before estimating the next largest. This is similar to the way successive interference cancellation algorithms work in the communications literature [9]. Figure 1 illustrates how this algorithm works. The smooth green line labeled simulated data spectrum is the actual spectrum to be estimated and the yellow "Fourier transform of the data" shows what a naive Fourier transform of the data would be (note the large amount of signal leakage at higher frequencies). The algorithm starts by fitting a first order polynomial to the irregularly sampled data (blue curve in the figure). The spectrum of this first order fit is calculated and only the spectrum within T dB of the maximum is kept ($T = 5$ for the example shown below). A residual is then calculated by subtracting the polynomial fit from the data. Let $\mathbf{x}_1$ be the raw data and $\hat{\mathbf{x}}_1$ be the first order fit. Then define the residual as

$$\mathbf{x}_2 = \mathbf{x}_1 - \hat{\mathbf{x}}_1. \tag{1}$$

This algorithm iterates as follows:

1) Fit a n^{th} order polynomial to the n^{th} residual signal $\mathbf{x}_n$. This fit is represented as $\hat{\mathbf{x}}_n$.
2) Calculate the spectrum of $\hat{\mathbf{x}}_n$ and add the top T dB of the spectrum to the previously calculated spectrum.
3) Calculate the $(n+1)^{th}$ residual as

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \hat{\mathbf{x}}_n. \tag{2}$$

The n^{th} order polynomial fit to each residual is shown in the figure. This algorithm maintains the $\mathcal{O}(N \log N)$ computational order by resampling the polynomial fitted data to a uniform sample rate with the same number of samples as the input data and using the FFT to calculate the spectrum. All other processing within this algorithm is $\mathcal{O}(N)$.

B. Performance of Algorithm

An example of the performance of this algorithm is illustrated in Figure 2. For this case, simulated data are generated by running Gaussian noise through a 4^{th} order digital Butterworth filter with a cutoff frequency of 0.005 Hz and a sample rate of 1 Hz. The data are then decimated randomly such that the average time between samples is 10 seconds and there are only 100 samples to calculate the spectrum. This small number of samples is a challenge to the Maximum-Likelihood based algorithms. As seen in this figure, the average performance of the algorithm is a close match to the truth down to about 55 dB below the peak.

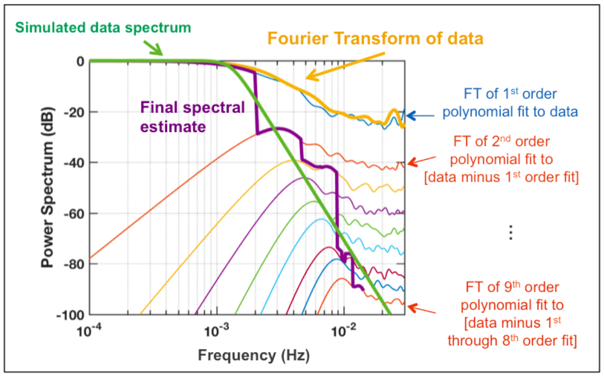

Fig. 1. Illustration of the fast successive spectral estimation algorithm. Each stage of the approximation is shown in relation to the final spectral estimate as well as the truth spectrum.

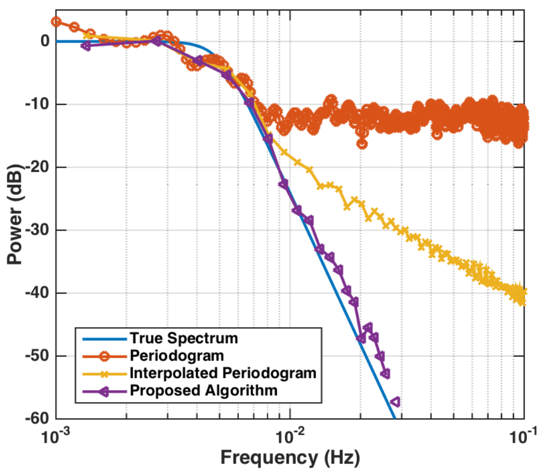

Fig. 2. Average estimated spectrum (average taken over 20 realizations). Solid line represents true spectrum (4^{th} order digital Butterworth filter with cutoff at 0.005 Hz), circles represents periodogram using the DFT directly on irregularly spaced samples, 'x's represents the periodogram of an interpolated data set (using FFT), and triangles represent proposed algorithm.

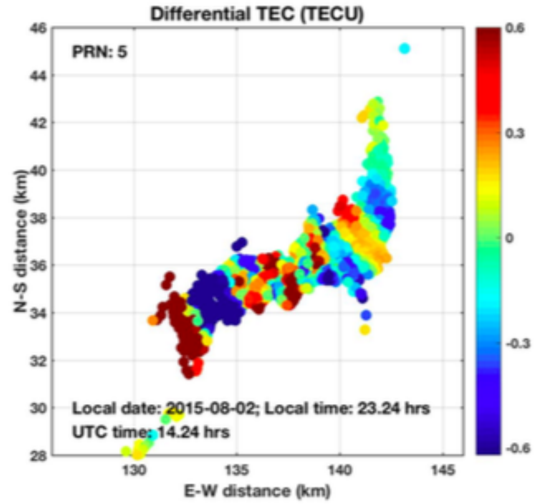

Fig. 3. Plot of TEC fluctuations as derived from GPS receiver data in Japan.

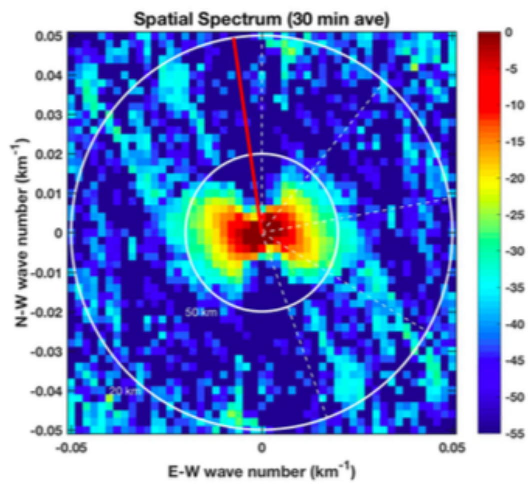

Fig. 4. Estimated 2-dimensional spectral density function for Japan on 2 August 2015.

III. Application

This algorithm is easily extensible to multi-dimensional data. For this, a d-dimensional polynomial is used to approximate the data at each stage and a d-dimensional FFT is used for computing the spectrum.

Currently, this algorithm is being used to estimate a 2-dimensional ionospheric irregularity spectral density function [6]. Data from large networks of GPS receivers in both Japan and California are being used to estimate the fluctuations in the total electron content (TEC) of the ionosphere (see Figure 3 for an illustration of the differential TEC over Japan on 2 August 2015). The spectral density of these fluctuations are estimated through the proposed algorithm and results are shown in Figure 4.

References

[1] P. Babu and P. Stoica, "Spectral analysis of nonuniformly sampled data – a review," *Digital Signal Processing*, 2010.

[2] P. Stoica and P. Babu, "Naximum-likelihood nonparametric estimation of smooth spectra from irregularly sampled data," *IEEE Trans. Sig. Proc.*, 2011.

[3] T. P. Bronez, "Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences," *IEEE Trans. Acoustics, Speech and Sig. Proc.*, 1988.

[4] J. Sward, J. Brynolfsson, A. Jakobsson, and M. Hansson-Sandsten, "Smooth 2-D frequency estimation using covariance fitting," in *European Signal Processing Conference*, 2014.

[5] P. Daniell, "On the theoretical specification and sampling properties of autocorrelated time-series," *Journal of Royal Statistical Society*, 1946.

[6] E. H. Lay, M. E. Light, P. A. Parker, and C. S. Carrano, "Determining ionospheric irregularity spectral density function from Japan GEONET," in *AGU Fall Meeting*, 2015.

[7] A. Jessup, W. Melville, and W. Keller, "Breaking waves affecting microwave backscatter 1. detection and verification," *J. Geophys. Res.*, 1991.

[8] L. Benedict, H. Nobash, and C. Tropea, "Estimation of tubulent velocity spectra from laser Dopper data," *Meas. Sci. Technol.*, 2000.

[9] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE Journal on Selected Areas in Communications*, 1994.