Carnegie Mellon University
Research Showcase @ CMU

Dissertations | Theses and Dissertations

6-2011

The Short Time Fourier Transform and Local Signals

Shuhei Okamura

Follow this and additional works at: http://repository.cmu.edu/dissertations

Recommended Citation
Okamura, Shuhei, "The Short Time Fourier Transform and Local Signals" (2011). *Dissertations.* Paper 58.

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Research Showcase @ CMU. It has been accepted for inclusion in Dissertations by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

CARNEGIE MELLON UNIVERSITY

THE SHORT TIME FOURIER TRANSFORM AND LOCAL SIGNALS

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

In

STATISTICS

by

SHUHEI OKUMURA

Department of Statistics

Carnegie Mellon University

Pittsburgh, Pennsylvania 15213

June, 2011

© Copyright by Shuhei Okumura 2011

All right reserved.

Abstract

In this thesis, I examine the theoretical properties of the short time discrete Fourier transform (STFT). The STFT is obtained by applying the Fourier transform by a fixed-sized, moving window to input series. We move the window by one time point at a time, so we have overlapping windows. I present several theoretical properties of the STFT, applied to various types of complex-valued, univariate time series inputs, and their outputs in closed forms. In particular, just like the discrete Fourier transform, the STFT's modulus time series takes large positive values when the input is a periodic signal. One main point is that a white noise time series input results in the STFT output being a complex-valued stationary time series and we can derive the time and time-frequency dependency structure such as the cross-covariance functions. Our primary focus is the detection of local periodic signals. I present a method to detect local signals by computing the probability that the squared modulus STFT time series has consecutive large values exceeding some threshold after one exceeding observation following one observation less than the threshold. We discuss a method to reduce the computation of such probabilities by the Box-Cox transformation and the delta method, and show that it works well in comparison to the Monte Carlo simulation method.

Acknowledgments

First and foremost, I would like to thank Professor Bill Eddy. His intelligence and insight made it possible for me to complete this thesis. He has been a patient mentor and provided me with helpful guidance throughout my graduate study. In spite of a huge number of projects and wide-ranging responsibilities, he always made time for me. I greatly benefited from his research group meetings as well, where I was given opportunities to listen to and participate in inspiring works and discussions. I would also like to express my gratitude to Professors Jelena Kovačević, Chad Schafer, and Howard Seltman for being on my committee and for their constructive feedback to help shape this thesis. Their guidance and support played an indispensable role in this work. I am indebted to them so much more than I can describe. A very special thanks goes to Professor Jianming Wang who was a visitor to the department during the 2007-08 academic year and introduced me to the topic of this thesis. I have learned very much from his passion and dedication to his work. I am deeply thankful that he was one of those people who would appear out of nowhere and leave with everlasting positive influence. I call him a ninja. I also thank the faculty, staff, and fellow students for wonderful learning opportunities and a great environment.

I appreciate the advice and support from Professors Anto Bagic, William Williams, William Hrusa, John D. Norton, Shingo Oue, Anthony Brockwell, John Lehoczky, Takeo Kanade, Hugh Young, Tanzy Love, Kaori Idemaru, Lori Holt, Namiko Kunimoto, Marios Savvides, Marc Sommer, and Yoko Franchetti, and also from Alexander Düring, Philip Lee, and Shigeru Sasao. Their wisdom and experience helped me nurture both in and outside school.

I would like to thank Professors Julia Norton, Eric Suess, and Bruce Trumbo for their support and for helping me learn and grow through irreplaceable experiences during my undergraduate study. Many meetings lasted for hours. Their passion and encouragement are unforgettable. They showed me by examples how statisticians can contribute to many different fields and how rewarding such life is. Professor Ronald Randles of the University of Florida cleared the sky by answering many questions on graduate school when I happened to be seated next to him on a bus tour in Minneapolis during the 2005 Joint Statistical Meetings. I am happy that I am still perfectly convinced that pursuing graduate study in statistics was the right decision, and I appreciate many people's valuable time and help along the way.

Finally, I would like to acknowledge my families and friends who continued to support me throughout many years and let me share fantastic times together. Many played tennis with me. I am amazed at how I have always been surrounded by truly warm, caring people. I am forever grateful for such blessings.

Contents

1 Introduction and Outline **1**

2 Definition and Computation of STFT **4**

2.1 DFT and STFT 4

2.2 In Matrix Forms 6

2.3 Recursive Formulae 9

2.4 Previous Work 10

3 STFT on a White Noise Time Series **13**

3.1 Definition of White Noise 13

3.2 Theoretical Properties of the STFT on White Noise 16

3.2.1 The Bivariate Distribution of $|A_k^t|^2$ and $|A_k^{t+h}|^2$ 19

3.3 An Example 21

4 STFT on a Global Signal **24**

4.1 Periodic Signal 24

4.2 General Signal With Fourier Representation 26

4.3 Leakage With Periodic Signals 28

4.3.1 An Integer Number Of Periods 29

4.3.2 A Non-Integer Number Of Periods 30

4.4 Kronecker Delta Function 31

4.5 Step Function And Ringing 32

5 STFT on a Simple Local Signal **34**

5.1 Periodic Signal 34

5.2 An Example 36

6 Detection By Marginal Distribution **39**

6.1 Data of a Local Signal With Noise 39

6.2 Sample Quantiles 42

6.3 Marginal Threshold 45

7 Detecting Local Signals By Considering the Time Dependency Structure Of the STFT Output Time Series **47**

7.1 By Using One-Step Prediction With A Bivariate MA Process and Identifying Large Residuals 48

7.2 By Considering the Probability Of Observing Consecutive Large Values Exceeding A Threshold 51

7.3 Gaussian Stationary Process 54

7.3.1 AR(p) 55

7.3.2 The Box-Cox Transformation 56

7.3.3 The Delta Method 59

7.3.4 By the Monte Carlo Simulation Method 62

7.3.5 Comparison of the Two Methods 63

8 Conclusion and Future Work **65**

A References **68**

List of Figures

3.1 The top two plots show complex-valued Gaussian white noise time series input. The bottom three show the (complex-valued) STFT output of the input with window size 10. No visually obvious pattern exists, except neighboring points are often similar. 22

3.2 The top row shows the two time series of the STFT output angle(A_2^t) and angle(A_3^t) computed from the example in Figure 3.1. The scatter plots in the middle row are one- and two-step functions of the time series, angle(A_2^{t-1}) against angle(A_2^t) and angle(A_2^{t-2}) against angle(A_2^t), respectively. The last row shows similar scatter plots for angle(A_3^{t-1}) against angle(A_2^t) and angle(A_3^{t-2}) against angle(A_2^t). We see that the cross-covariance functions are not appropriate measures for the dependence of these nonlinear time series. 23

5.1 A simple example: The top two plots are the complex-valued input, which has a cosine function in the middle in the real part and is zero-valued in the imaginary part. The bottom three plots show the (complex-valued) STFT output: squared modulus, real and imaginary parts. 38

6.1 The input is a complex-valued Gaussian white noise plus a real-valued periodic local signal. The top plot shows the real part and the bottom plot shows the imaginary part of the time series input. We will consider ways to detect this local signal in this chapter and next. 40

6.2 The squared modulus STFT output resulting from the input in Figure 6.1. The large values of $k = 2$ and 8 indicate the local signal. 41

6.3 The histograms of natural logarithm of the squared modulus STFT in Figure 6.2 for $k = 1, \ldots, 9$. We notive that the values larger than 3 occur at $k = 2$ and 8, which does not happen at other frequency indices, thus indicating the existence of a local periodic signal. 43

6.4 The sample quantile of log of the squared modulus STFT in Figure 6.2. Clearly, two frequency indices $k = 2$ (dashed line) and 8 (dotted line) have distributions different from the others (solid lines), indicating the existence of a local periodic signal. 44

6.5 The time series of natural logarithm of the squared modulus for $k = 2$ (dashed line) and $k = 8$ (dotted line), along with the log of $\mathrm{Exp}(\sigma_{RR}+\sigma_{II})$. We observe large values where the local signal exists. 46

7.1 The top two plots show the time series of Mahalanobis distance of the residuals computed from the one-step prediction function of the bivariate moving average process, along with 99 percentile of $\chi^2_{df=2}$. They show many small values between large values and thus are not helpful for finding local periodic signals. The bottom plot is their scatterplot, which says that the two time series are almost identical. 50

7.2 The choice of the transformation parameter λ for the Box-Cox transformation, applied to the squared modulus STFT time series of Gaussian white noise. In the top plot, λ values are plotted against the log-likelihood function. The marginal distribution of the time series before the transformation is exponential. The bottom plot shows that with choice of 0.27, we have approximately Gaussian marginal distribution, plotted along with the Gaussian distribution function with the maximum likelihood parameter estimates. 58

7.3 The residuals from AR(1) fitted to the Box-Cox transformed data in Figure 7.2, with the coefficient chosen by the delta method. The top plot shows the marginal distribution of the residuals, which is approximately Gaussian. The middle and bottomo plots show the autocorrelation function and partial autocorrelation function of the residuals, respectively, which show that the residuals are approximately white noise. They indicate that the Box-Cox transform and the delta method work reasonably well. 61

7.4 The conditional probability of observing s consecutive values exceeding the threshold q after one exceeding observation that follows one observation below the threshold, $\Pr(Y_{t+s} \geq q, \ldots, Y_{t+1} \geq q | Y_t \geq q, Y_{t-1} < q)$. Comparing the probabilities computed in two ways. One method uses the delta method and the Box-Cox transformation (solid line), while the other uses the Monte Carlo simulation (dashed line). This indicates that our approximation method works well. 64

Chapter 1

Introduction and Outline

In this thesis, I am going to examine the theoretical properties of the short time Fourier transform (STFT) in discrete time. The STFT is obtained by applying the Fourier transform by a fixed-sized, moving window to input series. We move the window by one time point at a time, so we have overlapping windows. I will present several theoretical properties of the STFT, applied to various types of complex-valued, univariate time series inputs, and their outputs in closed forms.

In particular, just like the discrete Fourier transform, the STFT's modulus time series takes large positive values when the input is a periodic signal. One main point is that a white noise time series input results in the STFT output being a complex-valued stationary time series and we can derive the time dependency structure such as the cross-covariance functions. As we will see in Chapter 2, the study of the STFT's time dependency structures has not been previously conducted.

Our primary focus will be the detection of local periodic signals. A local signal is a signal that appears only in parts of data. When the noise level is high, the detection of such signal is difficult with traditional methods such as the discrete Fourier transform since the discrete Fourier transform of frequencies of no interest can take large values as well as those of interest.

I will present a method to detect local signals by computing the probability that the squared modulus STFT time series has consecutive large values exceeding some threshold. We will discuss a method to reduce the computation of such probabilities by the Box-Cox transformation and the delta method, and show that it works well in comparison to the Monte Carlo simulation method. In this approximation computation, we will rely on the time dependency structure that we derive in Chapter 3.

We show the organization of this thesis and the summary of each chapter as follows. Chapter 2 introduces the discrete Fourier transform and the STFT with similarities and differences between them as well as the absolute summability condition that lets us define the DFT and the STFT. We will assume this absolute summability condition in this thesis. We will also show the STFT's computation, and a fast algorithm for it derived by Wang et al. (2009). We end Chapter 2 with literature survey. The subject of the STFT dates back to at least 1946 by Gabor, and there have been several papers. The STFT has been used in various applications, but only in exploratory ways. No research paper has investigated the time dependency structure of the STFT. This is precisely the aim of this thesis.

Chapter 3 investigates the theoretical properties of the STFT applied to a complex-valued white noise series. We will define the complex-valued white noise series by extending the definition of the real-valued white noise series in a straightforward manner. We will see that the STFT output forms a stationary time series. Our focus will be on the cross-covariance functions between different frequency indices because other time series properties such as the spectrum are functions of these cross-covariance functions.

Chapter 4 presents a variety of input series whose STFT output results in closed forms. We will show ways to reduce computation of the STFT. We will consider periodic signals, signals with simple Fourier representations, the Kronecker delta function, and the step function. We will also discuss the problem of leakage and ringing. Chapter 5 considers the STFT resulting from a local periodic signal, where we have a periodic signal between zero-valued regions. Such input will cause the STFT output to be tractable only under very limited

circumstances. We will present an example to show these points.

Chapters 6 and 7 take the opposite direction of using the STFT. That is, we use the STFT output to investigate the given input series with added noise. Chapter 6 illustrates preliminary and exploratory analysis with the STFT to detect a local signal. We will compare the marginal distributions of different frequency indices of the squared modulus STFT, and derive the threshold to evaluate the large values. The approach in Chapter 6, however, ignores the fact that the STFT is a stationary time series and has time dependency structure. This fact will be utilized in Chapter 7. We will evaluate the probability that multiple consecutive observations of the squared modulus STFT time series exceed some threshold, and use it to conclude if such observations can occur just as a consequence of a random, stationary process or a local signal is causing the non-stationarity. The complex form of the squared modulus STFT time series makes it difficult to compute such probabilities analytically, so we use the Monte Carlo simulation. There we also illustrate the usefulness of the Gaussian autoregressive process in computing such probabilities, and how to approximately transform our STFT into it by the Box-Cox transformation and the delta method. We compare the approximation method to the Monte Carlo method. Chapter 9 summarizes the thesis and points out problems that we do not address in this thesis.

Chapter 2

Definition and Computation of STFT

In this chapter, we define the short time discrete Fourier transform and show its computation. We begin with a brief review of the discrete Fourier transform that is the basis of the STFT and illustrate contrasts between the two methods' computation with matrix forms, and then show a fast algorithm to compute the STFT.

Throughout this thesis, we assume that we have a complex-valued, univariate, discrete time series $\{X_t\}_{t=0}^{M-1}$ of finite length M.

2.1 DFT and STFT

The discrete Fourier transform (DFT) is given by

$$D_k = \frac{1}{\sqrt{M}} \sum_{j=0}^{M-1} X_j \omega_M^{-jk} \ , \qquad k = 0, 1, \ldots, M-1 \tag{2.1}$$

where

$$\omega_M^{\ell} = \exp\left(\frac{i2\pi\ell}{M}\right) = \cos\left(\frac{2\pi\ell}{M}\right) + i\sin\left(\frac{2\pi\ell}{M}\right).$$

The DFT is basically a least square regression of the input on $\lfloor M/2 \rfloor$ periodic functions with different frequencies, and finding the coefficients of amplitudes and phases that will

result in a perfect fit (zero residual errors). When the time series is real-valued, we focus on the range $k = 0, \ldots, \lfloor M/2 \rfloor$ because the ignored range is just the complex conjugate of the focused range. We can recover the original time series by applying the inverse DFT to the DFT coefficients, where the inverse DFT is given by

$$X_j = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} D_k \omega_M^{jk} \ , \qquad j = 0, 1, \ldots, M-1. \tag{2.2}$$

The constant multiplier $1/\sqrt{M}$ in (2.1) and (2.2) is chosen because of its consistency in both equations, and thus is not particularly important. We can use 1 in either (2.1) or (2.2) and $1/M$ in the other. In this thesis we use $1/\sqrt{M}$ in both the DFT and the inverse DFT.

In order for the Fourier transform pair (2.1) and (2.2) to exist and be well-defined, we assume that the input series satisfies the absolute summability condition:

$$\sum_{j=0}^{M-1} |X_j| < \infty \tag{2.3}$$

throughout this thesis. This assumption lets us define the short time Fourier transform as below.

Now, the short time Fourier transform (STFT) is computed by applying the DFT to a continuous subset of the data with $N \leq M$ data points. We call this range the *window*. We first compute the DFT of $(X_0, \ldots, X_{N-1})$. Second, we move the window by one time index and the compute the DFT of $(X_1, \ldots, X_N)$. We repeat this procedure until the window covers the last N data points of the input and compute the DFT of $(X_{M-N}, \ldots, X_{M-1})$. Thus, we define the STFT;

$$A_k^t = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} X_{j+t-N+1} \omega_N^{-jk} \tag{2.4}$$

$$k = 0, 1, \ldots, N-1; \qquad\qquad t = N-1, \ldots, M-1.$$

We emphasize that each data point X_j is included in the computation of the DFT N times, and that the STFT is different from the *partitioned DFT* which partitions the input series X_t into multiple series of length N and computes the DFT in each segment, where each data point X_j is included in the computation of the DFT only once. Our STFT uses overlapping windows, while the partitioned DFT uses non-overlapping windows. This computational redundancy in the STFT provides us with some time dependency structure (depending on the input) that we will study in Chapter 3. Of course, when $N = M$, the STFT is equivalent to the DFT.

Now we have a complex-valued array $\mathcal{A}$ of size N-by-$(M - N + 1)$. $\{A^t\}_{t=N-1}^{M-1}$ can be seen as an N-dimensional complex-valued time series of length $(M-N+1)$. We think of the time index t in the horizontal direction and the frequency index k in the vertical direction of the array $\mathcal{A}$.

2.2 In Matrix Forms

To further clarify the distinction between the DFT (2.1) and the STFT (2.4), we present their computation in matrix forms.

Given the input $\{X_t\}_{t=0}^{M-1}$ of length M, the DFT can be written as:

$$\begin{aligned}
\begin{bmatrix} D_0 \\ D_1 \\ D_2 \\ \vdots \\ D_{M-1} \end{bmatrix} &= \frac{1}{\sqrt{M}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega_M^{-1} & \omega_M^{-2} & \cdots & \omega_M^{-1(M-1)} \\ 1 & \omega_M^{-2} & \omega_M^{-4} & \cdots & \omega_M^{-2(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_M^{-1(M-1)} & \omega_M^{-2(M-1)} & \cdots & \omega_M^{-(M-1)^2} \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{M-1} \end{bmatrix}. \\
&= \frac{1}{\sqrt{M}} \mathbf{F}_M \overrightarrow{X},
\end{aligned}$$

where $\mathbf{F}_M$ denotes the M-by-M Fourier transform matrix, and $\overrightarrow{X} = (X_0, X_1, \ldots, X_{M-1})^\intercal$.

Earlier, we described the STFT output matrix as a complex-valued array $\mathcal{A}$ of size N-by-$(M-N+1)$ with the time index t in the horizontal direction and the frequency index k in the vertical direction of the array $\mathcal{A}$. To be consistent with this, we can write

$$\mathcal{A} = \begin{bmatrix} A_0^{N-1} & A_0^{N} & A_0^{N+1} & \cdots & A_0^{M-1} \\ A_1^{N-1} & A_1^{N} & A_1^{N+1} & \cdots & A_1^{M-1} \\ A_2^{N-1} & A_2^{N} & A_2^{N+1} & \cdots & A_2^{M-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{N-1}^{N-1} & A_{N-1}^{N} & A_{N-1}^{N+1} & \cdots & A_{N-1}^{M-1} \end{bmatrix}$$

$$= \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega_M^{-1} & \omega_M^{-2} & \cdots & \omega_M^{-1(M-1)} \\ 1 & \omega_M^{-2} & \omega_M^{-4} & \cdots & \omega_M^{-2(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_M^{-1(M-1)} & \omega_M^{-2(M-1)} & \cdots & \omega_M^{-(M-1)^2} \end{bmatrix} \begin{bmatrix} X_0 & X_1 & X_2 & \cdots & X_{M-1-N+1} \\ X_1 & X_2 & X_3 & \cdots & X_{M-1-N+2} \\ X_2 & X_3 & X_4 & \cdots & X_{M-1-N+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{N-1} & X_N & X_{N+1} & \cdots & X_{M-1} \end{bmatrix}.$$

This representation, however, involves properly aligning the input $\{X_t\}_{t=0}^{M-1}$ and does not clearly show the difference between the DFT and the STFT. We consider another representation.

Our second approach is to represent the N-by-$(M-N+1)$ STFT matrix $\mathcal{A}$ as a column vector of length $N \times (M-N+1)$ by stacking one column after another from left to right. That is,

$$(A_0^{N-1}, A_1^{N-1}, \ldots, A_{N-1}^{N-1}, A_0^{N}, A_1^{N}, \ldots, A_{N-1}^{N}, A_0^{N+1}, A_1^{N+1}, \ldots, A_{N-1}^{N+1}, \ldots, \ldots, A_0^{M-1}, A_1^{M-1} \ldots, A_{N-1}^{M-1})^{\intercal} = \frac{1}{\sqrt{N}} \times$$

$$\begin{vmatrix}
1 & 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
1 & \omega_N^{-1} & \omega_N^{-2} & \cdots & \omega_N^{-1(N-1)} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \omega_N^{-1(N-1)} & \omega_N^{-2(N-1)} & \cdots & \omega_N^{-(N-1)^2} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 1 & 1 & \cdots & 1 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 1 & \omega_N^{-1} & \cdots & \omega_N^{-1(N-2)} & \omega_N^{-1(N-1)} & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 1 & \omega_N^{-1(N-1)} & \cdots & \omega_N^{-(N-2)(N-1)} & \omega_N^{-(N-1)^2} & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 1 & 1 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & \omega_N^{-1(N-3)} & \omega_N^{-1(N-2)} & \omega_N^{-1(N-1)} & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
0 & 0 & 1 & \cdots & \omega_N^{-(N-3)(N-1)} & \omega_N^{-(N-2)(N-1)} & \omega_N^{-(N-1)^2} & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 1 & \cdots & 1 \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & \omega_N^{-1} & \cdots & \omega_N^{-1(N-1)} \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & \omega_N^{-1(N-1)} & \cdots & \omega_N^{-(N-1)^2}
\end{vmatrix}
\begin{bmatrix}
X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{N-1} \\ X_N \\ X_{N+1} \\ \vdots \\ X_{M-N} \\ X_{M-N+1} \\ \vdots \\ X_{M-1}
\end{bmatrix}.$$

We see that $\mathbf{F}_N$ appears in the upper left $N \times N$ submatrix of the matrix, computing the DFT on the first N data points of the input X_t, i.e., $(X_0, \ldots, X_{N-1})$. Then in the next set of N rows (the $N+1^{\text{th}}$ row to the $2N^{\text{th}}$ row), $\mathbf{F}_N$ moves to the right by one and computes the DFT of the $(X_1, \ldots, X_N)$. And we repeat this process until we compute the DFT of $(X_{M-N}, \ldots, X_{M-1})$, which is shown as $\mathbf{F}_N$ in the bottom right corner of the matrix. All in all, this matrix contains a very large number of entries of 0's, and some pattern in which $\mathbf{F}_N$ appears. As one can imagine, there is a substantial amount of computational redundancy and there are ways to reduce it such as one described in Wang et al. (2009), which we consider in the next section.

2.3 Recursive Formulae

When the window moves by one time index, the region covered by the two window positions can be taken into consideration and be used to reduce computation, as shown in Wang et al. (2009);

$$A_k^{t+1} = \omega_N^k \left(A_k^t - \frac{1}{\sqrt{N}} X_{t-N+1} + \frac{1}{\sqrt{N}} X_{t+1} \right).$$

They also derived the computation when the window moves by $s \geq 1$ steps.

$$A_k^{t+s} = \omega_N^{ks} \left(A_k^t + \frac{1}{\sqrt{N}} \sum_{j=0}^{s-1} \left(X_{t+1+j} - X_{t-N+1+j} \right) \omega_N^{-jk} \right).$$

This can be also written as:

$$\begin{aligned} A_k^{t+s} &= \left[\cos\left(\frac{2\pi ks}{N}\right) + i \sin\left(\frac{2\pi ks}{N}\right) \right] \Bigg[\mathrm{Re}(A_k^t) + i\mathrm{Im}(A_k^t) \\ &\quad + \frac{1}{\sqrt{N}} \sum_{j=0}^{s-1} \left[\mathrm{Re}(X_{t+1+j} - X_{t-N+1+j}) + i\mathrm{Im}(X_{t+1+j} - X_{t-N+1+j}) \right] \\ &\qquad \left[\cos\left(\frac{2\pi jk}{N}\right) - i \sin\left(\frac{2\pi jk}{N}\right) \right] \Bigg], \\ &\quad \text{by the identity } \omega_N^\ell = \cos\left(\frac{2\pi \ell}{N}\right) + i \sin\left(\frac{2\pi \ell}{N}\right) \text{ for } \ell, N \in (-\infty, \infty). \end{aligned}$$

We can use this formula to further derive the real part $\mathrm{Re}(A_k^{t+s})$, the imaginary part $\mathrm{Im}(A_k^{t+s})$, and the squared modulus $|A_k^{t+s}|^2$:

$$\begin{aligned} \mathrm{Re}(A_k^{t+s}) &= \cos\left(\frac{2\pi ks}{N}\right) \mathrm{Re}(A_k^t) - \sin\left(\frac{2\pi ks}{N}\right) \mathrm{Im}(A_k^t) \\ &\quad + \frac{1}{\sqrt{N}} \Bigg[\sum_{j=0}^{s-1} \mathrm{Re}(X_{t+1+j} - X_{t-N+1+j}) \cos\left(\frac{2\pi k(j-s)}{N}\right) \\ &\quad + \mathrm{Im}(X_{t+1+j} - X_{t-N+1+j}) \sin\left(\frac{2\pi k(j-s)}{N}\right) \Bigg], \end{aligned}$$

$$
\begin{aligned}
\mathrm{Im}(A_k^{t+s}) &= \sin\left(\frac{2\pi ks}{N}\right)\mathrm{Re}(A_k^t) + \cos\left(\frac{2\pi ks}{N}\right)\mathrm{Im}(A_k^t) \\
&\quad + \frac{1}{\sqrt{N}}\left[\sum_{j=0}^{s-1} -\mathrm{Re}(X_{t+1+j} - X_{t-N+1+j})\sin\left(\frac{2\pi k(j-s)}{N}\right)\right. \\
&\quad \left. + \mathrm{Im}(X_{t+1+j} - X_{t-N+1+j})\cos\left(\frac{2\pi k(j-s)}{N}\right)\right], \qquad \text{and} \\
|A_k^{t+s}|^2 &= |A_k^t|^2 + \frac{\phi^2+\psi^2}{N} + \frac{2\phi}{\sqrt{N}}\mathrm{Re}(A_k^t) + \frac{2\psi}{\sqrt{N}}\mathrm{Im}(A_k^t) \qquad \text{where} \\
\phi &= \sum_{j=0}^{s-1} \mathrm{Re}(X_{t+1+j} - X_{t-N+1+j})\cos\left(\frac{2\pi kj}{N}\right) \\
&\quad + \mathrm{Im}(X_{t+1+j} - X_{t-N+1+j})\sin\left(\frac{2\pi kj}{N}\right) \qquad \text{and} \\
\psi &= \sum_{j=0}^{s-1} -\mathrm{Re}(X_{t+1+j} - X_{t-N+1+j})\sin\left(\frac{2\pi kj}{N}\right) \\
&\quad + \mathrm{Im}(X_{t+1+j} - X_{t-N+1+j})\cos\left(\frac{2\pi kj}{N}\right).
\end{aligned}
$$

Once we calculate A_k^t, i.e. the DFT of $(X_{t-N+1}, \ldots, X_t)$, we can use A_k^t and avoid directly calculating A_k^{t+s}, i.e., the DFT of $(X_{t+s-N+1}, \ldots, X_{t+s})$. Of course, if $s \geq N$, then the two DFT calculations are independent and we have the partitioned DFT, therefore these formulae do not reduce computing time (although they are still mathematically correct but have poor numerical properties). Later we will see that we are primarily interested in the squared modulus time series $\{|A_k^t|^2\}_t$. By taking advantage of computational redundancy of overlapping windows, these formulae can reduce computational time.

2.4 Previous Work

According to Vetterli et al. (2011), the earliest application of the STFT was by Dennis Gabor to analyze speech data (Gabor 1946). The STFT is also known under many names such as *the windowed Fourier Transform*, *the Gabor transform*, and *the local Fourier transform*. We will use the STFT exclusively in this thesis. To be precise, the STFT is actually a special

case of the Gabor transform, with the input series within each window position is multiplied by unequal weights c_j's to weigh the values at the center of the window more than the values near the window edges:

$$A_k^t = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} c_j X_{j+t-N+1} \omega_N^{-jk}. \tag{2.5}$$

This definition is more general than the STFT defined in (2.4), which can be achieved by setting $c_j = 1$ for all j's. In the field of signal processing, using weights for the DFT is common and called "tapering." Among a large number of proposed weight functions for c_j's, one of the most commonly used functions was introduced by Blackman and Tukey (1959):

$$c_j = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\frac{2\pi(j - \frac{N+1}{2})}{N}\right)\right] & \text{for } 0 \leq j \leq \frac{\alpha N}{2} \text{ and } N(1 - \frac{\alpha}{2}) \leq j \leq N \\ 1 & \text{otherwise} \end{cases}$$

for $\alpha \in [0, 1]$. Here we are focusing on the DFT, so $N = M$. Tapers like this result in smoothing neighboring DFT coefficients. In this thesis we will only focus the STFT on the case without tapering, thus $c_j = 1$ for all j's.

Some authors (e.g., Shumway and Stoffer 2006) use the partitioned discrete Fourier transform and call this the STFT. Again, the partitioned DFT partitions the input series X_t into multiple series of length N and computes the DFT in each segment, and is different from our STFT with overlapping windows. For the partitioned DFT, Percival and Walden (1993) mention a variety of methods to smooth neighboring DFT coefficients.

The STFT has been used in various applications such as diagnosing atherosclerosis (Latifoglu et al. 2009), optical metrology (Qian 2004), and the surface waves induced by moving ships (Wyatt 1988). Dirgenali et al. (2006) use the STFT for the electrogastrography data of stomach muscle to study the frequency of abnormalities in diabetic gastroparesis patients. Jiang and He (2009) apply the STFT to study the frequency change in power systems and electric devices, and confirm the usefulness of the approach by simulation. These authors, however, use the STFT in exploratory ways and do not provide thresholds to conclude sta-

tistical significance. This will be the focus in Chapters 6 and 7 of this thesis.

The STFT also opened questions from different perspectives. Avargel and Cohen (2010) consider applying the STFT to the output of some unknown nonlinear system in comparison to the STFT of the input, and model and identify the nonlinear system, although they use only the quadratically nonlinear systems. Xia (1998) analyzes the signal-to-noise ratio in the STFT for multicomponent signals in additive white noise, and shows how the SNR increases with respect to the number of components, the sampling rate, and the window size.

Some authors study more general conditions on the input series than the absolute summability condition under which the STFT exists such as bounded power signals and almost periodic functions (Partington and Unalmis 2001; Radha and Thangavelu 2009; Matusiak et al. 2010). Finding such conditions is beyond the scope of this thesis. For our purposes, it suffices to assume the absolute summability condition (2.3).

However, I am not aware of any research paper where the time dependency structure of the STFT is investigated. This is precisely the aim of this thesis. In Chapter 3, we will study the time dependency structure such as the cross-covariance functions of the STFT, for example, $E\{[|A_k^{t+h}|^2 - E(|A_k^{t+h}|^2)][|A_\ell^t|^2 - E(|A_\ell^t|^2)]\}$, resulting from a white noise time series input. No paper has ever investigated this simple theoretical property. Chapter 4 considers other types of input. Then later we will examine ways of using such results to find local features of a given time series input.

Chapter 3

STFT on a White Noise Time Series

In Chapter 2, we defined the short time Fourier transform (STFT) and examined its computation. In this chapter, we start to impose some assumptions on the input signal and consider its STFT output. In particular, in this chapter we assume a white noise time series input. Just like the discrete Fourier transform, the STFT takes a complex-valued input. We begin by defining the complex-valued white noise input.

We then point out five ways to look at the STFT output. Given a white noise input, the resulting STFT forms a multivariate stationary time series. We derive the cross-covariance functions of the component time series. Finally, an example will show that one of the ways to look at the STFT is not particularly informative.

3.1 Definition of White Noise

When the original time series X_t is a white noise time series, its STFT will form a multivariate time series of interest. A real-valued time series X_t is called white noise, if

- $\mathrm{E}[X_t] = 0 \ \forall t$
- $\mathrm{Var}[X_t] = \sigma^2$ (constant) $\forall t$, and
- $\mathrm{Cov}[X_i, X_j] = 0 \ \forall i \neq j$.

This does not assume any distribution. Since I have not found any definition of complex-valued white noise, I just define it here. The following definition is a straightforward extension of the above real-valued white noise. First, we assume a two-dimensional real-valued white noise $\mathrm{Re}(X_t)$ and $\mathrm{Im}(X_t)$. Second, we define the complex-valued time series by adding the first component to the second component multiplied by the imaginary number $i = \sqrt{-1}$, i.e., $X_t = \mathrm{Re}(X_t) + i\mathrm{Im}(X_t)$. Now we describe the relationship between the two dimensions at different time points. We assume that the following are constant $\forall t$, along with notations where they are non-zero:

- $\mathrm{E}[\mathrm{Re}(X_t)] = \mathrm{E}[\mathrm{Im}(X_t)] = 0$
- $\mathrm{Var}[\mathrm{Re}(X_t)] = \sigma_{RR}$
- $\mathrm{Var}[\mathrm{Im}(X_t)] = \sigma_{II}$
- $\mathrm{E}[\mathrm{Re}(X_t)\mathrm{Im}(X_t)] = \sigma_{RI}$
- $\mathrm{Cov}[X_i, X_j] = 0\ \forall i \neq j$, which means $\mathrm{Cov}[\mathrm{Re}(X_i),\mathrm{Re}(X_j)] = \mathrm{Cov}[\mathrm{Im}(X_i),\mathrm{Im}(X_j)] = \mathrm{Cov}[\mathrm{Re}(X_i),\mathrm{Im}(X_j)] = \mathrm{Cov}[\mathrm{Im}(X_i),\mathrm{Re}(X_j)] = 0$
- $\mathrm{E}[(\mathrm{Re}(X_t))^3] = \mu_{R^3}$
- $\mathrm{E}[(\mathrm{Im}(X_t))^3] = \mu_{I^3}$
- $\mathrm{E}[(\mathrm{Re}(X_t))^2(\mathrm{Im}(X_t))] = \mu_{R^2I}$
- $\mathrm{E}[(\mathrm{Re}(X_t))(\mathrm{Im}(X_t))^2] = \mu_{RI^2}$
- $\mathrm{E}[(\mathrm{Re}(X_t))^3(\mathrm{Im}(X_t))] = \mu_{R^3I}$
- $\mathrm{E}[(\mathrm{Re}(X_t))(\mathrm{Im}(X_t))^3] = \mu_{RI^3}$
- $\mathrm{E}[(\mathrm{Re}(X_t))^2(\mathrm{Im}(X_t))^2] = \mu_{R^2I^2}$
- $\mathrm{E}[(\mathrm{Re}(X_t))^4] = \mu_{R^4}$ and

- $\mathrm{E}[(\mathrm{Im}(X_t))^4] = \mu_{I^4}$.

Again, we do not assume any distribution. We assume constants for higher moments in the last nine lines than the real-valued white noise and provided the notations only because later we will compute $E\{[|A_k^{t+h}|^2 - E(|A_k^{t+h}|^2)][|A_\ell^t|^2 - E(|A_\ell^t|^2)]\}$. Otherwise, the first five lines in the above definition would have been sufficient. Of course, the real-valued white noise can be obtained by setting $\mathrm{Im}(X_t) = 0 \forall t$ and thus

$\sigma_{II} = \sigma_{RI} = \mu_{I^3} = \mu_{R^2I} = \mu_{RI^2} = \mu_{R^3I} = \mu_{RI^3} = \mu_{R^2I^2} = \mu_{I^4} = 0$.

We recall that given an input $\{X_t\}_{t=0}^{M-1}$, the STFT output matrix $\mathcal{A}$ with window size N is complex-valued and of size N-by-$(M-N+1)$. There are at least four ways of looking at one frequency index k (one row) of the STFT matrix $\mathcal{A}$,

(1) complex-valued $\{A_k^t\}_t$,

(2) the real part $\{\mathrm{Re}(A_k^t)\}_t$,

(3) the imaginary part $\{\mathrm{Im}(A_k^t)\}_t$, and

(4) the squared modulus $\{|A_k^t|^2\}_t$.

We note that each of these time series is stationary given a white noise time series input, real-valued or complex-valued. We have found the autocovariance function, cross-covariance function (between different k's), and the spectrum for each of the four time series. We will examine these properties in the next section. We note (1) is a univariate complex-valued moving average process; (2) and (3) are the widely-used univariate real-valued moving average processes (Brockwell and Davis 1991) with order $N-1$; and (4) is the sum of the squared real part and the squared imaginary part, and therefore is univariate and real-valued. It does not have a universal name. We will refer to it as *the squared modulus time series*. The squared modulus time series is our primary interest because the large values observed in it will indicate the existence of periodic signals as we will see in subsequent chapters.

Another time series is the phase of A_k^t, or the angle between the positive real-axis of the complex plane and the complex value, $\mathrm{angle}(A_k^t) = \mathrm{Arg}(A_k^t) = \mathrm{atan2}(\mathrm{Im}(A_k^t),\mathrm{Re}(A_k^t))$. This is a nonlinear time series. Known properties are: 1) the marginal distribution is uniform on

$[-\pi, \pi]$ when X_t is Gaussian; and 2) the cross-covariance (between two different frequency indices k's) functions are zero at lags greater than or equal to the window size N. The bivariate probability density function of (angle(A_k^t), angle(A_ℓ^{t+h})) is hard to find analytically, because it involves transforming 4 random variables with non-independence structure. Later we will look at one example of simulated data and its STFT angle time series in Figure 3.2, where we clearly see nonlinearity. In such a case, the cross-covariance functions are no longer appropriate measures for the dependence of nonlinear time series (Fan and Yao 2003). One problem is that the time series angle(A_k^t) "jumps" near the boundaries. For example, if $A_k^t = -1 + 0.01i$ and $A_k^{t+1} = -1 - 0.01i$, then angle(A_k^t) $= 3.131593$ and angle(A_k^{t+1}) $= -3.131593 \neq \text{Arg}(1+0.01i)+\pi = 3.151592 = -3.131593+2\pi$. Such distribution as angle(A_k^t) is known as a circular distribution. This phenomenon makes it hard to utilize this phase time series for detecting local signals, while other series (1)-(4) can do the task, as shown later.

3.2 Theoretical Properties of the STFT on White Noise

In this section, we will consider some theoretical properties of the short time Fourier transform, applied to the complex-valued white noise time series. We focuse on the cross-covariance functions of the time series (1)-(4) described in the previous section, because other theoretical properties are just simple functions of the cross-covariances.

Let $CCF_{Y,Z}(h)$ denote the cross-covariance function of two time series Y_t and Z_t at lag h, that is, $CCF_{Y,Z}(h) = E\{(Y^{t+h} - E[Y^{t+h}])(Z^t - E[Z^t])^*\}$, where $*$ denotes the complex conjugate. The autocovariance function of a time series Y_t is the CCF of itself, i.e., $ACF_Y(h) = E\{(Y^{t+h} - E[Y^{t+h}])(Y^t - E[Y^t])^*\} = CCF_{Y,Y}(h)$, so we only show the CCF's.

For each of the four types of time series (1)-(4), $CCF(h) = 0$ for $|h| \geq N$. This is intuitive because the two STFT windows for A_k^{t+h} and A_k^t do not overlap and we are just computing two independent discrete Fourier transform. Also, $E[A_k^t] = 0$, as the STFT is the

weighted average of zero-mean random variables, and $E[|A_k^t|^2] = \sigma_{RR} + \sigma_{II}$ $\forall t$ and k.

For $0 \leq h < N$,

$$
\begin{aligned}
CCF_{A_k,A_\ell}(h) &= E\{(A_k^{t+h} - E[A_k^{t+h}])(A_\ell^t - E[A_\ell^t])^*\} \\
&= \frac{1}{N} E\left[\left(\sum_{j=0}^{N-1} X_{j+t+h-N+1}\omega_N^{-jk}\right)\left(\sum_{j=0}^{N-1} X_{j+t-N+1}\omega_N^{-j\ell}\right)^*\right] \\
&\overset{(a)}{=} \frac{1}{N} E\left[\left(\sum_{j=0}^{N-1} X_{j+t+h-N+1}\omega_N^{-jk}\right)\left(\sum_{j=0}^{N-1} X^*_{j+t-N+1}(\omega_N^{-j\ell})^*\right)\right] \\
&\overset{(b)}{=} \frac{1}{N} E\left[\left(\sum_{j=0}^{N-1} X_{j+t+h-N+1}\omega_N^{-jk}\right)\left(\sum_{j=0}^{N-1} X^*_{j+t-N+1}\omega_N^{j\ell}\right)\right] \\
&\overset{(c)}{=} \frac{1}{N} E\left[\left(\sum_{j=0}^{N-h-1} X_{j+t+h-N+1}\omega_N^{-jk}\right)\left(\sum_{j=h}^{N-1} X^*_{j+t-N+1}\omega_N^{j\ell}\right)\right] \\
&\overset{(d)}{=} \frac{1}{N} \sum_{m=0}^{N-h-1} \omega_N^{-km+\ell(m+h)} E[X_{m+t+h-N+1}X^*_{m+t+h-N+1}] \\
&\overset{(e)}{=} \frac{\sigma_{RR}+\sigma_{II}}{N} \sum_{m=0}^{N-h-1} \omega_N^{-km+\ell(m+h)},
\end{aligned}
$$

where (a) follows because for two complex numbers $c_1, c_2 \in \mathbb{C}, (c_1c_2)^* = c_1^*c_2^*$; (b) because ω_N^{-m} for $m \in \mathbb{R}, (\omega_N^{-m})^* = \omega_N^m$; (c) because only those X_t's in both windows are non-zero upon cross multiplication and taking expectation; (d) because $\mathrm{Cov}[X_i, X_j] = 0$ $\forall i \neq j$; and (e) because $E[X_tX_t^*] = E[(\mathrm{Re}(X_t) + i\mathrm{Im}(X_t))(\mathrm{Re}(X_t) + i\mathrm{Im}(X_t))^*]$
$= E[(\mathrm{Re}(X_t) + i\mathrm{Im}(X_t))(\mathrm{Re}(X_t) - i\mathrm{Im}(X_t))] = E[\mathrm{Re}^2(X_t) + i\mathrm{Im}^2(X_t)] = \sigma_{RR} + \sigma_{II}$.

Similarly, we can find the cross-covariance functions for other time series (2) $\{\mathrm{Re}(A_k^t)\}_t$, (3) $\{\mathrm{Im}(A_k^t)\}_t$, and (4) $\{|A_k^t|^2\}_t$. For notational simplicity, we let

$$
c_k(m) = \cos\left(\frac{2\pi km}{N}\right) \text{ and } s_k(m) = \sin\left(\frac{2\pi km}{N}\right).
$$

Then, for $0 \leq h < N$,

$$
\begin{aligned}
CCF_{Re(A_k),Re(A_\ell)}(h) &= \frac{1}{N}\Bigg[\sigma_{RR}\sum_{m=0}^{N-h-1} c_k(m)c_\ell(m+h) + \sigma_{II}\sum_{m=0}^{N-h-1} s_k(m)s_\ell(m+h) \\
&\quad + \sigma_{RI}\sum_{m=0}^{N-h-1} c_k(m)s_\ell(m+h) + \sigma_{RI}\sum_{m=0}^{N-h-1} s_k(m)c_\ell(m+h)\Bigg] \\
CCF_{Im(A_k),Im(A_\ell)}(h) &= \frac{1}{N}\Bigg[\sigma_{RR}\sum_{m=0}^{N-h-1} s_k(m)s_\ell(m+h) + \sigma_{II}\sum_{m=0}^{N-h-1} c_k(m)c_\ell(m+h) \\
&\quad -\sigma_{RI}\sum_{m=0}^{N-h-1} c_k(m)s_\ell(m+h) - \sigma_{RI}\sum_{m=0}^{N-h-1} s_k(m)c_\ell(m+h)\Bigg] \\
CCF_{Re(A_k),Im(A_\ell)}(h) &= \frac{1}{N}\Bigg[-\sigma_{RR}\sum_{m=0}^{N-h-1} c_k(m)s_\ell(m+h) + \sigma_{II}\sum_{m=0}^{N-h-1} s_k(m)c_\ell(m+h) \\
&\quad +\sigma_{RI}\sum_{m=0}^{N-h-1} c_k(m)c_\ell(m+h) - \sigma_{RI}\sum_{m=0}^{N-h-1} s_k(m)s_\ell(m+h)\Bigg] \\
CCF_{Im(A_k),Re(A_\ell)}(h) &= \frac{1}{N}\Bigg[-\sigma_{RR}\sum_{m=0}^{N-h-1} s_k(m)c_\ell(m+h) + \sigma_{II}\sum_{m=0}^{N-h-1} c_k(m)s_\ell(m+h) \\
&\quad +\sigma_{RI}\sum_{m=0}^{N-h-1} c_k(m)c_\ell(m+h) - \sigma_{RI}\sum_{m=0}^{N-h-1} s_k(m)s_\ell(m+h)\Bigg] \\
CCF_{|A_k|^2,|A_\ell|^2}(h) &= -(\sigma_{RR}+\sigma_{II})^2 + \frac{1}{N^2}\Bigg[(N-h)(\mu_{R^4}+\mu_{I^4}+2\mu_{R^2I^2}) \\
&\quad +(N^2-N+h)((\sigma_{RR})^2+(\sigma_{II})^2+2\sigma_{RR}\sigma_{II}) \\
&\quad +4(\sigma_{RI})^2\sum_{p=0}^{N-h-1}\sum_{q=0}^{N-h-1} s_k(p-q)s_\ell(q-p) \\
&\quad +4\mu_{R^2I^2}\sum_{p=0}^{N-h-1}\sum_{q=0}^{N-h-1} s_k(p-q)s_\ell(p-q) \\
&\quad +2((\sigma_{RR})^2+(\sigma_{II})^2+2(\sigma_{RI})^2)\sum_{p=0,p\neq q}^{N-h-1}\sum_{q=0}^{N-h-1} c_k(p-q)c_\ell(p-q)\Bigg].
\end{aligned}
$$

These hold for both $k \leq \ell$ and $k \geq \ell$ and for any Gaussian or non-Gaussian input, as long as the moments up to the fourth are the same. For a real-valued time series, we can let $\sigma_{II} = \sigma_{RI} = \mu_{I^3} = \mu_{R^2I} = \mu_{RI^2} = \mu_{R^3I} = \mu_{RI^3} = \mu_{R^2I^2} = \mu_{I^4} = 0$. We can also compute the spectral density of a univariate time series Y_t with $f(\nu) = \sum_{h=-\infty}^{\infty} CCF_{y,y}(h)e^{-i2\pi\nu h}$,

and the cross spectrum between two series Y_t and Z_t, $f_{xy}(\nu) = \sum_{h=-\infty}^{\infty} CCF_{y,z}(h)e^{-i2\pi\nu h}$ for $-0.5 \leq \nu \leq 0.5$.

If the input is Gaussian white noise, then the marginal distribution of the time series (1)-(3) is also Gaussian with mean zero and variance derived above with CCF with $h = 0$, and the marginal distribution of the squared modulus time series (4) $\{|A_k^t|^2\}_t$ (for any k) is the exponential with mean $E[|A_k^t|^2] = \sigma_{RR} + \sigma_{II}$.

3.2.1 The Bivariate Distribution of $|A_k^t|^2$ and $|A_k^{t+h}|^2$

Here we assume Gaussian input. Since the STFT is a linear transformation, we can easily find the bivariate probability density functions whose covariances are derived above, except for the bivariate distribution of $(|A_k^t|^2, |A_k^{t+h}|^2)$ with $0 < h < N$. As mentioned above, the marginal distribution is exponential, so we have a bivariate exponential (or gamma) distribution. However, unlike the bivariate Gaussian, just knowing the covariance does not determine the bivariate density, as demonstrated by Krishnaiah and Rao (1961). Their discussion helps us derive the characteristic function. We cannot express the bivariate density in a closed form.

We are interested in the bivariate distribution of $(Z_1, Z_2) = (Y_1^2 + Y_2^2, Y_3^2 + Y_4^2)$, with $Y_1 = \mathrm{Re}(A_k^t)$, $Y_2 = \mathrm{Im}(A_k^t)$, $Y_3 = \mathrm{Re}(A_k^{t+h})$, and $Y_4 = \mathrm{Im}(A_k^{t+h})$, where the Y's are distributed as 4-dimensional multivariate Gaussian with mean vector $\mathbf{0}$ and covariance $\mathbf{V}$ shown earlier.

$$W = Y^{\mathbf{T}}Y = \begin{bmatrix} Y_1^2 & Y_1Y_2 & Y_1Y_3 & Y_1Y_4 \\ Y_1Y_2 & Y_2^2 & Y_2Y_3 & Y_2Y_4 \\ Y_1Y_3 & Y_2Y_3 & Y_3^2 & Y_3Y_4 \\ Y_1Y_4 & Y_2Y_4 & Y_3Y_4 & Y_4^2 \end{bmatrix}$$

has a 4-by-4 Wishart distribution (a random matrix) and its characteristic function (by definition) is

$$\varphi_W(T) = \mathbf{E}\left[\exp\{i\mathbf{tr}(T^{\intercal}W)\}\right] = |I_4 - 2iTV|^{-\frac{1}{2}}$$

where T is a 4-by-4 real-valued matrix. The first equality holds by the definition of the characteristic function for a random matrix in general, including the Wishart distribution. The second equality derives from the definition and the probability density function of the Wishart distribution. By setting

$$\Theta = \begin{bmatrix} t_1 & 0 & 0 & 0 \\ 0 & t_1 & 0 & 0 \\ 0 & 0 & t_2 & 0 \\ 0 & 0 & 0 & t_2 \end{bmatrix},$$

$$\begin{aligned} \varphi_W(\Theta) &= |I_4 - 2i\Theta V|^{-\frac{1}{2}} \\ &= \mathbf{E}\left[\exp\{i\mathbf{tr}(\Theta^\intercal W)\}\right] \\ &= \mathbf{E}\left[\exp\{i(Y_1^2 t_1 + Y_2^2 t_1 + Y_3^2 t_2 + Y_4^2 t_2)\}\right] \\ &= \mathbf{E}\left[\exp\{i(Z_1 t_1 + Z_2 t_2)\}\right] \\ &= \varphi_{Z_1,Z_2}(t_1, t_2). \end{aligned}$$

The last equality holds by the definition of the characteristic function of a bivariate distribution. Thus we have found the characteristic function of the bivariate distribution (Z_1, Z_2): $\varphi_{Z_1,Z_2}(t_1, t_2) = |I_4 - 2i\Theta V|^{-\frac{1}{2}}$. By the inversion formula we can express the probability density function of (Z_1, Z_2), although not in a closed form:

$$f_{Z_1,Z_2}(y_1, y_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i(z_1 t_1 + z_2 t_2)} \varphi_{Z_1,Z_2}(t_1, t_2) dt_1 dt_2 \qquad 0 < z_1, z_2 < \infty.$$

3.3 An Example

Here we look at the STFT of a white noise time series to support the claim made at the end of Section 3.1 that the STFT phase time series $\{\text{angle}(A_k^t)\}_t$ is hard to work with.

The two plots on the top of Figure 3.1 show a complex-valued Gaussian white noise time series of length 50. The real part and imaginary part are shown on separate plots, both with the time index on the x-axis. The white noise input was generated from the bivariate Gaussian distribution with $E[\text{Re}(X_t)] = E[\text{Im}(X_t)] = 0$, $\text{Var}[\text{Re}(X_t)] = \text{Var}[\text{Im}(X_t)] = 1$, and $E[\text{Re}(X_t)\text{Im}(X_t)] = 0.5$.

The three plots on the bottom of Figure 3.1 are the squared modulus, the real and imaginary parts of the resulting STFT with window size $N = 10$ (and thus $k = 0, \ldots, 9$). As expected, it is hard to see any pattern in the STFT, except that neighboring values are often similar, both vertically (across frequency indices) and horizontally (across time indices).

Figure 3.2 shows two STFT phase time series $\{\text{angle}(A_2^t)\}_t$ and $\{\text{angle}(A_3^t)\}_t$ and their one- and two-step functions. The two time series are bounded on $[-\pi, \pi]$ and show "jumps" as we discussed earlier. Each time series appears to show an increasing trend over time. When one observation $\{\text{angle}(A_2^t)\}_t$ is near π, the next observation $\{\text{angle}(A_2^t)\}_t$ is either near π again, or near $-\pi$. In the latter case, we suspect a jump. But there is no way to confirm such suspicion in the complex plane.

The four scatter plots are 1) $\text{angle}(A_2^{t-1})$ against $\text{angle}(A_2^t)$, 2) $\text{angle}(A_2^{t-2})$ against $\text{angle}(A_2^t)$, 3) $\text{angle}(A_3^{t-1})$ against $\text{angle}(A_2^t)$, and 4) $\text{angle}(A_3^{t-1})$ against $\text{angle}(A_2^t)$. They clearly show nonlinearity, that is, we cannot approximate by linearly regressing $\text{angle}(A_2^t)$ on $\text{angle}(A_2^{t-1})$. Thus they indicate that the cross-covariance functions are not appropriate measures for the dependence of these nonlinear time series. Therefore in this thesis we will not study the STFT phase time series.

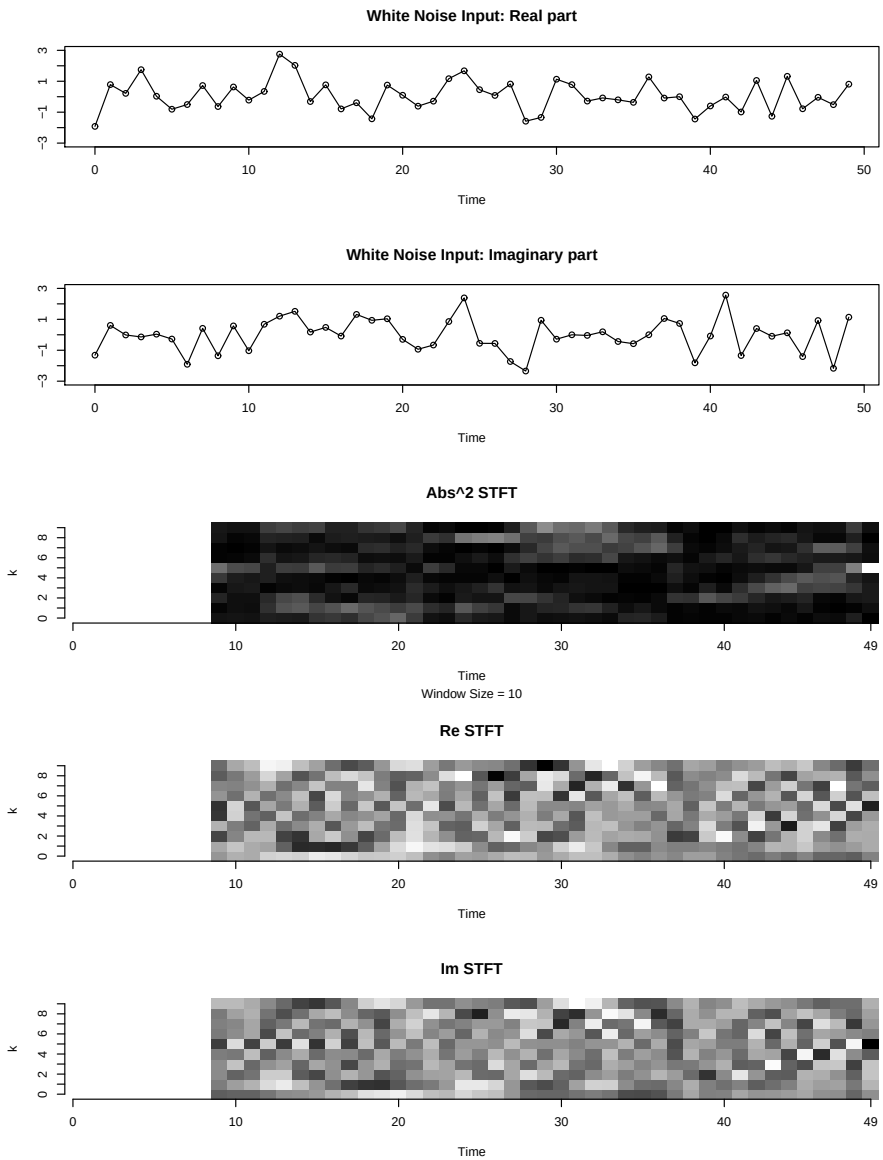

Figure 3.1: The top two plots show complex-valued Gaussian white noise time series input. The bottom three show the (complex-valued) STFT output of the input with window size 10. No visually obvious pattern exists, except neighboring points are often similar.

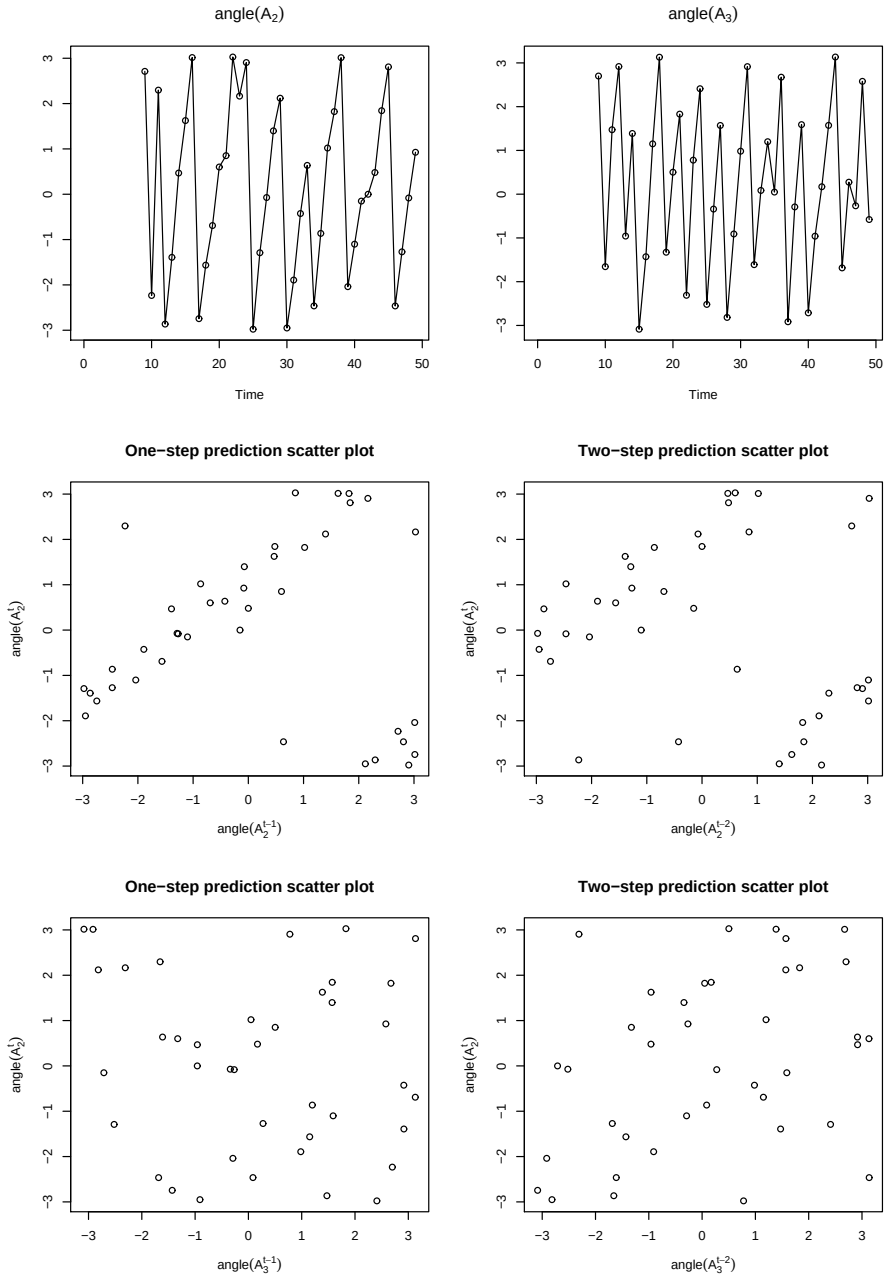

Figure 3.2: The top row shows the two time series of the STFT output angle(A_2^t) and angle(A_3^t) computed from the example in Figure 3.1. The scatter plots in the middle row are one- and two-step functions of the time series, angle(A_2^{t-1}) against angle(A_2^t) and angle(A_2^{t-2}) against angle(A_2^t), respectively. The last row shows similar scatter plots for angle(A_3^{t-1}) against angle(A_2^t) and angle(A_3^{t-2}) against angle(A_2^t). We see that the cross-covariance functions are not appropriate measures for the dependence of these nonlinear time series.

Chapter 4

STFT on a Global Signal

In this chapter, we examine the STFT resulting from data that have particular forms throughout time. They produce the STFTs in closed forms. In particular, periodic signals are going to be the focus in this thesis. We start with a simple periodic signal that does not result in a phenomenon called leakage and then consider more general and irregular inputs.

4.1 Periodic Signal

Suppose we have a complex-valued time series Y_t of length M with real-valued amplitudes A and B, the number of cycles L (not necessarily an integer) and real-valued phases ϕ_A and ϕ_B, where

$$Y_t = A\cos\left(\frac{2\pi Lt}{M} + \phi_A\right) + iB\cos\left(\frac{2\pi Lt}{M} + \phi_B\right) \text{ for } t = 0, 1, \ldots, M-1. \tag{4.1}$$

Let us call such signal with the same sinusoidal form throughout the time a *global signal*, (global in time) as opposed to a *local signal* (local in time) that will be seen in the next chapter. A real-valued input can be obtained simply by setting $B = 0$. Let k^* be the number of cycles within the STFT window. Suppose k^* is an integer for simplicity, thus

there will be no leakage caused by the STFT. We will shortly see what leakage means and consider cases when k^* is not an integer. When k^* is an integer, we have explicit forms of the resulting STFT, denoted by G_k^t instead of A_k^t (to be used in the next chapter), for $t = N-1, \ldots, M-1$,

$$
\begin{aligned}
G_{k^*}^t &= \frac{A\sqrt{N}}{2}\exp\left(i\left(\phi_A + \frac{2\pi k^*}{N}(t-N+1)\right)\right) + \frac{iB\sqrt{N}}{2}\exp\left(i\left(\phi_B + \frac{2\pi k^*}{N}(t-N+1)\right)\right) \\
&= \left[\frac{A\sqrt{N}}{2}\cos\left(\phi_A + \frac{2\pi k^*}{N}(t-N+1)\right) + \frac{-B\sqrt{N}}{2}\sin\left(\phi_B + \frac{2\pi k^*}{N}(t-N+1)\right)\right] \\
&\quad + i\left[\frac{A\sqrt{N}}{2}\sin\left(\phi_A + \frac{2\pi k^*}{N}(t-N+1)\right) + \frac{B\sqrt{N}}{2}\cos\left(\phi_B + \frac{2\pi k^*}{N}(t-N+1)\right)\right] \\
|G_{k^*}^t|^2 &= \frac{N(A^2+B^2)}{4} + \frac{ABN}{2}\sin(\phi_A - \phi_B) \\
G_{k^{**}}^t &= \left[\frac{A\sqrt{N}}{2}\cos\left(\phi_A + \frac{2\pi k^*}{N}(t-N+1)\right) + \frac{B\sqrt{N}}{2}\sin\left(\phi_B + \frac{2\pi k^*}{N}(t-N+1)\right)\right] \\
&\quad + i\left[\frac{-A\sqrt{N}}{2}\sin\left(\phi_A + \frac{2\pi k^*}{N}(t-N+1)\right) + \frac{B\sqrt{N}}{2}\cos\left(\phi_B + \frac{2\pi k^*}{N}(t-N+1)\right)\right] \\
|G_{k^{**}}^t|^2 &= \frac{N(A^2+B^2)}{4} + \frac{-ABN}{2}\sin(\phi_A - \phi_B).
\end{aligned}
$$

These results can be easily calculated simply by plugging in the signal Y_t in the STFT formula (2.4). G_k^t equals $\mathbf{0}$ for all t at any frequency index k other than k^* and $k^{**} = N - k^*$. Again, if k^* is not an integer, we will have a phenomenon called "leakage" caused by the STFT when there is a difference between the signal's frequency and the sampling frequency, which results in non-zero STFT at other k's than k^* and k^{**} (Cristi 2004), and we will see the resulting STFT later.

Sliding the window is the same as applying the DFT to the same signal with the phases ϕ_A and ϕ_B changing. Perhaps the two squared modulus time series staying constant as a function of time is intuitive because the squared modulus of the DFT measures the amplitude of the signal and ignores the phases. In contrast, the real-part and imaginary-part time series at each of k^* and k^{**} form sinusoidal signal outputs as the STFT window moves along.

4.2 General Signal With Fourier Representation

Now we consider an input signal of any form. By the inverse Fourier representation (2.2), any function satisfying the absolute summability condition (2.3), periodic or non-periodic, can be described as a linear combination of periodic functions ω's with different frequencies and complex-valued weight coefficients G's.

$$\text{Global Signal: } X_t = \frac{1}{\sqrt{M}} \sum_{a=0}^{M-1} G_a \omega_M^{at} \qquad \text{for } t = 0, 1, \ldots, M-1. \tag{4.2}$$

We can just plug X_t in the STFT formula (2.4) and look at the output. We will see that we can reduce the computation of the STFT of any input and that the STFT at time index t is:

$$A_k^t = \frac{1}{\sqrt{MN}} \sum_{\substack{a=1 \\ aN/M \notin \mathbb{N}}}^{M-1} G_a \omega_M^{a(t-N+1)} \frac{1-\omega_M^{aN}}{1-\omega_M^{a-kM/N}}. \tag{4.3}$$

We will examine the STFT at time index $t+N-1$ instead of at t, because of its computational and notational simplicity and because it is easy to change back to t.

$$A_k^{t+N-1} = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \omega_N^{-jk} X_{t+j} \tag{4.4}$$

$$= \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \omega_N^{-jk} \left[\frac{1}{\sqrt{M}} \sum_{a=0}^{M-1} G_a \omega_M^{a(t+j)} \right] \tag{4.5}$$

$$= \frac{1}{\sqrt{MN}} \sum_{j=0}^{N-1} \omega_N^{-jk} \sum_{a=0}^{M-1} G_a \omega_M^{a(t+j)} \tag{4.6}$$

$$= \frac{1}{\sqrt{MN}} \sum_{a=0}^{M-1} G_a \omega_M^{at} \sum_{j=0}^{N-1} \omega_N^{j(\frac{N}{M}a-k)} \tag{4.7}$$

$$\overset{(1)}{=} \frac{1}{\sqrt{MN}} \sum_{\substack{a=1 \\ aN/M \notin \mathbb{N}}}^{M-1} G_a \omega_M^{at} \sum_{j=0}^{N-1} \omega_N^{j(\frac{N}{M}a-k)} \tag{4.8}$$

$$\overset{(2)}{=} \frac{1}{\sqrt{MN}} \sum_{\substack{a=1 \\ aN/M \notin \mathbb{N}}}^{M-1} G_a \omega_M^{at} \frac{1-\omega_M^{aN}}{1-\omega_M^{a-kM/N}}. \tag{4.9}$$

(1) is true because when $a = 0$ and for a's such that $a \cdot \frac{N}{M}$ is an integer, the inner sum results in zero. And (2) holds because

$$\frac{aN}{M} \notin \mathbb{N} \quad \Rightarrow \quad \frac{1}{N}\left(\frac{aN}{M} - k\right) \notin \mathbb{N}$$

$$\Rightarrow \quad \sum_{j=0}^{N-1} \omega_N^{j(\frac{N}{M}a-k)} = \frac{1-\omega^{\frac{N}{M}a-k}}{1-\omega_N^{\frac{N}{M}a-k}} \overset{(3)}{=} \frac{1-\omega^{\frac{N}{M}a}}{1-\omega_N^{\frac{N}{M}a-k}} = \frac{1-\omega_M^{aN}}{1-\omega_M^{a-kM/N}},$$

where (3) is true because $\omega^{\frac{N}{M}a-k} = \omega^{\frac{N}{M}a-k}\omega^{-k} = \omega^{\frac{N}{M}a} \cdot 1 = \omega^{\frac{N}{M}a}$ for any $k \in \mathbb{N}$. Here we made use of the partial sum of a geometric series:

$$\sum_{j=0}^{N-1} \omega_N^{-jk} = \frac{1-\omega_N^{-kN}}{1-\omega_N^{-k}} = \frac{1-\omega^{-k}}{1-\omega_N^{-k}} \quad \text{(assuming } k/N \notin \mathbb{N}\text{)}. \tag{4.10}$$

Thus we established the equation (4.3). The equation (4.3) appears more complicated but we can reduce the computation compared to (4.5).

Of course, the computation of the STFT in (4.3) can be further reduced when many of the Fourier coefficients G's in (4.2) are equal to zero, as we will assume in section 4.3.1.

The previous section 4.1 considered cases where k^* is an integer, and we noted that we do not have leakage, meaning the STFT equal to zero at any frequency index k other than k^* and k^{**}, and we have the STFT at $k = k^*$ and k^{**} in a simple, closed form. The next section considers cases where k^* is not an integer and we have leakage.

4.3 Leakage With Periodic Signals

Section 4.1 considered an input of a global simple periodic signal where we had no leakage. There the STFT resulted in a closed-form periodic output at frequency indices at $k = k*$ and k^{**} and zeros at all other frequency indices. The assumption there was that the STFT window size N was chosen properly which would result in k^* complete cycles of the periodic signal. In this section we consider cases where such assumption is not satisfied.

Here we consider periodic signals (4.1) again, which, for the purpose of the Fourier representation (4.2), can be also written as

$$X_t = A\cos\left(\frac{2\pi Lt}{M} + \phi_A\right) + iB\cos\left(\frac{2\pi Lt}{M} + \phi_B\right) \qquad \text{for } t = 0, 1, \ldots, M-1 \tag{4.11}$$

$$= \frac{A}{2}\left[e^{i\phi_A}\omega_M^{Lt} + e^{-i\phi_A}\omega_M^{-Lt}\right] + \frac{iB}{2}\left[e^{i\phi_B}\omega_M^{Lt} + e^{-i\phi_B}\omega_M^{-Lt}\right] \tag{4.12}$$

$$= \left[\frac{A}{2}e^{i\phi_A} + \frac{iB}{2}e^{i\phi_B}\right]\omega_M^{Lt} + \left[\frac{A}{2}e^{-i\phi_A} + \frac{iB}{2}e^{-i\phi_B}\right]\omega_M^{-Lt} \tag{4.13}$$

But now we assume the window size N is chosen so that $\frac{LN}{M} \notin \mathbb{N}$, thus we have leakage. So we have non-zero STFT coefficients at *all* of the frequency indices, rather than only at $k = k^*$ and k^{**}.

In Section 4.3.1, we consider an input that has an integer number L of complete cycles in the whole input $\{X_t\}_{t=0}^{M-1}$ so that the signal's frequency matches with the sampling frequency, which results in a rather simple STFT output in a closed form even with leakage. In Section 4.3.2, we consider an input that has a non-integer number of complete cycles in the whole input $\{X_t\}_{t=0}^{M-1}$ so that the signal's frequency does not match with the sampling frequency, which still results in a simple STFT output in a closed form even with leakage.

4.3.1 An Integer Number Of Periods

When L is an integer in (4.11), that is, when the input has an integer number L of complete cycles in the whole input $\{X_t\}_{t=0}^{M-1}$, then we can represent the series with only two non-zero Fourier coefficients in (4.2). We can find the coefficients exactly:

$$\begin{aligned}
X_t &= \frac{1}{\sqrt{M}}\left[G_L\omega_M^{Lt} + G_{M-L}\omega_M^{(M-L)t}\right] \quad \text{for } t = 0, 1, \ldots, M-1 \\
&= \frac{1}{\sqrt{M}}\left[G_L\omega_M^{Lt} + G_{M-L}\omega_M^{-Lt}\right] \\
\text{Thus, } G_L &= \frac{A\sqrt{M}}{2}e^{i\phi_A} + \frac{iB\sqrt{M}}{2}e^{i\phi_B} \quad \text{and} \quad G_{M-L} = \frac{A\sqrt{M}}{2}e^{-i\phi_A} + \frac{iB\sqrt{M}}{2}e^{-i\phi_B}.
\end{aligned}$$

We found the two non-zero Fourier coefficients in a closed form. Now, starting with the STFT for a general signal (4.3), we plug in the above Fourier coefficients to find the STFT output:

$$\begin{aligned}
A_k^t &= \frac{1}{\sqrt{MN}} \sum_{\substack{a=1 \\ aN/M \notin \mathbb{N}}}^{M-1} G_a \omega_M^{a(t-N+1)} \frac{1-\omega_M^{aN}}{1-\omega_M^{a-kM/N}} \\
&= \frac{1}{\sqrt{MN}}\left[G_L\omega_M^{L(t-N+1)}\frac{1-\omega_M^{LN}}{1-\omega_M^{L-kM/N}} + G_{M-L}\omega_M^{(M-L)(t-N+1)}\frac{1-\omega_M^{(M-L)N}}{1-\omega_M^{(M-L)-kM/N}}\right] \\
&= \frac{1}{\sqrt{MN}}\left[G_L\omega_M^{L(t-N+1)}\frac{1-\omega_M^{LN}}{1-\omega_M^{L-kM/N}} + G_{M-L}\omega_M^{-L(t-N+1)}\frac{1-\omega_M^{-LN}}{1-\omega_M^{-L-kM/N}}\right] \\
&= \frac{Ae^{i\phi_A} + iBe^{i\phi_B}}{2\sqrt{N}}\omega_M^{L(t-N+1)}\frac{1-\omega_M^{LN}}{1-\omega_M^{-(\frac{kM}{N}-L)}} \\
&\quad + \frac{Ae^{-i\phi_A} + iBe^{-i\phi_B}}{2\sqrt{N}}\omega_M^{-L(t-N+1)}\frac{1-\omega_M^{-LN}}{1-\omega_M^{-(\frac{kM}{N}+L)}}.
\end{aligned} \tag{4.14}$$

This is non-zero at *all* frequency indices k's. We note that this is different from the case considered in Section 4.1 where we assumed no leakage and had the STFT equal to non-zero at only two frequency indices.

Thus, when the input periodic signal has an integer number of complete cycles, we can

find the STFT output in a simple closed form even when the window size N is chosen so that we have we have leakage, a more usual case when the signal frequency is unknown.

4.3.2 A Non-Integer Number Of Periods

When L is not an integer in (4.11), then there are more than two non-zero Fourier coefficients in (4.2), thus the Fourier representation (4.3) may not be efficient. However, we can still directly plug in and achieve the same result as that with L being an integer in (4.14). The result is the same, but the point is that ω_M^L and ω_M^{-L} do not belong to the Fourier frequencies because a's in G_a's are integers from 0 to $M-1$, and the inverse Fourier representation (4.2) does not produce the line (4.15) below but rather it can be derived in a direct way. We start with a real-valued input and then generalize the result to a complex-valued input.

When The Input Is A Real-Valued Periodic Function

Suppose the input is $\{X_t\}_{t=0}^{M-1}$ with length M of a periodic function with amplitude A, phase ϕ, frequency $\frac{L}{M}$ (L not necessarily an integer), and real-valued:

$$\begin{aligned} X_t &= A\cos\left(\frac{2\pi Lt}{M}+\phi\right) = \frac{A}{2}\left[e^{i(2\pi Lt/M+\phi)} + e^{-i(2\pi Lt/M+\phi)}\right] \\ &= \frac{A}{2}\left[e^{i\phi}\omega_M^{Lt} + e^{-i\phi}\omega_M^{-Lt}\right] \end{aligned} \tag{4.15}$$

This is of course the same as setting $B=0$ in (4.11). Then, by directly plugging in (4.15):

$$\begin{aligned} A_k^t &= \frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}\omega_N^{-jk}X_{j+t-N+1} = \frac{A}{2\sqrt{N}}\sum_{j=0}^{N-1}\omega_N^{-jk}\left[e^{i\phi}\omega_M^{L(j+t-N+1)} + e^{-i\phi}\omega_M^{-L(j+t-N+1)}\right] \\ &= \frac{A}{2\sqrt{N}}\left[e^{i\phi}\omega_M^{L(t-N+1)}\sum_{j=0}^{N-1}\omega_N^{-j(k-\frac{LN}{M})} + e^{-i\phi}\omega_M^{-L(t-N+1)}\sum_{j=0}^{N-1}\omega_N^{-j(k+\frac{LN}{M})}\right] \\ &= \frac{A}{2\sqrt{N}}\left[e^{i\phi}\omega_M^{L(t-N+1)}\frac{1-\omega^{-(k-\frac{LN}{M})}}{1-\omega_N^{-(k-\frac{LN}{M})}} + e^{-i\phi}\omega_M^{-L(t-N+1)}\frac{1-\omega^{-(k+\frac{LN}{M})}}{1-\omega_N^{-(k+\frac{LN}{M})}}\right], \end{aligned} \tag{4.16}$$

where the last equality uses the partial sum of a geometric series (4.10).

When The Input Is A Complex-Valued Periodic Function

Generalizing the input to the complex-valued case is straightforward with the use of the STFT's linearity property. That is, the STFT of $\{c_1 Y_t + c_2 Z_t\}_t$ for $c_1, c_2 \in \mathbb{C}$ is the same as the STFT of $\{Y_t\}_t$ multiplied by c_1 plus the STFT of $\{Z_t\}_t$ multiplied by c_2. This is a trivial consequence from the discrete Fourier transform. Now, given

$$X_t = A\cos\left(\frac{2\pi Lt}{M} + \phi_A\right) + iB\cos\left(\frac{2\pi Lt}{M} + \phi_B\right), \quad \text{(the same as before(4.11))}$$

$$\begin{aligned}
A_k^t &= \frac{A}{2\sqrt{N}}\left[e^{i\phi_A}\omega_M^{L(t-N+1)}\frac{1-\omega^{-(k-\frac{LN}{M})}}{1-\omega_N^{-(k-\frac{LN}{M})}} + e^{-i\phi_A}\omega_M^{-L(t-N+1)}\frac{1-\omega^{-(k+\frac{LN}{M})}}{1-\omega_N^{-(k+\frac{LN}{M})}}\right] \\
&\quad + i\frac{B}{2\sqrt{N}}\left[e^{i\phi_B}\omega_M^{L(t-N+1)}\frac{1-\omega^{-(k-\frac{LN}{M})}}{1-\omega_N^{-(k-\frac{LN}{M})}} + e^{-i\phi_B}\omega_M^{-L(t-N+1)}\frac{1-\omega^{-(k+\frac{LN}{M})}}{1-\omega_N^{-(k+\frac{LN}{M})}}\right] \\
&= \frac{Ae^{i\phi_A} + iBe^{i\phi_B}}{2\sqrt{N}}\omega_M^{L(t-N+1)}\frac{1-\omega^{-(k-\frac{LN}{M})}}{1-\omega_N^{-(k-\frac{LN}{M})}} \\
&\quad + \frac{Ae^{-i\phi_A} + iBe^{-i\phi_B}}{2\sqrt{N}}\omega_M^{-L(t-N+1)}\frac{1-\omega^{-(k+\frac{LN}{M})}}{1-\omega_N^{-(k+\frac{LN}{M})}}
\end{aligned} \tag{4.17}$$

Thus, we obtain (4.17), the same result as (4.14). We achieved (4.14) for L an integer and (4.17) for L not an integer. The end results are the same, but the forms that the input series takes are different between the two cases. For each case, there exists an STFT output in a fairly simple, closed form.

4.4 Kronecker Delta Function

For the rest of Chapter 4, we consider a more general input than periodic signals. Those are best handled by plugging in the input in the STFT definition (2.4), rather than by the use of the general representation by the inverse Fourier transform (4.3) as in the previous

sections.

Suppose we have an input series $\{X_t\}_{t=0}^{M-1}$ that is a Kronecker delta function:

$$X_t = \begin{cases} C, & \text{if } t = d \\ 0, & \text{otherwise} \end{cases}$$

where $C \in \mathbb{C}$.

(1) When the window does not include d, i.e., $t < d$ or $t > d + N - 1$, then the resulting STFT is zero for all k and for all t.

(2) When the window includes d, i.e., $d \leq t \leq d + N - 1$, then

$$\begin{aligned} A_k^t &= \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} X_{j+t-N+1} \omega_N^{-jk} \\ &= \frac{C}{\sqrt{N}} \omega_N^{-k(d-t+N-1)} \end{aligned}$$

The result is obtained just by plugging the input in the STFT formula (2.4). Especially, at $t = d + N - 1$,

$$A_k^{d+N-1} = \frac{C}{\sqrt{N}} \ \forall k.$$

4.5 Step Function And Ringing

Suppose we have an input series $\{X_t\}_{t=0}^{M-1}$ that is a step function:

$$X_t = \begin{cases} C, & \text{if } t \geq d \\ 0, & \text{otherwise} \end{cases}$$

where $C \in \mathbb{C}$.

(1) When the window is placed before d, i.e., $t < d$, then the resulting STFT is zero for

all k and for all t.

(2) When the window is placed after d, i.e., $t > d + N - 1$, then the resulting STFT is

$$A_k^t = \begin{cases} \frac{1}{\sqrt{N}} C \sum_{j=0}^{N-1} \omega_N^{-jk} = C\sqrt{N}, & \text{for } k = 0 \\ 0, & \text{otherwise.} \end{cases}$$

(3) Now, when the window covers d, i.e., $d \leq t \leq d + N - 1$, then

$$\begin{aligned} A_k^t &= \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} X_{j+t-N+1} \omega_N^{-jk} &= \frac{1}{\sqrt{N}} C \sum_{j=d-t+N-1}^{N-1} \omega_N^{-jk} \\ &= \frac{C}{\sqrt{N}} \sum_{j=0}^{t-d} \omega_N^{-k(j+d-t+N-1)} &= \frac{C}{\sqrt{N}} \omega_N^{-k(d-t+N-1)} \sum_{j=0}^{t-d} \omega_N^{-jk} \end{aligned}$$

So,

$$A_k^t = \begin{cases} \frac{C\omega_N^{-k(d-t+N-1)}}{\sqrt{N}} (t - d + 1) & \text{for } k = 0, \\ \frac{C\omega_N^{-k(d-t+N-1)}}{\sqrt{N}} \frac{1 - \omega_N^{-k(t-d+1)}}{1 - \omega_N^{-k}} & \text{otherwise.} \end{cases}$$

Again, the last equality uses the partial sum of a geometric series (4.10). Unlike (2) where the window was placed after d and we had the STFT resulting in a non-zero value at $k = 0$ only, this time the STFT coefficients are non-zero at all frequency indices. In general, the Fourier transform, which describes the input series as a linear combination of continuous functions, is not suitable for representing a discontinuous function like this step function. This phenomenon, which occurs when the DFT (or the STFT) is applied to a discontinuous input with a sudden change (the input does not have to be a constant function but can be anything) is known as "ringing" (Percival and Walden 1993). This is perhaps best explained by an example, as we will see in Section 5.2.

Chapter 5

STFT on a Simple Local Signal

Chapter 4 considered a variety of input signals $\{X_t\}_{t=0}^{M-1}$ that maintain the same form throughout the time index from 0 to $M-1$ without discontinuities, as in (4.1). In this chapter, we start imposing an assumption of discontinuities on the input, that is, we assume local signals. We will look at the STFT output of such local signals. and see that closed-form outputs exist under very limited conditions. We will present a simple example to illustrate general points. In subsequent chapters, we will use the STFT to detect the existence of such local signal.

5.1 Periodic Signal

Unlike global signals (4.1), local signals appear only at parts of the data between the starting point S and the ending point E $(0 \leq S < E \leq M-1)$. We assume the signal value to be zero where the signal is not present. Such a signal can be obtained by applying an indicator function to the above global signal Y_t (4.1) for $t = 0, 1, \ldots, M-1$:

$$X_t = I_{(S \leq t \leq E)} \cdot Y_t = I_{(S \leq t \leq E)} \cdot \left[A \cos\left(\frac{2\pi K t}{M} + \phi_A \right) + iB \cos\left(\frac{2\pi K t}{M} + \phi_B \right) \right]. \tag{5.1}$$

This representation of a simple local function X_t has seven parameters; amplitudes A and B, the number of cycles K, phases ϕ_A and ϕ_B, starting point S, and ending point E.

When applied to the zero-valued region at the beginning, the resulting STFT is zero at all the k's for $t = N-1, \ldots, S-1$, and also at the ending; for $t = E+N, \ldots, M-1$.

When the STFT window is only on the local signal, for $S^* \leq t \leq E$, where $S^* = S+N-1$, the STFT A_k^t is exactly equal to the STFT G_k^t applied to the global signal Y_t; sinusoidal signal outputs at $k = k^*$ and k^{**}, and zero at any other k.

Now, the STFT gets more complicated when the window covers both the zero-valued region at the beginning and the local signal; for $S \leq t < S^*$. The STFT results in non-zero at other k's as well. This phenomenon is called "ringing" (Percival and Walden 1993). Ringing occurs when the DFT is applied to a region with discontinuity, which in this particular case is the change from the zero constant to the periodic function. For any k and $1 \leq d \leq N-1$,

$$A_k^{S^*-d} = G_k^{S^*-d} - \frac{1}{\sqrt{N}} \sum_{j=0}^{d-1} Y_{S-d+j} \omega_N^{-jk}. \tag{5.2}$$

Similarly, when the window covers both the end of the local signal and the following zero-valued region (for $E < t \leq E+N-1$),

$$A_k^{E+d} = G_k^{E+d} - \frac{1}{\sqrt{N}} \sum_{j=0}^{d-1} Y_{E+d-j} \omega_N^{-(N-1-j)k}. \tag{5.3}$$

When N/k^* is an integer, simpler expressions exist at k^* (and k^{**}) and various time points. For $j = 0, 1, \ldots, k^*$,

$$A_{k^*}^{S^*-(N/k^*)j} = \left(1 - \frac{j}{k^*}\right) A_{k^*}^{S^*} = \left(1 - \frac{j}{k^*}\right) G_{k^*}^{S^*} \tag{5.4}$$

$$A_{k^*}^{E+(N/k^*)j} = \left(1 - \frac{j}{k^*}\right) A_{k^*}^{E} = \left(1 - \frac{j}{k^*}\right) G_{k^*}^{E}. \tag{5.5}$$

This shows that the STFT coefficients are proportional to the fraction of the window that

covers the signal and the zero-valued region at certain time points. In general, the more signal the window covers, the larger the STFT coefficients are in terms of absolute values and the closer they are to the STFT coefficients obtained when the window covers the signal only.

5.2 An Example

We provide a simple example to illustrate how the STFT works on a local signal. The time series $\{X_t\}_{t=0}^{49}$ in Figure 5.1 consists of zeros at the beginning and at the end, and a cosine function of length 20 with periodicity 4 and amplitude 5 in the middle from $t = 15$ to $t = 34$. This local signal can be described with the representation in Section 4.1 with $M = 50$, $A = 5$, $B = 0$, $K = 10$, $\phi_A = \phi_B = 0$, $S = 15$, and $E = 34$. We use a window size $N = 10$ that results in $k = 0, \ldots, 9$. The k that matches the signal's frequency is $k^* = 2$ (and thus $k^{**} = 8$). We examine the STFT in three paragraphs below.

(1) When the STFT window is on the zero valued region at the beginning and at the end (for $t = 9, \ldots, 14, 44, \ldots, 49$), the complex-valued STFT is zero at all the k's.

(2) When the STFT window is only on the cosine function (for $t = 24, \ldots, 34$), the STFT behaves exactly the same way as it does for a global signal: the squared modulus STFT is constant at $k = 2, 8$ and zero at all other k's over the region $|A^t|^2 = (0, 0, 62.5, 0, 0, 0, 0, 0, 62.5, 0)$, and the real and imaginary STFT produce sinusoidal signal outputs at $k = 2$ and 8, and equal to zero at all other k's.

(3) When the STFT window is on both a zero-valued region and the local cosine function (for $t = 15, \ldots, 23, 35, \ldots, 43$), we have $0 \leq |A_2^t| \leq 62.5$, and the more the window covers the cosine function, the higher the squared modulus is. As given by the closed-form expressions in (5.4) and (5.5), $A_2^{19} = 3.952847 - 0i = A_2^{24}/2 = (7.905694 - 0i)/2$ and $A_2^{39} = 3.952847 - 0i = A_2^{34}/2 = (7.905694 - 0i)/2$. In general, the expression for the resulting STFT in these regions does not simplify because of the ringing phenomenon.

The important observation is that (2) when the window is only on the cosine function, then the squared modulus time series $\{|A_2^t|^2\}_t$ and $\{|A_8^t|^2\}_t$ take large, positive values, as investigated in Section 4.1. This remark, in turn, can be used to indicate the existence of a local periodic signal. Thus we suspect the existence of a local periodic signal if we observe the squared modulus time series taking large positive values. This is exactly why we are primarily interested in the STFT's squared modulus time series. Chapter 6 presents a preliminary analysis, and Chapter 7 shows more formal procedures to recognize local signals.

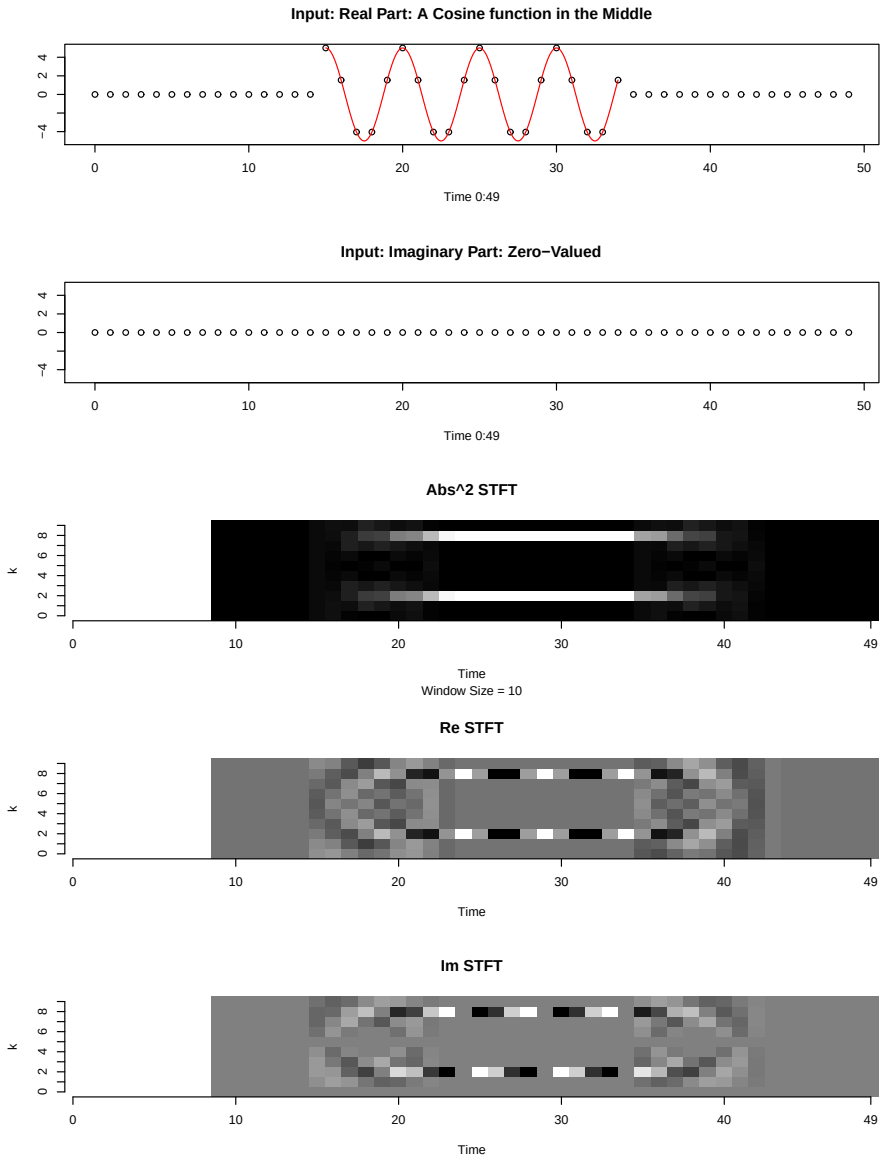

Figure 5.1: A simple example: The top two plots are the complex-valued input, which has a cosine function in the middle in the real part and is zero-valued in the imaginary part. The bottom three plots show the (complex-valued) STFT output: squared modulus, real and imaginary parts.

Chapter 6

Detection By Marginal Distribution

So far in this thesis we have discussed the short time Fourier transform and its output when we apply the STFT to various forms of inputs. We have also considered a simple local signal in the previous chapter. For the rest of this thesis, we will consider ways to detect a local signal, thus reversing the perspective we have taken. In this chapter we focus on the marginal distribution of STFT and ignore the time dependency structure. This chapter serves the role of exploratory and preliminary data analysis. In the next chapter we will present methods that take advantage of the time dependency structure of the STFT output as we considered in Chapter 3.

6.1 Data of a Local Signal With Noise

In order to illustrate exploratory and preliminary data analysis for detecting a local signal, we first present a simulated data set. The complex-valued input data $\{X_t\}_{t=0}^{499}$ in Figure 6.1 was generated by adding a global signal and a local signal. The global signal is complex-valued white noise time series with $\mathrm{Var}[\mathrm{Re}(X_t)] = \mathrm{Var}[\mathrm{Im}(X_t)] = 1$ and $\mathrm{Cov}[\mathrm{Re}(X_t),\mathrm{Im}(X_t)]$ $= 0.5$ The local signal is a real-valued periodic signal with amplitude $A = 2$ and has exactly 20 complete cycles from $t = 101, \ldots, 200$. This local signal can be represented by (4.1) with

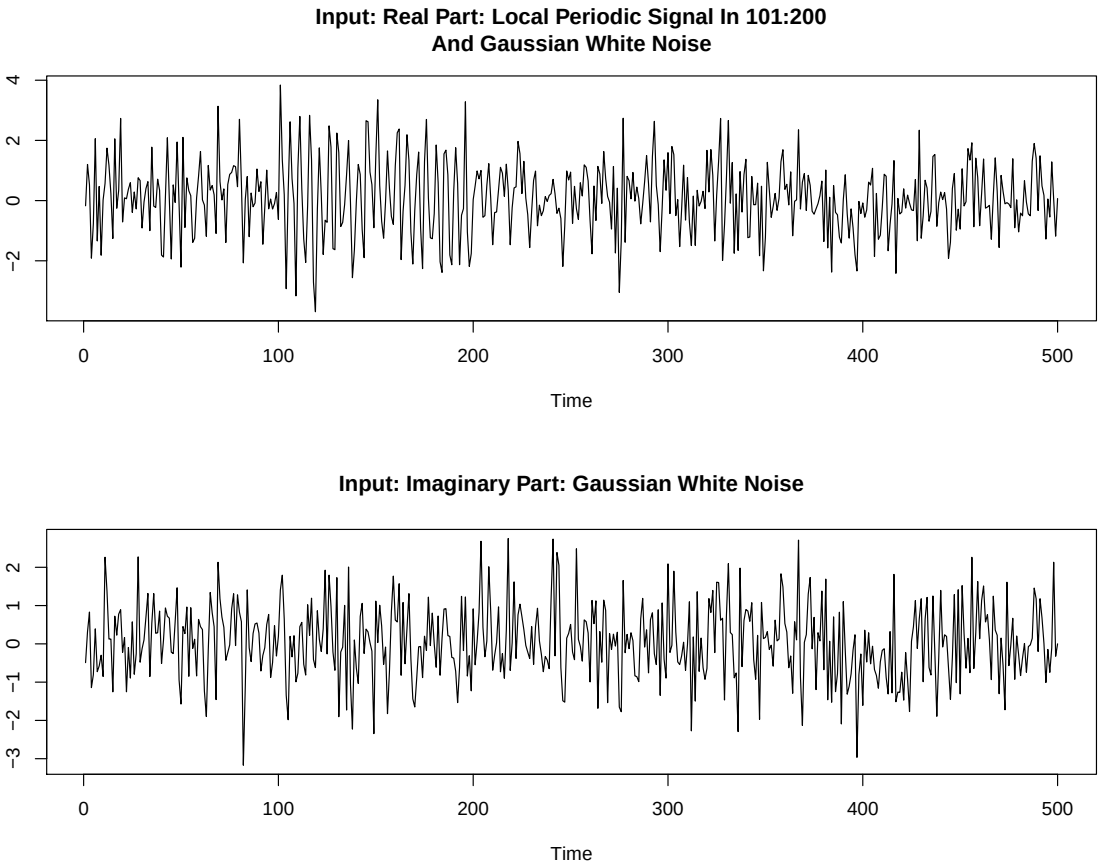

Figure 6.1: The input is a complex-valued Gaussian white noise plus a real-valued periodic local signal. The top plot shows the real part and the bottom plot shows the imaginary part of the time series input. We will consider ways to detect this local signal in this chapter and next.

$A = 2$, $B = 0$, $L = 100$, $\phi_A = \phi_B = 0$, $S = 101$, and $E = 200$. We will use this input series for this chapter and next in order to illustrate our analysis to detect a local periodic signal.

Figure 6.2 is the squared modulus STFT $\{|A_k^t|^2\}_t$ applied to the input with window size $N = 10$. The frequency indices $k = 2$ and 8 correspond to the frequency of the local periodic signal and they take large positive values when the window is near or covering the local signal, as described in Chapter 5.

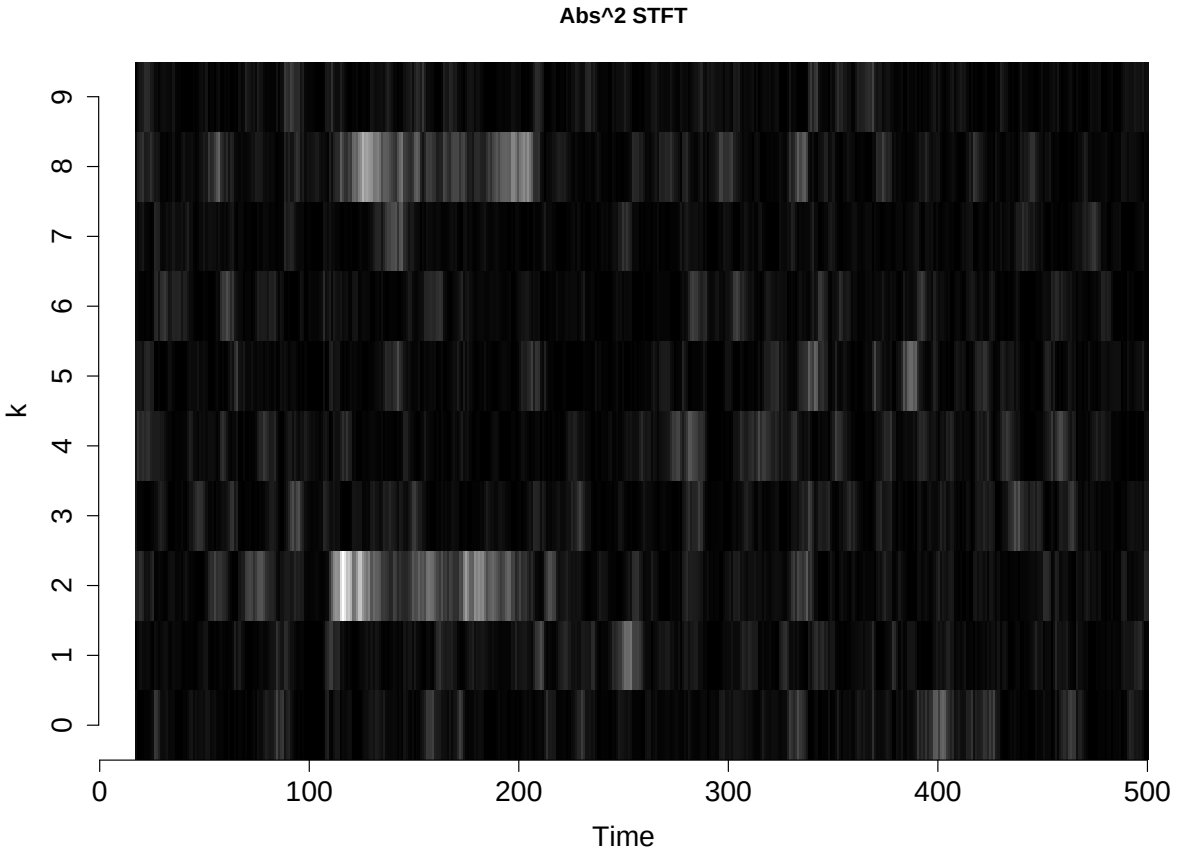

Figure 6.2: The squared modulus STFT output resulting from the input in Figure 6.1. The large values of $k = 2$ and 8 indicate the local signal.

6.2 Sample Quantiles

One simple way to detect a periodic local signal is to look at the the (squared) modulus of the STFT, which would show large values where such local signal exists. The squared modulus STFT at each frequency index is distributed as exponential when the input series is Gaussian white noise, as we saw in Section 3.2.1. Including local periodic signals would change this distribution with large positive values. Since we are interested in large values, we take natural logarithm. Figure 6.3 shows the log of these squared modulus STFT distributions. We notice large values occur at $k = 2$ and 8 (exceeding 3, which does not happen at other frequency indicies), and they are much larger than large values in other frequency indices. Thus, we would suspect the existence of a local signal at STFT frequency index 2.

Comparing distributions is perhaps much easier with the sample quantile plot. Figure 6.4 shows the sample cumulative distribution functions of the squared modulus STFT, while Figure 6.3 shows their histograms. In the sample quantile plot Figure 6.4 we can observe clearly that $k = 2$ (dashed line) and $k = 8$ (dotted line) have distributions different from the others (solid lines). The difference is caused because these two have larger positive values than others.

In this section we examined the distributions of the squared modulus STFT in terms of their marginal distribution and ignored the time dependency structure which we derived in Chapter 2. This approach may be too simple, but it is an important step in exploratory data analysis.

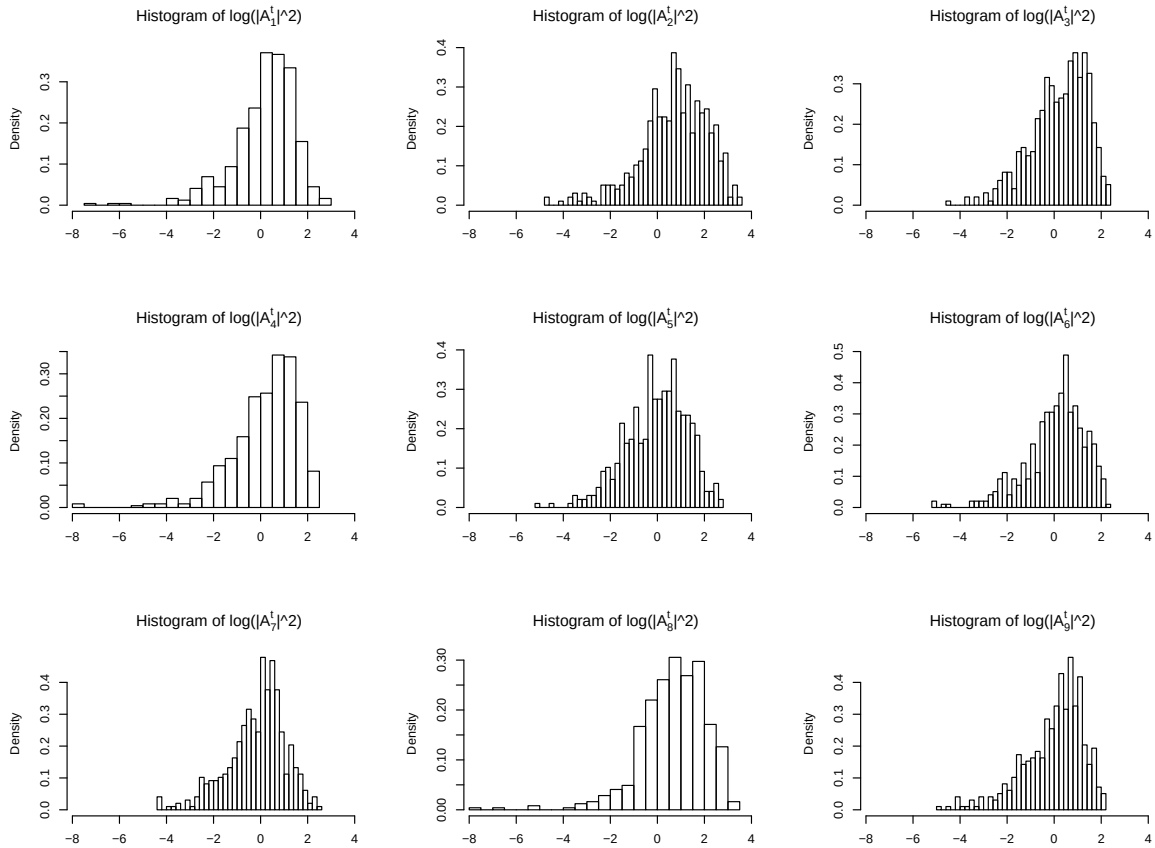

Figure 6.3: The histograms of natural logarithm of the squared modulus STFT in Figure 6.2 for $k = 1, \ldots, 9$. We notive that the values larger than 3 occur at $k = 2$ and 8, which does not happen at other frequency indices, thus indicating the existence of a local periodic signal.

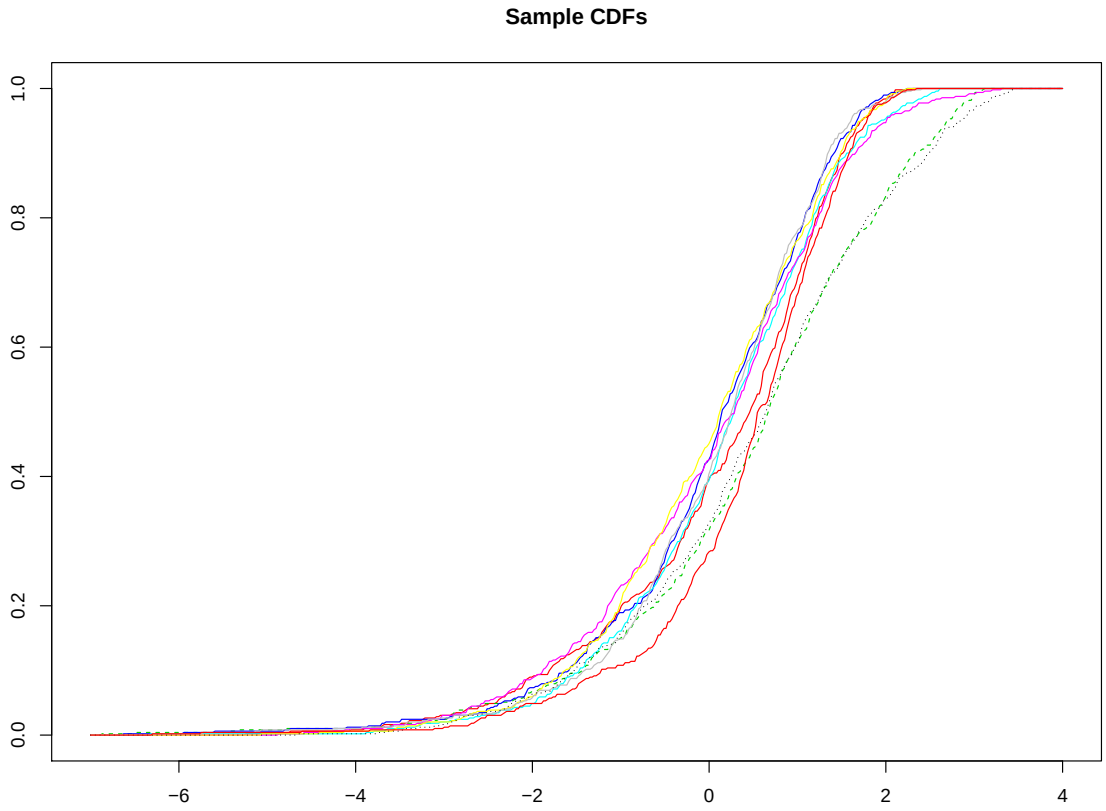

Figure 6.4: The sample quantile of log of the squared modulus STFT in Figure 6.2. Clearly, two frequency indices $k = 2$ (dashed line) and 8 (dotted line) have distributions different from the others (solid lines), indicating the existence of a local periodic signal.

6.3 Marginal Threshold

In the previous section, we considered exploratory and preliminary analysis of the squared modulus STFT to see which frequency indices have large positive observations and different distributions, which can suggest the existence of local periodic signals. Once we determine which frequency indices exhibit large observations, we can look deeper into at which time range they occur. Figure 6.5 shows the time series plots of $\{\log(|A_2^t|^2)\}_t$ (dashed line) and $\{\log(|A_8^t|^2)\}_t$ (dotted line). We look at the log of the squared modulus because we are concerned with large positive values. We notice that both series have spikes on multiple time ranges, and that the spike that lasts over the longest period of time is around the time range from 101 to 200, where the local periodic signal exists.

As we mentioned earlier, if we assume Gaussian input, squared modulus STFT is distributed as exponential with mean $\sigma_{RR}+\sigma_{II}$. Suppose for now that we know the parameters σ_{RR} and σ_{II}. Then, we can easily find out the 99 percentile of such distribution and set it as threshold. This threshold that we derive from the marginal distribution can be used to compare the time series with. Natural log of this threshold is the green horizontal line in Figure 6.5. As expected, both time series exceed the threshold around the time range from 101 to 200.

This approach of constructing a threshold from the marginal distribution is simple and intuitive, but it ignores the fact that the short time Fourier transform forms a stationary time series under the assumption of white noise input. In Chapter 3, we derived the theoretical properties such as the autocovariance functions $E\{[|A_k^{t+h}|^2 - E(|A_k^{t+h}|^2)][|A_k^t|^2 - E(|A_k^t|^2)]\}$. In the next chapter, we will present methods to detect local signals that incorporate the time dependency structure of the STFT.

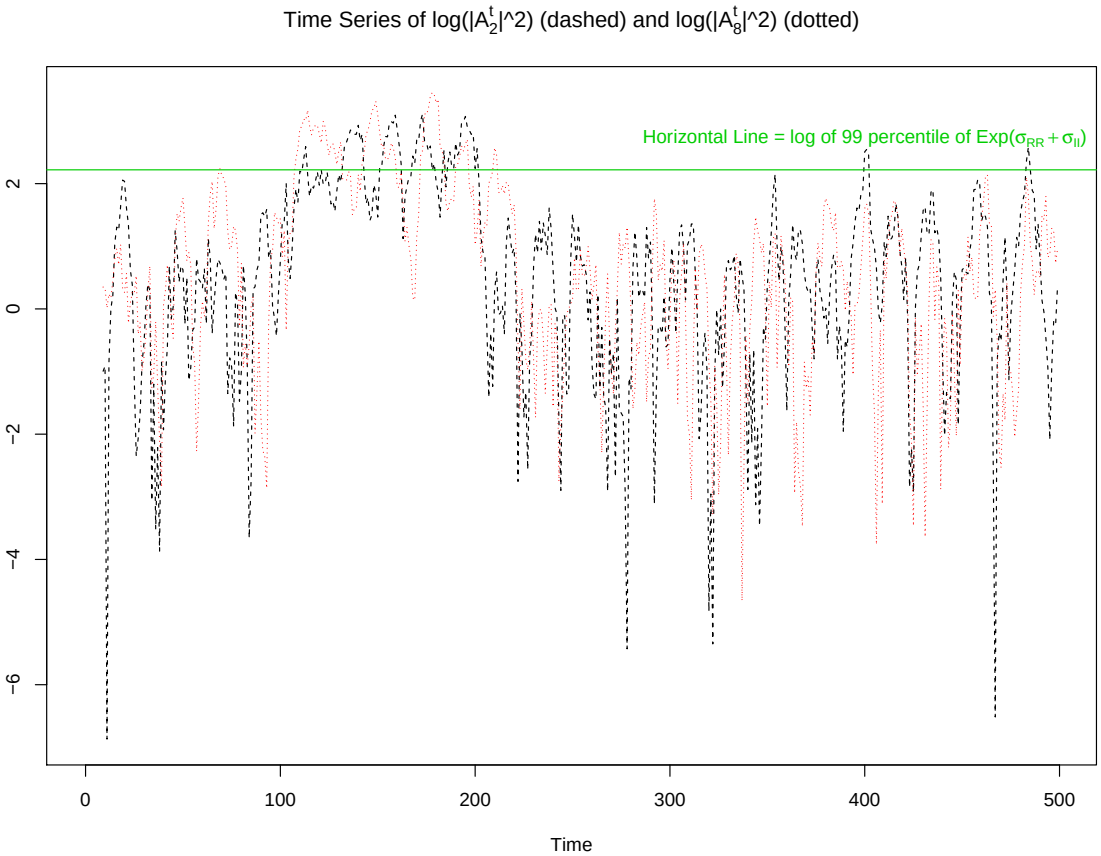

Figure 6.5: The time series of natural logarithm of the squared modulus for $k = 2$ (dashed line) and $k = 8$ (dotted line), along with the log of $\text{Exp}(\sigma_{RR} + \sigma_{II})$. We observe large values where the local signal exists.

Chapter 7

Detecting Local Signals By Considering the Time Dependency Structure Of the STFT Output Time Series

In the previous chapter, we presented exploratory and preliminary analysis methods to detect local periodic signals based on the squared modulus of the short time Fourier transform. That is because their large positive values indicate the existence of local signals. There we examined which STFT frequency indices have different marginal distributions from the others, and constructed a threshold from the marginal distribution to judge how large these values should be in order to conclude the existence of local signals. However, those methods ignore the fact that the STFT forms a stationary time series under the assumption of white noise input. In this chapter, we consider ways to detect local signals by taking advantage of the time dependency structure we studied in Chapter 3, using the same data from the previous chapter. In Section 7.1 we look at the STFT as a moving average process and consider using the large residuals as a sign of the existence of local signals. In Section 7.2 we

consider unusual lengths of consecutive large values that exceed some threshold and use them to detect local signals, and also examine approximation methods to reduce computation.

7.1 By Using One-Step Prediction With A Bivariate MA Process and Identifying Large Residuals

In Chapter 3 we saw that under the Gaussian white noise input assumption, for each k, the complex-valued time series $\{A_k^t\}_t$ is a complex-valued moving average process with order $N-1$. Now instead of treating this as a complex-valued univariate process, we can construct a real-valued bivariate moving average process $\{(\mathrm{Re}(A_k^t), \mathrm{Im}(A_k^t))\}_t$. That is, we look at the real part of the STFT and the imaginary part of the STFT individually, and since both are now real-valued, we can see them as a real-valued bivariate process. From the definition of the STFT (2.4), this can be written as

$$\begin{bmatrix} \mathrm{Re}(A_k^t) \\ \mathrm{Im}(A_k^t) \end{bmatrix} = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \begin{bmatrix} \cos\left(\frac{2\pi kj}{N}\right) & \sin\left(\frac{2\pi kj}{N}\right) \\ -\sin\left(\frac{2\pi kj}{N}\right) & \cos\left(\frac{2\pi kj}{N}\right) \end{bmatrix} \begin{bmatrix} \mathrm{Re}(X_{t-N+1+j}) \\ \mathrm{Im}(X_{t-N+1+j}) \end{bmatrix}, \quad k = 0, \ldots N-1. \tag{7.1}$$

Since this is a stationary series we can construct the one-step predictor of $(\mathrm{Re}(A_k^t), \mathrm{Im}(A_k^t))$ based on the past values $(\mathrm{Re}(A_k^{t-1}), \mathrm{Im}(A_k^{t-1})), (\mathrm{Re}(A_k^{t-2}), \mathrm{Im}(A_k^{t-2})), \ldots)$. And the one with the minimum mean squared error (Reinsel 1997) is

$$\begin{bmatrix} \widehat{\mathrm{Re}(A_k^t)} \\ \widehat{\mathrm{Im}(A_k^t)} \end{bmatrix} = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-2} \begin{bmatrix} \cos\left(\frac{2\pi kj}{N}\right) & \sin\left(\frac{2\pi kj}{N}\right) \\ -\sin\left(\frac{2\pi kj}{N}\right) & \cos\left(\frac{2\pi kj}{N}\right) \end{bmatrix} \begin{bmatrix} \mathrm{Re}(X_{t-N+1+j}) \\ \mathrm{Im}(X_{t-N+1+j}) \end{bmatrix}. \tag{7.2}$$

Under the Gaussian white noise input assumption, the residuals

$$\begin{bmatrix} \widehat{E_R^t} \\ \widehat{E_I^t} \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(A_k^t) \\ \mathrm{Im}(A_k^t) \end{bmatrix} - \begin{bmatrix} \widehat{\mathrm{Re}(A_k^t)} \\ \widehat{\mathrm{Im}(A_k^t)} \end{bmatrix} \tag{7.3}$$

are distributed as the bivariate Gaussian with mean vector $\mathbf{0}$ and covariance matrix

$$\mathbf{V}_E = \frac{1}{N}\mathbf{\Phi}^\intercal \mathbf{\Sigma} \mathbf{\Phi}, \tag{7.4}$$

where

$$\mathbf{\Phi} = \begin{bmatrix} \cos\left(\frac{2\pi k(N-1)}{N}\right) & -\sin\left(\frac{2\pi k(N-1)}{N}\right) \\ \sin\left(\frac{2\pi k(N-1)}{N}\right) & \cos\left(\frac{2\pi k(N-1)}{N}\right) \end{bmatrix} \qquad \text{and}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathrm{Var}[\mathrm{Re}(X_t)] & \mathrm{E}[\mathrm{Re}(X_t)\mathrm{Im}(X_t)] \\ \mathrm{E}[\mathrm{Re}(X_t)\mathrm{Im}(X_t)] & \mathrm{Var}[\mathrm{Im}(X_t)] \end{bmatrix}.$$

Now that the residuals are distributed as the bivariate Gaussian, a natural way to identify outlying observations is to compute the Mahalanobis distance:

$$M_t = \left(\begin{bmatrix} \widehat{E_R^t} \\ \widehat{E_I^t} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)^\intercal \mathbf{V}_E^{-1} \left(\begin{bmatrix} \widehat{E_R^t} \\ \widehat{E_I^t} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right). \tag{7.5}$$

This univariate observations are distributed as $\chi^2_{df=2}$. So the large positive values indicate outliers.

These residuals' Mahalanobis distance series are plotted in Figure 7.1 for $k = 2$ (top) and $k = 8$ (middle). These plots also show the 99 percentile of the $\chi^2_{df=2}$ distribution in the horizontal line. We do observe large positive values where the local signal exists. The bottom plot in Figure 7.1 is their scatterplot which shows that they are practically the same with correlation 0.93.

It was hoped that the Mahalanobis distance series M_t have large values where the local signal exists. However, there are many small values between large values in such range. It does not appear helpful to use the Mahalanobis distance series of the residuals from constructing a real-valued bivariate moving average process.

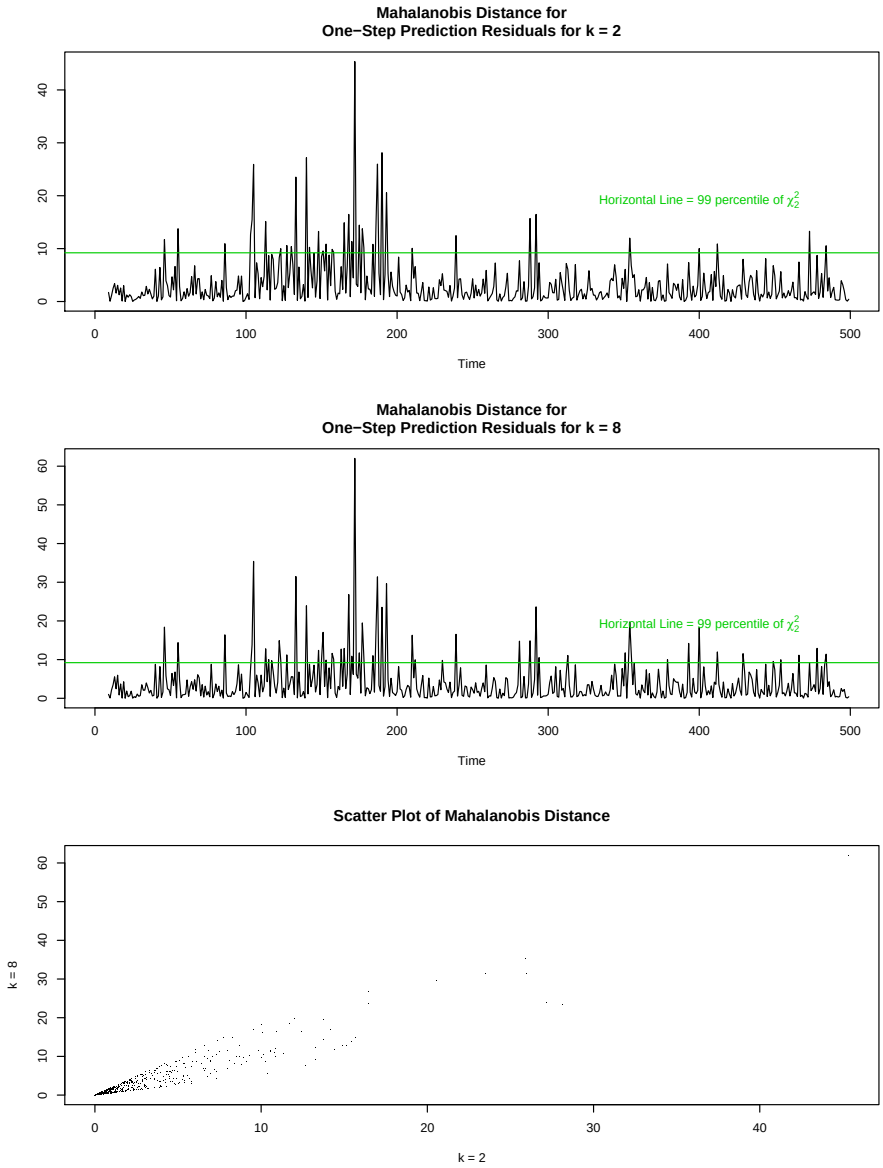

Figure 7.1: The top two plots show the time series of Mahalanobis distance of the residuals computed from the one-step prediction function of the bivariate moving average process, along with 99 percentile of $\chi^2_{df=2}$. They show many small values between large values and thus are not helpful for finding local periodic signals. The bottom plot is their scatterplot, which says that the two time series are almost identical.

7.2 By Considering the Probability Of Observing Consecutive Large Values Exceeding A Threshold

Here we will explore a slightly different approach to detect local signals. For each k, the STFT time series 1) complex-valued $\{A_k^t\}_t$, 2) $\{\mathrm{Re}(A_k^t)\}_t$, 3) $\{\mathrm{Im}(A_k^t)\}_t$, and 4) $\{|A_k^t|^2\}_t$ are stationary, which is often rephrased as *mean reverting*. A mean-reverting process fluctuates around its mean, so an unusually large or small observation tends to be followed by a more moderate observation. Of course, as a random process, it can be followed by an even more unusual observation. But the likelihood of consecutive unusual observations is very small. This is the basic principle we follow here. We focus on positive large values since $\{|A_k^t|^2\}_t$ produces positive large values at a local periodic signal with a matching frequency.

Suppose a stationary time series $\{Y_t\}_t$ observes a value larger than some threshold q (for example, $q = E[Y_t] + 2\sqrt{\mathrm{Var}[Y_t]}$) at some time point t after one observation below the threshold; $(Y_t \geq q, Y_{t-1} < q)$. We like to know the probability that the next observation exceeds the threshold again, i.e., $Pr(Y_{t+1} \geq q | Y_t \geq q, Y_{t-1} < q)$, and then the probability that the next observation exceeds the threshold once again, i.e.,
$Pr(Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t \geq q, Y_{t-1} < q)$. Continuing this way, we can find the probability that s consecutive observations exceed the threshold after observing one such observation following one below the threshold:

$$Pr(Y_{t+s} \geq q, \ldots, Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t \geq q, Y_{t-1} < q) \tag{7.6}$$

$$= \frac{Pr(Y_{t+s} \geq q, \ldots, Y_{t+1} \geq q, Y_t \geq q, Y_{t-1} < q)}{Pr(Y_t \geq q, Y_{t-1} < q)} \tag{7.7}$$

$$\begin{aligned} &= \left[\int_{-\infty}^{q} \int_{q}^{\infty} Pr(Y_t = y_0, Y_{t-1} = y_1) dy_0 dy_1 \right]^{-1} \\ &\quad \times \int_{-\infty}^{q} \int_{q}^{\infty} Pr(Y_{t+s} \geq q, \ldots, Y_{t+1} \geq q | Y_t = y_0, Y_{t-1} = y_1) Pr(Y_t = y_0, Y_{t-1} = y_1) dy_0 dy_1 \end{aligned} \tag{7.8}$$

We can discretize it if necessary or preferred:

$$\begin{aligned} &\approx \left[\sum_{j_0=1}^{B_0} \sum_{j_1=1}^{B_1} Pr(Y_t = y_{0,j_0}, Y_{t-1} = y_{1,j_1}) \right]^{-1} \\ &\quad \times \sum_{j_0=1}^{B_0} \sum_{j_1=1}^{B_1} Pr(Y_{t+s} \geq q, \ldots, Y_{t+1} \geq q | Y_t = y_{0,j_0}, Y_{t-1} = y_{1,j_1}) Pr(Y_t = y_{0,j_0}, Y_{t-1} = y_{1,j_1}) \end{aligned} \tag{7.9}$$

where $\{y_{0,1} \ldots y_{0,B_0}\}$ are B_0 discrete points on (q, ∞), and $\{y_{1,1} \ldots y_{1,B_1}\}$ are B_1 discrete points on $(-\infty, q)$.

Thus, if we observe a long stream of large values exceeding the threshold in the squared modulus STFT, then we can compute the probability above (7.6) and see how small it is to decide whether that time region is just following the stationarity assumption or something is happening to cause such a long series of consecutive observations exceeding the threshold.

(7.6), the probability that s consecutive observations exceed the threshold after observing one such observation following one below the threshold, can be made more complicated with $(\ell + 1)$ conditioned variables as follows, but this representation and computation is not easy

and we will just use (7.8) later.

$$Pr(Y_{t+s} \geq q, \ldots, Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t \geq q, Y_{t-1} < q) =$$

$$\left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{q} \int_{q}^{\infty} Pr(Y_t = y_0, Y_{t-1} = y_1, Y_{t-2} = y_2, \ldots, Y_{t-\ell} = y_\ell) dy_0 dy_1 dy_2 \ldots dy_\ell\right]^{-1}$$

$$\times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{q} \int_{q}^{\infty}$$

$$Pr(Y_{t+s} \geq q, \ldots, Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t = y_0, Y_{t-1} = y_1, Y_{t-2} = y_2, \ldots, Y_{t-\ell} = y_\ell)$$

$$\times \; Pr(Y_t = y_0, Y_{t-1} = y_1, Y_{t-2} = y_2, \ldots, Y_{t-\ell} = y_\ell) dy_0 dy_1 dy_2 \ldots dy_\ell. \tag{7.10}$$

The discretized version is

$$\left[\sum_{j_\ell=1}^{B_\ell} \cdots \sum_{j_2=1}^{B_2} \sum_{j_1=1}^{B_1} \sum_{j_0=1}^{B_0} Pr(Y_t = y_{0,j_0}, Y_{t-1} = y_{1,j_1}, Y_{t-2} = y_{2,j_2}, \ldots, Y_{t-\ell} = y_{\ell,j_\ell})\right]^{-1}$$

$$\sum_{j_\ell=1}^{B_\ell} \cdots \sum_{j_2=1}^{B_2} \sum_{j_1=1}^{B_1} \sum_{j_0=1}^{B_0}$$

$$Pr(Y_{t+s} \geq q, \ldots, Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t = y_{0,j_0}, Y_{t-1} = y_{1,j_1}, Y_{t-2} = y_{2,j_2}, \ldots, Y_{t-\ell} = y_{\ell,j_\ell})$$

$$\times \; Pr(Y_t = y_{0,j_0}, Y_{t-1} = y_{1,j_1}, Y_{t-2} = y_{2,j_2}, \ldots, Y_{t-\ell} = y_{\ell,j_\ell}). \tag{7.11}$$

where $\{y_{0,1} \ldots y_{0,B_0}\}$ are B_0 discrete points on (q, ∞), $\{y_{1,1} \ldots y_{1,B_1}\}$ are B_1 discrete points on $(-\infty, q)$, and for $2 \leq k \leq \ell$, $\{y_{k,1} \ldots y_{k,B_k}\}$ are B_k discrete points on $(-\infty, \infty)$.

This approach can be applied to any stationary time series input, not just white noise, because the STFT from any stationary input produces a stationary output. But the output time series is more complicated, although simulation will circumnavigate problems as we will see later. Here we focus on the Gaussian white noise input.

7.3 Gaussian Stationary Process

The conditional probabilities $Pr(Y_{t+s} \geq q, \ldots, Y_{t+1} \geq q | Y_t = y_0, \ldots, Y_{t-\ell} = y_\ell)$ are easy to compute when the time series $\{Y_t\}_t$ is a Gaussian stationary process. This is because the conditional distribution is also Gaussian. In this section, we will show the conditional mean and the conditional variance-covariance matrix of the conditional Gaussian distribution.

To begin with, we suppose $(Y_{t+s}, \ldots, Y_{t+2}, Y_{t+1}, Y_t, Y_{t-1}, Y_{t-2}, \ldots Y_{t-\ell})^\intercal$ is distributed as the $(s+1+\ell)$-dimensional Gaussian with a constant mean vector $\overrightarrow{\mu}$ and the variance-covariance matrix

$$\mathbf{V} = \begin{bmatrix} \gamma(0) & \cdots & \gamma(s-2) & \gamma(s-1) & \gamma(s) & \gamma(s+1) & \gamma(s+2) & \cdots & \gamma(s+\ell) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \gamma(s-2) & \cdots & \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \gamma(4) & \cdots & \gamma(\ell+2) \\ \gamma(s-1) & \cdots & \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(\ell+1) \\ \gamma(s) & \cdots & \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(\ell) \\ \gamma(s+1) & \cdots & \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(\ell+1) \\ \gamma(s+2) & \cdots & \gamma(4) & \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(\ell+2) \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(s+\ell) & \cdots & \gamma(\ell+2) & \gamma(\ell+1) & \gamma(\ell) & \gamma(\ell-1) & \gamma(\ell-2) & \cdots & \gamma(0) \end{bmatrix}$$

where $\gamma(h)$ is the autocovariance function at lag h, i.e., $\mathrm{E}[(Y_{t+h} - \mathrm{E}[Y_{t+h}])(Y_t - \mathrm{E}[Y_t])]$.

Now we partition μ and $\mathbf{V}$ as follows

$$\begin{aligned} \overrightarrow{\mu} &= \begin{bmatrix} \overrightarrow{\mu_1} \\ \overrightarrow{\mu_2} \end{bmatrix} \text{ with sizes } \begin{bmatrix} s \times 1 \\ (\ell+1) \times 1 \end{bmatrix} \text{ and} \\ \mathbf{V} &= \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \text{ with sizes } \begin{bmatrix} s \times s & s \times (\ell+1) \\ (\ell+1) \times s & (\ell+1) \times (\ell+1) \end{bmatrix}. \end{aligned}$$

Then the conditional distribution $(Y_{t+s}, \ldots, Y_{t+1} | Y_t = y_0, Y_{t-1} = y_1, \ldots, Y_{t-\ell} = y_\ell)$ is also the s-dimensional Gaussian with mean vector

$$\begin{aligned} \overrightarrow{\mu_c} &= \overrightarrow{\mu_1} + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\overrightarrow{y} - \overrightarrow{\mu_2}) \quad \text{where } \overrightarrow{y} = (y_0, y_1, \ldots, y_\ell)^\intercal \\ &\quad \text{and covariance matrix} \\ \mathbf{V}_c &= \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}. \end{aligned}$$

With modern computing power, the conditional probabilities

$$Pr(Y_{t+s} \geq q, \ldots, Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t = y_0, Y_{t-1} = y_1, Y_{t-2} = y_2, \ldots, Y_{t-\ell} = y_\ell)$$

are tractable when the time series $\{Y_t\}_t$ is a Gaussian stationary process. Schervish (1984) and Schervish (1985) present an algorithm to compute multivariate Gaussian probabilities with error bounds, which we can use to evaluate the above conditional probabilities.

7.3.1 AR(p)

If we further assume the time series $\{Y_t\}_t$ is a Gaussian stationary AR(p) process, the computation of the conditional probability becomes much simpler. This is because the conditional distribution

$$(Y_{t+s} \geq q, \ldots, Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t = y_0, Y_{t-1} = y_1, Y_{t-2} = y_2, \ldots, Y_{t-(p-1)} = y_{p-1})$$

is the same as the conditional distribution

$$(Y_{t+s} \geq q, \ldots, Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t = y_0, Y_{t-1} = y_1, Y_{t-2} = y_2, \ldots, Y_{t-\ell} = y_\ell)$$

for $\ell \geq p$. That is, additional knowledge of more than $(p-1)$ past values provides nothing (the Markov property). This is perhaps intuitive from the form: $Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + Z_t$ in

which only the p recent past values matter. Therefore we can set $\ell = p$ if the time series $\{Y_t\}_t$ is a Gaussian stationary AR(p) process.

However, in general, this result does not apply to MA(q) or ARMA(p,q) processes, where the larger ℓ is, the more information we have and the more concentrated the conditional probability becomes (Shumway and Stoffer 2006). Thus, we would like transform the time series of concern (the squared modulus STFT) into approximately a Gaussian AR process.

7.3.2 The Box-Cox Transformation

As we saw in Chapter 3, under the Gaussian white noise assumption, for fixed k, the time series $\{|A_k^t|^2\}_t$ is a non-Gaussian, non-linear stationary process. But we would like to make it approximately Gaussian AR(1) for computational simplicity, as shown above, by the Box-Cox transformation:

$$Y_t = \begin{cases} \dfrac{(|A_k^t|^2)^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log\left(|A_k^t|^2\right), & \text{if } \lambda = 0. \end{cases}$$

We have found the best $\lambda \approx 0.27$ based on simulation for $k = 2$ and 8. The top plot of Figure 7.2 shows the choice of λ against the Gaussian likelihood. We note that the likelihood is very flat around 0.27, which means that choices nearby would work as well. The bottom plot of Figure 7.2 is the histogram of the transformed squared modulus STFT. It displays the Gaussian probability density function in the red line. We see that the transformation works well for the simulated data. Strictly speaking, this transformed variable has a Weibull

distribution with probability density function

$$
\begin{aligned}
f_Y(y) &= (\sigma_{RR}+\sigma_{II})^{-\lambda}\left(\frac{y-\frac{-1}{\lambda}}{\lambda^{-1}(\sigma_{RR}+\sigma_{II})^{\lambda}}\right)^{\frac{1}{\lambda}-1}\exp\left(-\left(\frac{y-\frac{-1}{\lambda}}{\lambda^{-1}(\sigma_{RR}+\sigma_{II})^{\lambda}}\right)^{\frac{1}{\lambda}}\right) \\
&\quad \text{for} \quad \frac{-1}{\lambda} < y < \infty, \qquad \text{with} \\
E[Y_t] &= \frac{(\sigma_{RR}+\sigma_{II})^{\lambda}}{\lambda}\Gamma(1+\lambda)-\frac{1}{\lambda}, \qquad \text{and} \\
Var[Y_t] &= \frac{(\sigma_{RR}+\sigma_{II})^{2\lambda}}{\lambda^2}\left[\Gamma(1+2\lambda)-\Big(\Gamma(1+\lambda)\Big)^2\right], \qquad \text{where} \\
\Gamma(z) &= \int_0^{\infty} t^{z-1}e^{-t}dt \quad \text{(the Gamma function).}
\end{aligned}
$$

So we know the mean and variance of the transformed squared modulus STFT time series. Next we will approximately find the autocovariance of the transformed time series: $E[Y_{t+1}Y_t]$.

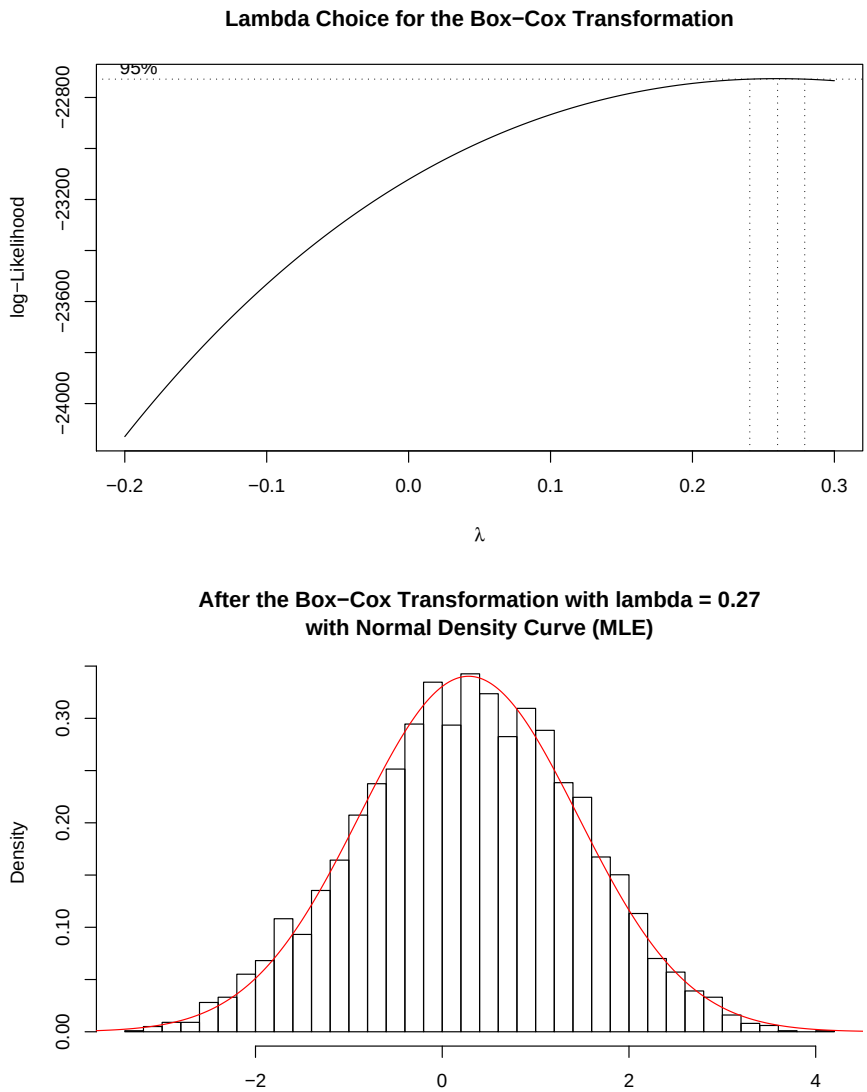

Figure 7.2: The choice of the transformation parameter λ for the Box-Cox transformation, applied to the squared modulus STFT time series of Gaussian white noise. In the top plot, λ values are plotted against the log-likelihood function. The marginal distribution of the time series before the transformation is exponential. The bottom plot shows that with choice of 0.27, we have approximately Gaussian marginal distribution, plotted along with the Gaussian distribution function with the maximum likelihood parameter estimates.

7.3.3 The Delta Method

We use the delta method to approximately find $E[Y_{t+1}Y_t]$ to specify the AR parameters that work for transforming the time series $\{|A_k^t|^2\}_t$ into approximately AR(1).

$$
\begin{aligned}
\text{First, let } f(x_1, x_2) &= \frac{x_1^\lambda - 1}{\lambda}\frac{x_2^\lambda - 1}{\lambda} \qquad \text{then we find the derivatives;} \\
f_1 &= \frac{\partial f}{\partial x_1} = \frac{x_1^{\lambda-1}}{\lambda}(x_2^\lambda - 1); \qquad f_2 = \frac{\partial f}{\partial x_2} = \frac{x_2^{\lambda-1}}{\lambda}(x_1^\lambda - 1) \\
f_{11} &= \frac{\partial^2 f}{\partial^2 x_1} = \frac{\lambda - 1}{\lambda}x_1^{\lambda-2}(x_2^\lambda - 1); \qquad f_{22} = \frac{\partial^2 f}{\partial^2 x_2} = \frac{\lambda - 1}{\lambda}x_2^{\lambda-2}(x_1^\lambda - 1) \\
f_{12} &= \frac{\partial^2 f}{\partial x_1 \partial x_2} = (x_1 x_2)^{\lambda-1}
\end{aligned}
$$

Letting $\mu = E[|A_k^{t+1}|^2] = E[|A_k^t|^2] = \sigma_{RR} + \sigma_{II}$ (as we saw in Chapter 3), we find the approximate expectation as follows;

$$
\begin{aligned}
E[f(|A_k^{t+1}|^2, |A_k^t|^2)] &\approx E\Bigg[f(\mu,\mu) + f_1(\mu,\mu)[|A_k^{t+1}|^2 - \mu] + f_2(\mu,\mu)[|A_k^t|^2 - \mu] \\
&\quad + \frac{1}{2!}f_{11}(\mu,\mu)[|A_k^{t+1}|^2 - \mu]^2 + \frac{1}{2!}f_{22}(\mu,\mu)[|A_k^t|^2 - \mu]^2 \\
&\quad + \frac{2}{2!}f_{12}(\mu,\mu)[|A_k^{t+1}|^2 - \mu][|A_k^t|^2 - \mu]\Bigg] \\
&= \left(\frac{\mu^\lambda - 1}{\lambda}\right)\left(\frac{\mu^\lambda - 1}{\lambda}\right) + 0 + 0 \\
&\quad + \frac{\lambda - 1}{2\lambda}\mu^{\lambda-2}(\mu^\lambda - 1)\mu^2 + \frac{\lambda - 1}{2\lambda}\mu^{\lambda-2}(\mu^\lambda - 1)\mu^2 \\
&\quad + (\mu\mu)^{\lambda-1}\mathrm{Cov}[|A_k^{t+1}|^2, |A_k^t|^2] \\
&= \left(\frac{\mu^\lambda - 1}{\lambda}\right)^2 + \frac{\lambda - 1}{\lambda}(\mu^{2\lambda} - \mu^\lambda) + \mu^{2\lambda-2}ACVF_{|A_k|^2}(1) = E[\widehat{Y_{t+1}Y_t}]
\end{aligned}
$$

where $ACVF_{|A_k|^2}(1)$ is the autocovariance function at lag 1 for the time series $\{|A_k^t|^2\}_t$, which can be found in Chapter 3.

Thus,

$$Cov[Y_{t+1}, Y_t] = E[Y_{t+1}Y_t] - (E[Y_t])^2 \approx E\widehat{[Y_{t+1}Y_t]} - (E[Y_t])^2 = Cov\widehat{[Y_{t+1}, Y_t]}.$$

Finally, we find the AR(1) coefficient

$$\widehat{\phi} = \frac{Cov[Y_{t+1}, Y_t]}{Var[Y_t]} \approx \frac{Cov\widehat{[Y_{t+1}, Y_t]}}{Var[Y_t]}.$$

We can calculate the residuals by

$$\epsilon_t = Y_t - \widehat{\phi} Y_{t-1} - E[Y_t](1 - \widehat{\phi}).$$

Figure 7.3 shows the histogram, ACF and PACF of the residuals from this AR(1) model fitted to the Box-Cox transformed data in Figure 7.2. It has some moderately large PACF's, but all in all this is evidence that the approximation works reasonably well with the choice of AR(1) parameter chosen by the delta method.

Probability of s Consecutive Large Values

Finally, we are at the stage where we can compute the probability of observing s consecutive values exceeding the transformed threshold $q^* = \frac{q^\lambda - 1}{\lambda}$:
$Pr(Y_{t+s} \geq q^*, \ldots, Y_{t+2} \geq q^*, Y_{t+1} \geq q^* | Y_t \geq q^*, Y_{t-1} < q^*)$. We can use (7.8) (with only the current observation Y_t and one past observation Y_{t-1} as the conditional variables) as the process is now approximately Gaussian AR(1). If we were to choose to approximate the series by an MA process or ARMA process we would have to include as much past values as possible to use (7.10) and complicate the computation of the conditional probability.

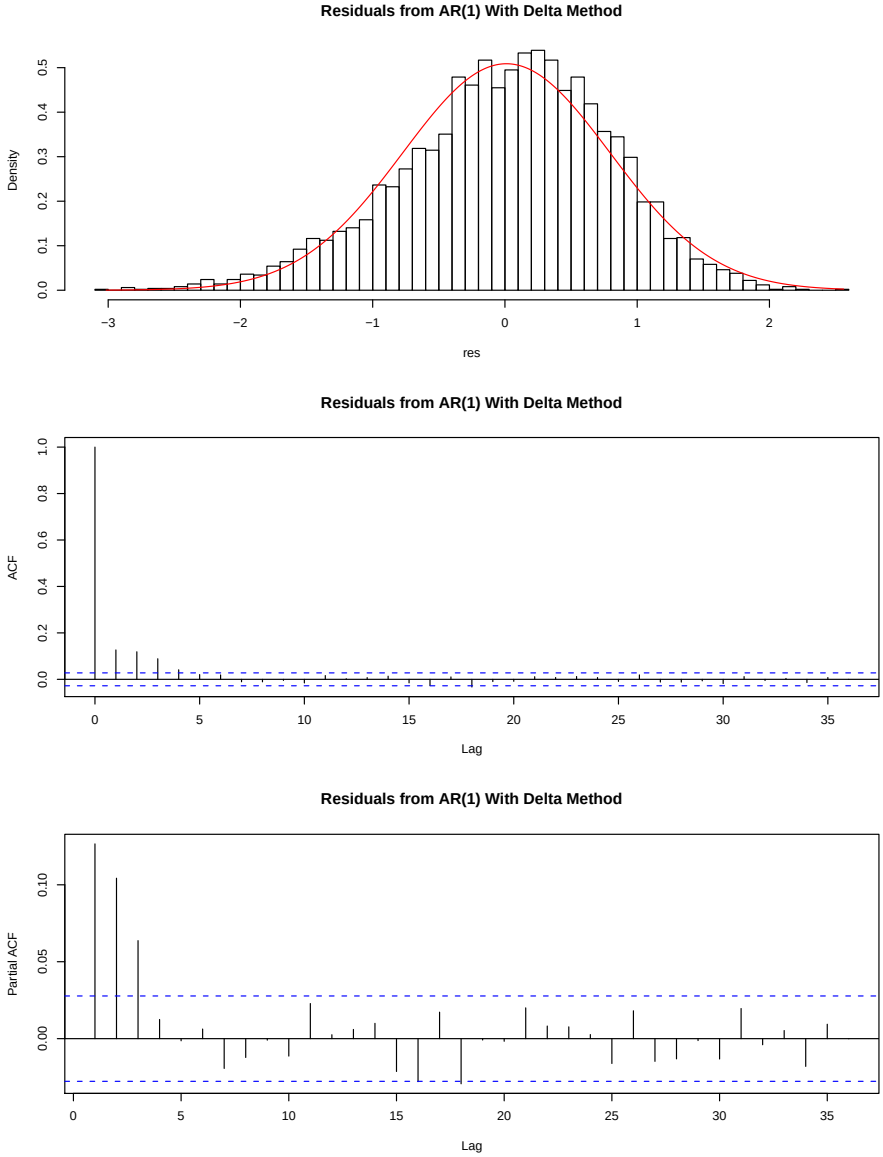

Figure 7.3: The residuals from AR(1) fitted to the Box-Cox transformed data in Figure 7.2, with the coefficient chosen by the delta method. The top plot shows the marginal distribution of the residuals, which is approximately Gaussian. The middle and bottomo plots show the autocorrelation function and partial autocorrelation function of the residuals, respectively, which show that the residuals are approximately white noise. They indicate that the Box-Cox transform and the delta method work reasonably well.

7.3.4 By the Monte Carlo Simulation Method

The conditional probability $Pr(Y_{t+s} \geq q, \ldots, Y_{t+2} \geq q, Y_{t+1} \geq q | Y_t \geq q, Y_{t-1} < q)$ can be found numerically by simulation. It can be computationally expensive, but does not rely on approximation or transformation. This can be used for any input, not just Gaussian white noise input. It works as follows:

1) Generate a long sequence of an input series under assumption.

2) Compute the squared modulus STFT.

3a) Find the first time point where the 2) exceeds the threshold q.

3b) If no observation is found, then go back to 1) with a longer time series input.

4) Record how many consecutive time points 2) exceeds the threshold q.

5) Go back to 1).

Repeat this until enough data are accumulated. Then measure the proportion of the data that are larger than or equal to s.

The computational amount increases as the window size increases and also as the threshold q increases because exceeding values are observed less often and we have to simulate long sequences of input series and compute their squared modulus STFT. In 3) and 4), of course we can use the second time point where the squared modulus exceeds the threshold again, after the first stream of exceeding values. We can continue to use the third time point and so forth. Using only the first stream of exceeding values as above wastes the remaining part of the squared modulus STFT and is just for illustration.

We emphasize that this method does not use the Box-Cox transformation or the delta method for approximation.

7.3.5 Comparison of the Two Methods

Here we compare the two methods. The first method uses the Box-Cox transformation and Gaussian AR approximation, while the second method uses the Monte Carlo simulation. The threshold q was chosen as the 95 percentile of the $\mathrm{Exp}(\sigma_{RR} + \sigma_{II})$ distribution. We use these two methods and compute the probability that s consecutive observations exceed the threshold after observing one such observation following one below the threshold.

The following table summarizes the probabilities up to 10 steps. The approximation method's probabilities decrease faster than the simulation method, but the two series are relatively similar. This shows that our method of computing the conditional probabilities with the Box-Cox transformation and the delta method works reasonably well.

step	0	1	2	3	4	5	6	7	8	9	10
AR(1)	1	0.444	0.199	0.092	0.042	0.020	0.009	0.004	0.002	0.001	0.000
Sim	1	0.581	0.244	0.116	0.073	0.026	0.018	0.007	0.003	0.002	0.000

Figure 7.4 plots this table with AR(1) in the solid line and simulation in the dashed line.

Thus, there is less than 1 percent of the probability of observing more than 7 consecutive large values exceeding the threshold q after one exceeding observation that follows one below it in the squared modulus STFT time series $\{|A_k^t|^2\}_t$, given that the input series is the Gaussian white noise. Therefore if we observe a longer stream of exceeding values than 7 time points, we suspect that there is a local periodic signal.

In this chapter, we examined methods to detect local signals. The first method uses transformation and approximation. And the second method uses the Monte Carlo simulation. The both methods can be used to compute the probability of observing s consecutive large values exceeding the threshold q after one exceeding observation that follows one below it in the squared modulus STFT time series $\{|A_k^t|^2\}_t$. The first method reduces the computational burden that the second method suffers from, but the two methods produce similar probabilities so we know that the first method works reasonably well.

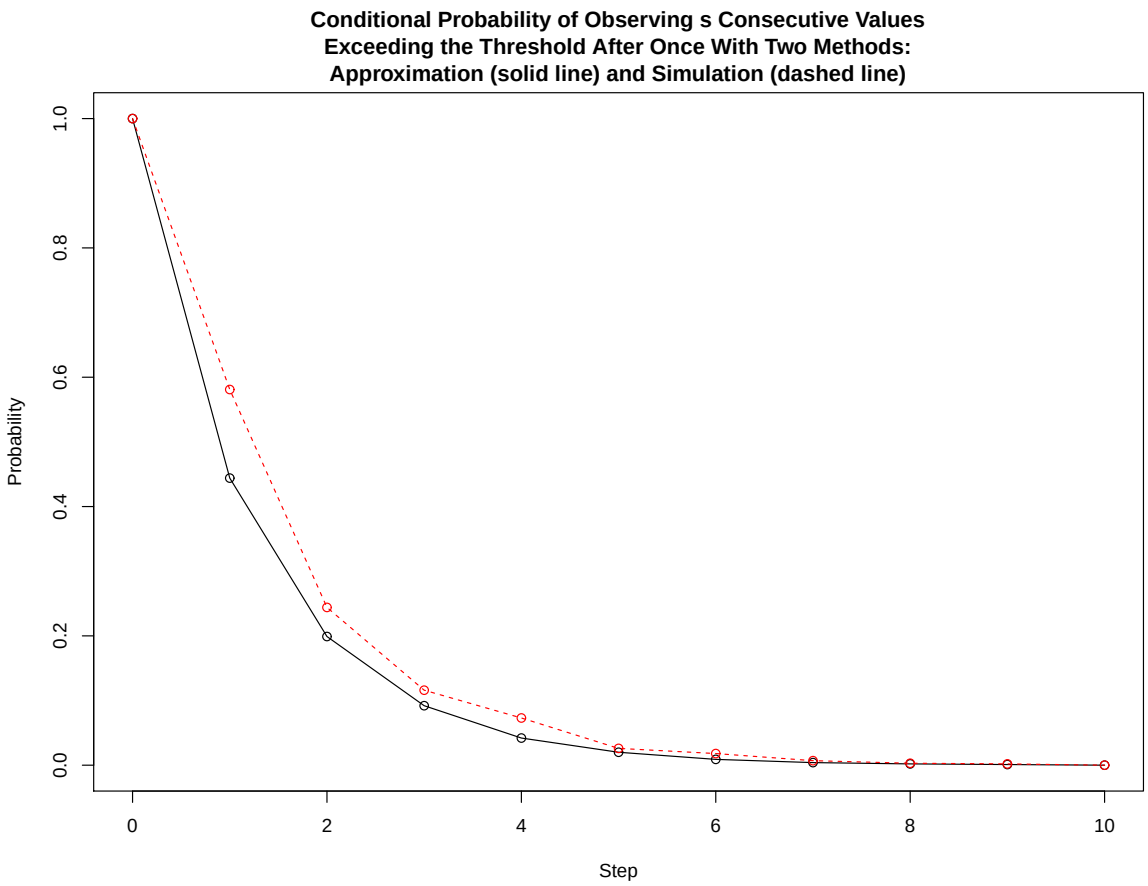

Figure 7.4: The conditional probability of observing s consecutive values exceeding the threshold q after one exceeding observation that follows one observation below the threshold, $\Pr(Y_{t+s} \geq q, \ldots, Y_{t+1} \geq q | Y_t \geq q, Y_{t-1} < q)$. Comparing the probabilities computed in two ways. One method uses the delta method and the Box-Cox transformation (solid line), while the other uses the Monte Carlo simulation (dashed line). This indicates that our approximation method works well.

Chapter 8

Conclusion and Future Work

The STFT computes the discrete Fourier transform many times with overlapping windows. In this thesis, I have shown several theoretical properties of the short time Fourier transform, applied to various types of complex-valued, univariate time series inputs. We showed the closed-form output from several kinds of input series. In particular, just like the discrete Fourier transform, the STFT's modulus time series takes large positive values when the input is a periodic signal. One main point is that a white noise time series input results in the STFT output being a complex-valued stationary time series and we can derive the time dependency structure such as the cross-covariance functions. Our primary focus was the detection of local periodic signals. We presented a method to detect local signals by computing the probability that the squared modulus STFT time series has consecutive large values exceeding some threshold after one exceeding observation following one observation less than the threshold. We discussed a method to reduce the computation of such probabilities by the Box-Cox transformation and the delta method, and showed that it works reasonably well in comparison to the Monte Carlo simulation method.

Originally, the concept of the STFT was brought to our attention by a visiting professor Jianming Wang from Tianjin Polytechnic University who was studying fast algorithms for the STFT with Professor William F. Eddy at the Department of Statistics at Carnegie

Mellon University during the 2007-2008 academic year. One of their research goals was to discover and investigate high frequency oscillations observed in epilepsy patients with the neuroscience device Magnetoencephalography (Wang et al. 2009). The Magnetoencephalography invasively measures extremely weak magnetic fields on human scalp more than 1000 times per second.

Many neural activities in human brain exhibit periodic signals. Some of them may be present and observable for a short period of time, say, 200 milliseconds, in an experiment of length, say, 2000 milliseconds. Such local signals may not be detected if we apply the discrete Fourier transform to the whole data of 2000 milliseconds because of the poor signal to noise ratio of the Magnetoencephalography. However, the STFT may be able to help us notice it. Then we can proceed to implement the method developed in Chapter 7 to see whether or not the data after the STFT is significantly non-stationary with a specified significance level such as 0.01. As the field of neuroscience and the Magnetoencephalography are relatively new, we may be able to discover local signals previously unknown.

One immediate challenge is the selection of the STFT window size N. As the Nyquist-Shannon sampling theorem states, the window size needs to be at least twice as large as the highest frequency of interest. So for example, if we set Magnetoencephalography's sampling rate at 1000 times per second (1 kHz) and are interested in a signal of 0.1 kHz (100 cycles per second), then we would need to set the window size at least 200 (milliseconds). But choosing N too large (better frequency resolution) would make it harder to locate the local signal and thus would result in less time resolution. With the STFT, we face a trade-off between time resolution and frequency resolution. It would also be difficult to choose an optimal window size that does not result in leakage as described earlier.

One issue with time series inputs is that it is unlikely to be a white noise time series. We rarely have such a scenario. Even if we did, it would be difficult to estimate the variance of the input series when the local signal is present and distorts the variance estimation. If we do not have a white noise time series input, then we would need to find a model that fits

the time series input well (this model specification alone can be a difficult, time-consuming task) and then simulate the input many times to implement the Monte Carlo simulation method described in Section 7.3.4 to find the probability of consecutive values exceeding the threshold after observing one, because we may not always be able to find a way to transform the STFT output to a Gaussian autoregressive model as we did for the Gaussian white noise input. Then finding such probabilities by simulation are computationally expensive.

The short time Fourier transform is a growing subject and it opened up many questions. We can expect successful applications and more development of it in the near future.

Appendix A

References

Alsteris, L. D., Paliwal, K. K. (2007), "Iterative reconstruction of speech from short-time Fourier transform phase and magnitude spectra," *Computer Speech And Language*, 21, 174-186.

Avargel, Y., and Cohen I. (2010), "Modeling and Identification of Nonlinear Systems in the Short-Time Fourier Transform Domain," *IEEE Transactions on Signal Processing*, 58, 291-304.

Blackman, R. B., and Tukey, J. W. (1959), *The Measurement of Power Spectra from the Point of View of Communications Engineering*, New York: Dover.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (Second Edition), New York: Springer-Verlag.

Cristi, R. (2004), *Modern Digital Signal Processing*, Pacific Grove, California: Brooks/Cole-Thomson Learning.

Dirgenali, F., Kara, S., and Okkesim, S. (2006), "Estimation of wavelet and short-time Fourier transform sonograms of normal and diabetic subjects' electrogastrogram," *Computers in Biology and Medicine*, 36, 1289-1302.

Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer-Verlag.

Gabor, D. (1946), "Theory of Communication," *J. IEEE*, 93, 429-457.

Jiang, Y. Q. , and He, Y. G. (2009), "Frequency estimation of electric signals based on the adaptive short-time Fourier transform," *International Journal of Electronics*, 96, 267-279.

Krishnaiah, P. R., and Rao, M. M. (1961), "Remarks on a Multivariate Gamma Distribution," *The American Mathematical Monthly*, 68, 342-346.

Latifoglu, F., Kara, S., and Imal, E. (2009), "Comparison of Short-Time Fourier Transform and Eigenvector MUSIC Methods Using Discrete Wavelet Transform for Diagnosis of Atherosclerosis," *Journal of Medical Systems*, 33, 189-197.

Matusiak, E., Michaeli, T., and Eldar, Y. C. (2010), "Noninvertible Gabor Transforms," *IEEE Transactions on Signal Processing*, 58, 2597-2612.

Partington, J. R., and Unalmis, B. (2001), "On the windowed Fourier transform and wavelet transform of almost periodic functions," *Applied and Computational Harmonic Analysis*, 10, 45-60.

Percival, D. B., and Walden, A. T. (1993), *Spectral Analysis for Physical Applications*, Cambridge: Cambridge University Press.

Qian, K. M. (2004), "Windowed Fourier transform method for demodulation of carrier fringes," *Optical Engineering*, 43, 1472-1473.

Radha, R., and Thangavelu, S. (2009), "Holomorphic Sobolev spaces, Hermite and special Hermite semigroups and a Paley-Wiener theorem for the windowed Fourier transform," *Journal of Mathematical Analysis and Applications*, 354, 564-574.

Reinsel, G. C. (1997), *Elements of Multivariate Time Series Analysis* (Second Edition), New York: Springer-Verlag.

Schervish, M. J. (1984), "Algorithm AS 195: Multivariate Normal Probabilities with Error Bound," *Journal of the Royal Statistical Society*, Series C, 33, 81-94.

Schervish, M. J. (1985), "Corrections: Algorithm AS 195: Multivariate Normal Probabilities with Error Bound," *Journal of the Royal Statistical Society*, Series C, 34, 103-104

Shumway, R. H., and Stoffer, D.S. (2006), *Time Series Analysis and Its Applications: With R Examples*, New York: Springer-Verlag.

Vetterli, M., Kovačević, J., and Goyal, V. K. (in press), *Fourier and Wavelet Signal Processing*, Retrieved Jan 13, 2011, from http://www.fourierandwavelets.org/book.pdf

Wang, J. M., Woods, B., and Eddy, W. F. (2009), "MEG, RFFTs, and the Hunt for High Frequency Oscillations," *Proceedings of the 2009 2nd International Congress on Image and Signal Processing.*

Wyatt, D. C. (1988), "Analysis of ship-generated surface-waves using a method based upon the local Fourier transform," *Journal of Geophysical Research-Oceans*, 93, 14133-14164.

Xia, X. G. (1998), "A quantitative analysis of SNR in the short-time Fourier transform domain for multicomponent signals," *IEEE Transactions on Signal Processing*, 46, 200-203.