12-2019

Period Estimation and Denoising Families of Nonuniformly Sampled Time Series

William Seguine
East Tennessee State University

Period Estimation and Denoising Families of Nonuniformly Sampled Time Series

---

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

---

by

William Seguine

December 2019

Jeff Knisley, Ph.D., Chair

Rodney Keaton, Ph.D.

Michele Joyner, Ph.D.

Keywords: nonuniformly sampled, consensus, spectrum, time series, least squares.

ABSTRACT

Period Estimation and Denoising Families of Nonuniformly Sampled Time Series

by

William Seguine

Nonuniformly sampled time series are common in astronomy, finance, and other areas of research. Commonly, these time series belong to a family of signals recorded from the same phenomenon. Period estimation and denoising of such data relies on periodograms. In particular, the Lomb-Scargle periodogram and its extension, the Multiband Lomb-Scargle, are at the forefront of time series period estimation. However, these methods are not without flaws. This paper explores alternatives to the Lomb-Scargle and Multiband Lomb-Scargle. In particular, this thesis uses regularized least squares and the convolution theorem to introduce a spectral consensus model of a family of nonuniformly sampled time series.

ACKNOWLEDGMENTS

I would like to thank my committee chairman, Dr. Knisley, for his assistance while working on this thesis. Without his guidance, understanding and patience, this thesis would not have been enjoyable. I would also like to thank the committee members; Dr. Joyner and Dr. Keaton, for their invaluable input in this process. Finally, my sincerest gratitude goes to my family whose love and encouragement kept me working.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 Nonuniformly Sampled Time Series

Time series are processes that evolve over time. These are denoted $\{x_0, x_1, \ldots, x_N\}$ where x_k is defined to be the state of the process at time t_k. They are commonly used in the financial world to model a the evolution of the price of an asset over a span of time [15].

In meteorology, time series can be used to study the effects of weather variations on diseases and pest populations in agriculture [2]. Using time series analysis, Garrett *et al.* in [2] found that crop yields could vary regardless of pest organism growth and variations in weather conditions. Time series are also used to study and predict earthquakes [10].

Time series are used in astronomy to study the intensity of electromagnetic radiation emitted from stars, which is called luminosity. In particular, Vanderplas *et al.* in [17] used data from RR Lyrae variable stars to illustrate the effectiveness of the Multiband Lomb-Scargle algorithm.

There are two distinct types of time series: uniformly sampled, and nonuniformly sampled. A uniformly sampled time series $\{x_0, x_1, \ldots, x_N\}$ is of the form

$$x_k = f(k\Delta t + a) \tag{1}$$

for some function f and fixed Δt and a. Because Δt is a fixed value, there is a uniform time interval between each sample measurement. An example of one such time series can be found in Figure 1 .

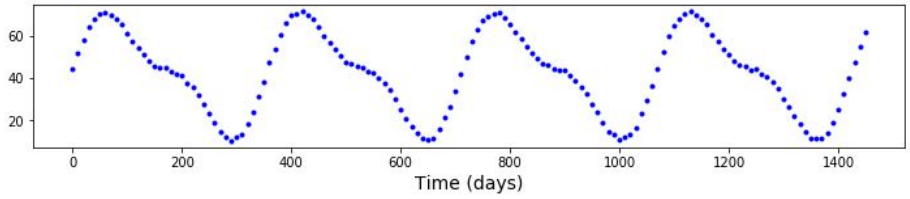

Figure 1: A Uniformly Sampled Time Series

If Δt in Equation 1 is a random variable with a nontrivial probability distribution (i.e. $\Delta t \in P$ a positive-valued probability distribution), the time series is said to be nonuniformly sampled (also called an unevenly sampled time series [13]). In astronomy, nonuniformly sampled time series may arise due the restrictions of the instrumentation used, or may arise due to environmental variations. In fields such as biology, there may be constraints on samples that lead to a time series becoming nonuniformly sampled [3].

Whether uniformly or nonuniformly sampled, a time series X_t can be written as

$$X_t = S_t + R_t \tag{2}$$

where S_t is said to be the signal and R_t is the noise component (or variation due to observational errors) [13]. The noise component R_t in equation (2) is assumed to be normally distributed with a finite variance and mean μ of zero [13]. We also assume that the signal is stationary, can be represented as a Fourier series, or both.

In finance, astronomy, biology, and similar fields of research, it may be necessary to conduct more than one observation or experiment of the same phenomenon for the same reason. This results in a set of similar time series as they are all samplings of a similar source. We call a set of such data a family of time series. It may also

be necessary to determine the period of S_t (if any) in order to make more accurate observations and predictions regarding data. The most recent development known for determining the period of times series (uniformly or nonuniformly sampled) is the Lomb-Scargle periodogram and its extension to a family of time series the Multiband Lomb-Scargle [17]. However, this approach is not without issues as discussed in the third section of this chapter.

1.2 Chi-Square Estimation

The chi-square estimation is an approach to estimating the maximum likelihood of a frequency in a periodogram. Employing this method requires the assumption that the missing values are distributed according to a chi-square distribution. Data corresponding to a single phenomenon is often in the form (t_j, x_j) where x_j is the recorded value of a phenomenon at time t_j. In the simplest case, S_t is a single sinusoid and the noise component R_t is zero.

We can thus assume the data is modeled by $\phi(t)$ given by

$$\phi(t) = A\sin(\omega t) + B\cos(\omega t). \tag{3}$$

We want to find the values of amplitudes and frequency A,B, and omega respectively such that if $\theta = [A, B, \omega]$, then

$$\theta = \text{argmin}_{\theta\in\Omega}\left(\sum_j |x_j - \phi(t_j;\theta)|^2\right) \tag{4}$$

where Ω is a set of admissible parameters In order to do this, consider the function

(called the chi-square comparison)

$$P(\omega) = \frac{1}{2} \sum_j (x_j - \phi(t_j; \theta))^2 . \tag{5}$$

By our assumption of the model of the data $\phi(t)$, this implies that x_j must also be of the form

$$x_j = A \sin(\omega t_j) + B \cos(\omega t_j) . \tag{6}$$

That is, the data (t_j, x_j) must be in the same form as the assumed model. In order to determine the optimal values for A and B, we multiply the above equation separately by $B \cos(\omega t_j)$ and $A \sin(\omega t_j)$. This results in the equation

$$Bx_j \cos(\omega t_j) = AB \sin(\omega t_j) \cos(\omega t_j) + B^2 \cos^2(\omega t_j) \tag{7}$$

as well as the associated equation

$$Ax_j \sin(\omega t_j) = AB \sin(\omega t_j) \cos(\omega t_j) + A^2 \sin^2(\omega t_j) . \tag{8}$$

From this, we have a system of equations with which the values of A and B can be computed explicitly in terms of $\sin(\omega t_j)$ and $\cos(\omega t_j)$. Consequently, we need only to estimate ω. The Lomb-Scargle periodogram employs the Chi-square estimation method as discussed in the next section.

1.3 The Lomb-Scargle Periodogram

The Lomb-Scargle periodogram was first detailed by Lomb in 1976 as a method for determining the period of nonuniformly sampled time series [16]. Those time series of particular interest to Lomb were those within the field of astronomy. This is not

to say, however, that the method is limited in use to astronomical data. The Lomb-Scargle periodogram is a common tool in biology as well. It has been employed by Glynn et al. to determine the periodicity of gene expression patterns of a particular species of protozoans [3]. It was also used to determine the underlying periodicity of body temperatures of organisms prior to hibernation and to determine the period of the metabolic cycle of crassulacean acid metabolism of certain plants [3].

Letting $\omega = 2\pi f$, the method begins with the classical periodogram (i.e. $P(f)$) for a time series $g(t)$ given in [16] by

$$P(f) = \frac{1}{N}\left|\sum_{n=1}^{N} g_n e^{-2\pi i f t_n}\right|^2, \tag{9}$$

where f is the frequency (i.e. the reciprocal of the period). Observe that this expression is similar to a squared expression of the Fourier Transform. While the two appear similar, they have different interpretations [16]. One can then apply the definition of the complex exponential, as given in [16], resulting in

$$P(f) = \frac{1}{N}\left(\sum_{n=1}^{N} g_n \cos(2\pi f t_n)\right)^2 + \frac{1}{N}\left(\sum_{n=1}^{N} g_n \sin(2\pi f t_n)\right)^2. \tag{10}$$

An issue with this method is the variation is unavoidable when conducting a statistical estimation of period [16]. In order to make this method invariant to time shifts in the data, Scargle introduced the general periodogram

$$P_S(f) = \frac{A^2}{N}\left(\sum_{n=1}^{N} g_n \cos\left(2\pi f\left(t_n - \tau\right)\right)\right)^2 + \frac{B^2}{N}\left(\sum_{n=1}^{N} g_n \sin\left(2\pi f\left(t_n - \tau\right)\right)\right)^2 \tag{11}$$

where $\frac{A^2}{N}$ and $\frac{B^2}{N}$ are defined by

$$\frac{A^2}{N} = \frac{1}{2\sum_{n=1}^{N}\cos^2\left(2\pi f\left(t_n - \tau\right)\right)} \quad \text{and} \quad \frac{B^2}{N} = \frac{1}{2\sum_{n=1}^{N}\sin^2\left(2\pi f\left(t_n - \tau\right)\right)}. \tag{12}$$

The shift τ is related to $P_s(f)$ by

$$\tau = \frac{1}{4\pi f} \tan^{-1}\left(\frac{\sum_{n=1}^{N} \sin\left(4\pi f t_n\right)}{\sum_{n=1}^{N} \cos\left(4\pi f t_n\right)}\right). \tag{13}$$

The result is the Lomb-Scargle periodogram given by

$$P_X(f) = \frac{1}{2}\left[\frac{\left(\sum_{n=1}^{N} g_n \cos\left(2\pi f\left(t_n-\tau\right)\right)\right)^2}{2\sum_{n=1}^{N}\cos^2\left(2\pi f\left(t_n-\tau\right)\right)} + \frac{\left(\sum_{n=1}^{N} g_n \sin\left(2\pi f\left(t_n-\tau\right)\right)\right)^2}{2\sum_{n=1}^{N}\sin^2\left(2\pi f\left(t_n-\tau\right)\right)}\right] \tag{14}$$

subject to the constraint

$$\tau = \frac{1}{4\pi f} \tan^{-1}\left(\frac{\sum_{n=1}^{N} \sin\left(4\pi f t_n\right)}{\sum_{n=1}^{N} \cos\left(4\pi f t_n\right)}\right). \tag{15}$$

Scargle also modified the work of Lomb et al. in order to reduce the effects of aliasing: an appearance of signals with higher frequencies appearing in the lower frequency area of the spectrum [13].

It was also shown that this modified method possessed useful properties. First, the behavior of the periodogram for nonuniformly sampled data is also that of uniformly sampled data [13]. That is, it does not matter whether or not the signal was sampled at regular time intervals. Also, the periodogram is equivalent to a least squares fitting of sinusoids to the original data [13]. To illustrate the Lomb-Scargle periodogram, we provide an example.

Example 1.1. *Consider the signal of period $\frac{1}{2}$ modeled by $f(t) = \cos(4\pi(t-1))$. We take a sample of one-hundred values of this signal as displayed in Figure 2. We then apply the definition of the Lomb-Scargle periodogram and plot the power as it changes with the frequency using the matplotlib, numpy, and scipy.signal libraries in Python. The periodogram for the example signal is displayed in Figure 3.*

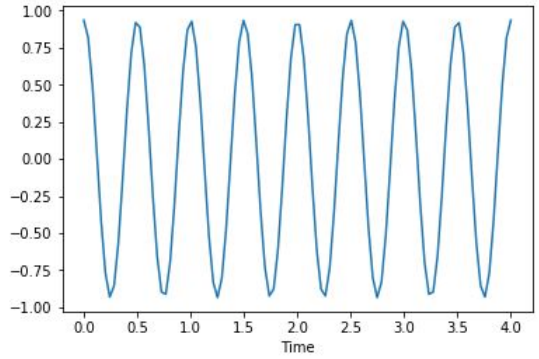

Figure 2: A Sampling of a Continuous Signal Given by $a\cos(4\pi(t-1))$

Notice that there is a peak in the power at the frequency value of 2*. From this periodogram then, we can estimate the period (calculated by taking the reciprocal of the frequency) to be* $\frac{1}{2}$ *in agreement with our predefined signal period. Also notice that the periodogram has peaks of lesser magnitude at other frequencies. Such phenomena are Type I errors called false alarms and can lead to the assumption that these frequencies are constituent frequencies of the original signal. Because false alarms are present with a single sinusoidal signal, it is believed that false alarms are also present with more complex signals.*

The MultiBand form of the Lomb-Scargle periodogram was detailed by Jacob VanderPlas and Željko Ivezic in an effort to estimate the period of a family of nonuniformly sampled time series data [17]. Here, a regularization is employed by first defining a diagonal parameter matrix Λ with M diagonal entries, then constructing a chi-square estimation by

$$\chi^2_\Lambda(\omega) = (y - X_\omega\theta)^T \Sigma^{-1} (y - X_\omega\theta) + \theta^T \Lambda\theta, \tag{16}$$

where θ is defined to be the magnitude of the parameters, Σ is the covariance matrix

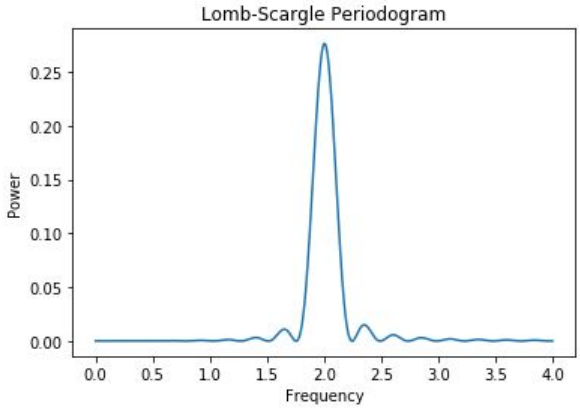

Figure 3: The Lomb-Scargle Periodogram for the Example Signal

of the noise, and X_ω is the design matrix [17] defined for a signal of length N by

$$X_\omega = \begin{bmatrix} \sin\omega t_1 & \cos\omega t_1 \\ \sin\omega t_2 & \cos\omega t_2 \\ \sin\omega t_3 & \cos\omega t_3 \\ \vdots & \vdots \\ \sin\omega t_N & \cos\omega t_N \end{bmatrix}. \tag{17}$$

The authors then define a floating-mean periodogram for a single band k by

$$P_N^k(\omega) = 1 - \frac{\chi^2_{min,k}(\omega)}{\chi^2_{(0,k)}}. \tag{18}$$

This definition is used to define the multiband form of the Lomb-Scargle. For a set of data containing B bands, the Lomb-Scargle estimator given in [17] is

$$P_N^{(0,1)} = \frac{\sum_{n=1}^{B} \chi^2_{min,k}(\omega)}{\sum_{n=1}^{B} \chi^2_{(0,k)}}. \tag{19}$$

1.4 Consensus Spectral Modeling

Alternatively, spectral consensus models form a consensus on the spectrum of a family of data by applying the convolution theorem in some manner. This implies

that products of Fourier Transforms of the data are used. Such methods of modeling are common in biology to identify commonalities in so-called hot-spots in nucleic acids, protein interactions, and electroencephalogram data [19]. An example of a consensus spectrum is found in Figure 4. Here the horizontal axis represents the

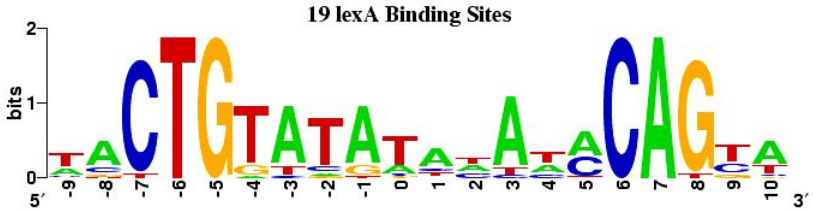

Figure 4: A Consensus Model of Transcription Binding Sites

different DNA transcription factor binding locations in the chromosomes of *E. coli*, the letters represent the four nucleotides found in DNA, and vertical axis represents the power of the consensus given in bits [11]. The power of each binding site together with the nucleotide is determined using a method known as DNA footprinting [11].

There are many benefits to employing spectral consensus modeling. First, the model estimates a common signal relatively independent of any added noise to the signal. Second, statistical methods such as bootstrapping and resampling of the data can be used to generate confidence intervals on the common spectrum. Finally, if the Fourier transform is to be employed in forming a consensus, the algorithm for forming a consensus is less computationally expensive than various other methods by using the Fast Fourier Transform algorithm.

It is not without flaws, however. Phase representation of the signal consensus is not stable and is often not reliable. That is, it is difficult and sometimes impossible, to

recover the phase of a signal after finding a consensus on the spectrum. As such, it is primarily used where the frequencies are the primary focus such as mass spectroscopy and applications in biology such as proteomics [19]. As there is nothing currently in the literature concerning period estimation for a family of nonuniformly sampled time series using consensus spectral models, it is the goal of this thesis to develop a consensus spectral model alternative to the Multiband Lomb-Scargle method. A primary issue that must be addressed is that the consensus spectrum is not a direct transform of any given signal but is, potentially, a minimum-norm least squares estimation.

2 BACKGROUND INFORMATION

2.1 Signal Spaces

A vector space is a nonempty set of elements called vectors together with a scalar field and binary operations of vector addition and multiplication of a vector by an element in the scalar field that satisfies the group criteria under vector addition [18]. That is, vector addition is associative; there exists an identity element so that any vector added to this element is the vector itself, and for every vector in the set there is an inverse element. It is also required that vector addition is commutative and multiplication of a vector by a scalar satisfies the distributive property [18]. An example of a vector space is the set $\mathbb{R}^n$ together with the scalar field of real numbers $\mathbb{R}$.

A couple of vector spaces that are of use in signal processing are the Hilbert spaces $\ell^p(\mathbb{Z})$ for $p = 1$ and $p = 2$. Here, $\ell^1(\mathbb{Z})$ is defined to be

$$\Big\{\langle \ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots \rangle | \sum_{i=-\infty}^{\infty} |x_i| < \infty \Big\}. \tag{20}$$

Similarly, the Hilbert space $\ell^2(\mathbb{Z})$ is defined to be

$$\Big\{\langle \ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots \rangle | \sum_{i=-\infty}^{\infty} |x_i|^2 < \infty \Big\}. \tag{21}$$

Of particular interest are the norms associated with these Hilbert spaces. These norms satisfy several properties. First, the norms satisfy the Triangle Inequality [12]. That is, for $x, y \in \ell^p(\mathbb{Z})$, we have

$$\|x - y\| \leq \|x\| - \|y\|. \tag{22}$$

The $\ell^p(\mathbb{Z})$ norms also exhibit the nonnegativity property which means that the norm of any element in this space is at least 0 where a norm of 0 is exclusive to the zero vector [12]. In addition, the norms defined on $\ell^p(\mathbb{Z})$ have positive homogeneity. From [12], this means that for any $a \in \mathbb{R}$ we have

$$\|ax\| = |a|\|x\|. \tag{23}$$

That is, we can factor out the scalar a from the norm with the condition that we take the absolute value. The signals for which we estimate the period are recorded over a finite time interval via a finite sampling rate. This means that the time series are elements of a Hilbert space, in particular $\ell^2(\mathbb{Z})$.

2.2 Fourier Transform

When analyzing a signal (i.e. a time series) it may be necessary or insightful to instead consider its frequency content. To do this, we employ the following definition.

Definition 2.1. *Let $x(t)$ be a time series of length N. Then the Fourier transform $\mathcal{F}(x)$ is defined to be*

$$\mathcal{F}(x) = \frac{1}{N}\sum_{k=0}^{N-1} x(k)e^{\frac{2\pi ikn}{N}}.$$

Note that this transform is a mapping of a time series or signal of length n from n-dimensional real space ($\mathbb{R}^n$) to its spectral domain. The Fourier transform has several properties that are useful when analyzing a signal. From [12], if $f(t)$ and $g(t)$ are both signals where $g(t) = \overline{f(-t)}$, then

$$\mathcal{F}(g(t)) = \overline{\mathcal{F}(f(t))}.$$

In particular, if $g(t) = f(-t)$, then the two signals have the same Fourier transform (i.e. the two signals are the same in the spectral domain). When developing our approach to period estimation, if two signals from the same phenomenon are measured in such a way that they are out of phase with each other (i.e. delay due to instrument readings etc, $g(t) = f(-t)$) then it is expected that the two have the same spectra. By this first property, we know that f and g indeed have the same spectra.

Another property of the Fourier transform that will be useful to our algorithm development in particular is what is called the convolution theorem. That is if $f(t)$ and $g(t)$ are signals for which $\mathcal{F}(f)$ and $\mathcal{F}(g)$ exist, then

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g) \tag{24}$$

where $*$ represents the convolution operator [12]. Here, the product $\mathcal{F}(f)\mathcal{F}(g)$ is performed component-wise. So, for $\mathcal{F}(f) = [\hat{x_1}, \hat{x_2}, \ldots, \hat{x_n}]$ and $\mathcal{F}(g) = [\hat{y_1}, \hat{y_2}, \ldots, \hat{y_n}]$, we have

$$\mathcal{F}(f)\mathcal{F}(g) = [\hat{x_1}\hat{y_1}, \hat{x_2}\hat{y_2}, \ldots, \hat{x_n}\hat{y_n}]. \tag{25}$$

Notice in Equation 25 that if either $\hat{x_k}$ or $\hat{y_k}$ are close to 0, the product $\hat{x_k}\hat{y_k}$ is also close to zero. We use this property in our algorithm development to form a consensus on the spectra of constituent signals of a family of time series as the products of components via the convolution theorem reinforce commonalities between spectra and diminish dissimilar components. To illustrate how the Fourier transform works, we give an example.

Example 2.2. *Consider the continuous signal of a sinusoidal wave given by*

$$a\cos(4\pi(t-1))$$

for some amplitude a, as seen in Figure 5.

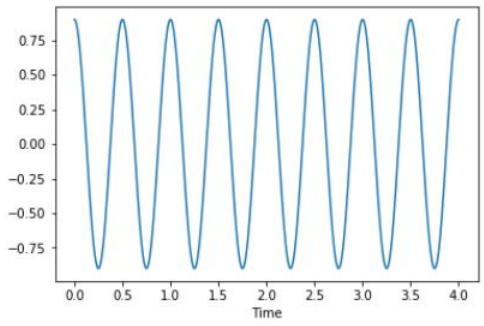

Figure 5: A Continuous Signal Given by $a\cos(4\pi(t-1))$

We create a time series by taking samples of the signal at uniform intervals as seen in Figure 6. Applying the definition of the Fourier Transform yields the spectrum

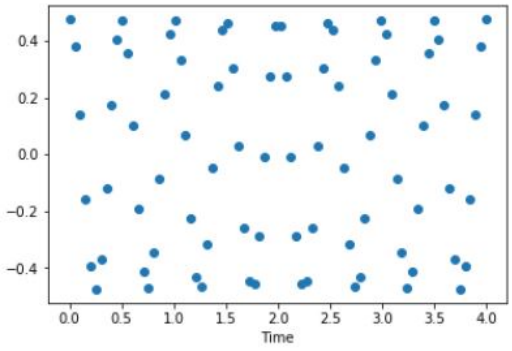

Figure 6: A Sample of 80 Values of the Continuous Signal

displayed in Figure 7.

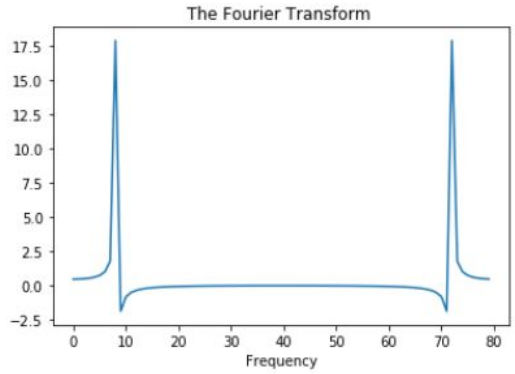

Figure 7: The Spectral Decomposition of the Time Series

Observe that the plot of the spectrum has two distinct peaks namely at $0 + T$ and $n - T$ for frequency T. Because the Fourier Transform exhibits such symmetry, we will only consider the first half of the spectrum of any time series. Also, from

the definition, the Fourier transform cannot be performed on nonuniformly sampled data as the sampling times of the signal may not coincide with the regularly-spaced intervals of the transform. This discrepancy is handled in Section 2 of Chapter 3.

2.3 Convolution

Convolution is a binary operation that takes two signals and produces a third signal related to the product of the two original signals. It can be used for feature extraction in images and is the basis of several kernel-based methods in the study of neural networks [14]. In [14], Simard *et al.* used convolution-based methods to filter images and increase or decrease color saturation for a grayscale image. For this thesis we use the definition of convolution given in [18] as detailed in Definition 2.3.

Definition 2.3. *For time series $x(t)$ and $h(t)$ of length n, the convolution of the two time series, denoted $x(t) * h(t)$, is defined by*

$$x(t) * h(t) = \sum_{k=1}^{n} x(k)h(t-k). \tag{26}$$

Notice that convolution is a product of the elements of $x(t)$ with a delayed, or time-shifted, copy of the signal $h(t)$. In signal processing, this definition of convolution can be used as a method of filtering a signal [14]. For example, if h has only a finite number of nonzero components and $x(\omega)$ is the spectrum of a time series, the convolution $x * h$ eliminates frequencies in the spectrum by multiplying by 0. This kind of filter is called a Finite Impulse Response (FIR) filter [18].

The convolution operation has several properties that are used in this thesis. First,

convolution, as given in [18], is associative, meaning for signals g, h, and k we have

$$g * (h * k) = (g * h) * k.$$

The second property of convolution used is commutativity [18]. This means that for all x and h, we have

$$x * h = h * x.$$

The final property is that of the Convolution Theorem detailed in Section 2.2. To illustrate the behavior of convolution in regards to periodic time series, we provide an example.

Example 2.4. *Suppose $x(t)$ is uniformly sampled from a continuous signal modeled by $x(t) = b\cos(2\pi(t-1))$, for a random number b as displayed in Figure 8.*

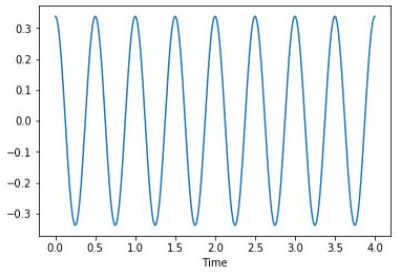

Figure 8: A Uniformly Sampled Signal Modeled by $x(t) = b\cos(2\pi(t-1))$

Similarly, suppose we have another signal $h(t)$ modeled by $h(t) = c\sin(4\pi(t-1))$ where c is a random number as displayed in Figure 9.

We use the definition of convolution given in Definition 2.3 to produce a new signal $(x * h)(t)$ as given in Figure 10.

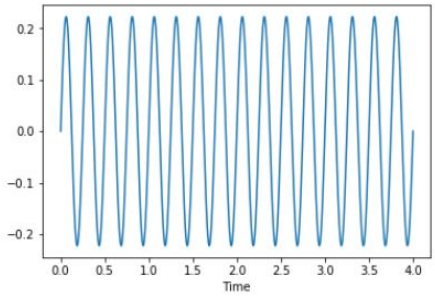

Figure 9: A Uniformly Sampled Signal Modeled by $h(t) = b\sin(4\pi(t-1))$

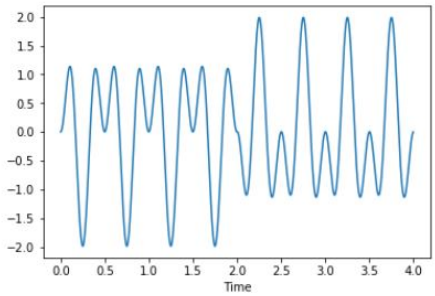

Figure 10: The Convolution of Signals Given in Figures 8 and 9 Respectively

2.4 Singular Value Decomposition and Moore-Penrose Pseudoinverse

When solving systems of equations, it becomes necessary to compute the inverse of a matrix. For example, the system of equations given by a coefficient matrix A in the equation $A\mathbf{x} = b$ can be solved by computing A^{-1} so that

$$\vec{x} = A^{-1}b. \tag{27}$$

However, if the matrix A is not invertible, whether underdetermined or overdetermined, the equation $A\mathbf{x} = b$ has no unique solutions. To demonstrate this, we provide an example.

Example 2.5. *Suppose that we are given a set of data in the form* (a_i, b_i) *and we want to calculate the line of best fit of these points, otherwise known as the regression line. To achieve this, we note that a general linear equation has two parameters: slope and y-intercept. Then, we form the system of linear equations* $A\boldsymbol{x} = b$ *where*

$$A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix} \tag{28}$$

and the associated column vector

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}. \tag{29}$$

If there are only two points to consider (say (a_1, b_1) *and* (a_2, b_2)*), then we have the system of equations given by*

$$\begin{bmatrix} 1 & a_1 \\ 1 & a_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{30}$$

The equation in (30) can be solved via augmenting the matrix A with the column vector b and using Gauss-Jordan elimination to produce the the reduced row-echelon form (RREF). From the RREF, we have that

$$x_1 = b_1 - a_2 \frac{b_2 - b_1}{a_2 - a_1},\ x_2 = \frac{b_2 - b_1}{a_2 - a_1}. \tag{31}$$

Notice that $b_2 - b_1$ *and* $a_2 - a_1$ *are the changes in the y and x coordinates respectively. This means that* x_2 *is the slope of the regression line and* x_1 *is the y intercept.*

The method of finding the coefficients outlined in the example does not work with three or more data points, however, as the matrix defined in Equation 28 will not have

an inverse. So, a different method must be used to solve linear systems of equations where the coefficient matrix is not invertible. To do this, we consider square roots of the eigenvalues of the covariance matrix A^TA called singular values [7]. If the matrix A is not a square matrix, we cannot find eigenvalues, but the covariance matrix is by definition a square matrix. With the definition of singular values, we discuss a method of finding these values for a matrix A. This is accomplished with Theorem 2.6 called the Singular Value Decomposition (SVD).

Theorem 2.6 (The Singular Value Decomposition [7])**.** *Let A be an $m \times n$ matrix. Then, there exists an $m \times m$ matrix U, an $n \times n$ matrix V, and a matrix Σ containing the singular values of A along the main diagonal, and 0 elsewhere such that*

$$A = U\Sigma V^T.$$

In Theorem 2.6, the columns of the U matrix are called the left singular vectors while the rows of V^T are called the right singular vectors [9]. The matrices U and V are also unitary matrices meaning that $U^TU = UU^T = I$ and $VV^T = V^TV = I$ [9]. This means that

$$A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T. \tag{32}$$

If a matrix A is full rank, the inverse can be calculated by computing the inverse of U, Σ, and V^T . Because U and V^T are unitary matrices, the transpose of the matrix is its inverse. Also, because Σ is a diagonal matrix, Σ^{-1} is computed by taking the reciprocal of each diagonal entry. The SVD is not limited only to finding coefficients of regression curves, as it has been used in applications such as latent semantic analysis, principal component analysis, and calculation of the Moore-Penrose pseudoinverse,

among others [8].

For noninvertible matrices, we have an alternative to the inverse called the Moore-Penrose pseudoinverse, and can be calculated from the SVD. Because the matrix is not invertible, it will have zero as an eigenvalue with multiplicity of at least one. This means that the matrix Σ is of the form

$$\Sigma = \left[\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array}\right], \tag{33}$$

where Σ_r is a diagonal matrix with size equal to the rank of the original matrix A [9]. With (33), the Moore-Penrose pseudoinverse is defined in [9] by

$$A^\dagger = V \left[\begin{array}{c|c} \Sigma_r^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] U^T. \tag{34}$$

In algorithm development, we use the Moore-Penrose pseudoinverse as it is a means of computing the inverse of a matrix if it is not full rank. In Theorems 3.3, 3.5, and 3.6, we have that all matrices are invertible. Also, calculation of $A^\dagger$ is computationally fast due to the algorithm developed by Courrieu in [1]. This algorithm also has an error in computation of less than 2×10^{-10} per coefficient [1].

2.5 Method of Least Squares

Similar to Example 2.5, suppose that we have a set of data in the form (x_i, y_i) as displayed in Figure 11 and we want to find a line that models the data. Another method of determining the coefficients for this line is the method of least squares. To do this, we assume the coefficients of the line of best fit, given by $a+mx$, are modeled by a column vector $\mathbf{x}$ such that

$$\mathbf{x} = \begin{bmatrix} a \\ m \end{bmatrix}. \tag{35}$$

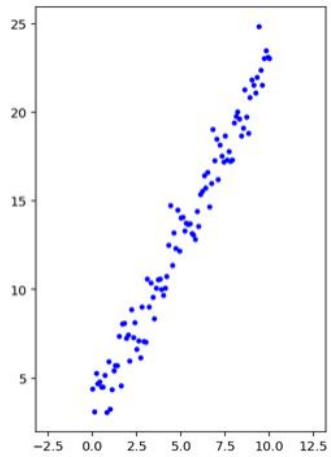

Figure 11: A Scatter Plot of Data Points of the Form (x_i, y_i)

The method of least squares considers the norm of the residuals, as described in [5] given by

$$\|A\mathbf{x} - \mathbf{b}\|^2, \tag{36}$$

where $\mathbf{b}$ is the column vector formed by the y coordinates of the data points and A is the matrix defined in (28) [9]. In order to compute the coefficients of the line, we must find a and m from (35) such that (36) is minimized, denoted a_0 and m_0 respectively. Notice that because the norm in (36) can be written as

$$\langle A\mathbf{x} - \mathbf{b}, A\mathbf{x} - \mathbf{b} \rangle, \tag{37}$$

we instead find the minimum of

$$(A\mathbf{x} - \mathbf{b})^T (A\mathbf{x} - \mathbf{b}). \tag{38}$$

It is shown in [9], that the solution to this minimization problem is given by

$$\mathbf{x} = A^{\dagger}\mathbf{b}, \tag{39}$$

where $A^{\dagger}$ is the Moore-Penrose pseudoinverse. Once these parameters are found, the equation of the line of best fit is given by $y = a_0 + m_0 x$ exemplified by Figure 12 indicated by the red line. In Figure 12 it is assumed that all data points have equal

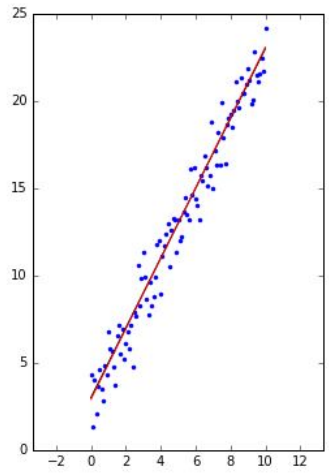

Figure 12: Data With the Line of Best Fit

importance or weight.

If the data exhibits high variation, it may be necessary to apply a weighting scheme in order to compute the coefficients of the line that better fits the data [4]. This is accomplished by assigning an importance, also called a weight, to the data points. So, possible outliers in the data receive a lower weight than the rest of the data. This results in the modification of (36) given by

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|W(A\mathbf{x} - \mathbf{b})\|^2, \tag{40}$$

where W is the matrix representation of the desired weighting scheme [6]. We may also impose a condition on the norm of the vector $\mathbf{x}$ called a regularization [4]. This

further modifies (36) to produce the regularized least squares method given by

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|W(A\mathbf{x}-\mathbf{b})\|^2 + \epsilon\|\mathbf{x}\|^2, \tag{41}$$

where ϵ is the weight assigned to the norm of $\mathbf{x}$ [4]. The method of regularized least squares is used in Chapter 3 to form a consensus on the spectrum of a family of nonuniformly sampled time series as we want a consensus spectrum that has a minimum norm.

3 RESULTS

3.1 Signal Testing and Benchmarking

In order to develop the algorithms, a number of tests were developed. As such, a signal testing laboratory was created using Python. The laboratory consisted of five baseline families of signals and fourteen other consensus test families that were deemed to adequately describe the majority of naturally-occurring phenomena. In both sets of tests, each test family consisted of thirty-two individual signals and frequencies used in all tests were chosen arbitrarily. The baseline tests were used as a proof of principle for each method and contained time series with no added noise. The first baseline test consisted of 32 identical signals of 7Hz. The second family of test signals decayed the signals from the first family exponentially. The remaining test families consisted of chirps at 16 Hz and a 3Hz chirp followed by a 7Hz chirp respectively.

Once a method had satisfactorily estimated the period for these tests, the second round of tests was administered. The first sets of these tests were standing waves consisting of a 3 Hz (or 7 Hz) carrier wave with added noise of magnitude 0.01 (and 0.1 respectively). Next, signals were fixed at either 1 Hz or 3Hz and a signal with a randomly generated frequency greater than 1 Hz, or 3 Hz respectively, was then superimposed upon the fixed carrier. The seventh family of test data was comprised of two signals that were allowed to vary in time and frequency. The latter half of the test families consisted of signals that had been amplified in some fashion. The first of these were amplified 3 Hz fixed signals with a signal of random frequency greater

than 3 Hz superimposed. Next, 16 Hz signals were amplified and transient signals of a random frequency between 5 Hz and 15 Hz was added. Following this test was the family of amplified 8 Hz moving signals with transient random signals added with frequency between 5Hz and 30 Hz. Finally, we considered the family of amplified 8 Hz moving signals with an added frequency-based amplified transient random signal of frequency between 5 Hz and 30 Hz and added pure 3 Hz signal. Upon sufficiently denoising and estimating the period for each of these fourteen test cases, a final test was administered consisting of two families of electroencephalographic data.

3.2 The Regularized Least Squares Approach

The first method considered is applying the least squares method to form a spectral consensus model. For this approach and the convolution approach discussed in the next section, it is necessary to make two assumptions. First, we must assume that the data is sampled at such a rate that either one value is recorded or no value is recorded during any given time partition. If instead of assuming that the data was sampled at such a rate, it is assumed that sampling times follow a random distribution it may be that two recordings are arbitrarily close together along the time axis as is possible when considering a probability distribution. As such, we only consider discrete sampling of signals as assumed. It must also be assumed that the data exhibits periodic behavior. Otherwise, the spectrum of each signal may not have a distinct set of peaks.

Because the data is sampled nonuniformly, each constituent signal in the family of signals contains missing values. In this case, we reorganize the signal as given in

the next definition.

Definition 3.1. *Let X be a nonuniformly sampled time series of length n with $n-m$ missing values where $n > m$, and σ be a permutation such that*

$$X = [x_{\sigma(1)}, x_{\sigma(2)}, \ldots x_{\sigma(m)}, x_{\sigma(m+1)}, \ldots, x_{\sigma(n-1)}, x_{\sigma(n)}]$$

and

$$x_{\sigma(m+1)}, x_{\sigma(m+2)}, \ldots, x_{\sigma(n-1)}, x_{\sigma(n)}$$

are missing values. That is to say, the permutation σ maps the missing values in the time series to the end. Then, the downsampling operator $P : \mathbb{R}^n \mapsto \mathbb{R}^m$ is defined to be

$$PX = [x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(m-1)}, x_{\sigma(m)}].$$

Before continuing to develop the approach, we prove a necessary result concerning the downsampling operator defined in Definition 3.1. Because the result of Theorem 3.3 contains a product of P with its Hermitian transpose, we first show what this product looks like.

Proposition 3.2. *Let P and X be as in definition 3.1. Then $PP^* = I_m$, and*

$$P^*P = \left[\begin{array}{c|c} I_m & 0 \\ \hline 0 & 0 \end{array}\right].$$

Proof. Let P and X be as in the definition. Then P has a matrix representation given by

$$P = \left[\begin{array}{c|c} I_m & 0 \end{array}\right], \tag{42}$$

where there are $n-m$ columns of zeros. Note that P can be represented by an $m \times n$ matrix. By the definition of Hermitian transpose, we have

$$P^* = \left[\begin{array}{c} I_m \\ \hline 0 \end{array}\right] \tag{43}$$

where there are $n-m$ rows of zeros. Note P^* is an $n \times m$ matrix. It follows that

$$PP^* = \left[\begin{array}{c|c} I_m & 0 \end{array}\right] \left[\begin{array}{c} I_m \\ \hline 0 \end{array}\right] = I_m. \tag{44}$$

Similarly,

$$P^*P = \left[\begin{array}{c} I_m \\ \hline 0 \end{array}\right] \left[\begin{array}{c|c} I_m & 0 \end{array}\right] = \left[\begin{array}{c|c} I_m & 0 \\ \hline 0 & 0 \end{array}\right]. \tag{45}$$

□

As stated in Definition 3.1, PX_j creates a new time series with the sample times that are not missing data in X_j. Suppose that the permutation σ_j acting on a time series X_j is of the form

$$X_j = [x_{\sigma(1)}, x_{\sigma(2)}, \ldots x_{\sigma(m)}, x_{\sigma(m+1)}, \ldots, x_{\sigma(n-1)}, x_{\sigma(n)}],$$

where $x_{\sigma(m+1)}, x_{\sigma(m+2)}, \ldots, x_{\sigma(n-1)}, x_{\sigma(n)}$ are missing values. If one replaces the missing data in X_j with zeros to form a new time series (say X_j'), so that

$$X_j' = [x_{\sigma(1)}, x_{\sigma(2)}, \ldots x_{\sigma(m)}, 0, 0, \ldots, 0, 0]$$

and if P_j is the downsampling operator for X_j, then

$$P_j X_j' = [x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(m-1)}, x_{\sigma(m)}] = P_j X_j. \tag{46}$$

We know from equation (46) that any sampled time in a time series that is missing a value can have 0 used as a placeholder value provided that the same P_j and σ_j are

used for both X_j and X_j'. This effectively converts a nonuniformly sampled time series to a uniformly sampled one provided the assumption is made that the sample times for X_j are not randomly distributed. As we have the necessary properties for the products of the downsampling operator P, we can now develop the means of finding the consensus spectral model for a family of nonuniformly sampled time series.

Theorem 3.3. *If $\{x_0, x_1, \ldots, x_m\}$ is a family of nonuniformly sampled time series of length n with $n-r$ missing values, where F is the Fourier Transform operator and P is the downsampling operator defined above, then the solution to the regularized least squares equation*

$$\sum_{j=1}^{m} \|P_j F^* y - x_j\|^2 + \epsilon \|y\|^2$$

is given by the expression

$$y = F\left(\sum_{j=1}^{n} P_j^* P_j + \epsilon I\right)^{-1} \sum_{j=1}^{n} {P_j}^* x_j.$$

Proof. Let P_j be the downsampling operator for time series x_j, F be the Fourier Transform operator, and y be the spectral consensus model for the family of time series. Also, let $\{x_1, x_2, \ldots, x_m\}$ be a family of nonuniformly sampled time series taken from the same phenomenon. Consider the regularized least squares expression given by

$$\sum_{j=1}^{m} \|P_j F^* y - x_j\|^2 + \epsilon \|y\|^2. \tag{47}$$

We then consider a perturbation of y. Let $\Delta \in \mathbb{R}^n$. Note that $y + \Delta$ is not the consensus spectral model with the minimum norm (from the definitions of perturbation and norm respectively). Then, by the definition of the norm we have

$$\sum_{j=1}^{m} (P_j F^* y - x_j + P_j F^* \Delta)^* (P_j F^* y - x_j + P_j F^* \Delta) + (y + \Delta)^* (y + \Delta). \tag{48}$$

Note that F^* is the Hermitian transpose of the Fourier Transform operator, etc. Performing the multiplication in equation (48), yields

$$\sum_{j=1}^{m}(y^*FP_j{}^* - x_j{}^* + \Delta^*FP_j{}^*)(P_jF^*y - x_j + P_jF^*\Delta) + y^*y + y^*\Delta + \Delta^*y + \Delta^*\Delta$$

$$= \sum_{j=1}^{m}(y^*FP_j{}^* - x_j{}^*)(P_jF^*y - x_j) + (y^*FP_j{}^* - x_j{}^*)P_jF^*\Delta + \Delta^*FP_j{}^*(P_jF^*y - x_j)$$

$$+y^*y + y^*\Delta + \Delta^*y + \Delta^*\Delta + O(\Delta^2). \tag{49}$$

The definition of the norm implies that the right side of equation (49) reduces to the expression

$$\sum_{j=1}^{m} \|P_jFy - x_j\|^2 + \epsilon\|y\|^2 + (y^*FP_j{}^* - x_j{}^*)P_jF^*\Delta + \Delta^*FP_j{}^*(P_jF^*y - x_j)$$

$$+y^*\Delta + \Delta^*y. \tag{50}$$

From the definition of an argmin, we have

$$\sum_{j=1}^{m} C + 2\,\mathrm{Re}((y^*FP_j{}^* - x_j^*)P_jF^*\Delta) + 2\epsilon\,\mathrm{Re}(y^*\Delta) + O(\Delta^2) \tag{51}$$

where $C = \|P_jF^*y - x_j\|^2 + \epsilon\|y\|^2$. This can be condensed to produce

$$\sum_{j=1}^{m} C + 2\,\mathrm{Re}(\Delta^*(FP_j{}^*(P_jF^*y - x_j) + \epsilon y) \tag{52}$$

Allowing Δ to approach 0 yields the equation

$$\sum_{j=1}^{m}(FP_j{}^*(P_jF^*y - x_j) + \epsilon y) = 0. \tag{53}$$

Distributing the $FP_j{}^*$ in Equation (53) and rearranging, we have

$$\sum_{j=1}^{m}(FP_j{}^*P_jF^* + \epsilon I)y = FP_j{}^*x_j. \tag{54}$$

We can factor out the Fourier and inverse Fourier transform respectively. This produces the equation

$$F\left(\sum_{j=1}^{m} {P_j}^* P_j + \epsilon I\right) F^* y = \sum_{j=1}^{m} F {P_j}^* x_j. \tag{55}$$

We multiply both sides of the equation on the left by the inverse Fourier Transform (i.e. F^*) This results in the following equation

$$\left(\sum_{j=1}^{m} {P_j}^* P_j + \epsilon I\right) F^* y = \sum_{j=1}^{m} {P_j}^* x_j. \tag{56}$$

From Proposition 3.2, we have that the matrix formed by the summation

$$\sum_{j=1}^{m} {P_j}^* P_j = \left[\begin{array}{c|c} C & 0 \\ \hline 0 & 0 \end{array}\right]. \tag{57}$$

Since $\epsilon > 0$ is strictly positive and the product ${P_j}^* P_j$ is of the form given in Proposition 3.2, all diagonal entries of the diagonal matrix are strictly positive. This means that the matrix

$$\sum_{j=1}^{m} {P_j}^* P_j + \epsilon I \tag{58}$$

is positive definite. Because the matrix in Equation (58) is positive definite, the inverse exists. Then, we solve for y producing

$$y = F\left(\sum_{j=1}^{m} {P_j}^* P_j + \epsilon I\right)^{-1} \sum_{j=1}^{m} {P_j}^* x_j. \tag{59}$$

□

Notice that this method relies heavily on the definition of the downsampling operator P. It is this operator that creates a new time series from signals with missing data. This allows us to perform the Fourier transform in order to analyze and form a consensus on the spectra of the family of signals. We can modify the consensus spectral model if we consider different families of signals.

Definition 3.4. *If $\{x_1, x_2, \ldots, x_m\}$ is a family of time series over a time interval $[a, b]$ with partition $\{t_0, t_1, \ldots, t_n\}$ such that $t_0 < t_1 < \cdots < t_n$, and for all $j \in \{1, 2, \ldots, m\}$ and $k \in \{1, 2, \ldots, n\}$, $x_j(t_k)$ is not a missing value, then the family is said to cover the time interval.*

So a family of time series covers a time interval if all of the sample times in the interval have at least one recorded value from a constituent signal. If a family of signals exhibits this behavior, the consensus spectral model in Equation (59) can be modified.

Corollary 3.5. *If a family of nonuniformly sampled time series $\{x_1, x_2, \ldots, x_m\}$ covers a sampling time $[a, b]$, then the consensus spectral model in Theorem 3.3 is given by*

$$y = F\left(\sum_{j=1}^{m} P_j{}^* P_j\right)^{-1} \sum_{j=1}^{m} P_j{}^* x_j. \tag{60}$$

Proof. As in Theorem 3.3, let P_j be the downsampling operator for time series x_j, F be the Fourier Transform operator, and y be the spectral consensus model for the family of time series. Also, let $\{x_1, x_2, \ldots, x_m\}$ be a family of nonuniformly sampled time series taken from the same phenomenon. Consider the regularized least squares expression given by

$$\sum_{j=1}^{m} \|P_j F^* y - x_j\|^2 + \epsilon \|y\|^2. \tag{61}$$

From Theorem 3.3 we have the consensus spectral model is

$$y = F\left(\sum_{j=1}^{m} P_j{}^* P_j + \epsilon I\right)^{-1} \sum_{j=1}^{m} P_j{}^* x_j. \tag{62}$$

From Proposition 3.2, we have that the matrix formed by the summation

$$\sum_{j=1}^{m} P_j{}^* P_j = \left[\begin{array}{c|c} C & 0 \\ \hline 0 & 0 \end{array}\right], \tag{63}$$

where C is an invertible matrix of the form

$$\begin{bmatrix} c_{t_1} & 0 & 0 & \dots & 0 \\ 0 & c_{t_2} & 0 & \dots & 0 \\ 0 & 0 & c_{t_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & c_{t_m} \end{bmatrix},$$

and each c_{t_k} is the count of nonuniformly sampled signals with a recorded value at time t_k. From our assumption that the family of time series covers the interval, however, all sampling times have at least one signal for which a value is recorded. This implies that the right-hand side Equation (63) is

$$\begin{bmatrix} c_{t_1} & 0 & 0 & \dots & 0 \\ 0 & c_{t_2} & 0 & \dots & 0 \\ 0 & 0 & c_{t_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & c_{t_n} \end{bmatrix}. \tag{64}$$

Since $\epsilon > 0$ is strictly positive, all diagonal entries of the diagonal matrix are strictly positive. Because the family of signals covers the time interval, we can allow ϵ to approach 0 while maintaining the invertibility of the matrix on the left-hand side in equation (63). Allowing ϵ to approach 0 in this fashion gives us the equation

$$y = F\left(\sum_{j=1}^{m} {P_j}^* P_j\right)^{-1} \sum_{j=1}^{m} {P_j}^* x_j. \tag{65}$$

□

If the family of time series does not cover the sampling interval, the matrix in Equation (63) is positive semi-definite and therefore may not be invertible. That is to say, if there is a 0 in the diagonal of the matrix, the determinant is therefore 0 and the matrix is not invertible. However, the addition of the term ϵI, for any $\epsilon > 0$, makes the matrix positive definite and invertible. This is because the diagonal entries that previously were 0, now have an ϵ instead.

3.3 The Convolution Approach

The approach involving the convolution theorem follows from a corollary of the least squares consensus model from Theorem 3.3. If instead of individual signals from a family, we consider pair-wise convolutions of signals, the result is Corollary 3.6.

Corollary 3.6. *If a family of signals* $\{x_1, x_2, \ldots, x_m\}$ *covers a time interval [a,b] and* $\{(P_i^* x_i * P_k^* x_k)\}$ *is the family of all pair-wise convolutions of* $x_i, x_k \in \{x_1, x_2, \ldots, x_n\}$ *with 0 substituted for missing values, then the spectral consensus model y is given by*

$$y = F\left(\sum_{j=1}^{m} {P_j}^* P_j\right)^{-1} \sum_{i,k=1}^{m} (P_i^* x_i * P_k^* x_k)$$

Proof. Let S be a family of nonuniformly sampled times series that covers a time interval [a,b]. Because S covers [a,b], the family defined by $\{(P_i^* x_i * P_k^* x_k) | x_i, x_k \in S\}$ also covers the interval [a,b]. From Equation (65) together with

$$\sum_{i,k=1}^{m} P_i^* x_i * P_k^* x_k \tag{66}$$

gives us that the consensus spectral model is given by

$$y = F\left(\sum_{j=1}^{m} {P_j}^* P_j\right)^{-1} \sum_{i,k=1}^{m} (P_i^* x_i * P_k^* x_k). \tag{67}$$

□

Note that if the Fourier transform operator commutes with the inverse matrix in Equation (67), then

$$F\left(\sum_{j=1}^{m} {P_j}^* P_j\right)^{-1} \sum_{i,k=1}^{m} (P_i^* x_i * P_k^* x_k) = \left(\sum_{j=1}^{m} {P_j}^* P_j\right)^{-1} \sum_{i,k=1}^{m} (F P_i^* x_i)(F P_k^* x_k). \tag{68}$$

The equivalence in Equation (68) can be thought of as the version of the convolution theorem for a family of nonuniformly sampled time series.

From Corollary 3.6 we can develop the computation methods necessary to determine the efficacy of this algorithhm using the Python computing language together with the *numpy, pandas, scipy,* and *matplotlib* libraries. Each test family described in Section 1 of Chapter 3 is in the form of a .dat file. So, the data is read into the program and entered into a Pandas data frame. From there, each signal is sampled via an exponential distribution found in the *numpy.random* library so that, on average, each signal is missing 20 percent of the total length in values. An example of a signal sampled this way can be found in Figure 13.

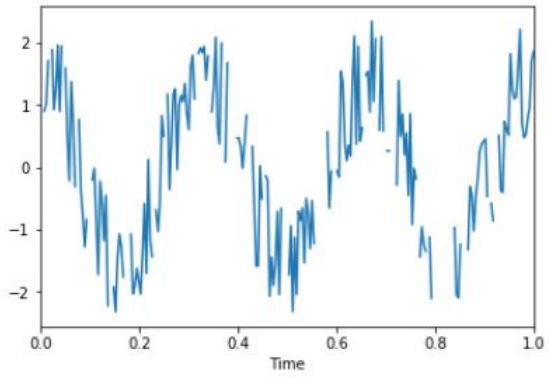

Figure 13: A Nonuniformly Sampled Signal from Consensus Test 3.

The signal exhibited in Figure 13 is from the third family of consensus test data which consisted of signals with a frequency of 3 Hz and added noise of magnitude 0.1. In particular the signal is the sixteenth in the family. Before applying the Fourier transform, we note that because the signals are sampled nonuniformly, and each individual signal contains a number of missing values. From Theorem 3.3 and Equation (46), we can replace all missing values with 0 and apply the Fourier Transform. For the sake of computation speed, we use the Fast Fourier transform function in the *scipy* library. Upon transforming the data via the Fast Fourier transform, we have a set of spectra of the constituent signals. An example of this can be seen in Figure 14.

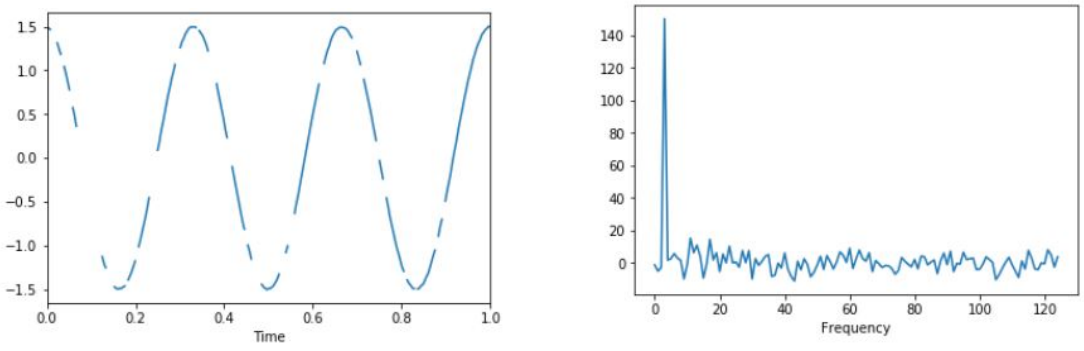

Figure 14: A Signal and its Fourier Transform After Replacing Missing Values.

Next, the Convolution Theorem is applied to the spectra to form a consensus model. For the test cases described in Section 1, it was found that applying the convolution theorem twice (i.e. convolving three spectra together sequentially) produced a consensus that denoised signals as in the Figure 15. The inverse Fourier Transform is then applied to the individual consensus models and are, for the sake of simplicity, plotted to ensure the method has successfully denoised the signal. An example of one

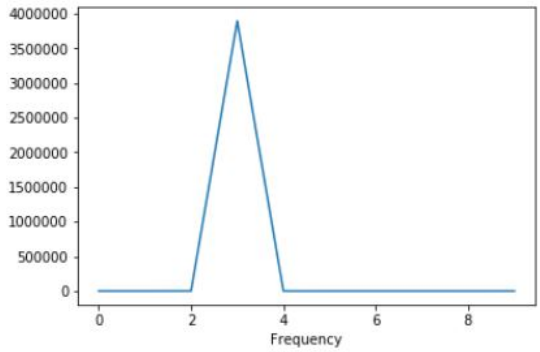

Figure 15: A Consensus Spectrum Formed Via the Convolution Theorem.

such plot is displayed in Figure 16.

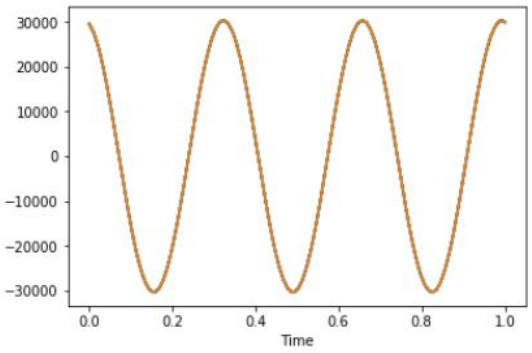

Figure 16: Inverse FFT Applied to Spectra.

4 CONCLUSIONS AND FUTURE RESEARCH

4.1 Conclusions

The Multiband Lomb-Scargle periodogram is the most recent development in period estimation for a family of nonuniformly sampled time series. To do this, the method relies upon regularization by defining a parameter matrix Λ, a matrix Σ defined to be the covariance matrix of the noise distributions for each band, and a design matrix X_ω [17]. These matrices are next used to produce the Chi-square estimation as given in Equation (19) for a family of signals. This method is prone to issues such as aliasing and unavoidable variation due to the statistical nature of the periodogram as described in [13] and [16].

It is the goal of this thesis to develop an alternative to the Lomb-Scargle periodogram using consensus spectral modeling. This is accomplished by first assuming that the time series are sampled in such a rate as either one value is recorded at any given sampling time or no value is recorded. This allows us to define a downsampling operator P that maps a signal of length n with $n - r$ missing values to a signal of length r with no missing values. The product of this operator and its Hermitian transpose (P^*P) is a key component in the development of the consensus spectral model shown in Theorem 3.3.

It is shown in Corollary 3.5 that if a family of time series covers a time interval, the consensus spectral model given in Theorem 3.3 can be modified by letting ϵ approach zero. It is also shown in Corollary 3.6 that if we consider the set of all pair-wise convolutions of nonuniformly sampled time series, instead of the family itself, the

consensus spectral model in Theorem 3.3 can be modified further.

4.2 Future Research

In Equation (68), it is a direct consequence that if the Fourier operator F commutes with the matrix given by (63), we have a form of the convolution theorem for a family of nonuniformly sampled time series. However, F may or may not have this property for a given family of signals. For future research, it may be analyzed under what circumstances the Fourier transform operator has the commutative property desired. We may also analyze time series exhibiting behaviors not of those described in Section 1 of Chapter 3 in order to better ascertain the limitations of the methods developed in this thesis.

BIBLIOGRAPHY

[1] P. Courrieu. "Fast Computation of Moore-Penrose Inverse Matrices". In: *CoRR* abs/0804.4809 (2008). arXiv: 0804.4809. URL: http://arxiv.org/abs/0804.4809.

[2] K.A. Garrett et al. "The effects of climate variability and the color of weather time series on agricultural diseases and pests, and on decisions for their management". In: *Agricultural and Forest Meteorology* 170 (2013). Agricultural prediction using climate model ensembles, pp. 216–227. ISSN: 0168-1923. DOI: https://doi.org/10.1016/j.agrformet.2012.04.018. URL: http://www.sciencedirect.com/science/article/pii/S016819231200158X.

[3] E. F. Glynn, J. Chen, and A. R. Mushegian. "Detecting Periodic Patterns in Unevenly Spaced Gene Expression Time Series Using LombfffdfffdfffdScargle Periodograms". In: *Bioinformatics* 22.3 (Nov. 2005), pp. 310–316. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bti789. eprint: http://oup.prod.sis.lan/bioinformatics/article-pdf/22/3/310/18529532/bti789.pdf. URL: https://doi.org/10.1093/bioinformatics/bti789.

[4] L. Hoegaerts et al. "Subset Based least squares subspace regression in RKHS". In: *Neurocomputing* 63 (2005). New Aspects in Neurocomputing: 11th European Symposium on Artificial Neural Networks, pp. 293–323. ISSN: 0925-2312. DOI:

https://doi.org/10.1016/j.neucom.2004.04.013. URL: http://www.sciencedirect.com/science/article/pii/S0925231204003273.

[5] H. A. L. Kiers. "Weighted least squares fitting using ordinary least squares algorithms". In: *Psychometrika* 62.2 (June 1997), pp. 251–266. ISSN: 1860-0980. DOI: 10.1007/BF02295279. URL: https://doi.org/10.1007/BF02295279.

[6] R. Kuc. *Introduction to Digital Signal Processing*. McGraw-Hill, Inc., 1988. ISBN: 0070355703.

[7] D.C. Lay. *Linear Algebra and Its Applications, 3rd Edition*. Pearson Education Inc., 2006, pp. 470–488.

[8] M. Mollenhauer et al. "Singular Value Decomposition of Operators on Reproducing Kernel Hilbert Spaces". In: *arXiv preprint arXiv:1807.09331* (2018). arXiv: 1807.09331 [math.FA].

[9] S. Orfanidis. "SVD, PCA, KLT, CCA, and All That". In: *Optimum Signal Processing* (2007), pp. 332–525.

[10] K. Rehfeld et al. "Comparison of correlation analysis Techniques for Irregularly Sampled Time Series". In: *Nonlinear Processes in Geophysics* 18.3 (2011), pp. 389–404. DOI: 10.5194/npg-18-389-2011. URL: https://www.nonlin-processes-geophys.net/18/389/2011/.

[11] K. Robison, A. M. McGuire, and G. M. Church. "A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia

coli K-12 genome11 Edited by R. Ebright". In: *Journal of Molecular Biology* 284.2 (1998), pp. 241–254. ISSN: 0022-2836. DOI: https://doi.org/10.1006/jmbi.1998.2160. URL: http://www.sciencedirect.com/science/article/pii/S002228369892160X.

[12] W. Rudin. *Real and Complex Analysis*. Tata McGraw-hill education, 2006.

[13] J. Scargle. "Studies in Astronomical Time Series Analysis. II - Statistical Aspects of Spectral Analysis of Unevenly Spaced Data". In: *The Astrophysical Journal* 263 (Jan. 1983). DOI: 10.1086/160554.

[14] P. Simard et al. "Boxlets: a fast convolution algorithm for signal processing and neural networks". In: *Advances in Neural Information Processing Systems*. 1999, pp. 571–577.

[15] R. S Tsay. "Financial Time Series". In: *Wiley StatsRef: Statistics Reference Online* (2014), pp. 1–23.

[16] J. T. VanderPlas. "Understanding the Lomb-Scargle Periodogram". In: *The Astrophysical Journal Supplement Series* 236.1 (May 2018), p. 16. DOI: 10.3847/1538-4365/aab766. URL: https://doi.org/10.3847%2F1538-4365%2Faab766.

[17] J. T. VanderPlas and Ž. Ivezic. "Periodograms For Multiband Astronomical Time Series". In: *The Astrophysical Journal* 812.1 (Oct. 2015), p. 18. DOI:

10.1088/0004-637x/812/1/18. URL: https://doi.org/10.1088%2F0004-637x%2F812%2F1%2F18.

[18] M. Vetterli, J. Kovačević, and V. K Goyal. *Foundations of signal processing*. Cambridge University Press, 2014.

[19] A. Young. "A Consensus Model for Electroencephalogram Data Via the S-Transform". MA thesis. East Tennessee State University, May 2012.

Appendix: Python Code

```
#!/usr/bin/env python
# coding: utf-8


# In[ ]:



### code used to demonstrate the convolution corollary



get_ipython().run_line_magic('matplotlib', 'inline')
from matplotlib import pyplot as plt


import numpy as np
import pandas as pd
from numpy.random import exponential, randint
from scipy import signal
from scipy.fftpack import fft, ifft
from scipy import linalg
```

```
get_ipython().run_line_magic('ls', 'Tests')
##specify the test family of signals and read in as
##a pandas dataframe
FN = 'PureSignal1'
Signals = pd.read_csv( f'Tests\{FN}.dat', sep = '\t',
                       header = None).T

## label the signals and create  a copy of the data
##for sampling nonuniformly
Signals.columns = ["sig%s" % (i+1) for i in range(32)]
Signals.index = Signals.index / (len(Signals.index) - 1)
Signals_UnifSamples = Signals
Signals_NonUnfSmpls = Signals.copy()

## On average, every 20% of signal is missing
NumberMissingAve = len(Signals)*0.2

## sample the time series nonuniformly
for i in range(Signals.shape[1]):
    missing = exponential(len(Signals)/NumberMissingAve,
```

```
                                    int(NumberMissingAve) )
    missing = np.cumsum( np.maximum(missing,1) ).astype(int)
    Signals_NonUnfSmpls.iloc[missing[ missing < len(Signals)],
                                    i] = np.nan

## Fills missing values in each signal with 0
Patched=Signals_NonUnfSmpls.fillna(0)
Patched_mat=Patched.values

## creates matrices that will contain the spectrum of each signal
## and pairwise convolutions of signals respectively
Four_mat=np.zeros(Patched_mat.shape,np.complex64)
times=Patched.index
Conv_matr=np.zeros(Four_mat.shape, np.complex64)

##Computes the Fourier transform and convolutions
for i in range(len(Patched_mat[0,:])):
    Four_mat[:,i]=fft(Patched_mat[:,i])
for i in range(32):
    Conv_matr[:,i]=Four_mat[:,i]*Four_mat[:,i-1]*Four_mat[:,i-2]
```

```
## Takes in the ifft of the consensus spectral model for the
## family and plots an overlay of the signals
for col in range(32):
    plt.plot(times, ifft(Conv_matr[:,i]).real)
```

VITA

WILLIAM SEGUINE

Education:	M.S. Mathematics, East Tennessee State University Johnson City, Tennessee 2019 B.S. Computational Applied Mathematics, ETSU, Johnson City, Tennessee 2017 A.S. Gen. Science, Pellissippi State Community College, Knoxville, Tennessee, 2013
Professional Experience:	Graduate Teaching Assistant, ETSU, Johnson City, Tennessee, Fall 2019 Graduate Assistant, Tutor, ETSU Johnson City, TN, Spring 2018–Spring 2019