BAYESIAN SPECTRUM ESTIMATION OF UNEVENLY SAMPLED NONSTATIONARY DATA

Yuan Qi[†], *Thomas P. Minka*[‡], *and Rosalind W. Picard*[†]

[†]MIT Media Laboratory, Cambridge, MA, 02139, U.S.A.
[‡]Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

ABSTRACT

Spectral estimation methods typically assume stationarity and uniform spacing between samples of data. The non-stationarity of real data is usually accommodated by windowing methods, while the lack of uniformly-spaced samples is typically addressed by methods that "fill in" the data in some way. This paper presents a new approach to both of these problems: we use a non-stationary Kalman filter within a Bayesian framework to jointly estimate all spectral coefficients instantaneously. The new method works regardless of how the signal samples are spaced. We illustrate the method on several data sets, showing that it provides more accurate estimation than the Lomb-Scargle method and several classical spectral estimation methods.

1. INTRODUCTION

Spectrum estimation has been a classical research topic in signal processing communities for decades. Many approaches have been proposed ranging from modified periodogram, AR model based estimation, to the MUSIC algorithm [1]. Though all these algorithms have their own advantages, they all have two basic limitations: first , they work only for evenly sampled signals; second, they have the same stationarity assumption of the signal.

For unevenly sampled signals, the Lomb-Scargle periodogram is widely used [2, 3]. The Lomb-Scargle periodogram models the data as a single stationary sinusoid wave. It is based on Maximum likelihood estimation, which is later given an Bayesian interpretation by Bretthorst [4].

2. A BAYESIAN FRAMEWORK FOR NONSTATIONARY SPECTRUM ESTIMATION

In this section, we introduce a new Bayesian framework for estimating the nonstationary spectrum of a given signal. This framework does not assume any short time stationarity of the signals which by contrast classical spectrum estimation approaches are based on. The method works both for evenly and unevenly sampled data.

For the spectrum estimation problem, we observe the data $\mathbf{x}$: $\mathbf{x} = [x_1, x_2, \ldots, x_i, \ldots, x_N]^T$, where x_i is sampled at time t_i. When the data is unevenly sampled, $\mathbf{t} = [t_1, \ldots, t_N]^T$ contains useful information for spectrum estimation. We model the data as

$$x_i = a_{i0} + \sum_{j=1}^{M} a_{ij}\sin(2\pi f_j t_i) + b_{ij}\cos(2\pi f_j t_i) + v_i \quad (1)$$

$$\text{for } i = 1, \ldots, N.$$

where v_i is a noise variable. The number and value of frequency bands, M and f_j, could be chosen based on our prior knowledge. Later in this paper, we simply choose all the frequency bands to be equally spaced. Both a_{ij} and b_{ij} have real values. Note that for a nonstationary signal, a_{ij}, b_{ij}, and v_i depend on the sampling time t_i.

The use of a_{ij} and b_{ij} allows the signal to have a nonstationary amplitude $\sqrt{a_{ij}^2 + b_{ij}^2}$ and a changing phase $\arctan(\frac{b_{ij}}{a_{ij}})$ for the j^{th} frequency band at time t_i.

For equation (1), we define

$$\mathbf{s}_i = [a_{i0}, a_{i1}, a_{i2}, \ldots, a_{iM}, b_{i1}, b_{i2}, \ldots, b_{iM}]^T \quad (2)$$

$$\mathbf{c}_i = [1, \sin(2\pi f_1 t_i), \ldots, \sin(2\pi f_M t_i), \cos(2\pi f_1 t_i), \ldots, \cos(2\pi f_M t_i)] \quad (3)$$

For the nonstationary spectrum estimation, our goal is estimating the state vector $\mathbf{s}_i$ instantaneously at the sampling time t_i. To this end, we assume that the hidden states $\mathbf{s}_1 \dots \mathbf{s}_N$ form a Markov chain that emits a time series of observation $x_1 \dots x_N$:

$$\begin{aligned} \mathbf{s}_i &= \mathbf{s}_{i-1} + \mathbf{w}_i & (4) \\ x_i &= \mathbf{c}_i \mathbf{s}_i + v_i & (5) \end{aligned}$$

where $\mathbf{w}_i$ is the process noise at the sampling time t_i, and v_i is the observation noise at t_i. Based on the data we have, we can model the process and observation noises by Gaussian distributions or heavy-tailed non-Gaussian distributions. However, using non-Gaussian distributions will invokes the use of numerical approximation techniques in the inference procedure.

According to this model, the joint distribution of hidden states and observations can be computed as

$$p(\mathbf{s}_{1:N}, \mathbf{x}_{1:N}) = p(\mathbf{s}_1)p(x_1|\mathbf{s}_1)\prod_{i=2}^{N} p(\mathbf{s}_i|\mathbf{s}_{i-1})p(x_i|\mathbf{s}_i) \quad (6)$$

where $\mathbf{s}_{1:N} = [\mathbf{s}_1, \dots, \mathbf{s}_N]^T$ and $x_{1:N} = [x_1, \dots, x_N]^T$ denotes collections of states and observations from time t_1 to t_N.

The filtering distribution $p(\mathbf{s}_i|x_{1:i})$ can be sequentially estimated as follows

$$\begin{aligned} p(\mathbf{s}_i|x_{1:i-1}) &= \int_{\mathbf{s}_{i-1}} p(\mathbf{s}_i|\mathbf{s}_{i-1})p(\mathbf{s}_{i-1}|x_{1:i-1}) & (7) \\ p(\mathbf{s}_i|x_{1:i}) &= \frac{p(\mathbf{x}_i|\mathbf{s}_i)p(\mathbf{s}_i|x_{1:i-1})}{p(x_i|x_{1:i-1})} & (8) \end{aligned}$$

Then the spectrum at time t_i can be summarized by the mean of $p(\mathbf{s}_i|x_{1:i})$.

3. SPECTRUM ESTIMATION BY KALMAN FILTERING

3.1. Algorithm

If we use linear Gaussian models in equations (4) and (5):

$$\begin{aligned} \mathbf{w}_i &\sim \mathcal{N}(\mathbf{0}, \Gamma_i) & (9) \\ v_i &\sim \mathcal{N}(\mathbf{0}, \sigma^2), & (10) \end{aligned}$$

then $p(\mathbf{s}_i|x_{1:i-1})$ is also Gaussian, and we can use Kalman filtering to efficiently update these probabilities. To deal with the uneven sampling, we set

$$\Gamma_i = \mathbf{Z}(t_i - t_{i-1}); \quad (11)$$

where $\mathbf{Z}$ is a pre-defined constant matrix, which we say more about below, and $\Gamma_0 = 0$.

Denote $\mathbf{m}_i$ and $\mathbf{V}_i$ as the mean and covariance matrix of $p(\mathbf{s}_i|x_{1:i})$. We have the following Kalman filtering update [5, 6] equations:

$$\begin{aligned} \mathbf{m}_i &= \mathbf{m}_{i-1} + \mathbf{K}_i(x_i - \mathbf{c}_i\mathbf{m}_{i-1}) & (12) \\ \mathbf{V}_i &= (\mathbf{I} - \mathbf{K}_i\mathbf{c}_i)\mathbf{P}_{i-1} & (13) \end{aligned}$$

where

$$\begin{aligned} \mathbf{P}_{i-1} &= \mathbf{V}_{i-1} + \Gamma_{i-1} & (14) \\ \mathbf{K}_i &= \mathbf{P}_{i-1}\mathbf{c}_i^T(\mathbf{c}_i\mathbf{P}_{i-1}\mathbf{c}_i^T + \sigma^2)^{-1} & (15) \end{aligned}$$

Note that we have a nonstationary Kalman filtering algorithm; both $\mathbf{c}_i$ and Γ_{i-1} vary with time.

The recursions start off with

$$\begin{aligned} \mathbf{m}_1 &= \mathbf{m}_0 + \mathbf{K}_1(x_1 - \mathbf{c}_1\mathbf{m}_0) & (16) \\ \mathbf{V}_1 &= (\mathbf{I} - \mathbf{K}_1\mathbf{c}_1)\mathbf{V}_0 & (17) \\ \mathbf{K}_1 &= \mathbf{V}_0\mathbf{c}_1^T(\mathbf{c}_1\mathbf{V}_0\mathbf{c}_1^T + \sigma^2)^{-1} & (18) \end{aligned}$$

where $\mathbf{m}_0$ and $\mathbf{V}_0$ are pre-defined hyper-parameters for the prior distribution $p(\mathbf{s}_0)$, which we say more about below.

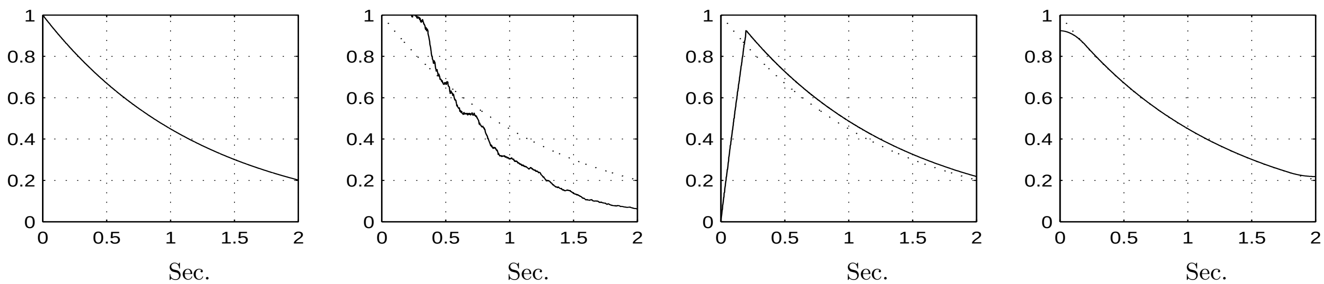
If we want to utilize not only the past information, but also the future information in the data set to estimate the spectrum, we may want to compute $p(\mathbf{s}_i|x_{1:N})$ where $x_{1:N}$ is the whole data set. As a well-known technique, Kalman smoothing can be employed to compute this posterior distribution [5].

3.2. Model Parameters and Hyperparameters

As a Bayesian method, this new algorithm allows us to incorporate prior knowledge into the estimation. First, if we have no information about the frequency content of the data, we may set Γ_i or more exactly $\mathbf{Z}$ in equation (11) to be a scaled identity matrix: $\mathbf{Z} = z\mathbf{I}$, where the positive, real-valued amplitude z controls the time variability of the amplitude of the estimated frequencies. Second, we may use a noninformative prior distribution $p(\mathbf{s}_0)$ by assigning $\mathbf{V}_0$ to be a scaled identity matrix, and $\mathbf{m}_0$ to be a zero vector. Third, if we think the data might only contain some known frequencies, we can represent our belief by assigning $\mathbf{m}_0$ to be a zero vector, and setting $\mathbf{V}_0$ to be a diagonal matrix with small variances for the 0 elements in $\mathbf{m}_0$. These are a few of the ways in which prior knowledge can be easily incorporated to guide the estimation.

3.3. Conquer Aliasing by Unevenly Sampling

In [7], G. L. Bretthorst showed that a generalization of the DFT can handle the case when data is unevenly sampled, resulting in a much larger effective bandwidth



(a) True Amplitudes (b) Lomb-Scargle (c) Kalman (d) Kalman Smoothing

Figure 1: True and estimated amplitudes for an unevenly sampled signal that contains one 125 Hz sinusoid modulated with an exponentially fast decaying amplitude.

than when the DFT is used on evenly sampled data. For spectrum estimation by Kalman filtering, the similar effect of unevenly sampling holds: the critical frequency beyond which aliasing occurs may be almost infinite for unevenly sampled data.

Let us first consider the reason why aliasing exists in the Kalman filtering spectrum for evenly sampled data. When the data are evenly sampled, we have

$$t_i = i\Delta t, \quad \text{for } i = 1, \ldots, N. \tag{19}$$

$$f_c = \frac{f_s}{2} = \frac{1}{2\Delta t} \tag{20}$$

where Δt is the time interval between two samples, f_s is the sampling frequency, and f_c is the cut-off frequency determined by the Nyquist criterion.

For simplicity, consider only the cosine basis in $\mathbf{c}_i$ defined in equation (3). If the frequency components in $\mathbf{c}_i$ are equally spaced between $-f_c$ and $-3f_c$, then $\mathbf{c}_i$ equals

$$\begin{aligned}
&[\cos(2\pi(-f_c+f_1)i\Delta t), \ldots, \cos(2\pi(f_c)i\Delta t) \\
&\quad \cos(2\pi(f_c+f_1)i\Delta t), \ldots, \cos(2\pi(3f_c)i\Delta t)] &(21)\\
&= [\cos(2\pi(-f_c+f_1)i\Delta t), \ldots, \cos(2\pi f_c i\Delta t) \\
&\quad \cos(2\pi(-f_c+f_1)i\Delta t + 2\pi(2f_c)i\Delta t), \ldots, \\
&\ldots, \cos(2\pi f_c i\Delta t + 2\pi(2f_c)i\Delta t)] &(22)\\
&= [\cos(2\pi(-f_c+f_1)i\Delta t), \ldots, \cos(2\pi f_c i\Delta t) \\
&\quad \cos(2\pi(-f_c+f_1)i\Delta t), \ldots, \cos(2\pi f_c i\Delta t)] &(23)
\end{aligned}$$

The frequencies corresponding to the repeated elements in $\mathbf{c}_i$ will have the same probabilities in Kalman filter using a non-informative prior. Thus, the spectral coefficients in the range from $-f_c+f_1$ to f_c are the same as those in the range from f_c+f_1 to $3f_c$. In other words, aliasing happens.

Now if the data are unevenly sampled, the time intervals between two samples may differ. Denote the largest common factor of all t_i's as $\Delta t'$. Then it follows $t_i = k_i\Delta t'$, for $i = 1, \ldots, N$, where k_i is an integer. For evenly sampled data, $k_i = 0, 1, \ldots, N-1$. For unevenly sampled data, k_i may start from a very large number.

Then we define the Nyquist cut-off frequency for irregular sampled data f_c' as $\frac{1}{2\Delta t'}$. Aliasing can still occur if $\mathbf{c}_i$ contains frequencies larger than f_c'. Note that $\Delta t'$ is less than or equal to the smallest time interval between data points. When the sampling is random, $\Delta t'$ may be as small as the numerical resolution of the system. For example, if t_i is stored by a 32 bit number, $\Delta t'$ will be around 2^{-32} and f_c' will be around 2^{31} Hz.

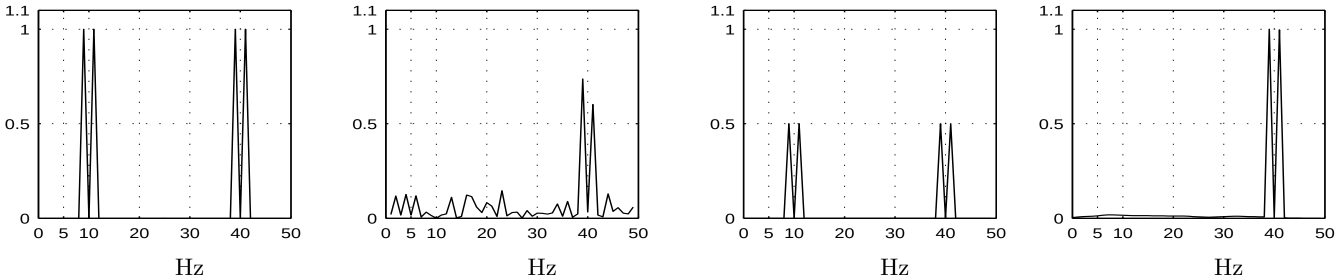
In other words, when the data are randomly sampled, or unevenly sampled in a well-designed way, use of this Kalman filtering spectrum estimation results in almost infinite f_c' and an infinite effective bandwidth that is essentially aliasing free.

4. EXPERIMENTS AND DISCUSSIONS

We test our algorithm and compare it with the Lomb-Scargle periodogram, which is widely used for spectrum estimation of unevenly sampled data, as well as with several other classical methods for evenly sampled data.

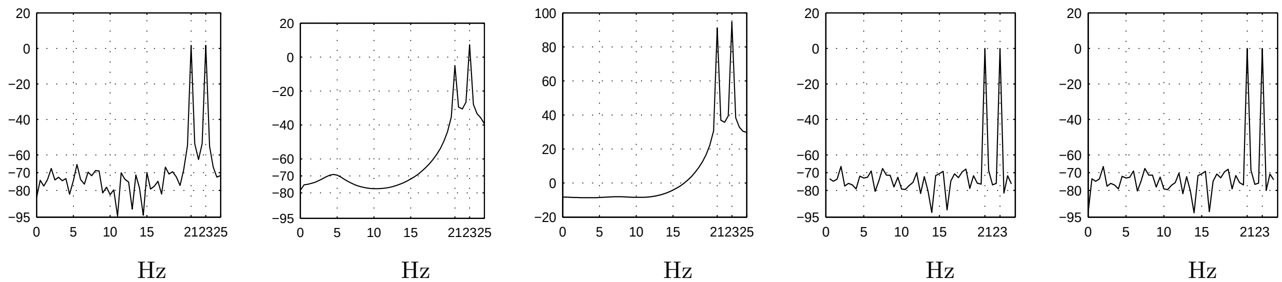
For the first evaluation, we synthesize an unevenly sampled signal that contains one 125Hz sinusoid wave modulated with an exponentially fast decaying amplitude. We compare the Lomb-Scargle periodogram, Kalman filtering, and Kalman smoothing. For the Lomb-Scargle periodogram, we use a short window size of 60 data points, with 59 points of overlap; less overlap yields visible "blocking" effects. For Kalman filtering and smoothing, we set $\mathbf{Z}$ to be a scaled identity matrix ($z = 1000$) in equation (11), and assign a noninformative prior on $p(\mathbf{s}_0)$ ($\mathbf{V}_0 = 1 \times 10^{10}$).

The true amplitude and the estimated amplitudes of 125 Hz components are plotted in figure 1. Except for the initialization (0.2 seconds) for Kalman filtering, the mean square error of the estimated amplitudes along the time axis is 0.0016; for Kalman smoothing, the mean square error is 0.0000080. For the Lomb-Scargle periodogram, the mean square error is 0.0384.



(a) Lomb-Scargle, even (b) Lomb-Scargle, uneven (c) Kalman, even (d) Kalman, uneven

Figure 2: Lomb-Scargle periodogram and Kalman Filtering spectra for a signal $\mathbf{x} = \sin(2\pi 39\mathbf{t}) + \sin((2\pi 41\mathbf{t})$ sampled 100 times over 2 seconds, with samples either evenly or randomly (unevenly) spaced.



(a)Modified Periodogram (b) Burg's algorithm (c) MUSIC algorithm (d) Lomb-Scargle (e) Kalman

Figure 3: Comparison of different spectral estimation algorithms. The signal is the sum of 21 and 23 Hz real sinusoid waves and white noise with variance 0.001. The signal is evenly sampled at 50 Hz over 2 seconds.

Kalman filtering and smoothing yield accurate estimates of the frequency and fast decaying amplitude, while the Lomb-Scargle periodogram fails to track the changing amplitude. Also, the Lomb-Scargle periodogram contains much more sidelobes than in the spectrograms obtained from Kalman filtering and smoothing. This is partly because Kalman filtering and smoothing joint estimate all the frequency bands and thus have the "explaining-away' effect: if the signal is well explained by one or some of the frequency bands, the influence of other frequency bands will be reduced. On the other hand, for the Lomb-Scargle periodogram, there is no interaction between the estimation of different frequency bands.

Next, we show (figure 2) that unevenly sampling removes aliasing in the traditional sense of bandlimiting to f_c. Results (a) and (c) show aliasing with even sampling; (b) and (d) do not show any aliasing, despite sampling the signal that contains 39 and 41 Hz waves at average 50Hz. Again, Kalman filtering approach outperforms the Lomb-Scargle periodogram. The Lomb-Scargle periodogram is based on a single frequency data model. Thus, its estimation of two close frequencies interfers each other, which in turn affects the spectrum estimation accuracy. In addition, comparing (c) and (d), we see that the estimation of Kalman filtering has an amplitude conservation property, i.e., the estimated amplitudes are equally distributed in the true and aliasing frequencies in (c).

Finally, we compare several classical methods with Lomb-Scargle and Kalman methods, on evenly sampled data (figure 3). For the modified periodogram, we use a Hamming window. For the Burg's algorithm, we choose a 6^{th} order AR model. For the MUSIC algorithm, we set the the signal subspace dimension to be 4. Note that the Y axes in figure 3 are in logarithm scale.

5. CONCLUSION

This paper has proposed a Bayesian Kalman-filter based method for spectrum estimation. Motivated by the need to handle unevenly-sampled noisy non-stationary data, we find that some of these problems (namely, the unevenly sampled nature) are actually advantageous in some sense. Our new method jointly estimates all the amplitudes and phases of frequency bands of interest instantaneously without the use of windowing, and is easily able to accomodate prior information about noise and signal structure. It is shown to provide outstanding frequency resolution, even on small data sets.

6. REFERENCES

[1] M. H. Hayes, *Statistical digital signal processing and modeling*, John Wiley and Sons, Inc, 1996.

[2] N. R. Lomb, "Least-squares frequency analysis of unevenly spaced data," *Astrophysical and Space Science*, pp. 447–462, 1976.

[3] J. D. Scargle, "Studies in astronomical time series analysis ii. statistical aspects of spectral analysis of unevenly sampled data," *Astrophysical Journal*, pp. 835–853, 1982.

[4] G. L. Bretthorst, "Bayesian spectrum analysis and parameter estimation," in *Lecture Notes in Statistics*, 48. Springer-Verlag, 1988.

[5] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, N.J., 1979.

[6] T. P. Minka, "From hidden markov models to linear dynamical systems," Tech. Rep. 531, Vision and Modeling Group of Media Lab, MIT, 1998.

[7] G. L. Bretthorst, "Nonuniform sampling: Bandwidth and aliasing," in *Maximum Entropy and Bayesian Methods in Science and Engineering*, 2000, pp. 1–28.