THE ESTIMATION OF HIGH-CONTRAST POWER SPECTRA VIA ITERATED WHITENING

by:

Jared McBride

A Dissertation submitted to the Faculty of the

GRADUATE INTERDISCIPLINARY PROGRAM IN APPLIED MATHEMATICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2023

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Jared McBride
titled:

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Kevin Lin (Aug 17, 2023 13:38 PDT)
Kevin Lin
Date: Aug 17, 2023

Shankar Venkataramani (Aug 18, 2023 02:43 GMT+5.5)
Shankar Venkataramani
Date: Aug 18, 2023

Joseph C. Watkins
Joseph C Watkins
Date: Aug 17, 2023

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Kevin Lin (Aug 17, 2023 13:38 PDT)
Kevin Lin
Dissertation committee chair
Mathematics
Date: Aug 17, 2023

Acknowledgments

Here at the culmination of my formal education I am grateful for the opportunity to reflect on the countless individuals who have made this accomplishment possible. I will only be able to highlight just a few of the very important examples and mentors in my life. First, I would like to mention my parents Don and Sheryl McBride, apart from the tremendous sacrifice in raising me they demonstrated unwavering encouragement and support. I am grateful for the hard work of my Father who set an example of scholarship and hard work in getting his own PhD. My mother more recently helps often and in various ways in looking after the little ones throughout my education here in Tucson.

While in junior high and high school there were many teachers that inspired me including Craig Hansen and Alan Vanthournout. At Brigham Young University, I was very impressed with Dr. Christopher Grant, how always had time for me, and Dr. John Dallon, who believed in me and was ever encouraging. Here at the University of Arizona, I have grown very much under the advisement and mentorship of Dr. Kevin Lin and Dr. Shankar Venkataramani. I am also very grateful for Dr. Joe Watkins and his recent contributions to this work.

The final summer in which I worked on this was a very busy one and I am very grateful to all the family and friends that helped us to get through it. I am particularly grateful to my Father- and Mother-in-law who hosted my family for the summer. I am grateful to my little ones Emeline, Daniel, Lily, and Grace for their support and prayers on my behalf. They have proven and invaluable source of joy being playing with them has always been refreshing and for me. I am grateful for the myriad sacrifices made by my dear wife, Erin, who helped proofread this dissertation, and without whose support, it would not have been possible to grow our family and achieve this degree at the same time.

Table of Contents

Table of Contents—*Continued*

Table of Contents—*Continued*

List of Figures

List of Figures—*Continued*

List of Figures—*Continued*

List of Tables

Abstract

Power spectra are a fundamental tool in data analysis, signal processing, and linear prediction and control. Many who seek to estimate power spectrum are obliged to do so with very little theoretical information of the underlying process and prefer accurate estimates which require as little effort as possible. This work focuses on estimation of the power spectrum of time series data from dynamical systems and stochastic differential equations (SDEs) and attempts to satisfy the preferences above. The method, called iterated whitening (IW) spectral estimation, iteratively builds inexpensive filters that progressively ameliorate the data until the resulting modified data is suitable for accurate spectral estimation. This spectral estimate is then post-processed to return an accurate estimate of the original data.

Time series from dynamical systems and SDEs often possess a very large dynamic spectral range which makes them difficult to estimate cheaply and accurately. IW provides a solution to this difficulty. In this dissertation, I discuss some of the issues that the Bartlett estimator has in approximating these "high-contrast" spectra by deriving expanded bias and variance formulas. I also showcase another technique for improving a spectral estimator using the method on control variates from Monte Carlo Markov chain theory.

I apply these methods to some time series from a solution to the Kuramoto-Sivashinsky equation which is in spatiotemporal chaos.

Chapter 1

Introduction

Loosely speaking, the power spectrum of a stationary (or approximately stationary) process tells one how much of the total "energy" of a signal is contributed by a particular "frequency" (see Chapter 2). In practice the estimation of the power spectrum of a stochastic process is often computed from data, i.e. samples or measurements from that underlying stochastic process. Frequency information from data can be very useful for a number of reasons. In physics, for example, different chemical elements, when excited, emit radiation of different frequencies, a sort of fingerprint. This is used to determine the chemical makeup of substances based on the radiation they emit or absorb. So, broadly speaking, frequency information of a signal can assist in the identification of the source of the data. For instance, it can help determine if data measured from an instrument is a real observation or if it represents a glitch or measurement noise [1].

The power spectrum also shows up in linear estimation theory and data driven model reduction. Wiener filtering is a method of optimal (in the sense of least squares) linear estimation that can be used in model reduction[1]. In order to construct the Wiener filter, one needs to compute or estimate power spectra. It was while working on this sort of model reduction that I first became interested in power spectrum estimation. The power spectra needed to be estimated in my work in model reduction turned out to possess a property that made them very difficult to estimate using classical methods. Namely, their spectra varied widely over many orders of magnitude, a property which I refer to as high-contrast and which I will discuss in greater detail in Chapter 3. In this study we focused on the estimation of power spectra of processes from chaotic dynamical systems and stochastic differential equations (SDEs). Such processes commonly exhibit high-contrast power spectra.

In this work, I present a method of power spectrum estimation that builds on a particular family of estimators and allows for much higher accuracy at frequencies with very low power. The method, called iterated whitening (IW) spectral estimation, iteratively builds inexpensive filters that progressively ameliorate the data until the resulting modified data is suitable for accurate

[1]For a full account of how the Wiener filter may be applied to data driven model reduction, I refer the reader to the work of Dr. Kevin Lin and Dr. Fei Lu [2].

spectral estimation. This spectral estimate is then post-processed to return an accurate estimate of the original data. I also discuss another method which improves the accuracy of spectral estimates by reducing their variance using a variance reduction technique from Markov chain Monte Carlo (MCMC) theory.

The remainder of the dissertation goes as follows. In Chapter 2, I review relevant mathematical concepts necessary for a thorough discussion of the IW power spectrum estimation procedure. Chapter 3 introduces a property of power spectrum, mentioned earlier, which I call high-contrast. I analyze the bias and variance properties and use the results to better demonstrate the difficulties in estimating high-contrast power spectra with the Bartlett estimator, in particular. This analysis is more detailed than what is ordinarily presented in standard sources, e.g., [3, 4]. The chapter also provides two examples of high-contrast spectra and illustrates the difficulties in estimating their power spectrum. After discussing some further preliminaries, in Chapter 4, I introduce the IW procedure. In Chapter 5, I provide a numerical study of IW and include a runtime analysis of all the estimators previously discussed. Chapter 6 discusses the MESA estimate. Chapter 7 reports another method that improves the Bartlett type estimator. This is a method that applies the variance reduction technique of control variates from Monte Carlo theory to spectrum estimation.

Chapter 8 provides a further demonstration of the IW procedure by applying it to data from the Kuramoto-Sivashinsky equation [5]. And finally, Chapter 9 contains a concluding discussion.

Chapter 2

Background on spectral estimation

In this chapter, I provide some background material that pertains to the development of the IW power spectrum estimate. I begin with a review of stationary stochastic processes and their autocovariances; these are central to the concept of a power spectrum. I then formally define the power spectrum and discuss some facts about it. The theory behind IW draws heavily on some signals and systems theory. As such, I include some useful facts from that field. I finish the chapter by reviewing in some detail the three common methods Bartlett's method, Welsh's method, and MESA.

2.1 Brief summary of mathematical framework

Let $X = (X_n,\ n \in \mathbb{Z})$ be a collection of random variables parameterized by $\mathbb{Z}$. We assume that X has finite moments of up to fourth order, meaning that for at least $r = 4$, $\mathbb{E}|X_n|^r < \infty$ for all $n \in \mathbb{Z}$. When this parameter n is associated with time, which we assume[1] throughout, X is referred to as a time series. For a time series X, define $\mu_n = \mathbb{E}X_n$ to be the mean at time n. Also define[2]

$$C[n, m] = E[(X_n - \mu_n)(X_m - \mu_m)^*] \tag{2.1}$$

to be the autocovariance of X at times n, m. Here * denotes the complex conjugate. For simplicity of explanation, I assume that X is a scalar process and therefore C is a scalar-valued function; however, all of what is discussed in this work may be generalized to the multichannel case where X is a vector-valued process. Observe that if X where vector-valued * would denote the conjugate transpose and C would be matrix-valued.

A time series is *wide-sense stationary* (WSS) if (1) the mean μ_n at time n does not depend on n, that is, it is constant, and (2) the autocovariance $C[n, m]$ depends only on the *lag* given by $n - m$

[1] This assumption is not necessary. If one were seeking the power spectrum in spatial frequencies this would all still apply. But consistent with the literature, I make this assumption.

[2] I find the practice of using square brackets when the argument of a function is intended to be discrete a nice convention and use it throughout.

that is

$$C[n,m] = C[n-m], \qquad \forall n,m \in \mathbb{Z}. \tag{2.2}$$

A time series is *strict-sense stationary* or simply *stationary* if for any finite collection of times $n_1,\ldots,n_k,\nu \in \mathbb{Z}$, the random vectors $(X_{n_1},\ldots,X_{n_k})$ and $(X_{n_1+\nu},\ldots,X_{n_k+\nu})$ have the same distribution. Wide-sense stationarity is a weaker condition than strict-sense stationarity and can be thought of as second order stationarity meaning that the first and second moments are invariant under time shifts. With strict-sense stationarity, on the other hand, all finite moments are invariant under time shifts. In this work we assume all processes are WSS. We will see that this assumption is necessary for there to be a power spectrum at all.

The autocovariances C of any WSS process X has the following properties. For all lags[3] n, $C[n] = C[-n]^*$. This can be worked out from the definition. And so, if X is real then so is C and we have $C[n] = C[-n]$, meaning C is even. The final property of autocovariances presented here is that for any positive integer k, any collection of times $n_1,\ldots,n_k \in \mathbb{Z}$, and any complex numbers $z_1,\ldots,z_k \in \mathbb{C}$ then

$$\sum_{j=1}^{k}\sum_{\ell=1}^{k} z_j C[n_j - n_\ell] z_\ell^* \geq 0. \tag{2.3}$$

This can be shown by observing that for the random variable $Y = \sum_{j=1}^{k} z_j X_{n_j}$

$$\begin{aligned} 0 \leq \mathrm{var}(Y) &= \sum_{j=1}^{k}\sum_{\ell=1}^{k} z_j \mathrm{cov}(X_{n_j}, X_{n_\ell}) z_\ell^* \\ &= \sum_{j=1}^{k}\sum_{\ell=1}^{k} z_j C[n_j - n_\ell] z_\ell^*. \end{aligned}$$

Now, because it is such a common situation in practice [6], I also assume that the autocovariances decay exponentially, meaning

$$|C[n]| < K\rho^n \qquad \text{for some } K \text{ and some } 0 < \rho < 1. \tag{2.4}$$

Given any WSS stochastic process $X = \big(X_n, n \in \mathbb{Z}\big)$, with autocovariances C, its *power spectrum*

[3]The input of the autocovariances of a WSS process will be referred to as *lags* or *time lags*, just as the index of X is referred to as *time* and, as we will see, the input of the power spectrum is *frequency*.

S is defined by

$$S(\omega) = \sum_{n=-\infty}^{\infty} C[n]e^{-in\omega}. \tag{2.5}$$

This is a Fourier series with the autocovariance sequence as the Fourier coefficients. I will refer to it as the discrete-time Fourier transform (DTFT) of C (as in [7]). The convergence of the power spectrum can be guaranteed under certain conditions such as having the autocovariances decay exponentially as mentioned above. This function is real-valued on the unit circle since, $C[-n] = C^*[n]$. Also, it can be shown that $S(\omega) \geq 0$ for all ω. This is a consequence of (2.3) and can be seen as follows. For any positive integer N and any $a \in \mathbb{C}$, pick $n_j = j$ and $z_j = ae^{-i(j-1)\omega}$ for $j = 1, \ldots, N$. Then

$$0 \leq \frac{1}{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N} ae^{-i(j-1)\omega}C[j-\ell]a^*e^{i(\ell-1)\omega} \tag{2.6}$$

$$= a\frac{1}{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N} C[j-\ell]e^{-i(j-\ell)\omega}a^* \tag{2.7}$$

$$= a\frac{1}{N}\sum_{k=-N}^{N} (N-|k|)C[k]e^{-ik\omega}a^* \tag{2.8}$$

$$= a\sum_{k=-N}^{N} C[k]e^{-ik\omega}a^* - a\frac{1}{N}\sum_{k=-N}^{N} |k|C[k]e^{-ik\omega}a^* \tag{2.9}$$

Then as $N \to \infty$ the first term in (2.9) becomes $S(\omega)|a|^2$ while the second term goes to zero since if the autocovariances decay exponentially, the sum converges. The proof is presented in a way that easily shows that if $S(\omega)$ were matrix valued[4], it would be positive semi-definite.

To better familiarize the reader with the power spectrum, there are a few facts that can provide intuition. First, observe that, employing the inverse Fourier transform

$$\text{var}(X) = C[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)e^{i\omega\cdot 0}\, d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)\, d\omega. \tag{2.10}$$

The variance of the time series is equal the integral of the power spectrum divided by 2π. So that the spectrum, in a sense, gives a distribution of the variance of the process over the space of frequencies. This next fact describes how one can generate a Gaussian time series with a given power spectrum. Let $S(\omega)$ be a nonnegative definite function on $[-\pi, \pi]$, let W_ω be a Wiener process over $[-\pi, \pi]$,

[4]This would be the case if the terms in X were vector-valued; the so-called multi-channel case.

now let

$$X_n = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{in\omega} \sqrt{S(\omega)}\, dW_\omega, \qquad n \in \mathbb{Z}. \tag{2.11}$$

It can be shown that the power spectrum of $X = (X_n)$ recovers S, that is $S_X(\omega) = S(\omega)$. For clarity the power spectrum will be subscripted with the process that it is spectrum of, and similarly for the autocovariances. To see that X defined from 2.11) does indeed have the spectrum S, first observe that X has mean zero for each n since W_ω is mean zero for each ω. Now we look at the covariances of X,

$$\mathbb{E}X_n X_m^* = \mathbb{E}\left(\frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{in\omega} \sqrt{S(\omega)}\, dW_\omega\right)\left(\frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{-im\omega} \sqrt{S(\omega)}\, dW_\omega\right) \tag{2.12}$$

$$= \mathbb{E}\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)\omega} S(\omega)\, d\omega \qquad \text{(Itô, see e.g. [8])} \tag{2.13}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)\omega} S(\omega)\, d\omega =: C[n-m]. \tag{2.14}$$

And so, X is stationary and, clearly, the DTFT of C will be S. The relation under (2.11) demonstrates that each term of the time series is some combination of quantities of different frequencies and that the power spectrum tells you how to weight each quantity according to the frequency.

Lastly, equation (2.11) provides a means for sampling from a process $X = (X_n; n \in \mathbb{Z})$ such that X is Gaussian and $S_X(\omega) = S(\omega)$.

2.2 Signals and systems theory: filtering

I now discuss concepts from signals and systems theory that will be useful in understanding IW. To begin with, a time series is often referred to as a *signal*. Following this convention, I will use the term *signal* and *time series* interchangeably. It is sometimes productive to consider (together with the power spectrum) the *z-spectrum* $\overline{S}(z)$[5] of a signal X. This is defined as

$$\overline{S}(z) = \sum_{n=-\infty}^{\infty} C[n] z^{-n}, \qquad \text{where, again,} \qquad C[n] = \mathbb{E}X_n X_0^*, \tag{2.15}$$

[5]Here and elsewhere I use convention that if a function is the DTFT of a sequence, the z-transform of that sequence will have the same name as the DTFT but with a bar over it. We will see this again when the transfer function is introduced.

This is the so-called *z-transform*[6] of the autocovariance sequence. Observe how the z-spectrum relates to the original spectrum,

$$S(\omega) = \overline{S}\left(e^{in\omega}\right). \tag{2.17}$$

With this in hand, we can move on. A *system* is an operator that maps signals to signals. A *linear, time-invariant* (LTI) system $\mathcal{L}$ maps signals to signals linearly, and in such a way that it commutes with the shift operator. That is, if the terms of the input are shifted by some amount, the terms of the output will be shifted by that same amount (if $\mathcal{L}X = Y$ then $\mathcal{L}\left(X_{\cdot+k}\right) = Y_{\cdot+k}$). An important result in the theory of LTI systems is that such a system $\mathcal{L}$ can be represented by convolution with the sequence $\mathcal{L}\delta$. Here $\delta = (\ldots, 0, 1, 0, \ldots)$, with the one at the index zero position, is the impulse signal. It can easily be shown using both the linearity and time-invariance of $\mathcal{L}$ that for $\ell = \mathcal{L}\delta$ and a signal X,

$$\mathcal{L}X = (\ell * X) \tag{2.18}$$

where

$$(\ell * X)_n = \sum_{j=-\infty}^{\infty} \ell_j X_{n-j}. \tag{2.19}$$

The sequence ℓ is called the *time domain representation* or *impulse response* of $\mathcal{L}$. Nonzero terms in the impulse response are sometime referred to as *taps*. This system is represented in frequency domain by

$$L(\theta) = \sum_{j=-\infty}^{\infty} \ell_j e^{-ij\theta}. \tag{2.20}$$

This is called the *frequency response* of $\mathcal{L}$ and is the DTFT of ℓ. The z-transform of the impulse response of the system,

$$\overline{L}(z) = \sum_{j=-\infty}^{\infty} \ell_n z^{-n}, \tag{2.21}$$

I will refer to as the *transfer function* of the filter[7]. The system $\mathcal{L}$ is often referred to as a *filter*, and the application of a system to a signal is accordingly referred to as *filtering* by that system.

A filter is said to be *stable* if the output of a bounded input is bounded. This is commonly

[6]The z-transform of a sequence ℓ is defined to be

$$L(z) = \sum_{n=-\infty}^{\infty} \ell_n z^{-n}. \tag{2.16}$$

[7]Some texts use the terms frequency response and transfer function interchangeably, while still others use the term transfer function to mean the impulse response.

referred to as BIBO (bounded-input bounded-output) stability. It can be shown that an LTI filter is stable if and only if its impulse response is absolutely summable [6, Sec. 6.2.2]. An upshot of stability, because of the summability of the impulse response, is that the transfer function converges on the unit circle.

A filter is said to be *causal* if its impulse response ℓ is zero for negative indices or lags[8]. This means that the filter output at time n did not make use of any term of the input after time n. Such filters allow for so-called online filtering in which one does not have access to any terms of the input beyond the past and present, say because the data is registered in real time. A causal filter $\mathcal{L}$, therefore, has a frequency response and transfer function of the forms

$$L(\omega) = \sum_{j=0}^{\infty} \ell_j e^{-ij\omega} \qquad \text{and} \qquad \overline{L}(z) = \sum_{j=0}^{\infty} \ell_j z^{-j}. \tag{2.22}$$

As a result, the transfer function $\overline{L}$ convergences absolutely at every point on and outside of the unit circle and is analytic there also. A filter that is not causal is said to be *noncausal*.

A filter can some times have an inverse. If it exists, this is a system that maps the output of the original system back to its input. A filter is said to be *minimum phase* if it is causal and stable and its inverse exists and is also causal and stable. From the above discussion, this means that its transfer function, therefore, is analytic on and outside of the unit circle and its multiplicative inverse[9] is also analytic on and outside of the unit circle. In the case in which the transfer function of a minimum phase system is known to be rational it must, necessarily, have all its zeros and poles within the unit circle.

We conclude this subsection with a very useful fact. This subject of filtering is pertinent to power spectra because filtering by a linear time-invariant filter effects the power spectrum of the signal in predictable ways. The point of IW is to filter the data so that the power spectrum is increasingly amenable to accurate estimation. It can be shown that the power spectrum of the filtered process $\ell * X$ is

$$S_{\ell * X}(\omega) = L(\omega) S_X(\omega) L^*(\omega) \tag{2.23}$$

($L(\omega)$ as in (2.20). For the reason mentioned above, we will use this result extensively.

[8]The indices of the impulse response of a filter are sometimes referred to as lags, as with the autocovariances. This makes sense when we consider the role of the index j in the sum defining the convolution in (2.19). If j is positive, the term of the input series is lagged by j in the jth term of the sum and receives the coefficient ℓ_j.

[9]If a system is stable and invertable, the transfer function of its inverse will be the multiplicative inverse of its transfer function.

2.3 Methods for spectral estimation

Now we discuss the estimation of the power spectrum of a WSS process given a finite length realization. I focus on three techniques for spectral estimation that are particularly useful in that they are cheap and scale fairly well to large problems of the sort we are interested in. They are Bartlett's method and Welsh's method [9], which are discussed in this chapter and Maximal Entropy Spectral Analysis (MESA) which I discuss in Chapter 6. The reason for delaying MESA to another chapter is that it is fundamentally different from the former two methods, which are both periodogram methods. These periodogram methods are more commonly used [10] than MESA.

These techniques are chosen because they are not only quick and inexpensive to implement but perform well given very little knowledge of the spectrum we are estimating. Which is apropos in the context mentioned above because, in general, very little can be known about the spectra of processes from dynamical systems and SDEs. It is the Bartlett estimate that we focus on improving. There are many other methods of estimating power spectra from data and I will briefly discuss these here. These other methods are not considered in this study because, though they can achieve a high level of accuracy, they (1) generally require information about the process from which the data is sampled, and (2) require effort from the user in tuning a number of parameters and optimizing other modeling choices. These points will be elaborated on below. The three methods discussed here are sometimes suggested as a pilot step whose output informs the initial stages of one of the more involved spectral analyses named above [3, Sec. 7.4][11].

Other techniques of spectral estimation include: multitaper [12], least-squares spectral analysis [13], singular spectrum analysis [14, 15], maximal likelihood estimates [16], and AutoRegressive Moving Averages (ARMA) fitting [16]. There are still other techniques that can be used, for instance, a reversible jump Markov chain Monte Carlo [17] procedure can be devised to estimate the power spectrum. This is similar to maximum likelihood estimates, and requires a likelihood function.

Since we will be mainly interested in the estimation of power spectra from data, let $x = \big(x_n, n = 1, \ldots, N\big)$ be a finite-time realization of the WSS process X, x will often be referred to as a *snapshot*. The techniques to follow use this data to approximate the power spectrum of X.

We begin with the periodogram, which is a significant player in both the methods of Bartlett and Welsh, which are often referred to as periodogram based estimates, whose work we then discuss.

2.3.1 The periodogram

One way to estimate the power spectrum is to appeal directly to the definition (2.5) of the power spectrum by empirically estimating the autocovariances and taking the discrete Fourier transform (DFT). Perhaps the most straightforward estimate of the autocovariances is given by

$$\tilde{C}'_X[n] = \frac{1}{N-|n|} \sum_{k=1\vee(1-n)}^{(N-n)\wedge N} x_{k+n} x_k^*. \tag{2.24}$$

This, however, is poor for large lags n since in this case we are estimating something that is naturally small but doing so with relatively few $(N-|n|)$ terms in the average. And so, the relative error can be quite high. One way to improve the estimate is to divide each sum by N as in

$$\tilde{C}_X[n] = \frac{1}{N} \sum_{k=1\vee(1-n)}^{N\wedge(N-n)} x_{k+n} x_k^*. \tag{2.25}$$

Changing the denominator introduces what is known as *bias*, which is the difference between the expected value of an estimator (when the data are thought of as random variables) and the true value that is being estimated. However, The larger denominator reduces the variance so that the overall relative error is reduced. This leads us to what is known as the *periodogram*,

$$\tilde{S}_X^{\text{per}}(\omega) = \sum_{n=-(N-1)}^{N-1} \tilde{C}_X[n] e^{-in\omega}. \tag{2.26}$$

It can be shown by a brief calculation that:

$$\begin{aligned}
\hat{S}_X^{\text{per}}(\omega) &= \sum_n \tilde{C}_X[n] e^{-in\omega} \\
&= \sum_n \frac{1}{N} \sum_j x_{n+j} x_j^* e^{-in\omega} \\
&= \frac{1}{N} \sum_k \sum_j x_k e^{-ik\omega} x_j^* e^{ij\omega} \qquad (k = n+j) \\
&= \frac{1}{N} |\hat{x}(\omega)|^2
\end{aligned}$$

where

$$\hat{x}(\omega) = \sum_{n=1}^{N} x_n e^{-in\omega} \tag{2.27}$$

is the DTFT of the snapshot x. By virtue of the fast Fourier transform (FFT), this estimate can be computed very efficiently[10]. The periodogram is asymptotically unbiased [3, Sec. 6.2.2 p. 417], meaning as the length of the time series N goes to infinity, the expectation of the estimate converges to the true spectrum. But it is inconsistent [3, Sec. 6.2.2 p. 425] (for a very nice discussion on this see pp. 429-431 of [3]), because the variance does not vanish as $N \to \infty$. In practice, the periodogram gives very rough estimates of the power spectrum.

2.3.2 Bartlett family of estimators

Recognizing the variance of the periodogram cannot be controlled by making N large, one could apply a variance reduction technique as follows. This procedure was purposed by M. Bartlett [18]. Split the snapshot $x = (x_n,\ n = 1, \ldots, N)$ into K sections of length L (so that $K = \lfloor N/L \rfloor$), $x^{(j)} = (x_n,\ n = L(j-1)+1, \ldots, Lj)$ for $j = 1, \ldots, K$ and compute the periodogram of each section. Then average these to get

$$\tilde{S}_X^{\text{Bart}}(\omega) = \frac{1}{K}\sum_{j=1}^{K}\frac{1}{L}|\hat{x}^{(j)}(\omega)|^2 = \frac{1}{K}\sum_{j=1}^{K}\tilde{S}_{L,j}^{\text{per}}(\omega) \tag{2.28}$$

where $\tilde{S}_{L,j}^{\text{per}}(\omega)$ is the periodogram of the j^{th} section (of length L) of the original snapshot x. The hope being that this allows us to control the variance since, loosely speaking,

$$\text{var}\left(\tilde{S}^{\text{Bart}}(\omega)\right) \approx \frac{1}{K}\text{var}\left(\tilde{S}_{L,j}^{\text{per}}(\omega)\right). \tag{2.29}$$

The relation (2.29) above is approximate, since in general the periodograms $\tilde{S}_{L,j}^{\text{per}}(\omega)$ are not independent. As it turns out, (2.28) is essentially equivalent [3, pp. 439-40] to

$$\tilde{S}_X^{\text{Bart}}(\omega) = \sum_{n=-L}^{L}\Lambda_L^{\text{Bart}}[n]\cdot\tilde{C}_X[n]e^{-in\omega} \tag{2.30}$$

where

$$\Lambda_L^{\text{Bart}}[n] = \begin{cases} 1-\frac{|n|}{L}, & |n| \le L \\ 0, & \text{otherwise} \end{cases} \tag{2.31}$$

[10] It should be noted that this provides values of the periodogram only on a regular grid, say ($\omega_j = \frac{2\pi}{N}$, $j = 0, \ldots, N-1$) where N is the length of the snapshot.

is the so-called *Bartlett lag window*. This function has the property that $\Lambda_L^{\text{Bart}}[n] = 0$ when $|n| \geq L$, for that reason L is referred to as the *truncation parameter*. Following this framework, many other so-called *lag windows* have been developed. A common one which we use almost exclusively is the *Parzen lag window*,

$$\Lambda_L^{\text{Parz}}(u) = \begin{cases} 1 - 6(u/L)^2 + 6(|u|/L)^3, & |u| \leq L/2, \\ 2(1 - |u|/L)^3, & L/2 \leq |u| \leq L, , \\ 0, & |u| > L \end{cases} \tag{2.32}$$

named after E. Parzen [19]. For a good discussion on lag window selection, see [3, Sec. 7.5]. The Parzen window is generally a good choice when little about the true spectrum is known.

The *generalized Bartlett* estimator $\tilde{S}_X^{\text{GB}}(\omega)$ is formed by replacing the lag window $\Lambda_L^{\text{Bart}}[n]$ in (2.30) with any of the many lag windows developed today. In practice it can be computed efficiently using FFT. This can be done by first using an FFT-based convolution to estimate the autocovariance[11], point-wise multiplying this by a preferred lag window and then return to frequency space by again employing FFT.

2.3.3 Welsh and tapering

A method of spectral estimation similar to Bartlett, at least in spirit, was developed by Peter Welsh around 1967 [9]. Similar to how the Bartlett method was formulated, this method involves dividing the input signal into possibly overlapping segments. Each segment is then point-wise multiplied by a *data window*[12] or taper. Tapering the data to zero at the ends reduces correlation between overlapping segments. The periodogram of each windowed segment is computed, and all of these are then averaged to obtain the spectral estimate.

In more detail the procedure goes as follows. First, divide the original snapshot $x = (x_n,\ n = 1, \ldots, N)$ into N_{en} (possibly overlapping) segments of length L_{seg}. Let $M = \lfloor (N - L_{\text{seg}})/(N_{\text{en}} - 1) \rfloor$ be the index offset between the start of consecutive segments. Written out explicitly, let

$$x^{(j)} = (x_n,\ n = M(j-1) + 1, \ldots, M(j-1) + L_{\text{seg}}) \qquad \text{for } j = 1, \ldots, N_{\text{en}}. \tag{2.33}$$

[11] The use of FFT here is not necessary, only expeditious.

[12] As opposed to Bartlett's method, in which the autocovariances are point-wise multiplied by a lag window.

The next step is to window (or taper) each segment by point-wise multiplying the data by the window. As with Bartlett's method, a number of tapers have been proposed. For example, the Tukey-Hanning taper is a data window in common use. It is a combination of the Hanning window and a rectangular window, allowing for a smooth transition between the two. The formula for the Tukey-Hanning taper is as follows:

$$w[n] = \begin{cases} \frac{1}{2}\left[1 - \cos\left(\frac{2\pi(n-1)}{L_{\text{seg}}\alpha}\right)\right], & \text{if } 1 \leq n < \frac{\alpha L_{\text{seg}}}{2}, \\ 1, & \text{if } \frac{\alpha L_{\text{seg}}}{2} \leq n < L_{\text{seg}}\left(1 - \frac{\alpha}{2}\right), \\ \frac{1}{2}\left[1 - \cos\left(\frac{2\pi(L_{\text{seg}}-n}{L_{\text{seg}}\alpha}\right)\right], & \text{if } L_{\text{seg}}\left(1 - \frac{\alpha}{2}\right) \leq n \leq L_{\text{seg}}, \end{cases} \tag{2.34}$$

here n is the sample index of the segment, ranging from 1 to L_{seg}, $0 \leq \alpha \leq 1$ is a parameter controlling the width of the taper. When $\alpha = 0$ the taper corresponds to the rectangular window and $\alpha = 1$, it corresponds to the Hanning window. Call the result $x_w^{(j)}$. After that, form the (modified) periodogram of each of these and average. So, that

$$\hat{S}^{\text{Welsh}}(\omega_k) = \sum_{j=1}^{N_{\text{en}}} \frac{1}{L_{\text{seg}}} |\hat{x}_w^{(j)}|^2. \tag{2.35}$$

For a list of data-windows together with insights on picking one that is suitable to a particular situation see [20].

Generally speaking Welsh's method performs better than Bartlett's method and is more widely used in practice [21]. However, one advantage of Bartlett's method over Welsh's is that it does better with shorter snapshots as (in its generalized form) segmenting the snapshot and analyzing each segment is not required [21, Sec. 10.7][13].

In this chapter we have reviewed some signals and systems theory and looked at three power spectrum estimators that are significant in the literature. In the next chapter we will focus primarily on the Bartlett estimator as we discuss the limitation it exhibits with high-contrast spectra.

[13] In her book, what Silvia Maria Alessio calls the *Blackman-Tukey* estimator is what I refer to as the generalized Bartlett or just Bartlett estimator.

Chapter 3

Estimation of high-contrast spectra

In this chapter we look at processes with what I call high-contrast spectra and analyze how the two periodogram based estimators perform on two simple examples with spectra of this type. To begin with, I define the *spectral dynamic range* of a process to be the quantity

$$\max_{\omega} \left\{ \log_{10} \left(S(\omega) \right) \right\} - \min_{\omega} \left\{ \log_{10} \left(S(\omega) \right) \right\}, \tag{3.1}$$

where S is the power spectrum. Here, the spectral dynamic range is measured in decades but can also be measured in decibels (dB), in which case, the values would be multiplied by ten[1]. For some context pertaining to the magnitude of this quantity, I draw from the literature. In September 1982 there was a special issue of the Proceedings of IEEE that focused on power spectrum estimation. In one of the many great articles in that issue, David Thomson [12] wrote the following about his experience with the spectral dynamic range, "I have frequently been told by time series analysts that spectra with ranges of over 40 or 50 dB approach the pathological. In contrast, my personal experience has been that when data are carefully collected and analyzed, spectra from physical origins rarely have less than 50 dB range. I have also experienced some situations when perfectly reasonable communications problems led to a desire to estimate spectra with ranges from 160 to over 200 dB." [12]. With that in mind, for a spectrum, by *high-contrast* I generally mean that the spectral dynamic range is more than 10 decades (100 dB).

As shall be seen, Bartlett estimators have difficulty accurately estimating high-contrast spectrum. To be more specific, frequencies of sufficiently low power are all overestimated to around the same power, giving the appearance of a threshold or floor below which the estimator cannot accurately estimate power. This phenomenon, which I will discuss in some detail below, I call the *estimator floor*. In the next section, I demonstrate for two examples, with high-contrast spectra, the performance of the two estimators heretofore discussed. I then focus the discussion on the estimator floor and discuss it from two perspectives. The first involves *spectral leakage* which is discussed briefly. The second deals with the bias and variance. For this perspective, I focus on

[1] For example, 5 decades = 50 dB.

the Bartlett estimator and develop expansions in the parameters L (truncation parameter) and N (length of the snapshot) of bias and variance for that estimator. I demonstrate a few instances of how these theoretical results match what we empirically observe. This result anticipates that a similar analysis can be done for the Welsh estimator as well but this is left for future work.

3.1 Some examples

We now look at two toy examples of processes with high-contrast spectra. These will be used throughout the remainder of the dissertation.

3.1.1 Gaussian process with a Gaussian spectrum

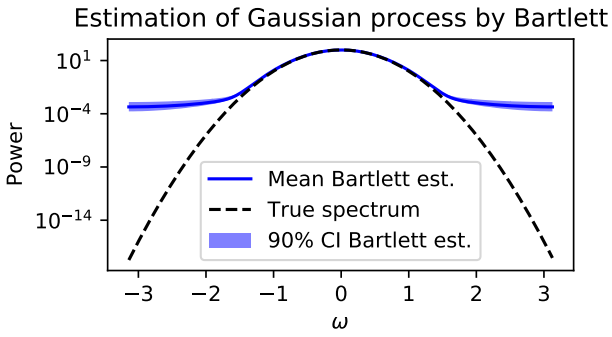
The first example we consider is a Gaussian process $X = (X_n, n \in \mathbb{Z})$ with the spectrum

$$S(\omega) = Ce^{-\frac{\omega^2}{2\sigma^2}} = 10^{-2x^2+2} \tag{3.2}$$

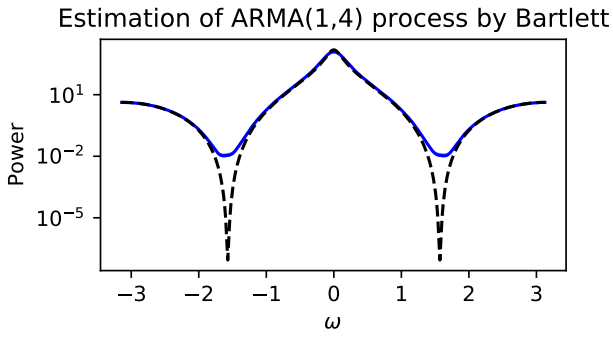
over the interval $[-\pi, \pi]$, (here $C = 100$ and $\sigma = (2\sqrt{\log 10})^{-1} \approx 0.33$) this has a spectral dynamic range of ≈ 20 decades (197 dB)[2]. For our investigation, the process $X = (X_n; n \in \mathbb{Z})$ such that X is Gaussian and $S_X(\omega) = S(\omega)$ was sampled using a standard transformation technique (see, for example, [22]). This process was contrived both for its simplicity and its large spectral dynamic range.

For each of the two estimators described in the previous chapter, the true spectrum together with the mean and the 90% confidence intervals for the estimated spectra of 100 realizations of this process are shown in Figures 3.1a and 3.1c. In each case the snapshot had length $N = 40{,}000$. For the Bartlett estimator, the Parzen lag window with truncation parameter $L = 50$ was used. For Welsh, the Tukey-Hanning data window was used with $\alpha = 0.1$. The data was divided into $N_{\text{en}} = 20$ segments of $L_{\text{seg}} = 3{,}000$ terms each. It can be seen in Figure 3.1a that the Bartlett estimate shows an estimator floor at around 10^{-4}, effecting frequencies $|\omega| > 1.5$. The Welsh estimate, in Figure 3.1c, clearly also has an estimator floor but it is much lower at around 10^{-10}.

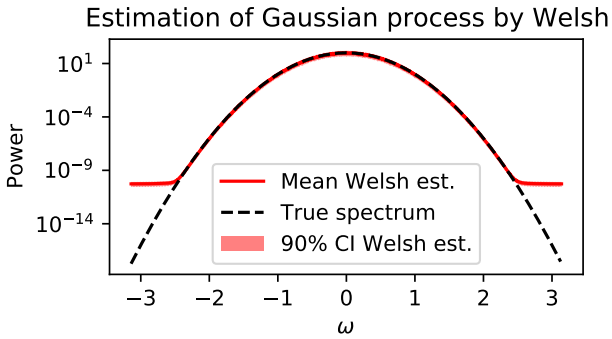
[2] $\log_{10}(\max_\omega S(\omega)) - \log_{10}(\min_\omega S(\omega)) \approx 19.74$.



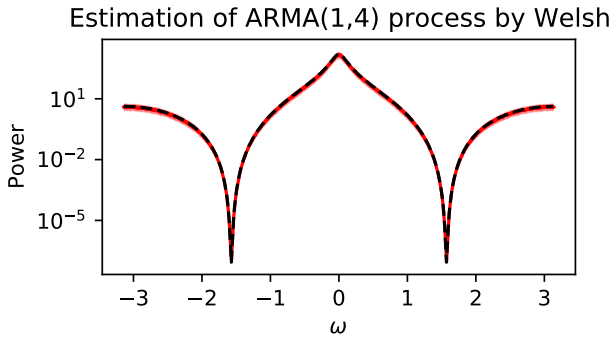
(a) Mean Bartlett estimate with Parzen lag window ($L = 50$) of the Gaussian process.



(b) Mean Bartlett estimate with Parzen lag window ($L = 50$) of the ARMA(1,4) process.



(c) Mean Welsh estimate with Tukey-Hanning Data window ($N_{\text{en}} = 20$, $L_{\text{seg}} = 3{,}000$) of the Gaussian process.



(d) Mean Welsh estimate with Tukey-Hanning Data window ($N_{\text{en}} = 20$, $L_{\text{seg}} = 3{,}000$) of ARMA(1,4) process, solid red.

Figure 3.1: Bartlett and Welsh power spectrum estimates of the Gaussian and ARMA(1,4) examples. The mean and 90% confidence interval of the spectral estimates of 100 independent realizations for each process (Gaussian and ARMA(1,4)) and for each estimator (Bartlett with Parzen lag window and Welsh with Tuky-Hanning data window). In all cases a snapshot of length $N = 40{,}000$ was used. In each plot the true spectrum is the dashed black curve.

Parameter	Value
pole	0.9
zeros	$\pm 0.99i$, $\pm 0.99i$
σ^2	1

Table 3.1: The parameters of the ARMA(1,4) process.

3.1.2 ARMA(1,4)

For the second example, I first introduce the term white noise. A *white noise* process is a process $e = (e_n,\ n \in \mathbb{Z})$ in which each e_n is a Gaussian random variable with zero mean and variance σ^2 and is independent of all the rest. Our second example is what is called an AutoRegressive Moving Averages (ARMA) process. An ARMA(p,q) process is a stochastic process $X = (X_n,\ n \in Z)$ that can be written in the form

$$\sum_{j=0}^{p} a_j X_{n-j} = \mu + \sum_{k=0} e_{n-k} \tag{3.3}$$

where $a = (a_j,\ j = 0, \ldots, p)$ and $b = (b_k,\ k = 0, \ldots, q)$ are the *autoregressive* and *moving averages* coefficients, respectively, μ is a constant (which we take to be zero) and e_n is a white noise process (with variance σ^2). The polynomials $\sum_{k=0}^{q} b_k z^k$ and $\sum_{j=0}^{p} a_j z^j$ are called the characteristic polynomials of the moving averages part and the autoregressive part, respectively (see, e.g. [16]). The specific process in this example was formulated by carefully picking the zeros of the characteristic polynomials so as to give a large spectral dynamic range. The parameters are reported in Table 3.1.

With this choice of poles we get autoregressive coefficients $a = (1, -0.9)$ and moving averages coefficients $b = (1, 0, 2 \cdot 0.99^2, 0, 0.99^4)$, so that X may be represented by (cf. equation (3.3))

$$(X * a)_n = (\epsilon * b)_n, \tag{3.4}$$

where ϵ is a white noise process of unit variance. By equation (2.23) we can compute the spectrum of X using (3.4) to be, with $r = 0.99$ and $r' = 0.9$,

$$S(\omega) = \frac{1 + 4r^4 + r^8 + 4(r^2 + r^6)\cos(2\omega) + 2r^4\cos(4\omega)}{1 + (r')^2 - 2r'\cos(\omega)}, \tag{3.5}$$

which has a spectral dynamic range of 10.25 decades (102.49 dB).

A plot of the true spectrum of this process, together with the mean and 90% confidence intervals of the power spectrum estimates of 100 realizations of this process is given in Figures 3.1b and 3.1d. Most of the parameters and modeling choices for the three estimators that were used to estimate the Gaussian process above were found suitable in estimating the spectrum of the ARMA(1,4) process and remained unchanged.

As with the Gaussian example, the Bartlett estimate demonstrates an estimator floor, this time at around 10^{-2} (effecting frequencies $1.25 < |\omega| < 2$). The Welsh estimate this time does not exhibit an estimator floor.

3.2 Estimator floor for Bartlett estimators

Figure 3.1 provides a demonstration of the aforementioned estimator floor for periodogram based estimates. There are, at least, two perspectives that may be taken to view the estimator floor. One involves the concept of *spectral leakage*, which notes that the lag window plays a role in "smearing" spectral power across power estimates over all frequencies. The other perspective is through the bias and variance, observing that the lag window and true spectrum to be estimated affect the bias and variance, and thus the overall error of the estimate.

In this section I briefly discuss spectral leakage but then set up for a much more in-depth discussion on the bias and variance perspective of the estimator floor. In the next section, through the statistical properties of the bias and variance, I hope to provide what I think may be a novel way of analyzing the estimator floor. Unfortunately, no conceptual bridge is made between these two perspectives here. However, I believe this endeavor to be rewarding future work.

3.2.1 Spectral Leakage

Recall that the Bartlett estimate may be written as follows.

$$\tilde{S}(\omega) = \sum_{n=-N+1}^{N-1} \Lambda_L[n]\tilde{C}[n]e^{-in\omega}. \tag{3.6}$$

Now, observe that this is the DTFT of the point-wise product of two sequences. By the convolution theorem for DTFT (see, e.g.[7]), (3.6) is equivalent to the convolution of the DTFT of each of the

two sequences, namely

$$\tilde{S}(\omega) = \int_{-\pi}^{\pi} \hat{S}^{\text{per}}(\theta) W_L(\omega - \theta)\, d\theta \tag{3.7}$$

where

$$\hat{S}^{\text{per}}(\theta) = \sum_{n=-(N-1)}^{(N-1)} \tilde{C}[n] e^{-i\theta n} = \frac{1}{N} \left| \sum_{n=1}^{N} x_n e^{i\theta(n-1)} \right|^2 \tag{3.8}$$

is the periodogram and

$$W_L(\theta) = \frac{1}{2\pi} \sum_{n=-L}^{L} \Lambda_L[n] e^{-i\theta n} \tag{3.9}$$

is the *spectral window* or the frequency response on the lag window $\Lambda_L[n]$. Figure 3.2 plots both the Parzen spectral window and periodogram of the Gaussian example. Notice that, roughly speaking, the side lobes of the Parzen spectral window don't go much below 10^{-5} over the whole frequency space $[-\pi, pi]$. This means that within the integral under (3.7) estimating the power at any frequency ω, there is a part over the interval $[\theta_1, \theta_2]$, say, which contains the peak so that

$$\tilde{S}(\omega) > \int_{\theta_1}^{\theta_2} \hat{S}^{\text{per}}(\theta) W_L(\omega - \theta) \approx (\theta_2 - \theta_1) \times \text{near peak power} \times 10^{-5} \tag{3.10}$$

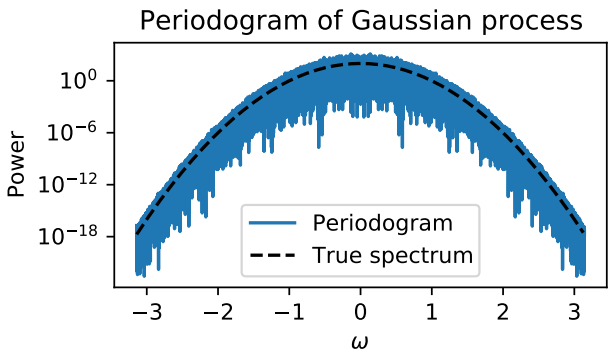
Now, the peak power of the periodogram is roughly greater than 500 over the interval $[-0.01, 0.0.1]$ and so, there is a contribution of roughly $0.02 \times 500 \times 10^{-5} = 10^{-4}$ to the approximate power estimated at any frequency. Note that the integrand is strictly positive so there is no possibility for cancellation. I stress that this argument for $\tilde{S}(\omega) > 10^{-4}$ will apply to any frequency ω. This agrees with Figure 3.2c which indeed has an estimator floor of about 10^{-4}. So, the area and the rate of decay of the side lobes of the spectral window controls, to some extend, the amount of spectral leakage.

This is an important concept to consider when selecting a lag window. Many lag windows have been developed since the 1950s that seek to minimize the area of the side lobes. These sorts of considerations are covered thoroughly in [3].

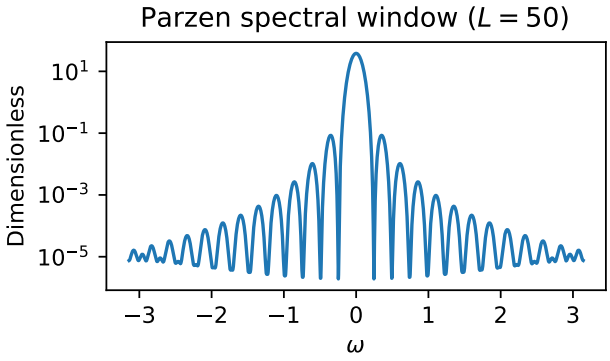
Welsh, as has been shown, also suffers from an estimator floor and a similar explanation can be found in the concept of spectral leakage.

3.2.2 Asymptotic bias and variance formulas

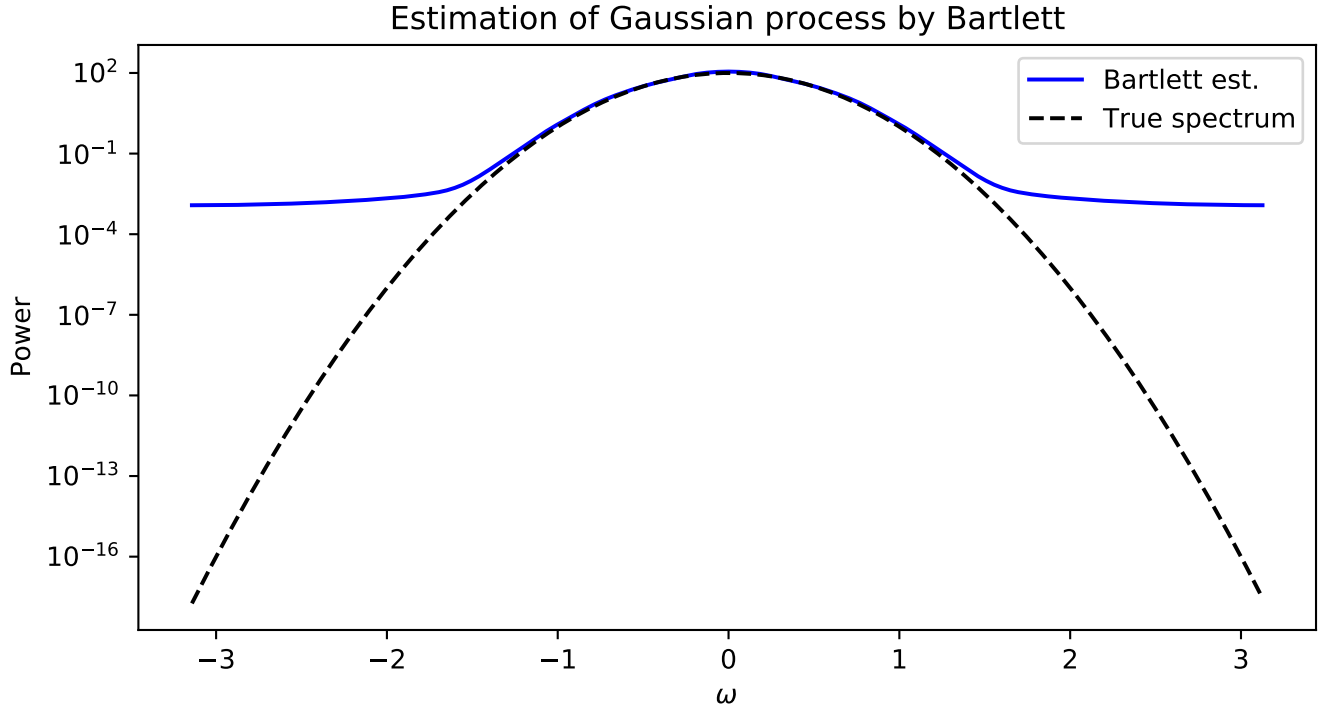
The purpose of the analysis is, ultimately, to provide understanding and support for the effectiveness of the IW procedure to be discussed later. With this in mind, an analysis of the Welsh estimator or



(a) The periodogram of a single snapshot of the Gaussian process, $N = 40{,}000$.



(b) The Parzen spectral window, $L = 50$.



(c) The Bartlett estimate of the snapshot whose periodogram is in Figure 3.2a. This is the convolution of the two plots above, Figure 3.2a and Figure 3.2b.

Figure 3.2: An illustration of spectral leakage demonstrated with the Gaussian example in which the Parzen lag window ($L = 50$) is used.

the MESA estimator is not necessary to our purpose, since IW can be formulated around Bartlett's method.

In 1957, E. Parzen [23] developed formulas for the asymptotic bias and variance of Bartlett estimators to leading order. They are

$$\mathrm{var}(\tilde{S}_X^{\mathrm{GB}}(\omega)) \sim K_1 \frac{L}{N} S_X^2(\omega) \qquad (\omega \neq 0, \pm\pi) \tag{3.11}$$

$$\mathrm{bias}(\tilde{S}_X^{\mathrm{GB}}(\omega)) \sim \frac{K_2}{L^2} S_X''(\omega), \tag{3.12}$$

where K_1 and K_2 are constants that depend only on the windowing function. Notice, in (3.12), from the second derivative, that the greater the curvature, the greater the bias. Also notice the familiar bias-variance trade off. By increasing L, bias can be made smaller at the expense of variance, which increases, and vice versa. And finally, observe in (3.11) the more precise analog of (2.29, noting $\frac{1}{K} \approx \frac{L}{N}$).

These formulas are, in fact, leading order terms of a (nonasymptotic) expansion of bias and variance in N and L. These leading order terms are (by themselves) good approximations when N and L are sufficiently large. However, it seems that, in practice, that assumption can be difficult to realize and as a result the estimate exhibits, empirically, bias and variance that are not accurately modeled by the asymptotic formulas. Said differently, these classical formulas do not match the analogous empirical estimates. This not a critique on the asymptotic formulas found in the literature, just a statement that there does seem to be a place for formulas that do not assume $L \ll N$. What I do in the next section is develop some of the higher order expansions of the bias and variance in N and L to get a fuller picture of what affects the bias and variance for more general values of N and L.

3.3 Higher order expansions of bias and variance

In this section, I state the results of an expansion of the bias and variance formulas for the Bartlett estimator. It extends the work of Parzen [23] which gives (3.12) and (3.11). For an in-depth development of the formulas to follow see, Appendix A.

3.4 Bias

Theorem 1 provides a few additional terms of the higher order expansion for bias. For a full explanation see Appendix A. Before stating the theorem I would like to briefly explain the parameters referenced therein. The main idea is to form an expansion of the lag window generator of the form

$$\lambda(u) = 1 - \lambda_1^{(r)}|u|^r - \lambda_2^{(r')}|u|^{r'} + O(|u|^{r''}) \tag{3.13}$$

where the precise definitions of the parameters r, r', r'', $\lambda_1^{(r)}$, and $\lambda_2^{(r')}$ (all depending solely of λ) are given in Appendix A.1. Another important player in the Theorem 1 is what Parzen referred to as the "generalized r^{th} derivative," namely

$$S^{(r)}(\omega) = \sum_{n=-\infty}^{\infty} |n|^r C[n] e^{-in\omega}, \tag{3.14}$$

where C is the true autocorrelation function of the process X. I now state the theorem.

Theorem 1. *Let X be a WSS process with autocovariances that decay exponentially at rate ρ (as in (2.4)) and with snapshot x of length N. And suppose $\tilde{S}(\omega)$ is the Bartlett estimate with lag window in scaled parameter form with truncation parameter L, the lag window generator λ being even with $\lambda(0) = 1$, then the bias may be expanded as follows.*

$$\begin{aligned} b(\omega) = &- \frac{\lambda_1^{(r)}}{L^r} S^{(r)}(\omega) - \frac{\lambda_2^{(r')}}{L^{r'}} S^{(r')}(\omega) && \text{(3.15a)} \\ &- \frac{1}{N} S^{(1)}(\omega) + \frac{\lambda_1^{(r)}}{NL^r} S^{(r+1)}(\omega) + \frac{\lambda_2^{(r')}}{NL^{r'}} S^{(r'+1)}(\omega) \\ &+ O\left(\frac{1}{L^{r''}}\right) + O\left(\frac{1}{NL^{r''}}\right) + O(\rho^N) && \text{(3.15b)} \end{aligned}$$

where r and $\lambda_1^{(r)}$ are as in (A.3), r' and $\lambda_2^{(r')}$ are as in (A.6), r'' is as in (A.8) and $S^{(r)}$ is the generalized r^{th} derivative.

Observe that the generalized derivatives have no dependence on the lag window and the lag window only effects r, r', r'', $\lambda_1^{(r)}$, and $\lambda_2^{(r')}$. For convenience, I have computed these quantities for a few common lag windows and displayed them in Table 3.2. All of the windows in Table 3.2 are in scaled parameter from.

By way of example, I report the bias for the Bartlett window and the Parzen window in Corol-

Lag window	r	$\lambda_1^{(r)}$	r'	$\lambda_1^{(r')}$	r''
Bartlett	1	1	-	-	-
Daniell	2	$\frac{\pi^2}{6}$	4	$-\frac{\pi^4}{120}$	6
General Tukey	2	$a\pi^2$	4	$-\frac{a\pi^4}{24}$	6
Parzen	2	6	3	-6	-
Bartlett-Priestly	2	$\frac{\pi^2}{10}$	4	$-\frac{\pi^4}{40}$	6
Blackman	2	$\frac{41}{1000}\pi^2$	4	$-\frac{89\pi^4}{1200}$	6

Table 3.2: Characteristic exponents and their associated coefficients for the first and second order and the third characteristic exponent for 6 common spectral windows.

laries 1.1 and 1.2, respectively.

Corollary 1.1. *Supposing all the conditions of Theorem 3 are met, if the Bartlett window is used, the bias of this estimate is*

$$b(\omega) = -\left(\frac{1}{L} + \frac{1}{N}\right) S^{(1)}(\omega). + \frac{1}{NL} S^{(2)}(\omega) + O(\rho^N) \tag{3.16}$$

Corollary 1.2. *Supposing all the conditions of Theorem 3 are met. If the Parzen window is used, the bias of this estimate is*

$$b(\omega) = -\frac{6}{L^2} S^{(2)}(\omega) + \frac{6}{L^3} S^{(3)}(\omega) - \frac{1}{N} S^{(1)}(\omega) + \frac{6}{NL^2} S^{(3)}(\omega) - \frac{6}{NL^3} S^{(4)}(\omega) + O(\rho^{L/2}). \tag{3.17}$$

A demonstration As a concrete example I now demonstrate how these higher orders terms manifest themselves. Consider the spectrum from Section 3.1.1, namely,

$$S(\omega) = Ce^{-\frac{\omega^2}{\sigma}} = 10^{-2x^2+2} \tag{3.18}$$

As in Figure 3.1a, I generated 100 realizations and estimated the power spectrum of each using the generalized Bartlett procedure with the Parzen lag window. From this I compute the average and subtract from that the known spectrum to produce an empirical estimate of the bias. Figure 3.3a shows three curves (1) the empirical bias, (2) the asymptotic formula for the bias to leading order

(the application of equation (3.12), and (3) near-exact formula for bias found under (3.17), excluding the final term. Observe that for these values of $N = 40{,}000$ and $L = 50$, the asymptotic formula does not predict the bias in this estimator. The reason for this is that $L = 50$ is not sufficiently small compared to $N = 40{,}000$. However, if we consider terms from Corollary 3.2, which does not assume L is much smaller than N, we are able to match the empirical bias. In this example the term $S^{(3)}(\omega)$, in Corollary 3.2, dominates for frequencies greater than about 1.5 in absolute value. The size of the $S^{(3)}(\omega)$ term is a function of the original spectrum.

We now turn our attention to the variance formula.

3.5 Variance

I now report the higher order expansion of the variance of the Bartlett estimate. And for a full explanation of the proof, again, please see Appendix A.2.

Theorem 2. *Let X be a WSS process with autocovariances that decay exponentially at rate ρ (as in (2.4)), so that the power spectrum S is smooth. Let x be a snapshot of X of length N. And suppose $\tilde{S}(\omega)$ is the Bartlett estimate with lag window in scaled parameter form with truncation parameter L, the lag window generator λ being even with $\lambda(0) = 1$, and let W_L be the spectral window. Then the variance of $\tilde{S}(\omega)$ may be expanded as follows.*

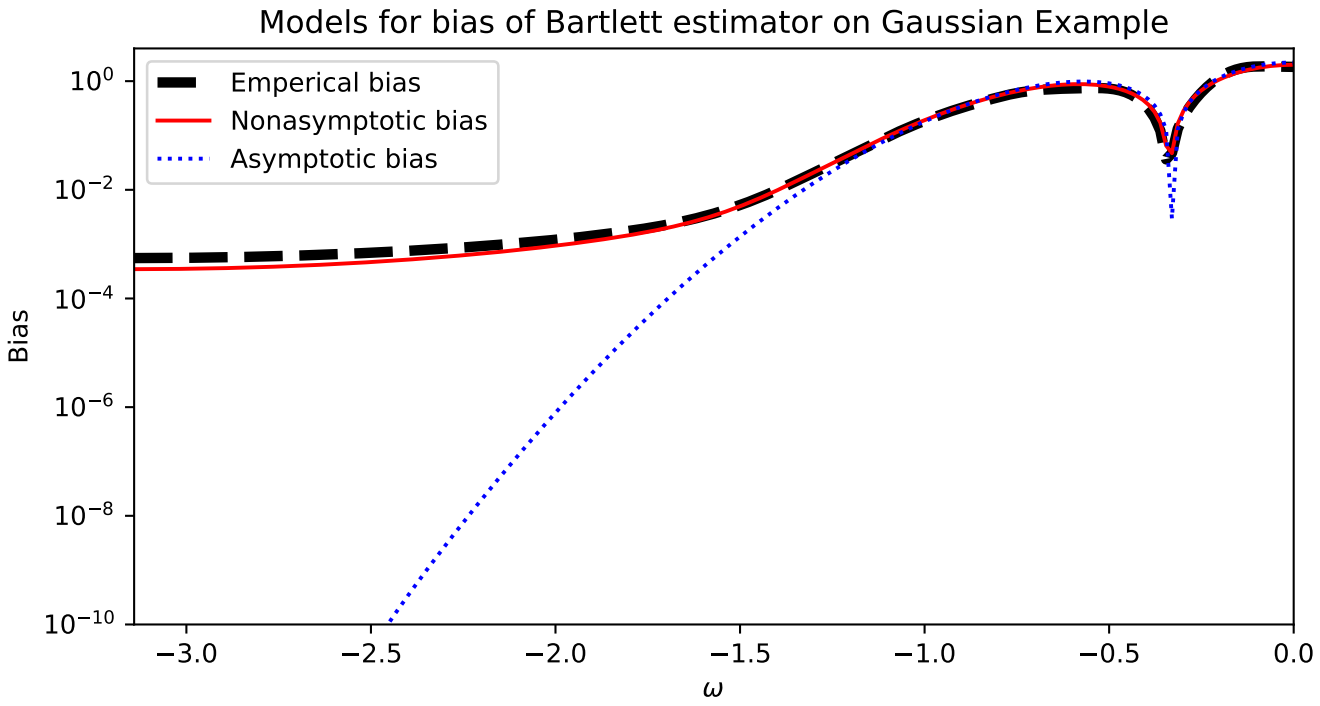
$$\begin{aligned}\operatorname{var}\left(\hat{S}(\omega)\right) =& (1+\delta_{\omega,0,\pi})\frac{2\pi L}{N}S^2(\omega)\int_{-1}^{1}\lambda^2(u)\,du + \frac{C_{\lambda,L}}{N} + \frac{Err(L,\omega)}{N} && (3.19)\\ &\frac{4\pi S(\omega)C^{\pm}_{S,W,L}(\omega)}{N^2} + \frac{4\pi K_{S,W,L}(\omega)}{N^2 L} + o(1/N^2L) && (3.20)\\ &+ k_e\left(\frac{S^2(\omega)}{N} + \frac{2C_{S,W}S(\omega)}{NL} + o(1/NL)\right) && (3.21)\end{aligned}$$

where the coefficients $C_{\lambda,L}$, $Err(L,\omega)$, $C^{\pm}_{S,W,L}(\omega)$, $K_{S,W,L}(\omega)$, k_e and $C_{S,W}$ are all defined in Appendix A.2.

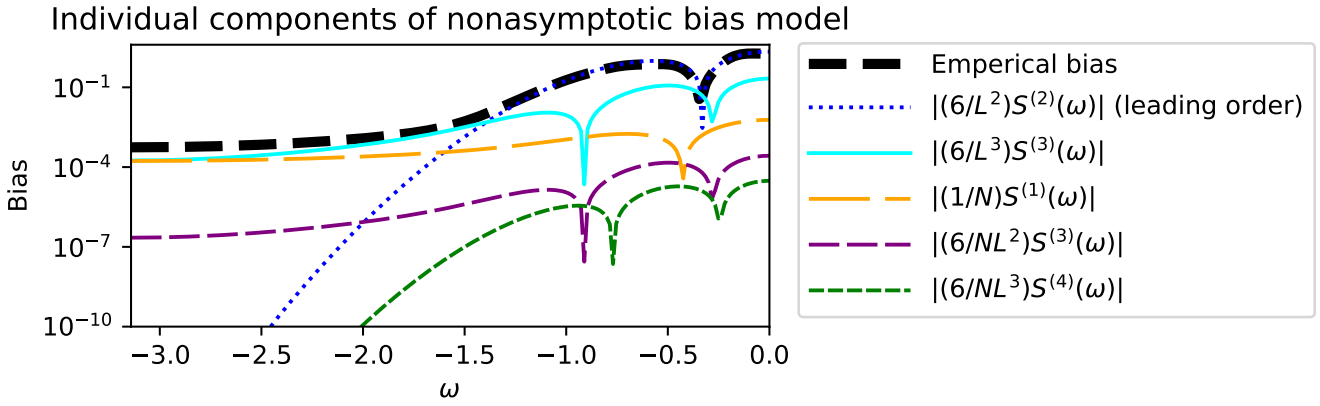
The parameter k_e vanishes if the process X is Gaussian.

3.6 Numerical study of errors

In this section, I demonstrate that for frequencies away from the estimation floor the leading order formulas found under (3.12) and (3.11) accurately model the scaling of the bias and variance. At the



(a) Empirical bias with leading order and high order formula. Three curves for the absolute bias for the Bartlett estimator. (1) The empirical bias computed from an ensemble of 100 snapshots of length $N = 40{,}000$, Thick black dashed line (2) Nonasymptotic bias (5-term full model) solid red line, (3) asymptotic bias (leading order) blue dotted line. Observe that the leading order asymptotic mode does not accurately predict the bias of the estimator for frequencies less than -1.3. Whereas the full bias model does. accurately predict the bias of the estimator through out the spectrum.



(b) Empirical bias with leading order and high order formula. Each of the five components of the full bias model. Observe that because L is not large enough to subdue the bias driven by the third generalized derivative term, which is comparable to the first generalized derivative term.

Figure 3.3: Demonstration of bias expansion. Here is a demonstration of how the terms that are higher order than the leading order term may dominate bias when L and N are not sufficiently large.

Parameter	Set of values
L	10; 20; 40; 80; 160; 320
N	1,029; 2,025; 4,375; 8,505; 16,807; 32,805

Table 3.3: The values of L and N for each of the two "runs" of the study of the scaling of the bias and variance of the IW power spectrum in L and N.

estimator floor, however, the scaling of the bias and variance cannot be modeled by these formulas. Rather the empirical scaling at the estimator floor suggests that one of the higher order term in the above expansions dominate. This suggests to me a the presence of an relation between the accuracy of the estimate and its ability for its statistical information to be predicted by (3.12) and (3.11). Such a relation may make for interesting work in the future.

3.6.1 Setup

This study considers the scaling in L and N of the empirical bias and variance of the estimated power spectrum for the Gaussian example from Section 3.1.1 and the ARMA(1,4) example from Section 3.1.2. For each of the two processes 1, 600 snapshots were sampled. Each of length $N_{\max} = 32{,}805$, Bartlett with the Parzen window was then used to estimate the spectrum of each snapshot with the following values for L and N. for a total of 30 different different spectral estimates for each snapshot.

3.6.2 Results

Figures 3.4 and 3.5 show the results of the study. The plots demonstrate the scaling of the bias and variance for the Bartlett estimate at and off of the estimator floor. I will go through each of figure in turn. Figure 3.4 pertains to the Gaussian process. Figure 3.4a shows the scaling of the bias at ($\omega = -\pi$) and off ($\omega = 0.13$) the estimator floor. Below that, in Figure 3.4b, is shown the scaling of the variance at the same two frequencies as for the plots of the bias scaling above. At the estimator floor the scaling of both bias and variance cannot be modeled by the leading order term of their respective formulas from Theorems 1 and 2. Rather, we see that at the estimator floor other terms in those expansions dominate (as was shown in the demonstration in Section 3.4, see Figure 3.3).

At the floor, the scaling for bias can be fit fairly well be a linear combination of the terms L^{-2},

L^{-3}, N^{-1}, $N^{-1}L^{-2}$, $N^{-1}L^{-3}$, the terms from (3.17). Off the floor we see the $1?L^2$ scaling with no dependence on N. Notice that as L increase the bias plot becomes more erratic. This is reasonable considering only 1600 independent realizations were used and the variance is more than 100. So, that one standard deviation of the bias (computed as the mean of the 1600 spectral estimates) would be around 0.25, which is around the order of the bias for the larger values of L. This make it not unreasonable to suspect that the $1/L^2$ scaling persists as L increases.

For the scaling of the variance, first, Figure 3.4b demonstrate that at the floor the variance is dominated by a term with basically no dependence on L, specifically a term proportional to $1/N^2$. Suggesting that at the floor the higher order term in $1/N^2$ dominates the L/N. However, off the floor we do see the asymptotic scaling from (3.11).

Figure 3.5 is the same as Figure 3.4 in all respects except that it for the ARMA(1,4) process. However, the same conclusion hold as for the Gaussian example.

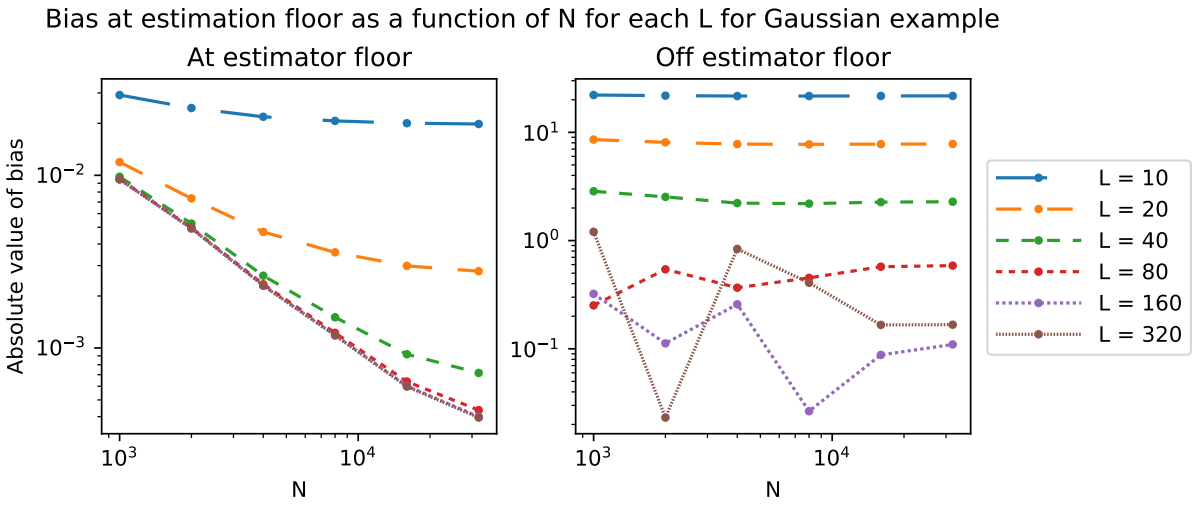
3.6.3 Discussion

We have just seen an expansion of bias and variance in N and L with out assuming $1 \ll L \ll N$. I have not been able to find these expansions anywhere in the literature. We have also seen the possible relation between an estimators accurate spectral estimation and its conformity to the leading order scaling formula. This was suggested by the scaling being dominated by a higher order term at the estimator floor and the leading order term off the estimator floor.

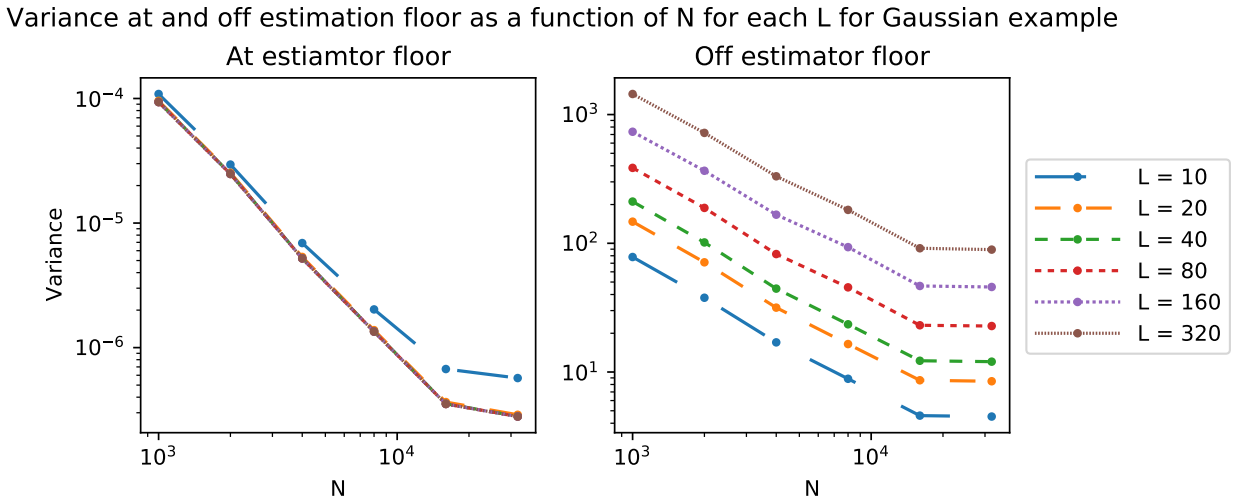
In the demonstration, the Parzen window was used. However, hopefully it was clear given Theorem 1 together with Table 3.3 that altering the window will not change the performance of the Bartlett estimate in a way that will significantly reduce the estimator floor.

As was mentioned, it seems that for high-contrast spectra N and L may need to be prohibitively large before the above formulas accurately model the bias or variance of the estimator.

What I have learn from this analysis together with the numerical study associated with it is that when L and N are not sufficiently large the bias and variance are complex functions of the structure of spectrum being estimated and the lag window. I have provided an analysis that attempts to better expose the complexities of these relations.

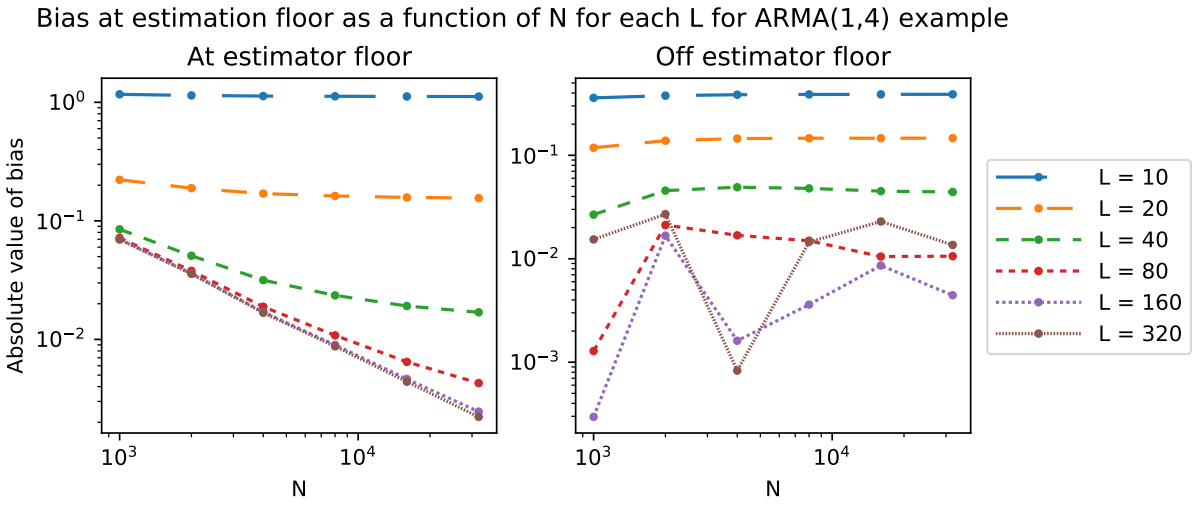


(a) Scaling of Bias of Bartlett estimate for Gaussian example. On the left is the scaling for the empirical bias for the Bartlett estimate of the Gaussian process for a frequency at the estimator floor ($\omega \approx -\pi$). On the right is the scaling for the estimator at a frequency off the estimator floor the estimator floor ($\omega \approx 0.13$).

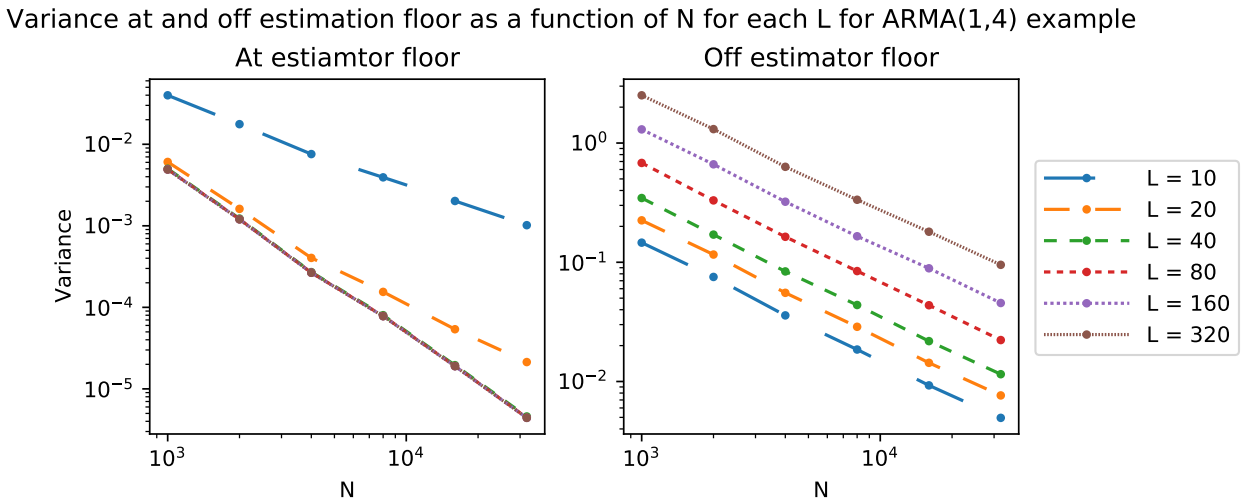


(b) Scaling of variance of Bartlett estimate for Gaussian example. On the left is the scaling for the empirical variance for the Bartlett estimate of the Gaussian process for a frequency at the estimator floor ($\omega \approx -\pi$). On the right is the scaling for the estimator at a frequency off the estimator floor the estimator floor ($\omega \approx 0.13$).

Figure 3.4: Scaling of Bias and variance of Bartlett estimate for Gaussian example. The results of the second run (1600 realizations) showing empirical bias and variance as a function of N for each L for the Gaussian process. Observe that for both examples there is a significant difference between the scaling of both the bias and variance at the estimator floor and off of it. Interestingly, off the estimator floor the scaling approximates that of the leading order terms of the bias and variance formulas, respectively. For bias at the estimator floor, we see a complex scaling that can be modeled by a linear combination of terms for the expansion in Theorem 1. While, off the estimation the plots suggest the $1/L^2$ scaling from (3.12). Similarly, the variance off the floor suggests the leading order term from (3.11) dominates. But at the floor, it appears that the $1/N^2$ term is dominating.



(a) Scaling of bias of Bartlett estimate for ARMA(1,4) example. On the left is the scaling for the empirical bias for the Bartlett estimate of the ARMA(1,4) process for a frequency at the estimator floor ($\omega \approx \pi/2$). On the right is the scaling for the estimator at a frequency off the estimator floor the estimator floor($\omega \approx 3$).



(b) Scaling of variance of Bartlett estimate for ARMA(1,4) example. On the left is the scaling for the empirical variance for the Bartlett estimate of the ARMA(1,4) process for a frequency at the estimator floor ($\omega \approx \pi/2$). On the right is the scaling for the estimator at a frequency off the estimator floor the estimator floor ($\omega \approx \pi/2$).

Figure 3.5: Scaling of Bias and variance of Bartlett estimate for ARMA(1,4) example. The results of the second run (1600 realizations) showing empirical bias and variance as a function of N for each L for the ARMA(1,4) process. We see that these results match those of the Gaussian process in Figure 3.4.

Chapter 4

Iterated whitening

I now discuss the IW method itself. It is a means of estimating power spectrum with a large spectral dynamic range. It is a nonparametric method that can achieve good results with not very many user decisions; I will talk about this more in Section 4.2.3. The way this is done is by iteratively filtering (whitening more specifically) the data to make the resulting spectrum more amenable to estimation and then use (2.23) to reshape the accurate estimate of the spectrum of the filtered data to the spectrum of the original data.

In this chapter I present the method of IW. Before I do, however, I will review a little more signals and systems theory, namely, the concept of spectral factorization and the closely related tools of modeling and whitening filters.

4.1 Power spectrum estimation and filtering

The next two subsections are additional background for the IW method.

4.1.1 Spectral factorization and modeling and whitening filters

Observe that the spectrum of a white noise process e is

$$S_e(\omega) = \sigma^2. \tag{4.1}$$

Suppose, for a process X (WSS), that there is a filter with frequency response $L(\omega)$ such that

$$S_X(\omega) = |L(\omega)|^2. \tag{4.2}$$

If $\ell = (\ell_n,\ n \in \mathbb{Z})$ is the impulse response of the filter and e is white noise with unit variance, then the process $\tilde{X} = \ell * e$ will have the same spectrum as X, since by (2.23),

$$S_{\tilde{X}}(\omega) = S_{\ell * e}(\omega) = |L(\omega)|^2 S_e(\omega) = |L(\omega)|^2 = S_X(\omega). \tag{4.3}$$

And so, $\tilde{X}$ *models* X in the sense that they have the same spectra, and therefore autocovariance information. We could also say that ℓ (more directly L) reshapes the flat spectrum of e to that of X. The system represented by ℓ is referred to as a *modeling* filter (sometimes a *reshaping* filter).

The factorization in (4.2) is guaranteed to exist since $S_X(\omega) \geq 0$. A straight forward example is

$$L(\omega) = \sqrt{S_X(\omega)}. \tag{4.4}$$

The filter represented in the frequency domain by $L(\omega)$ is, therefore, a modeling filter for the process X. In general the filter represented by (4.4) is noncausal (from Section 2.2) and may have a (bilaterally) infinite impulse response ℓ, since its Fourier series might only be expressible as

$$L(\omega) = \sum_{j=-\infty}^{\infty} \ell_j e^{-ij\omega}, \tag{4.5}$$

with $\ell_j \neq 0$ for infinitely many j. In practice this filter could be truncated and approximated by

$$L(\omega) \approx \sum_{j=-M}^{M} \ell_j e^{-ij\omega}. \tag{4.6}$$

Though, M may need to be quite large for the approximation to be of any use.

For a special class of factors, which the literature simply designates *spectral factors*, it is more natural to speak in terms of z-transforms rather than DTFTs. Bearing that in mind, $L(\omega) = \overline{L}(e^{i\omega})$ is a *spectral factor* if

$$\overline{S}_X(z) = \overline{L}(z)\overline{L}^*(z^{-*}) \tag{4.7}$$

and $\overline{L}(z)$ is minimum phase (from Section 2.2). Here $\overline{S}_X(z)$ is the z-spectrum of X and $z^{-*} = 1/z^*$ is a notation convention I will use. The existence of a spectral factor is guaranteed provided the so-called Paley-Wiener condition, namely,

$$\int_{-\pi}^{\pi} \log S_X(\omega)\, d\omega > -\infty \tag{4.8}$$

is met [24, 25, 26]. This factorization is not unique since for any spectral factor L of S_X, right-multiplication by a complex number on the unit circle U produces another spectral factor of S_X, since $(UL)(UL^*) = LL^* = S_X$. However, by imposing an additional condition, it may be made unique among spectral factors. If we require $L(0) = \overline{L}(1)$ to be real and nonnegative then L is

unique and I refer to this as the *canonical spectral factor* of S_X. The function $\overline{L}(z)$ is to be regarded as the transfer function[1] of what will then be a modeling filter. The reason this class of spectral factors is significant is that the resulting modeling filter is stable and causal and the inverse filter is also stabel and causal, by the definition of $\overline{L}$ being minimum phase.

In the usual case where the true power spectrum is unknown it is reasonable, if one seeks a causal modeling filter, to estimate the spectrum as a Laurent polynomial[2]. Such, for example, would be the case if a Bartlett estimate were used. There are a number of methods that may be employed to factor such functions (see e.g. [27]). It can be shown that a Laurent polynomial P of order m

$$P(z) = \sum_{n=-m}^{m} c_n z^{-n} \tag{4.10}$$

may be factored exactly such that

$$P(z) = \overline{L}(z)\overline{L}^*(z^{-*}) \qquad \text{and} \qquad \overline{L}(z) = \sum_{n=0}^{m} \ell_n z^{-n}. \tag{4.11}$$

Observe that given a spectral factor L, one can then recover the impulse response of the modeling filter by inverse Fourier,

$$\ell_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} L(\theta)e^{in\theta}. \tag{4.12}$$

Recognizing that the modeling filter takes in a white noise process and returns a process of spectrum S_X, it is clear that feeding a process of spectrum S_X into the inverse of the modeling filter will return a white noise process (that is, a process with unit power spectrum). This is manifestly true when we consider equation (2.23) together with the fact that the transfer function of the inverse filter is the multiplicative inverse of the transfer function of the filter. If we let $\mathcal{L}$ be a modeling filter for X with frequency response $L(\omega)$ and let $\mathcal{W} = \mathcal{L}^{-1}$ be the inverse filter with frequency response $W(\omega)$ and impulse response $w = (w_n,\ n \in \mathbb{Z})$ then

$$S_{w*X} = W(\omega)S_X(\omega)W(\omega)^* = S_X(\omega)|L(\omega)|^{-2} = S_X(\omega)/S_X(\omega) = 1. \tag{4.13}$$

[1] Recall the distinction between the transfer function and frequency response of a filter.

[2] A Laurent polynomial is an expression of the following form,

$$\sum_{n=-m}^{m} c_n z^{-n} \qquad \text{where } c_n^* = c_{-n}. \tag{4.9}$$

So that the filter computed from the inverse of the modeling filter returns a white noise process when X is passed through it. This is why the inverse of the modeling filter is called a *whitening* filter. *Whitening* a process is to pass it through a whitening filter.

4.1.2 Prefiltering

Possibly first suggested by Press and Tukey in 1956 [28][3, Sec. 7.4.1], prefiltering is a method for enhancing the performance of a power spectrum estimator. Some of the most difficult spectra to approximate using Bartlett's method are those with sharp or narrow peaks. Narrow is meant in the sense that the width of the peak at half power is relatively small. I will explain why this is and what I mean by small. Loosely speaking, the narrower the peak, the greater that the resolution of the spectral window must be to resolve it. This is problematic since the resolution of the spectral window depends inversely on L, the truncation parameter, and as we saw, increasing the truncation parameter increases variance. In turn, in order to keep variance at a tolerable level the length N of the snapshot needs to be very large, often prohibitively large. So, the narrower the peak the larger L must be to resolve it and the larger N must be to control the variance.

One way to proceed then is to smooth out the peak in the power spectrum. As we have seen we can affect the power spectrum by filtering the data appropriately. One might construct a filter based on some *a priori* knowledge of the power spectrum if such knowledge is available. Once that is done the power spectrum of the filtered data can be estimated then post processed to account for the affect of filtering to obtain an estimate of the original power spectrum. The theory hangs on the very useful result under (2.23). The result is applied as follows. Given the process X and some filter with impulse response $g = (g_j,\ j \in \mathbb{Z})$ then

$$S_X(\omega) = \frac{S_{g*X}(\omega)}{|G(\omega)|^2} \tag{4.14}$$

where G is the frequency response of the filter. The potential effectiveness of this proceedure comes from the fact, from Chapter 3, that the bias and variance of a Bartlett estimator depends of the true power spectrum and the smoother the spectrum the lower the bias. Hence, $S_{g*X}(\omega)$ may have a smaller bias.

The best case scenario is that the filter g approximates a whitening filter obtained by some pilot spectral estimate. In that case this procedure is usually referred to as *prewhitening*. This form of prefiltering is sometimes done by a very crude whitening filter of 2 taps and is called *AR(1)*

whitening. An AR(1) estimate of the process is very cheap to produce and needs not require a pilot power spectrum estimate. This may be what one would want to do if the time series is very short, also. Which brings me to an important word on caution.

There is a danger in making filters for the purpose of prefiltering too long. To see this let $g = (g_j,\ j = -M, \ldots, 0, \ldots, M')$ be the prefiltering filter and $x = (x_n.\ n = 1, \ldots, N)$ be the snapshot. Then the snapshot y of the filtered process $Y = g * X$ might be computed as

$$y_n = \sum_{j=-M\vee(n-N)}^{M'\wedge(n-1)} g_j x_{n-j} \qquad \text{for } n = 1, \ldots, N \tag{4.15}$$

However, the first term y_1 of the snapshot is clearly not a sample from $\sum_{j=-M}^{M'} g_j X_{1-j}$ (X is assumed to be bilaterally infinite) but rather is a sample from $\sum_{j=-M}^{0} g_j X_{1-j}$. The term y_1 from (4.15) I refer to as being *partially whitened*. And so, a truncation is necessary to ensure that y is a snapshot of the right (stationary) process. My experience has been that not truncating partially whitened terms from the approximately whitened snapshot can contribute significant bias to the estimated power spectrum of the whitened process. That is why I suggest, for approximately whitened data, using

$$y_n = \sum_{j=-M}^{M'} g_j x_{n-j} \qquad \text{for } n = M'+1, \ldots, N-M. \tag{4.16}$$

And so, whitening exacts a cost of $M + M'$ data points. And so very little is lost in performing AR(1) whitening. It is also faster to have short whitening filter. It will be seen in Section 5.2 that filtering the snapshots by convolution commonly constitutes the leading order in the runtime analysis

One last point before delving into IW. For the true power spectrum of a WSS process X and a filter with impulse response g it is the case that

$$S_{g*X}(\omega) = |G(\omega)|^2 S_X(\omega), \tag{4.17}$$

again, G is the frequency response of the filter. However, because of the limitations of a spectral estimate (denoted by the tilde), I wish to emphasis the following fact.

$$\tilde{S}_{g*X}(\omega) \qquad \text{and} \qquad |G(\omega)|^2 S_X(\omega) \qquad \text{may differ significantly.} \tag{4.18}$$

This is of central importance in the upcoming theory of IW spectral estimation. One reason for this can be noted from the analysis in Sections 3.4 and 3.5. In particular under (3.15) we see that the bias depends heavily on the true power spectrum and as (4.17) suggests, the true spectrum of a filtered process is changed by filtering. So, the bias in $\tilde{S}_{g*X}(\omega)$ and the bias in $\tilde{S}_X(\omega)$ may be very different with no guarantee that multiplication by $|G(\omega)|^2$ will abate this difference. And with that fact firmly in mind, we now discuss the IW procedure.

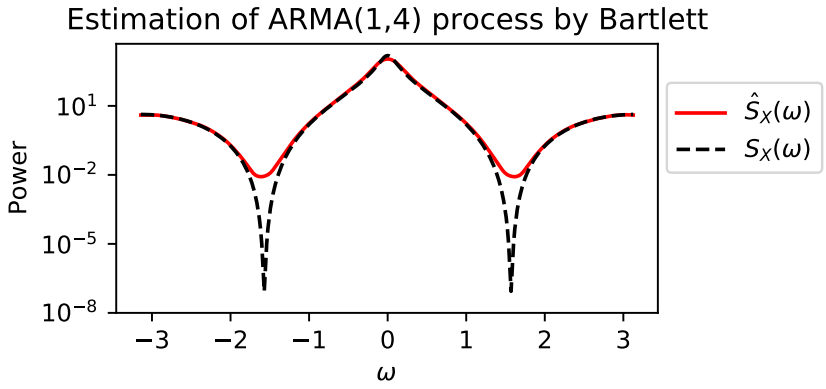
4.2 Iterated whitening

This section now presents the main novel contribution of the dissertation.

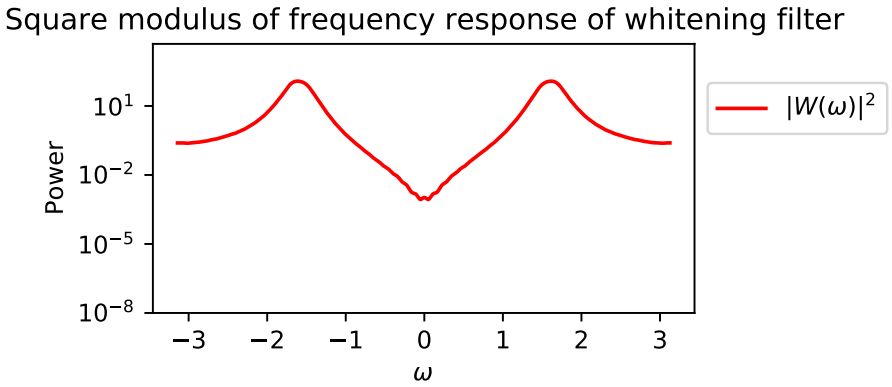
4.2.1 Theoretical underpinnings

Suppose I estimated the power spectrum of some process X from some snapshot x using a particular estimator and used that estimate to construct a whitening filter w as mentioned above. Further suppose I whitened the snapshot X by w to get $Y = w * X$. What would happen if I then used the same spectral estimator that was used to then estimate the spectrum of Y from $y = w * x$? Would it come out flat? The answer is that it depends on the accuracy of the original estimate. Consider the ARMA(1,4) example from Section 3.1.2 whose power is estimated by the Bartlett estimate from a snapshot, as shown in figure 4.1a. Observe the estimator floor in contrast to the deep troughs at $\pm\pi/2$ of the true power spectrum. In 4.1b, we see the modulus squared of the frequency response of the whitening filter ($|W(\omega)|^2$) drawn with the same y-axis as above. Its shape, rather than being the multiplicative inverse of the true spectrum as a true whitening filter would, is the multiplicative inverse of the estimated spectrum that spawned it. This demonstrates that for a whitening filter estimated by a spectral estimate its ability to whiten is limited according to the information in the estimated spectrum.

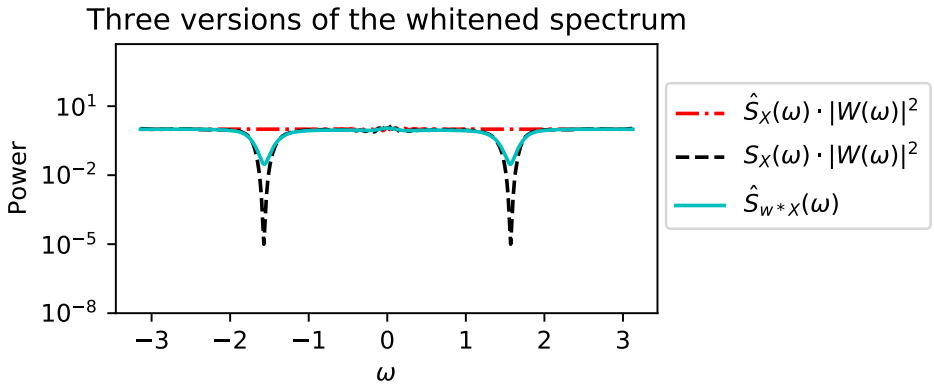
In Figure 4.1c, we see three different versions then of what might be considered the spectrum of the whitened process. If the power spectrum estimate was exact, these would all coincide. First, note that $|W(\omega)|^2$ multiplied by the original power spectrum estimate (dot-dashed red) is everywhere one. This is to be expected because the numerical spectral factorization and point-wise inversion of the estimated spectrum is usually done with a high degree of accuracy. Next, in Figure 4.1c is the true spectrum multiplied by $|W(\omega)|^2$. This is exactly $S_Y(\omega)$. Note that the very low parts of $S_X(\omega)$, were elevated to those of $S_Y(\omega)$) this must always be the case if spectrum estimator has a



(a) The spectrum of the ARMA(1,4) process estimated by Bartlett together with the true spectrum.



(b) Square modulus of the frequency response of the whitening filter. This determines how it will shape the *true* spectrum of the process passed through this filter.



(c) Three different versions of the spectrum of the whitened process. First is the modulus squared of the whitening filter ($|W(\omega)|^2$) multiplied by the spectrum that spawned it (dot-dashed red). Next is the true spectrum multiplied by $|W(\omega)|^2$ (dashed black). Lastly is the estimated spectrum of the filtered process (solid cyan).

Figure 4.1: An illustrated study of the effects of whitening on the true power spectra and estimated power spectra involved. Note that the range of the axes of each plot are the same.

estimator floor strictly below one. For if it does, the modulus square of the frequency response of the resulting whitening filter $|W(\omega)|^2$ will be greater than one at the frequencies associated with the estimator floor so that by multiplication, the true spectrum $S_X(\omega)$ will increase at those frequencies. Notice that, overall, $S_Y(\omega)$ is smoother than its predecessor particularly at every frequency that was accurately estimated by the original power spectrum estimate. The final curve in Figure 4.1c is the estimated spectrum of the filtered process. Most significantly this estimated spectrum is not flat, and the relatively "large" deviations from unity indicate frequencies where the original spectral estimate were inaccurate. To summarize this discussion I wish to highlight three key observations. Let X be the original process whose spectrum S_X is to be estimated by $\hat{S}_X$. Let w be the whitening filter computed from $\hat{S}_X$, and let $Y = w * X$ with true spectrum S_Y and approximated spectrum $\tilde{S}_Y$.

1. Frequencies ω where the original estimate are accurate ($\hat{S}_X(\omega)$ is close to S_X) will have a true "whitened" spectrum $S_Y(\omega)$ close to unity.

2. If, for some frequency ω associated with the estimator floor. The power spectrum estimate at that frequency is less than one, $\hat{S}_X(\omega) < 1$, then the true spectrum of the whitened process will be strictly greater than the true spectrum of the original process, $S_Y(\omega) > S_X(\omega)$

3. Deviation from unity of the estimated "whitened" spectrum $\hat{S}_Y$ suggest the original power spectrum estimate $\hat{S}_X$ is inaccurate.

4.2.2 The iterated whitening algorithm

I now present the method of IW. This is a procedure which may, in principle be applied to any spectral estimate, however, because we have the theoretical understanding of how the bias and variance of Bartlett estimates relate to the true spectrum we applied this procedure to estimator estimator. Given a snapshot $x = (x_n,\ n = 1, \ldots, N)$, truncation parameter L, and frequency grid size N_ω

1. Compute the empirical autocovariances $\tilde{C}_X[n]$ of x,

$$\tilde{C}_X[n] = \frac{1}{N} \sum_{\max(1,1-n)}^{\min(N,N-n)} x_{i+n} x_i^* \qquad \text{for } n = -L, \ldots, L. \tag{4.19}$$

2. Window the empirical autocovariances $\tilde{C}_X[n]$ with a windowing function $\Lambda_L[n]$ such as Parzen or Blackman, for example, to get $\Lambda_L[n] \cdot \tilde{C}_X[n]$, relabel this $\tilde{C}_X[n]$ and compute the spectral estimate

$$\tilde{S}_X(\omega) = \sum_{n=-L}^{L} \tilde{C}_X[n] e^{-in\omega}. \tag{4.20}$$

3. Extract a modeling $m^{(1)} = (m_n^{(1)},\ n = -L, \ldots, L)$ and whitening filter $w^{(1)} = (w_n^{(1)},\ n = -L, \ldots, L)$ from the (windowed) empirical autocovariances $\tilde{S}_X(\omega)$.

4. Using the whitening filter $w^{(1)}$, whiten the snapshot x and estimate the power spectrum $\tilde{S}_{w^{(1)} * X}$. (This may involve computing the autocovariances of the whitened process and windowing it.)

5. Inspect or otherwise test the whitened spectrum $\tilde{S}_{w^{(1)} * X}$ to determine if it is flat. If it is, proceed to step 6. If it is not flat, repeat steps 3-5 on the whitened process $w^{(1)} * x$ which can now be relabeled x. Repeat, until the whitened process $w^{(j)} * x$ has a flat spectrum. When it does proceed to step 6.

6. Compute the frequency response $H^{(j)}(\omega) = \sum_{n=0}^{L} m^{(j)} e^{-in\omega}$ for each modeling filter computed in step 3. Then point-wise multiply their square moduli to obtain the final spectral estimate $\tilde{S}_X^{\mathrm{IW}}(\omega)$ of the given time series

$$\tilde{S}_X^{\mathrm{IW}}(\omega) = \prod_j \left| H^{(j)}(\omega) \right|^2. \tag{4.21}$$

A few remarks If the spectral dynamic range is small then no iterations are require and the algorithm produces exactly the Bartlett estimate with the given lag window or, more generally, returns whatever spectral estimate was factored to produce the modeling filter.

One must remember to truncate the partially filtered terms of the "whitened" data. Failing to do so can add significant bias in the estimated spectrum of the whited process.

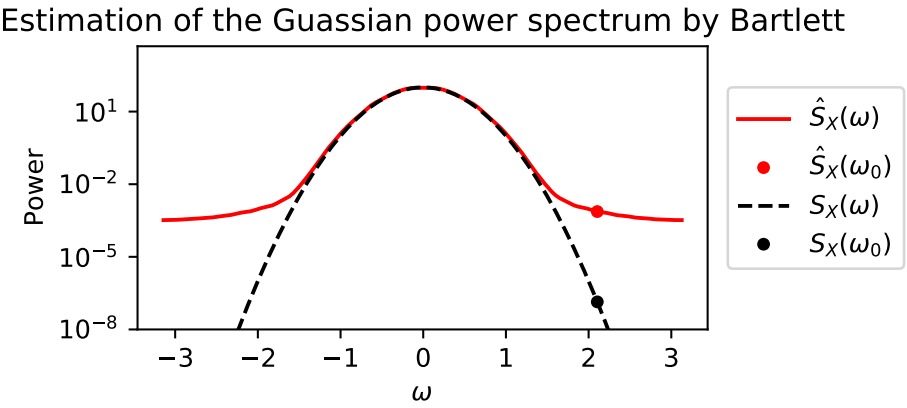
I would like to, now, provide a few insights to help us understand how this algorithm works. First consider the initial spectral estimate, which I have mind to be a Bartlett estimate, but any one will do. Suppose the spectrum to be estimated is high-contrast. There will be some frequencies for which the estimate is accurate and some for which it is not. Consider the case that the inaccurate estimate of the power for a frequency was too large, such as with the frequency ω_0 in Figure 4.2a, which

shows the Bartlett estimate of the familiar Gaussian example for Section 3.1.1. In my experience it is indeed most often the case that the estimator will over estimate low powers of spectra rather than under estimate high powers of spectra. Should the associated modeling filter for this step be utilized it will impart (or reshape) an excess of power to that frequency. Remember that the square modulus of the modeling filter is the spectral estimate. Similarly, the whitening filter (Figure 4.2b) will under whiten the spectrum at that frequency. But since the estimated power of the spectrum is still low (in the example in Figure 4.2a the estimated power is $\hat{S}_X(\omega_0) \approx 0.0012$ (≈ -29.3 dB)). So that the whitening filter will impart a gain or a factor of $|W(\omega_0)|^2 \approx 857 \approx 1/0.0012$ (or ≈ 29.3 dB) at that frequency. Figure 4.2b displays the square modulus of the whitening filter (of the first and only iteration in this demonstration). It is the graph of the factor by which S_X is multiplied to achieve S_{w*X}. This gain is quite a bit below the $\approx 7.2599 \cdot 10^6$ (≈ 68.6 dB) necessary to truly whiten the spectrum at that frequency (see 4.2c) but, as they say, it's a start. And after only one iteration is enough to bring the power of the "whitened" spectrum at this frequency close to the estimator floor of the estimator and so we expect the power of the original spectrum will be fairly well approximated at this frequency after only one iteration of the IW estimate.

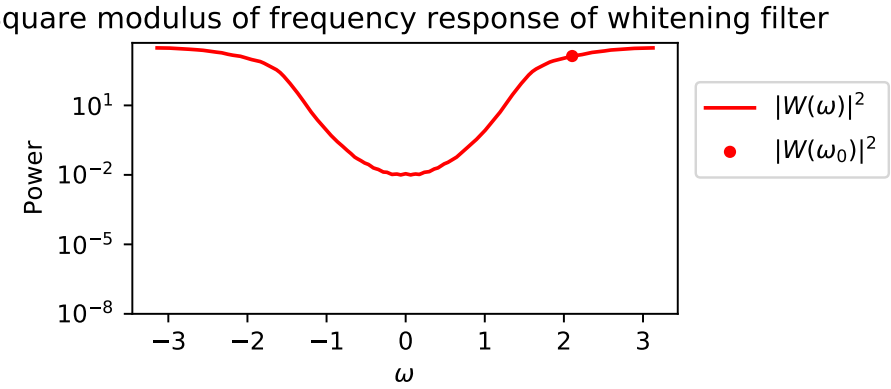
4.2.3 Numerical considerations and causal vs. noncausal

I would like to discuss the differences between the causal and noncausal (recall Section 2.2) approaches to IW. these terms where First of all, recall that a whitening filter is not unique and among the different ways of factoring an estimated power spectrum I have found two that posses particular advantages. The first is the canonical spectral factorization (seen under (4.7)). One advantage to this factorization is that when the power spectrum is in the form of a Laurent polynomial (as in (4.11)) of a particular order the spectral factor may be expressed exactly as a trigonometric polynomial of the same order. Meaning there is no truncation necessary when going from the spectral factor to the impulse response of the modeling filter. Furthermore, the modeling filter is guaranteed to be stable and causal and to have a stable, causal inverse. Appendix B discusses in great detail an efficient method for computing the canonical spectral factor of a power spectrum in the form of a Laurent polynomial (which will always be the case for a Bartlett power spectrum estimate). Unfortunately, going from the impulse response of the modeling filter to the impulse response of the whitening filter will generally require truncation.

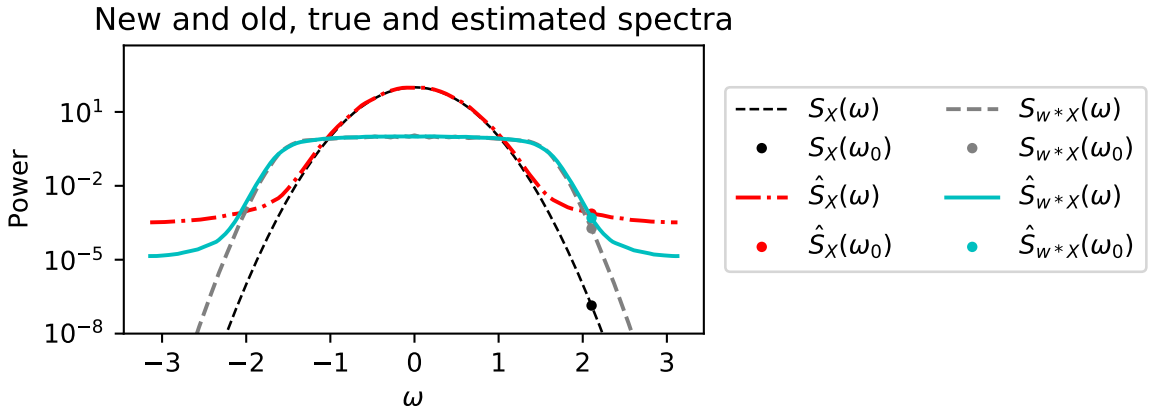
The second is the frequency-wise square root (seen under 4.4)). Its principal advantage is that it



(a) Bartlett estimate of Gaussian examples. Specifically it is the estimated spectrum of the example Gaussian process by a Bartlett estimator (solid red) together with the true Gaussian power spectrum (dashed black). The ordinates at ω_0 have been indicated by the dot.



(b) Frequency response of whitening filter modulus squared. The square modulus of the frequency response of the whitening filter with the point at ω_0 is plotted.



(c) Original and whitened spectrum with Bartlett estimates. Plotted here is the original true spectrum (thin dashed black) and the original spectral estimate (dot-dashed) red), from Figure 4.2a, plotted with the new ("whitened") true spectrum (dashed gray) and estimated spectrum (solid cyan). All associated ordinates at ω_0 have been emphasized

Figure 4.2: Demonstration of IW. As in Figure 4.1a, I illustrate a study of the effects of whitening on the true power spectra and estimated power spectra involved, but this time a particular point, at $\omega_0 = 2.1$ is emphasized to better see how the "whitened" spectrum is more amenable to estimation.

is very simple to implement (particularly when compared with the numerical spectral factorization discussed in Appendix B). A possible disadvantage is the the noncausal filter may be less stable. Through the course of experimenting on IW power spectrum estimation with certain data (the Kuramoto-Sivashinsky data to be discussed in Section 8.2) it was found that the noncausal whitening version exhibited some stability issues. It turns out that these particular issues disappeared when an odd frequency grid for the power spectrum estimates estimates were used.

I conclude with an open question that may be worth investigating. In my experimentation the causal whitening filter at each step was designed to have $L+1$ taps where the noncausal filter had $2L+1$ taps. The upshot of this is that the noncausal filter expended twice as many data points per iteration as did the causal filter. It would be interesting to compare the performance of these implementation in which the cost of data points per iterations was equal.

4.2.4 Demonstration

Below, are demonstrations of this method being used to estimate the spectra of the two running examples introduced in sections 3.1.1 and 3.1.2. The plots demonstrate the effectiveness of this technique. The data consisted of 100 independent snapshots from each process each of 40,000 samples. Figure 4.3 shows the true spectrum of the Gaussian process together with the mean of all 100 spectral estimates at each of six stages or iterations of the IW procedure with the Bartlett estimate. The 90% confidence intervals for each stage are also displayed. Then in Figure 4.4 we see the true spectrum of the ARMA(1,4) process together with the mean of the 100 power spectrum estimates taken at six stages or iterations of the IW procedure again with the Bartlett estimate. Again, the 90% confidence intervals are included. A close-up of the deep trough is provided.

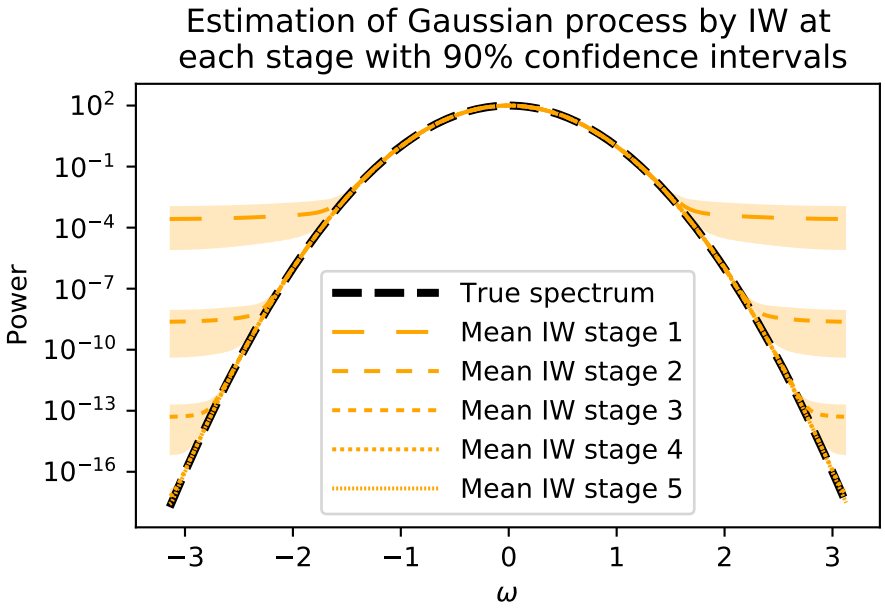


Figure 4.3: Power spectrum estimate of Gaussian example by IW. Each iteration of the IW estimate is displayed for the estimation of the Gaussian process example together with the 90% confidence interval for each estimate.

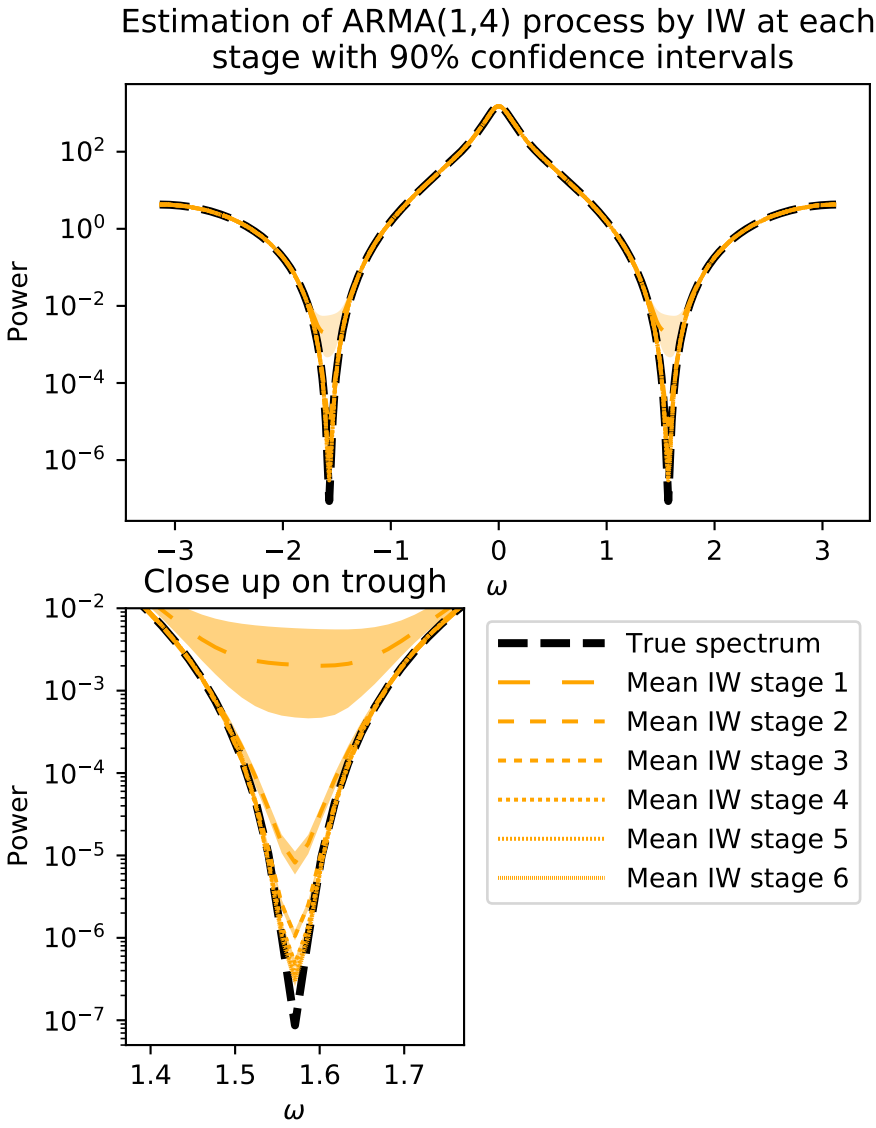


Figure 4.4: Power spectrum estimate of ARMA(1,4) example by IW. Here each iteration of the IW estimate is displayed for the estimation of the ARMA(1,4) process example together with the 90% confidence interval for each estimate.

Chapter 5

Analysis of iterated whitening

To provide a more full understanding of the method of IW, I study numerically how the bias and variance scale with the main parameters that can be chosen, the length N of the snapshot and the truncation parameter L of the lag window that is used. I hope this will help users of the method to be able to make better choices for these parameters. I leave to future work the theoretical investigation of these scalings. I conclude the chapter with a runtime analysis of the foregoing methods.

5.1 Numerical error analysis

I now describe the experiments that were run and the results.

5.1.1 Setup

This study, like the study in Section 3.6, investigates the scaling in L and N of the empirical bias and variance of the estimated power spectrum for the Gaussian example from Section 3.1.1 and the ARMA(1,4) example from Section 3.1.2. The estimators being studied are the Bartlett and noncausal IW estimates. The values of L and N range over same values as before in Section 3.6, which I reproduce here for convenience in Table 5.1.

As before, for each example, 1600 realizations of length $N_{\max} = 32{,}805$ were used and for each a total of 30 power spectrum estimates where computed (one for each pair of L and N). For the

Parameter	Set of values
L	10; 20; 40; 80; 160; 320
N	1,029; 2,025; 4,375; 8,505; 16,807; 32,805

Table 5.1: Values of L and N for each of the two "runs" of the study of the scaling of the bias and variance of the IW power spectrum in L and N.

Bartlett estimate, I used the Parzen window as usual. For the IW estimate, I used the noncausal implementation with a each time doing 10 iterations.

5.1.2 Results

We now compare the IW estimate to the Bartlett estimate at the estimator floor. Then, we look at the scaling off the estimator floor and conclude with a plot of the bias and variance for a fixed N, but varying L, each plotted against frequency. By the phrase "at the floor," I mean we are looking at the bias and variance of the power spectrum estimate at $\omega = -\pi$ for the Gaussian process and $\omega = \pi/2$ for the ARMA(1,4) process. These represent the lowest powered frequency of the true spectrum and consequently the frequency whose power is likely most difficult to estimate. When I say "off the floor" I mean $\omega = 0$ for the Gaussian process and $\omega = \pi/4$ for the ARMA(1,4) process. In the plots shown here, some of the points are missing. This is because the after 10 iterations, each iteration taking $2L$ data points per whitening step the snapshots run out of points (see Section 4.1.2).

At the estimator floor IW demonstrated a significant improvement over Bartlett. For the Gaussian process, Figure 5.1a show bias be very low, below 10^{-16} for each N and L. The ARMA process in Figure 5.2a demonstrates similar behave though it is not so low. Although, relative to the power of their respective spectra these seem to be on par with each other. Similar gains in accuracy are demonstrated towards the variance as seen in Figures 5.1b and 5.2b

Away from the estimator floor, we see a surprising change in performance. In Figure 5.3, we see that the bias for the Bartlett estimate for low values of L follows a distinct structure, which decays as L increases. At the same time, the variance plot reports that variances increase as L increases. Note also that the estimated standard deviation of the bias, given the associated variance estimate and the fact that the bias was computed using 1600 independent samples, is around the order of magnitude of the estimated bias. This suggest erratic behavior is because of the fluctuations in the data and may be overcome by increasing the number of realizations. This I did going from 100 to 1,600 realizations. However it appears that still may not be enough. Based on this data, we conclude that off the floor, the bias either does not vanish as L or $N \to \infty$, or does so very slowly.

For the variance, the right-hand plot in Figure 5.4b provides compelling evidence that the variance decrease in N and increase in L. This agrees with the right-hand plot in Figure 5.3b.

Figure 5.5 contains a profile of the bias and variance for three values of L and $N = 32,\ 805$

for the noncausal IW estimate. Here we can see that bias and variance has a slight dependence on L but only, it seems for the noncausal IW estimate. The Figure 5.5 also, therefore, demonstrates something about the distinction of between the causal and noncausal IW estimates.

We conclude from the data available that the IW estimates scales very well in bias and variance at the estimator floor. Meaning, these values can be controlled by varying N and L. However, off the floor the variance and bias of the IW estimator do not appear to be controllable by N or L. Whether they can be removed by other means will require further study.

An interesting thing can be observed by comparing the bias plots, Figures 5.5b, and 5.5a, with Figure 5.6a. In the latter we recognize a scalar multiple of the absolute value of the second derivative of the Gaussian (with an estimation floor, see Figure 3.3). Where in the (two) former, it can easily be verified that those are scalar multiples of the original spectrum. It can also be seen that the variance profile (across frequencies) for IW (causal and noncausal) approximates as multiple of the square of the original spectrum, though in Bartlett case of the Bartlett estimator this too has a floor.

5.2 Runtime analysis

I will now analyze the runtime for the algorithms discussed so far:

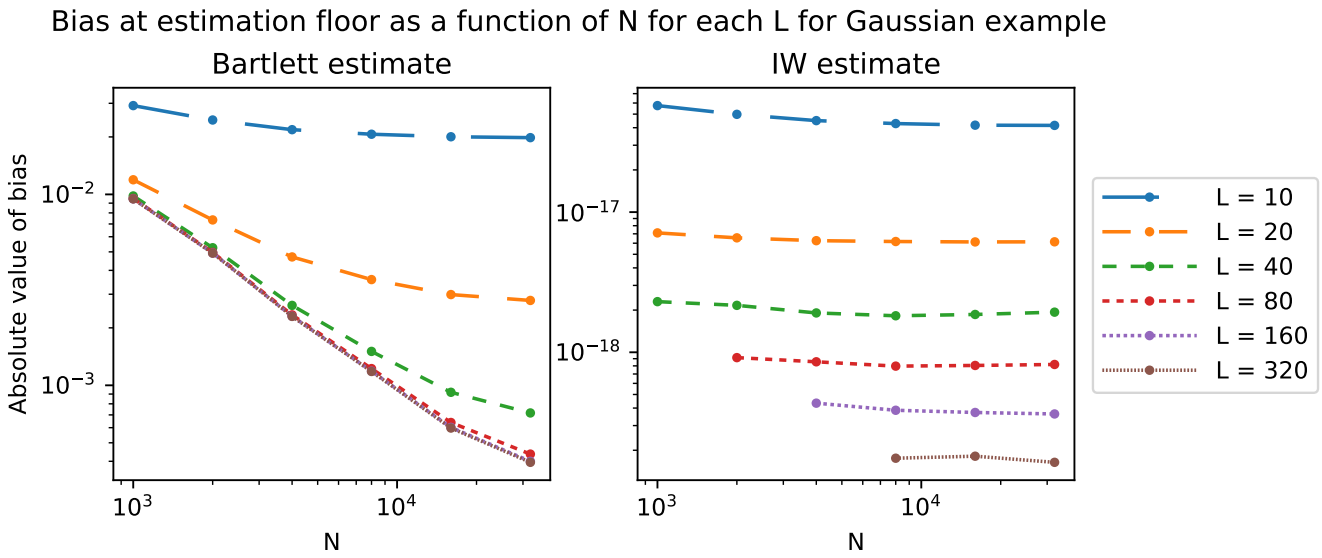
5.2.1 Bartlett

For Bartlett we have

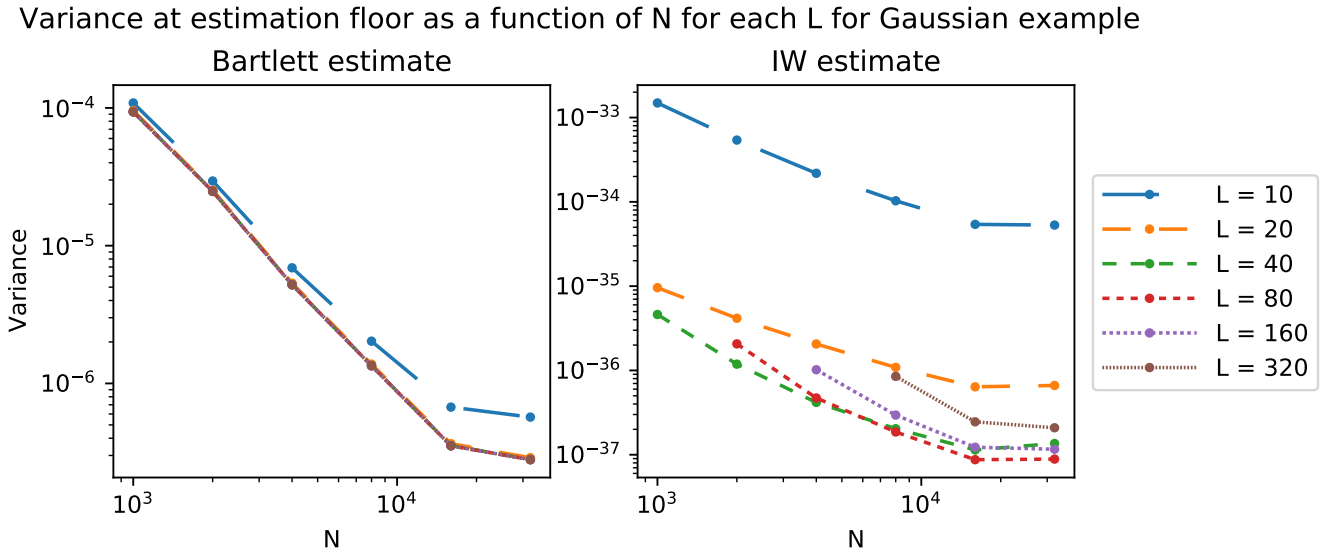
$$O(d^2LN) + O(d^2N_{\mathrm{f}} \log N_{\mathrm{f}}) \tag{5.1}$$

The main operations of these methods are computing the autocovariances and taking the its FFT. For the autocovariances we compute the sum found under (2.25). This can be computed two ways: (1) directly or (2) as a convolution using FFT. The direct approach involves the sum of $2L$ terms each term being a sum of a different size, namely, $N - |n|$ (n is the lag of the term)). This turns out to be $O(d^2LN)$ flops. If one uses a convolution we are looking at $O(d^2N \log(N))$ flops. It should be noted that if N is the product of small primes the "$\log(N)$" in that expression can be reduced.

The FFT is used to go from time domain to frequency domain and depends on the number of points in the desired frequency discretization N_{f} and is simply $O(d^2N_{\mathrm{f}} \log N_{\mathrm{f}})$. My observation has been that, for the most part, the resulting spectral estimate is insensitive to N_{f}. However, choosing

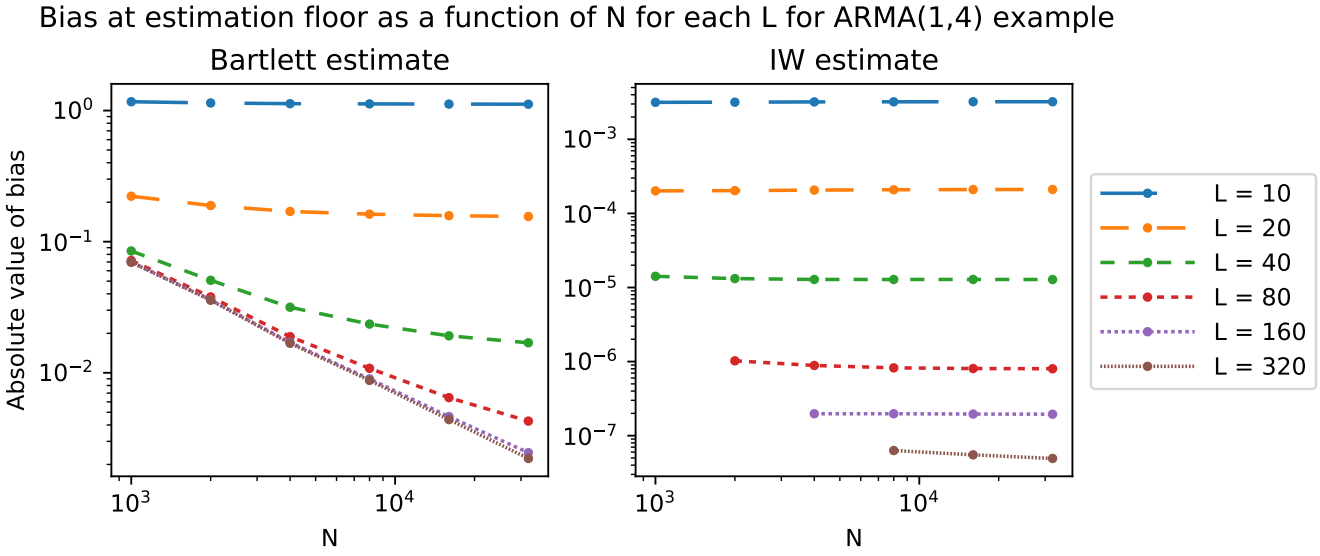


(a) Bias scaling at the estimator floor for the Bartlett estimator (left) and for the IW (noncausal) estimate (right). They are estimating the Gaussian power spectrum.

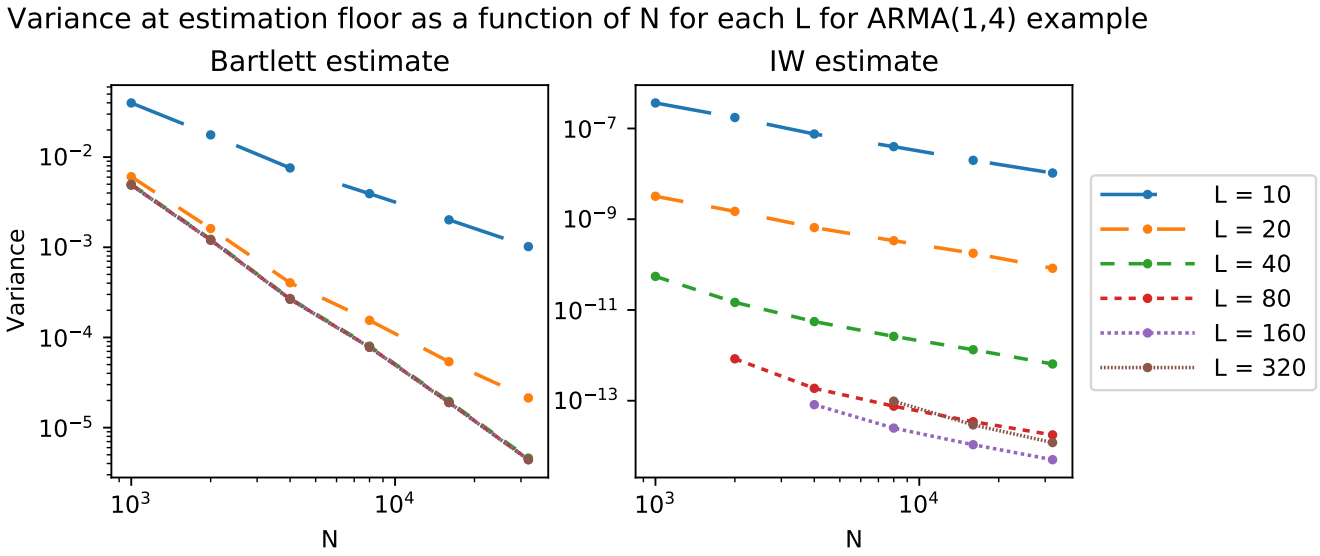


(b) Variance scaling at the estimator floor for the Bartlett estimator (left) and for the IW (noncausal) estimate (right). They are estimating the Gaussian power spectrum.

Figure 5.1: Scaling of bias and variance for the ARMA(1,4) example off the floor. Observe that for the bias Figure 5.1a Bartlett (on the right) again shows a complex scaling in L and N. The IW estimate, on the other hand, shows a very small bias that clearly has little if any N dependence. The variance plot for Bartlett, we have already seen. For IW the variance plot show a curious scaling, but at 10^{-35} one could image any number of things affecting those values.

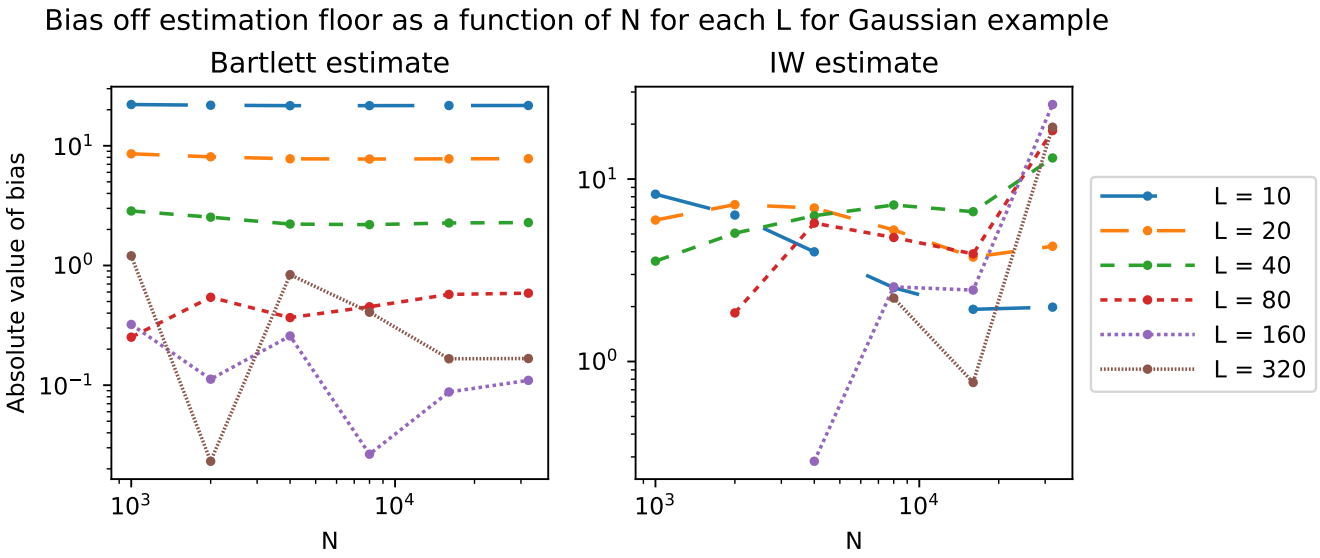


(a) Scaling of bias off the estimator floor for the ARMA(1,4) example

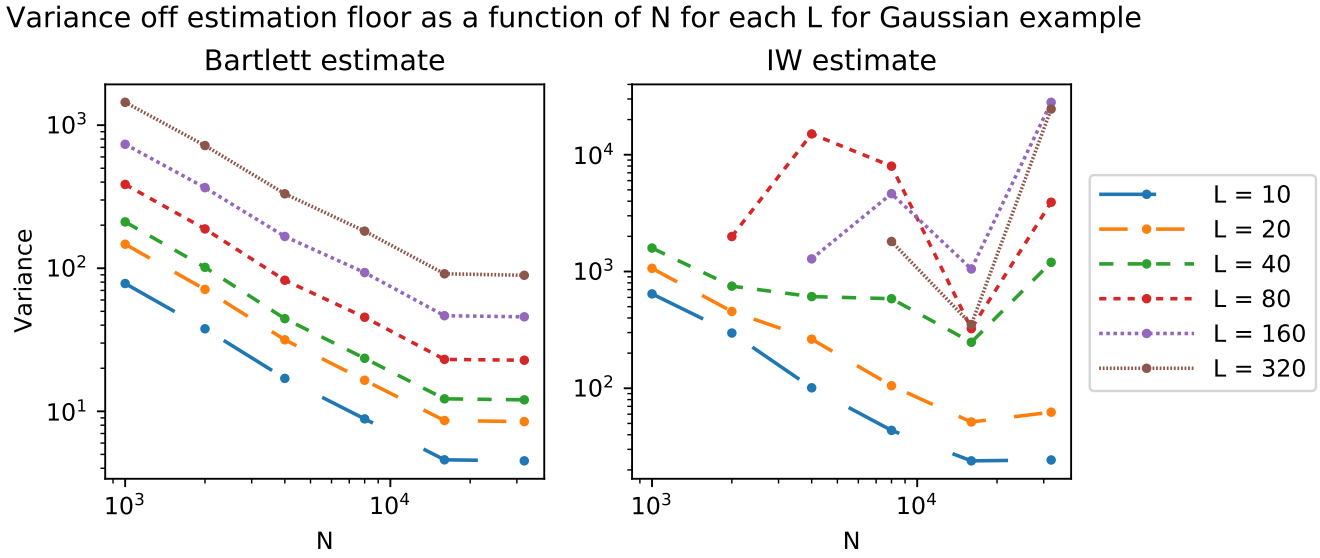


(b) Scaling of variance off the estimator floor for the ARMA(1,4) example

Figure 5.2: Scaling of bias and variance for the ARMA(1,4) example off the floor. For the bias of the IW estimate there is as with Bartlett no apparent dependence on N and the bias clearly decreases as L increases (also as with the Gaussian plot). For the variance We recognize, somewhat, the pattern exhibited in the for the Gaussian counterpart in Figure 5.1. Both of these suggest a variance decaying (perhaps slowly) as N grows.

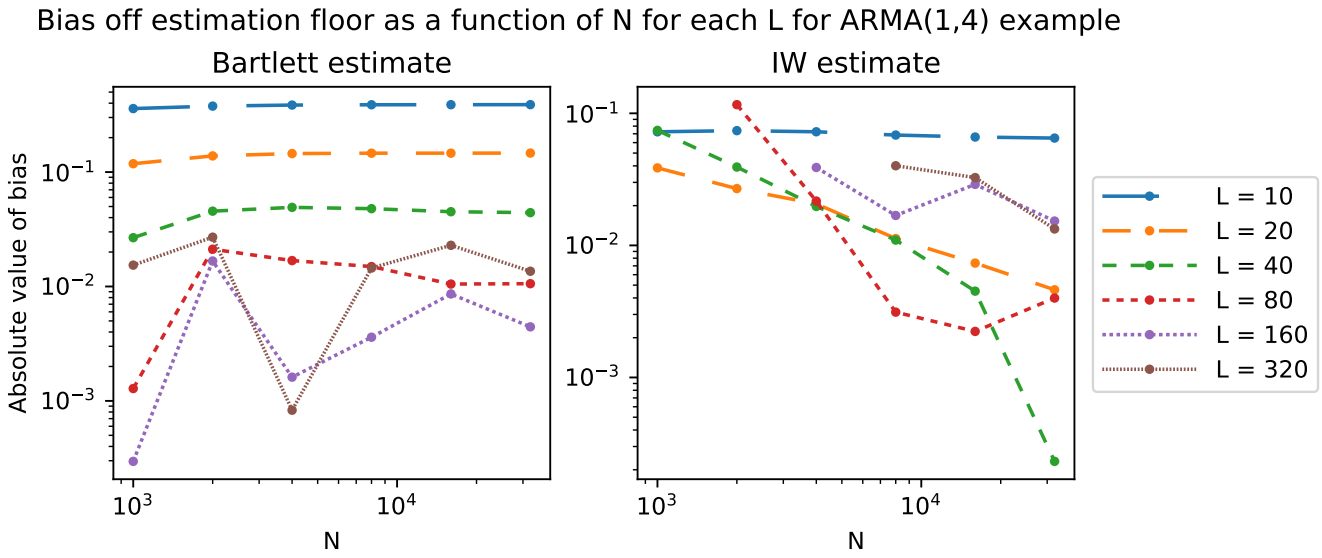


(a) Scaling of bias off the estimator floor for the Gaussian Example.

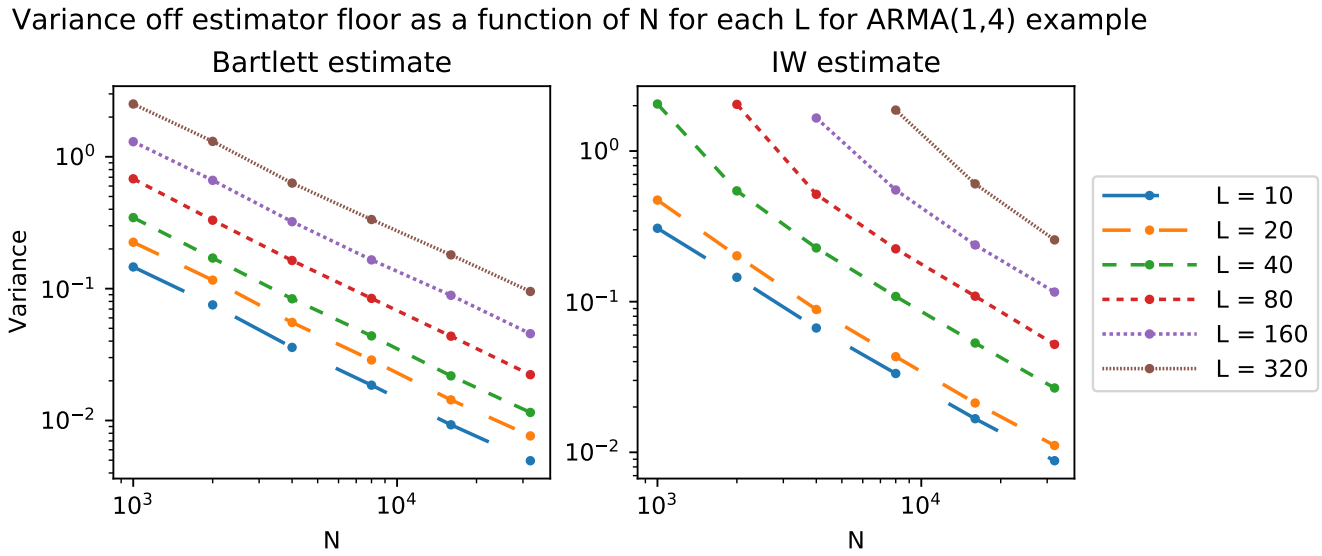


(b) Scaling of variance off the estimator floor for the Gaussian Example.

Figure 5.3: Scaling of bias and variance for the Gaussian example off the floor. In both of these plots bias plots it is important to notice the magnitude of the variance, that of IW is greater than for Bartlett while the bias appears to be roughly the same. For the variance we see that it increase in L as expected, in that L is generally used to control bias.

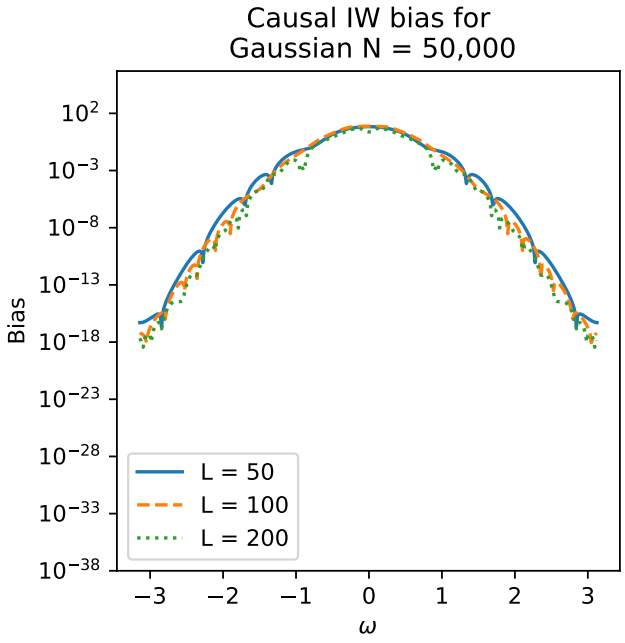


(a) Scaling of bias for the ARMA(1,4) example. These are estimated off the estimator floor. It shows the Bartlett estimate, of the ARMA(1,4) power spectrum, (left) and for the IW (noncausal) estimate (right).

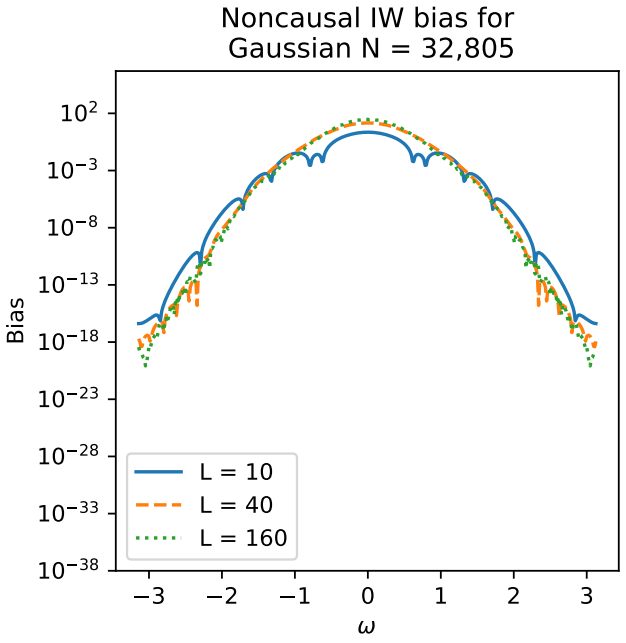


(b) Scaling of variance for the ARMA(1,4) example. These are estimated off the estimator floor. It shows the Bartlett estimate, of the ARMA(1,4) power spectrum, (left) and for the IW (noncausal) estimate (right).

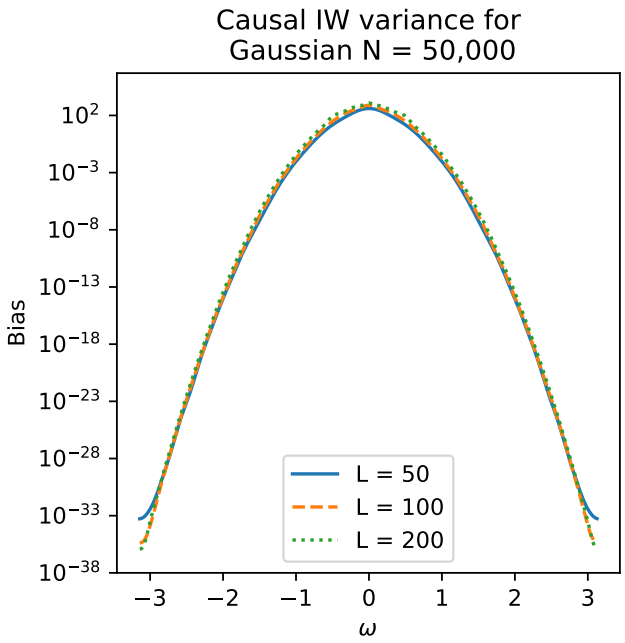
Figure 5.4: Scaling of bias and variance for the ARMA(1,4) example.



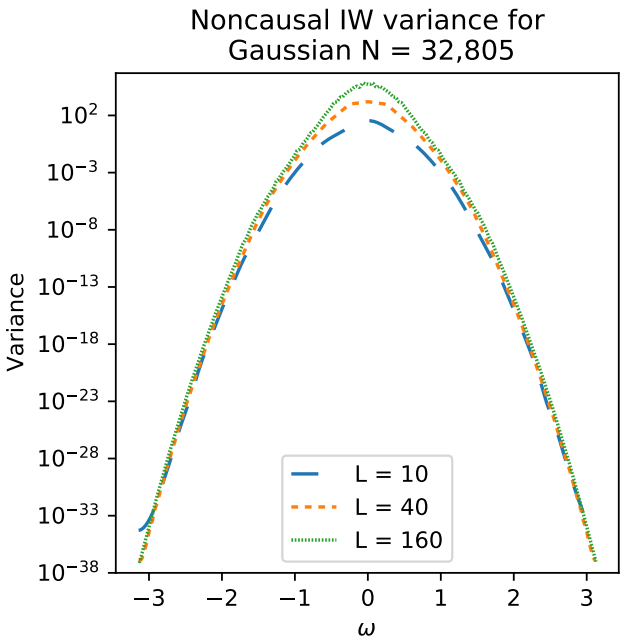
(a) Empirical bias of the causal IW estimate. IW went through a total of 10 iterations and was computed three times each for a different value of L but a fixed $N = 32,805$



(b) Empirical bias of the noncausal IW estimate. IW went through a total of 10 iterations and was computed three times each for a different value of L but a fixed $N = 32,805$

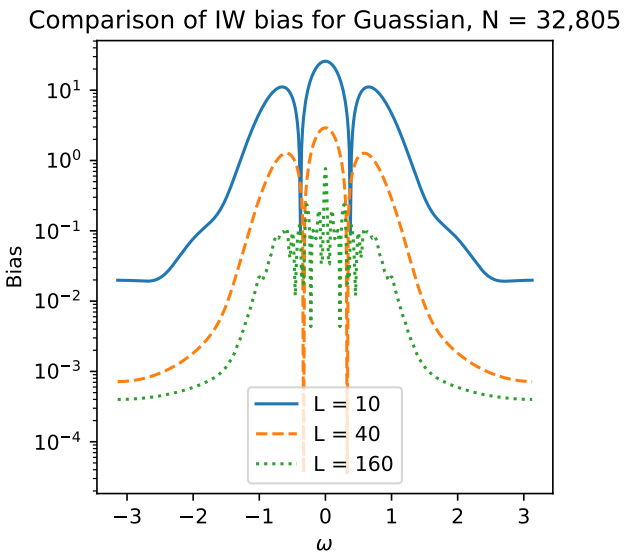


(c) Empirical variance of the causal IW estimate. The IW preformed 10 iterations. The variances were computed for three values of L and a fixed $N = 50,000$

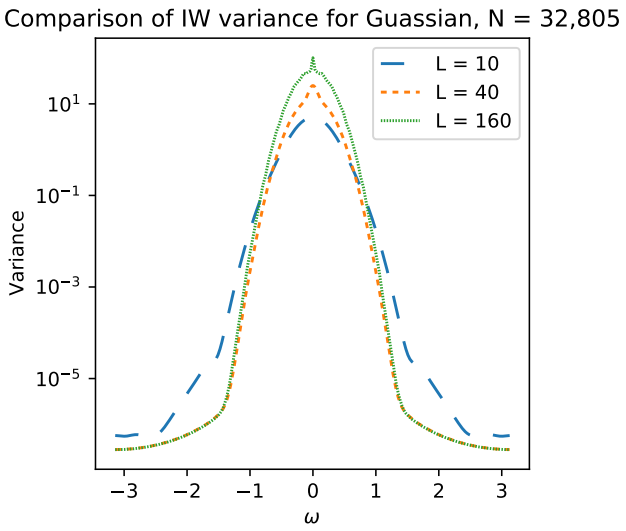


(d) Empirical variance of the noncausal IW estimate. The IW estimate went though 10 iterations. Three values of L and $N = 32,805$

Figure 5.5: Bias and variances as a function of frequency for three values of L and a fixed N. Observe that there is a slight dependence on L in both the bias and variance.



(a) Empirical bias of Bartlett estimate for three values of L and a fixed $N = 32,805$



(b) Empirical variance of Bartlett estimate for three values of L and a fixed $N = 32,805$

Figure 5.6: Bias and variance of Bartlett estimate. The bias and variance of the Bartlett estimate of an ensemble of 1600 realizations of the example Gaussian process for three values of L.

N_{f} to be odd seems to improve stability when it is an issue (for noncausal IW). Additionally, if $N_{\mathrm{f}} < 2L + 1$ then L effectively becomes L', the greatest integer such that $2L' + 1 \leq N_{\mathrm{f}}$. This is because the whitening filter is computed using FFT on the grid of size N_{f}. However, as far as I can tell, picking $N_{\mathrm{f}} > 2L + 1$ yield no improvement to the estimate, only a finer frequency grid on which the spectrum is estimated.

5.2.2 Welsh

For its accuracy, welsh is a very simple estimator to implement. Recall that for the welsh estimate, one

1. splits the time series,
2. point-wise multiplies each segment by the taper, $d^2 N_{\mathrm{en}} L_{\mathrm{seg}} = O(d^2 N)$ flops,
3. takes the FFT of each segment $O(d^2 N_{\mathrm{en}} L_{\mathrm{seg}} \log L_{\mathrm{seg}}) = O(d^2 N \log L_{\mathrm{seg}})$ flops, and
4. averages these, $O(d^2 N_{\mathrm{en}} L_{\mathrm{seg}}) = O(d^2 N)$.

So, that the leading order

$$O(d^2 N \log L_{\mathrm{seg}}) \tag{5.2}$$

5.2.3 IW

Now for IW we get

$$O(d^2mLN) + O(d^3mLN_{\mathrm{SF}}) + O(d^2mN_f \log(N_f)) + O(d^3mN_f) \tag{5.3}$$

At each iteration we compute a Bartlett estimate. This gives us the square of the modulus of the frequency response of the modeling filter. And so, we factor this, the choice of computing a causal or noncausal filter. For a noncausal filter, it requires $O(d^2N_f)$ flops for the matrix square root at each frequency. The next step is getting the impulse response of the whitening filter. To do this take the frequency-wise inverse of this, another $O(d^3N_f)$, and then compute the impulse response. This is found by taking the inverse FFT which is $O(d^3N_f \log N_f)$. This will be truncated to $2L+1$ terms so again $N_f = 2L+1$ is a good choice. For the causal filter. We revert to the autocovariances that were computed to form the original Bartlett estimator. These are fed into the CKMS algorithm (discussed in Appendix B, which requires $d^2L^2N_{\mathrm{SF}}$ flops in total, producing the impulse response of the modeling filter. From here we compute the frequency response of the modeling filter. This can be done using FFT at a cost of $O(d^2N_f \log N_f)$, again. The whitening filter can then be obtained by frequency-wise inversion ($O(d^3N_f)$) and then computing the impulse response by inverse FFT, yet again. To get the spectral estimate we take the modulus square of the frequency response of the modeling filter at each frequency.

For the autocovariances there is a difference between the single channel and multichannel case. In single channel one can use FFT to produce the autocovariance. In multichannel one could still use FFT to compute the convolution-like sum but the two transforms would need matrix multiplication at every frequency, (and it would have to be padded so as to account for aliasing). Alternatively, one could use a more direct implementation of (2.25) The would require $2L+1$ sums of roughly $N - |n|$ sums and products.

5.2.4 Discussion

The comparison of the runtime is straight forward since these methods are very similar. As expected IW spectral estimation takes a multiple of m, the number of iterations, times as many flops as Bartlett.

Chapter 6

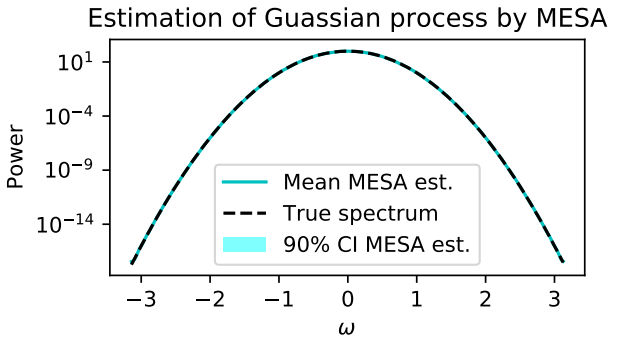
Maximal entropy spectral analysis

Maximum entropy spectral analysis (MESA) is an idea formulated by J. P. Burg [29, 30] in the 60s and 70s. It applies Jayne's concept of maximum entropy estimation [31] to power spectra (see also [32]). Burg noticed in the Bartlett estimate two assumptions which he found morally problematic. First, in using the Bartlett estimate one assumes the autocovariances of lag greater than the truncation parameter (L above) to be zero. Second, the approximations of the autocovariances that are used are modified by scaling. A nice quality of maximum entropy estimation is that it is maximally non-committal about missing data (see e.g. [30, 33, 10]). So, Burg proposed spectral estimation using the principle of maximum entropy together with *unaltered* estimates of the autocovariances while not assuming anything about the autocovariances not estimated. This resulted in the following variational problem for determining $\hat{S}(\omega)$ given the autocovariance sequence.

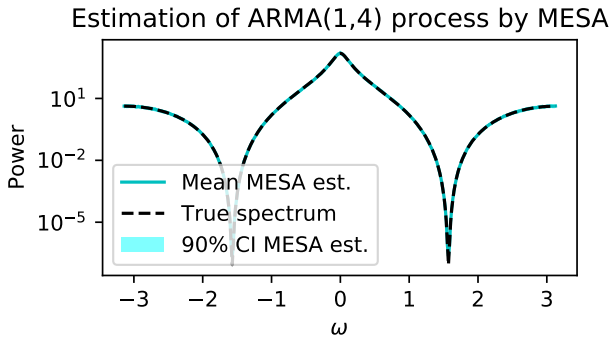
$$\begin{aligned} \max \quad & \int_0^\pi \log S(\omega)\, d\omega \\ \text{s.t.} \quad & \frac{1}{2\pi}\int_0^\pi S(\omega) e^{i\omega n}\, d\omega = \tilde{C}[n] \qquad \text{for } n = -p, \ldots, p, \end{aligned}$$

where p is some positive integer and $\tilde{C}[n]$ is the empirical autocovariance. As it turns out, the MESA estimator so obtained is the same as what one would get from fitting an AutoRegressive process of order p (AR(p)) to the data by least squares and then computing the spectrum of the AR(p) process [30].

A common way to fit an AR process to data is to estimate the autocovariances and solve the associated Yule-Walker equations, which relate the autoregressive coefficients and variance of the driving white noise to the autocovariances of the process (see e.g. [16, Chs. 17 and 21]). This can be done efficiently by the Levinson-Durbin Algorithm [16, 6]. The algorithm is recursive and starts by estimating an AR(1) process for the data. Using that together with the so-called reflection coefficient, found using the autocovariances data, the approximating AR(2) process can be found and so on. Burg [30] proposed what he called the modern-Levinson-Durbin algorithm in which, at each step the reflection coefficient is approximated directly from data, rather than using the empirical autocovariances. Either way, this process provides an ensemble of AR models of orders



(a) Mean MESA estimate ($p_{\max} = 100$, using the CAT criterion) of Gaussian process, solid cyan.



(b) Mean MESA estimate ($p_{\max} = 400$, using the CAT criterion) of ARMA(1,4) process, solid cyan.

Figure 6.1: The mean and 90% confidence interval of the spectral estimates of 100 independent realizations for each process (Gaussian and ARMA(1,4)) and for each estimator (Bartlett with Parzen lag window, Welsh with Tuky-Hanning data window, and MESA). In all cases a snapshot of length $N = 40{,}000$ was used. In each plot the true spectrum is the dashed black curve.

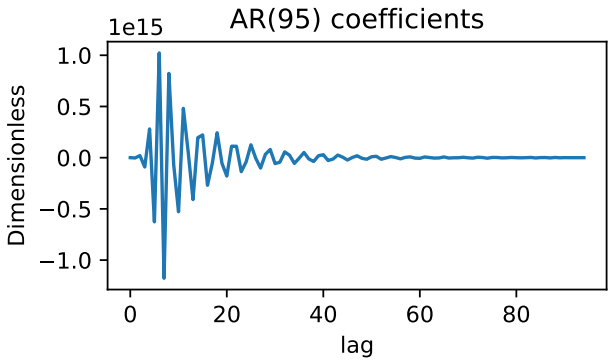
1 through $p_{\max}$ ($p_{\max}$ chosen by the user). Usually an information criterion (for standard examples see [10]) is used to select an order p and the MESA spectral estimate is given by

$$\tilde{S}_X^{\text{MESA}}(\omega) = \frac{\sigma_p^2}{|A(\omega)|^2} \tag{6.1}$$

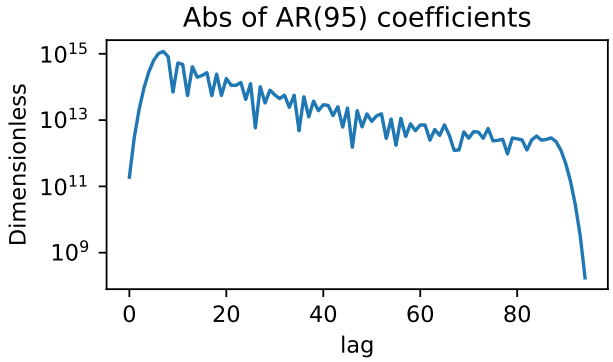
where σ_p^2 is the variance of the driving white noise process in the AR(p) model, that model's coefficients are $a_1, a_2, \ldots, a_p$ from whence we define

$$A(\omega) = 1 + \sum_{j=1}^{p} a_j e^{-in\omega}. \tag{6.2}$$

To demonstrate the effectiveness of The MESA estimate we use the same set up as in Figure 3.1, that is we find the MESA spectral estimate for 100 realizations of the Gaussian process and also for the ARMA(1,4) process. The mean together with the 90% confidence intervals are plotted (though they are so narrow as to hardly be distinguishable). The the Gaussian process $p_{\max} = 100$, but for the ARMA(1,4) a better chioce was $p_{\max} = 400$ to better accommodate the narrow trough of the ARMA(1,4) spectrum. Observe that for these examples MESA exhibits no estimator floor and accurately estimates the spectrum for all frequencies.



(a) Mean MESA estimate ($p_{\max} = 100$, using the CAT criterion) of Gaussian process, solid cyan.



(b) Mean MESA estimate ($p_{\max} = 400$, using the CAT criterion) of ARMA(1,4) process, solid cyan.

Figure 6.2: An example of a autoregressive filter with values that vary over many orders of magnitude an approximately alternate in sign.

6.1 Numerical issues estimating high-contract spectra by MESA

As John Burg himself observed [30], MESA does not allow for spectral leakage and so there is no estimator floor. However, for spectrum with a large spectral range there is a potential for numerical issues if one is not careful. As it turns out the, coefficients of a single autoregressive model obtained by fitting the data can vary by many orders of magnitude. To make matters worse these coefficients can also exhibit a near alternating behavior. To illustrate this point I draw on an example that I have only alluded to but will discuss in full detail in Chapter 8. The process comes from the study of a chaotic partial differential equation (Kuramoto-Sivashinsky). Figure 6.2).

The issue can be over come but it requires either higher precision computing or numerical delicacy. Either way it can add significantly to the cost of implementing this method.

6.2 Runtime analysis

For MESA we have

$$O(d^2 p_{\max}^2 N) + O(d^3 p_{\max}^2). \tag{6.3}$$

Burgs extension of the Levison-Durbin algorithm (in which the reflection coefficient is computed directly from data) requires filtering at every step. Recall the algorithm recursively computes AR(p) coefficients for the data for each order $p = 1, \ldots, p_{\max}$. So that that each step requires

$$O(d^2 p_{\max} N) + O(d^3 p_{\max}) \tag{6.4}$$

where the first term is from filtering the second terms of is from a matrix square root at every step. Doing this at each step then imparts another factor of $p_{\max}$.

Chapter 7

Method of control variates for spectral estimation

This work was also developed as a straightforward way to improve a power spectrum estimate. Initially, I thought that the estimator floor was primarily driven by the variance. This is because the bias, at least according to the leading order asymptotic formula under (A.16) suggested that the bias would be very small, especially considering that the contribution bias makes to the mean square error is the bias squared. And so, it was reasoned that using a variance reduction technique would be an effective way to improve the estimate. Recall that the Bartlett estimator was motivated by the simple variance reduction technique of sample mean, which proved to be very effective. In this chapter we just apply a more sophisticated variance reduction technique.

The chapter is organized as follows. First I discuss a variance reduction technique from Monte Carlo theory, the method of control variates. Then I explain how it can be used in improving the Bartlett estimators for spectral estimation. The Chapter conclude with a demonstration of the method applied to the running Gaussian example from Section 3.1.1.

7.1 Method of control variates

I will now give a brief exposition of the method of control variates. Suppose we want to estimate an expectation $\mu = \mathbb{E}X$, of some random variable X. This may be done by taking n IID samples X_i of X and taking the average $\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Clearly, under these assumptions, $\tilde{\mu}$ is an unbiased estimate of μ. The variance of this estimate is

$$\operatorname{var}(\tilde{\mu}) = \frac{\operatorname{var}(X)}{n}. \tag{7.1}$$

Now suppose we can sample from another random variable Y that is correlated with X and has zero mean. Notice that the random variable $X - \alpha Y$ also has expectation μ, for any α. The variance

can be computed as follows,

$$\begin{aligned}\mathrm{var}(X-\alpha Y) &= \mathbb{E}(X-\alpha Y-\mu)(X-\alpha Y-\mu)^* \\ &= \mathbb{E}(X-\mu)(X-\mu)^* - \alpha\mathbb{E}Y(X-\mu)^* - \mathbb{E}(X-\mu)Y^*\alpha^* + \alpha\mathbb{E}YY^*\alpha^* \\ &= \mathrm{var}(X) - 2\mathcal{R}\{\alpha\mathrm{cov}(Y,X)\} + |\alpha|^2\mathrm{var}(Y) \end{aligned} \tag{7.2}$$

Observe that the variance of $X-\alpha Y$ depends on α quadratically and can, in fact, be minimized by taking

$$\alpha = \frac{\mathrm{cov}(Y,X)^*}{\mathrm{var}(Y)} = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)}. \tag{7.3}$$

Now, let us consider the estimate

$$\tilde{\mu}^{\mathrm{cv}} = \frac{1}{n}\sum_{i=1}^{n} X_i - \alpha Y_i. \tag{7.4}$$

Where Y_i are n IID samples of Y. This is an unbiased estimate whose variance is

$$\mathrm{var}(\tilde{\mu}^{\mathrm{cv}}) = \frac{1}{n}\mathrm{var}(X-\alpha Y) \tag{7.5}$$

If the constant α is chosen according to equation (7.3) then we can optimally reduce the variance of the estimate $\tilde{\mu}^{\mathrm{cv}}$. With this particular choice of α we have

$$\mathrm{var}(X-\alpha Y) = \mathrm{var}(X) - \frac{|\mathrm{cov}(X,Y)|^2}{\mathrm{var}(Y)} = \left(1-|\rho_{XY}(0)|^2\right)\mathrm{var}(X) \tag{7.6}$$

And so,

$$\mathrm{var}(\tilde{\mu}^{\mathrm{cv}}) = \frac{1-|\rho_{XY}(0)|^2}{n}\mathrm{var}(X) = \left(1-|\rho_{XY}(0)|^2\right)\mathrm{var}(\tilde{\mu}). \tag{7.7}$$

7.2 Method of control variates in spectral estimation

In this section I describe how the method of control variates may be used to improve the accuracy of a power spectrum estimate. The object we are estimating is the power spectrum at each frequency. To only apply the sample mean technique again on the Bartlett estimator would only have the effect of altering the lag window. And as was pointed out in Chapter 4 this does very little to alter the estimator floor. So, we are obliged to use a different technique. To apply the method of control

variates we, as with the sample mean technique, split the snapshot into sections. The question now becomes what should we use as the control variates? A clever choice is the whitened power spectrum of each section. It is certainly true that these are correlated with the power spectrum of the associate section. But whiten spectra will not be mean zero, even asymptotically. This is because as L, N increase, the Bartlett power spectrum estimate, which is a consistent estimator, will produce better and better whitening filters. Asymptotically the whiten spectrum will approach unity at each frequency. One way to get around this is to take the log of whitened spectrum. This is not unreasonable because spectra are naturally analyzed on the log scale. These control variates will asymptotically be mean zero. And for finite L we would only expect a small amount of bias to be added. We will take the log of the power spectrum in each section as well. In effect we are investigating an variance reduced estimate of $\log(S_X)$. Once we compute that estimate we will take the exponential of the estimate to recover an estimate of S_X. With that in mind we proceed with the log of the whitened spectrum as the control variates.

Given a snapshot $x = (x_n,\ n = 1, \ldots, N)$, and any power spectrum estimation technique which will be referred to as SpecEst, we follow these steps.

1. Divide the full snapshot x into K segments $(x^{(k)})_{k=1}^{K}$, where

$$x^{(k)} = (x_j,\ j = M(k-1)+1, \ldots, Mk). \tag{7.8}$$

   and $M = \lfloor N/K \rfloor$.

2. For each segment k, $k = 1, \ldots, K$, estimate

   (a) the spectrum $\tilde{S}^{(k)} = \text{SpecEst}\left[x^{(k)}\right]$ and

   (b) the whitened spectrum $\tilde{W}^{(k)}$.

   For the whitened spectrum, this will involve estimating the whitening filter $\tilde{w}^{(k)}$, which may be done causally or noncausally, and then filtering to get the whitened data $Y^{(k)} = X^{(k)} \star \tilde{h}^{(k)}$ then estimate its power spectrum

$$\tilde{W}^{(k)} = \text{SpecEst}\left[Y^{(k)}\right]. \tag{7.9}$$

3. For each spectrum and whitened spectrum, take the logarithm $(\log \tilde{S}^{(k)})_{k=1}^{K}$ and $(\log \tilde{W}^{(k)})_{k=1}^{K}$.

4. Compute α by dividing the covariance of $\tilde{S}^{(\cdot)}$ and $\tilde{W}^{(\cdot)}$ by the variance of $\tilde{W}^{(\cdot)}$,

$$\alpha = \frac{\operatorname{cov}_k \left(\log \tilde{S}^{(k)}, \log \tilde{W}^{(k)} \right)}{\operatorname{var}_k \left(\log \tilde{W}^{(k)} \right)}. \tag{7.10}$$

   The subscript k denotes that we want the covariance and variance to be taken over the variable k.

5. Compute, this time using the full time series X, an estimate of the spectrum $\tilde{S} = \text{SpecEst}\,[X]$ and the whitened spectrum

$$\tilde{W} = \text{SpecEst}\,[Y] = \text{SpecEst}\,[X \star h] \tag{7.11}$$

   where h is the whitening filter of the full time series.

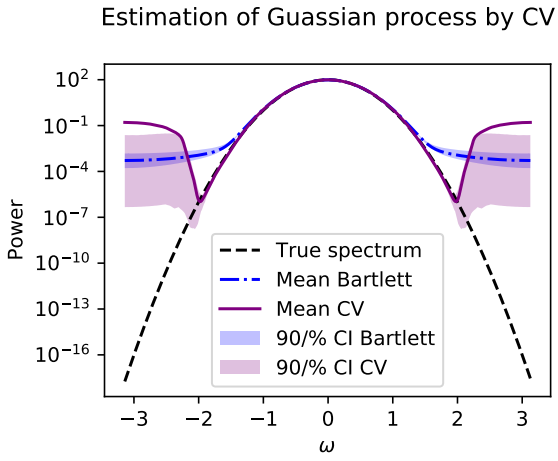
6. Compute the final spectral estimate as

$$\tilde{S}^{\text{CV}} = \exp\left(\log \tilde{S} - \alpha \log \tilde{W} \right). \tag{7.12}$$

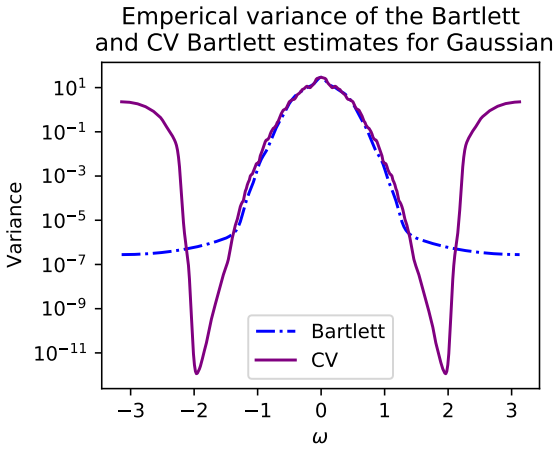
I also provide a description of the procedure in the form of an algorithm written in pseudo-code.

To demonstrate method let us consider again the familiar Gaussian example from Section 3.1.1. Again with a snapshot length of $N = 40{,}000$. We take 100 realizations and apply then method of control variates in conjunction with the Bartlett spectrum estimate. Figure 7.1a shows the mean estimate together with the 90% confidence interval. These have been plotted on top of the Bartlett estimate with its 90% confidence interval. Observe in Figure 7.1a that the power spectrum estimate follows the true power spectrum several decades (of power) farther than the Bartlett estimate.

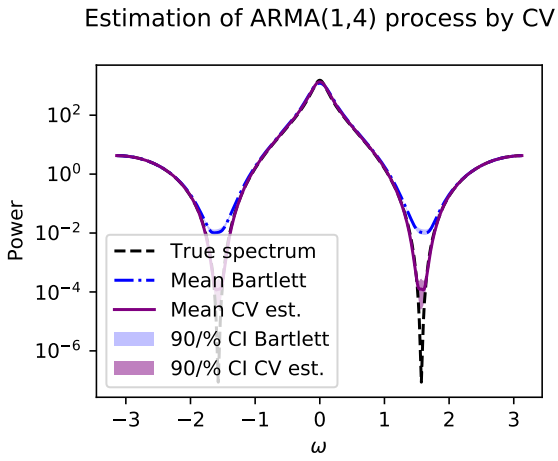
Figure 7.1a displays the Bartlett estimate with the control variate variance reduction technique applied. For convince, the plot also contains the Bartlett estimate with it. A plot of the of the variance of the two estimates also provided in Figures 7.1b and 7.1d. Observe the reduction that occurred in the variances.



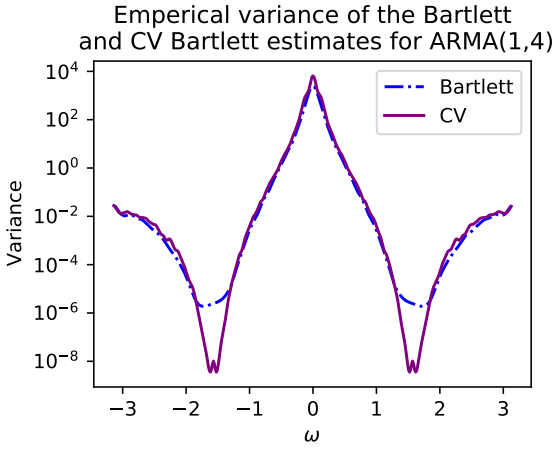
(a) Method of control variates applied to the Gaussian example together with the Bartlett estimate and the true spectrum. Observe that the power spectrum estimate of Bartlett is improved by CV over several decades then after a point the estimate becomes less accurate.



(b) Variance of the Bartlett power spectrum estimate and that of the CV estimate. Observe the reduction of the variance and note how that is associated with a better estimate at those frequencies.



(c) Method of control variates applied to the ARMA(1,4) example.



(d) The variance of the Bartlett power spectrum estimate and that of the CV estimate. Observe the reduction of the variance and note how that is associated with a better estimate at those frequencies.

Figure 7.1: Estimation of example spectra by CV. The plot on the right displays the analytic (dashed red) and approximated (blue) power spectra of the process Gaussian using the Parzen-Bartlett estimator, together with 90% confidence interval computed from 100 independent realizations. The plot on the left displays the estimated power spectrum using the PB estimator of the whitened Gaussian process together with its 90% confidence interval. The "true" spectrum of the whitened process is also displayed (dashed red) this was formed by dividing the true spectrum of the Gaussian process by the averaged PB estimate from the plot on the right.

Algorithm 1 SpectrumCV

procedure SpecCV(Process(N),Estimator,WF,M,N,N')

- Estimate α

for $i = 1$ to M **do**

$X^{(i)} \leftarrow$Process(N')

$S^{(i)} \leftarrow \mathrm{Estimator}(X^{(i)})$

$w^{(i)} \leftarrow \mathrm{WF}(X^{(i)}, S^{(i)})$

$W^{(i)} \leftarrow X^{(i)} \star w^{(i)}$

$E^{(i)} \leftarrow \mathrm{Estimator}(W^{(i)})$

$\alpha \leftarrow \dfrac{\mathrm{cov}(\{\log S^{(i)}\}_{i=1}^{M}, \{\log E^{(i)}\}_{i=1}^{M})}{\mathrm{var}(\{\log E^{(i)}\}_{i=1}^{M})}$

- Estimate $\overline{\log E}$

$\overline{\log E} \leftarrow \frac{1}{M}\sum_{i=1}^{M} \log E^{(i)}$

- Estimate power spectrum S_f^{cv}

$X_f \leftarrow$Process(N)

$S_f \leftarrow \mathrm{Estimator}(X_f)$

$w_f \leftarrow \mathrm{WF}(X_f, S_f)$

$W_f \leftarrow X_f \star w_f$

$E_f \leftarrow \mathrm{Estimator}(W_f)$

$S_f^{\mathrm{cv}} \leftarrow \exp(\log S_f - \alpha(\log E_f - \overline{\log E}))$

Chapter 8

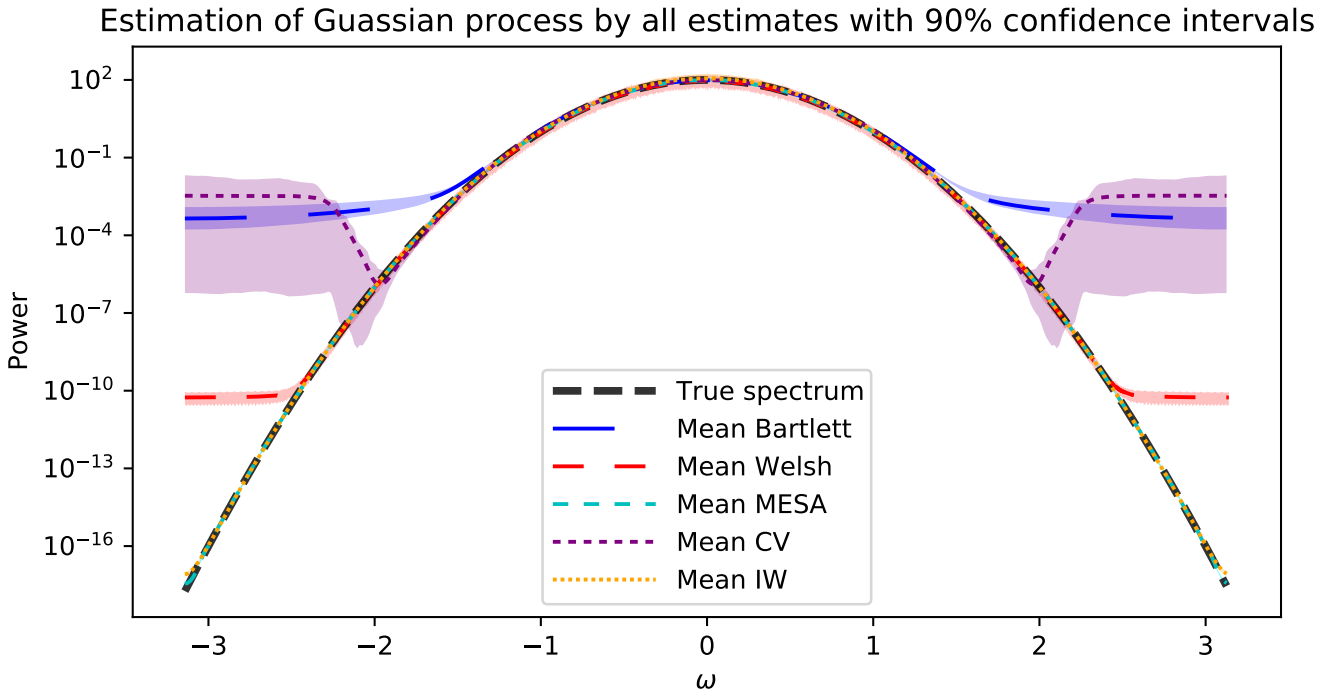
Examples and applications

In this chapter I go through three examples demonstrating the effectiveness of IW power spectrum estimation and compare its performance with the other methods mentioned. I first look at the Gaussian and ARMA(1,4) examples that we are already familiar with. Then, I look at a example from the dynamicals systems literature, namely the Fourier modes of a numerical solution to the Kuramoto-Sivashinsky equation.

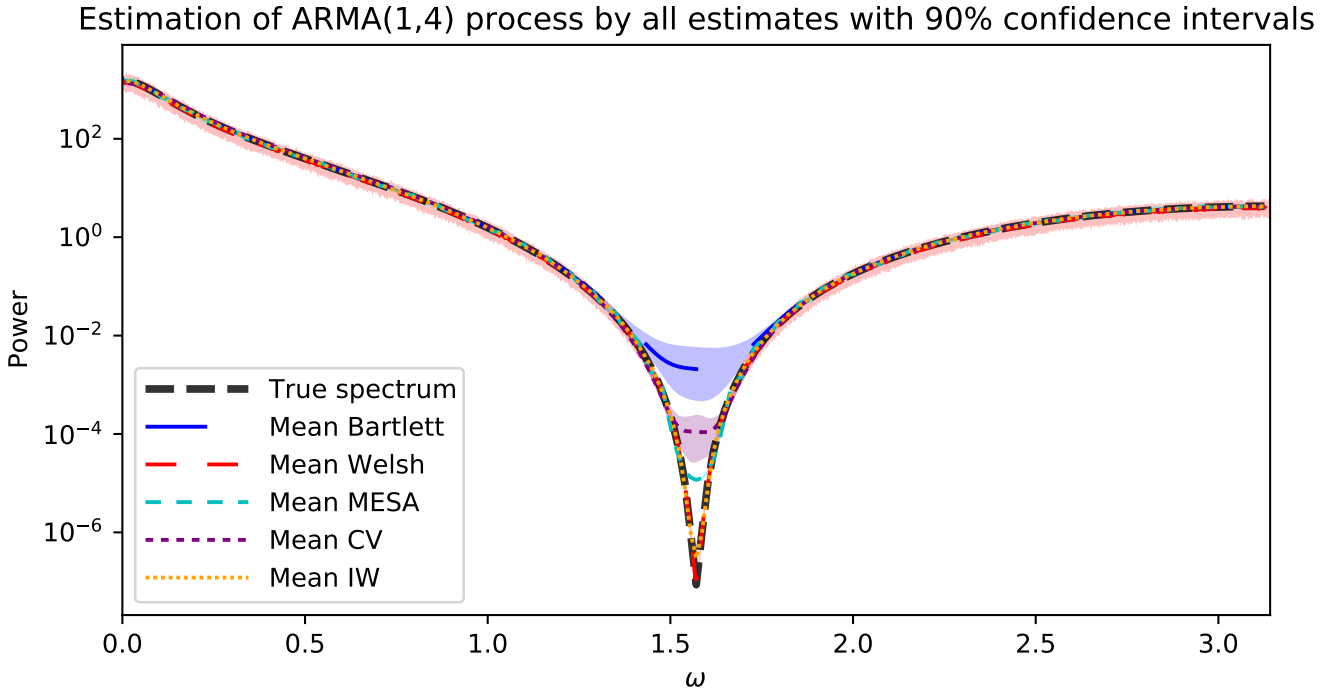
8.1 Gaussian and ARMA(1,4)

Here I return to the example introduced in section 3.1.1. As before we use a time series of $N = 40{,}000$ points and sample 100 realizations. For each realization I estimate the power spectrum using (1) Bartlett's method with the Parzen window, (2) Welsh's method with the Tukey-Hanning window, (3) MESA, (4) the method of control variates and (5) Iterated Whitening. I then compute the mean and 90% confidence interval for these estimated power spectra. Figure 8.1a shows the result whitening filter applied to the Gaussian example. The Bartlett, Welsh, and Bartlett CV estimates all process an estimator floor and fail to accurately estimate the power spectrum over all frequencies, MESA and IW on the other hand are able to estimate the power spectrum of the some 20 decades of contrast. Further more they do this with a very narrow 90% confidence interval. It is difficult to see but the confidence interval of the MESA estimate is much narrower than that of IW.

This study was also performed on the ARMA(1,4) process and the results can be seen in Figure 8.1b. The display for this process shows only the right-hand side of the plot where the frequencies are positive. This is to see more clear the though that the graph possesses. The spectrum is even since the data is real. For this plot Bartlett exhibits the highest bias and variance near $\pi/2$ which is the minimizer of the spectrum. The just below that is the control variate estimate Followed by Welsh, which still exhibits and estimator floor. The final two estimates MESA and IW are nearly on top of each other. Both with very narrow 90% confidence intervals.



(a) Power spectrum estimate of the Gaussian spectrum estimated 5 different ways. Complete details for the estimation can be found in the text.



(b) Power spectrum estimate of the ARMA(1,4) spectrum estimated 5 different ways. Complete details for the estimation can be found in the text.

8.2 Kuramoto-Sivashinsky equation

The next is example we discuss comes from the numerical solution of the Kuramoto-Sivashinsky (KS) equation. The Kuramoto-Sivashinsky equation, given by

$$u_t + uu_x + u_{xx} + u_{xxxx} = 0, \tag{8.1}$$

for $t \in [0, \infty)$, $x \in [0, L]$ with periodic boundary conditions. This is a prototypical model of spatiotemporal chaos and has been studied extensively (see e.g. [5]). Written in terms of its Fourier coefficients u_k, it becomes a system of ordinary differential equations in time

$$\dot{u}_k = -\frac{iq_k}{2} \sum_{\ell} u_\ell u_{k-\ell} + (q_k^2 - q_k^4)u_k, \qquad q_k = \frac{2\pi k}{L} \tag{8.2}$$

which we solve using the usual method of fourth order exponential time differencing (ETDRK4) from [34]. In this work I take $L = 2\pi/\sqrt{0.085}$. The initial condition $g(x) = u(x, 0)$ is randomized in that I took a random linear combination of the functions $\cos(2\pi n/L \cdot x)$ for $n = 1, \ldots, 5$. The time step I used in the solver is 0.001. In order to ensure that the system is at an equilibrium distribution (so it is stationary) I ran the solver 10^8 time steps and saved every one hundredth sample, and discarded the first half of the data. This amounts to a snapshot of 500,000 samples a time step of 0.1. This was done 40 times so that I have 40 realizations. Figure 8.2 shows an example solution to the KS equation.

And for Mesa, p_{max} was set to 100 and the Criterion Autoregressive Transfer function (CAT) [10] was employed for the selection of the order p.

The true spectrum for these Fourier modes is not know, so one needs to be careful when assessing the accuracy of the estimate. We consider the 3rd and 5th Fourier modes and estimate the power spectrum of these.

The mean and 90% confidence interval of the estimated spectra is plotted in Figure 8.3 Plots the power spectrum estimates for the 3rd Fourier mode of the KS solution.

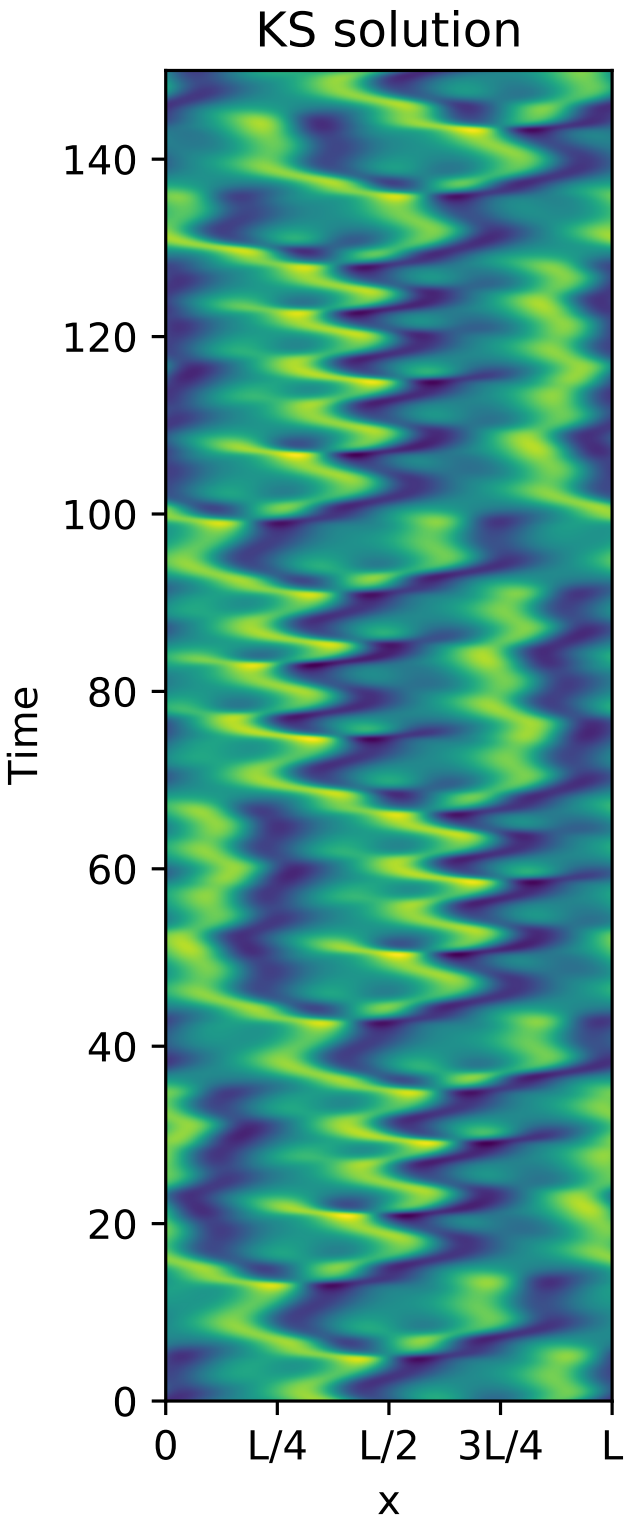
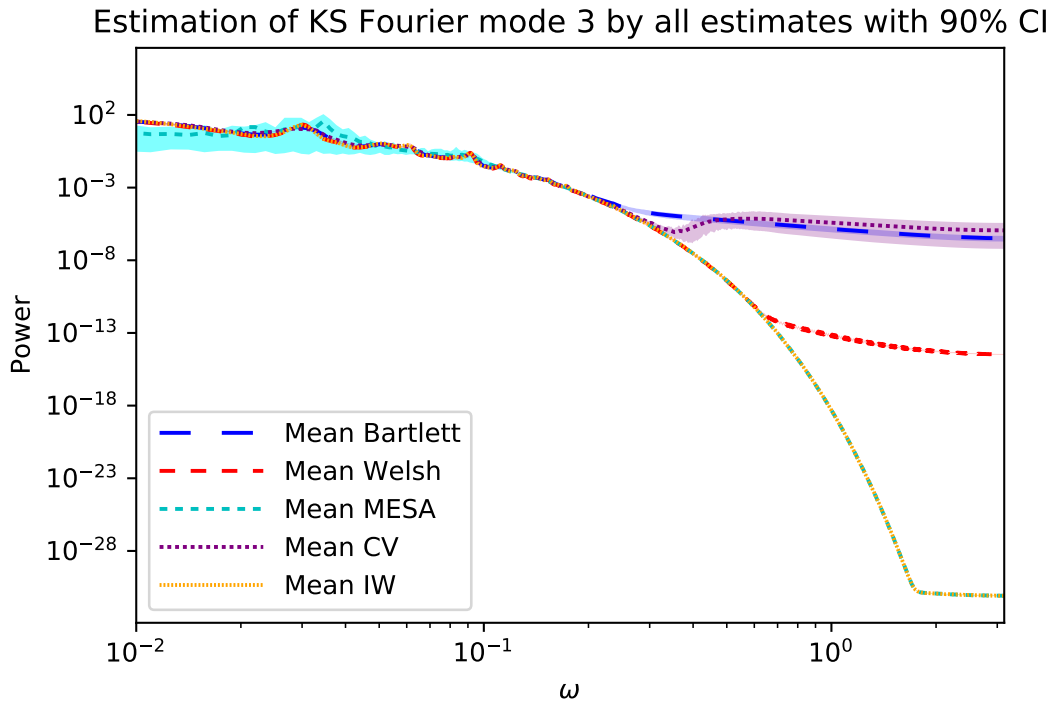
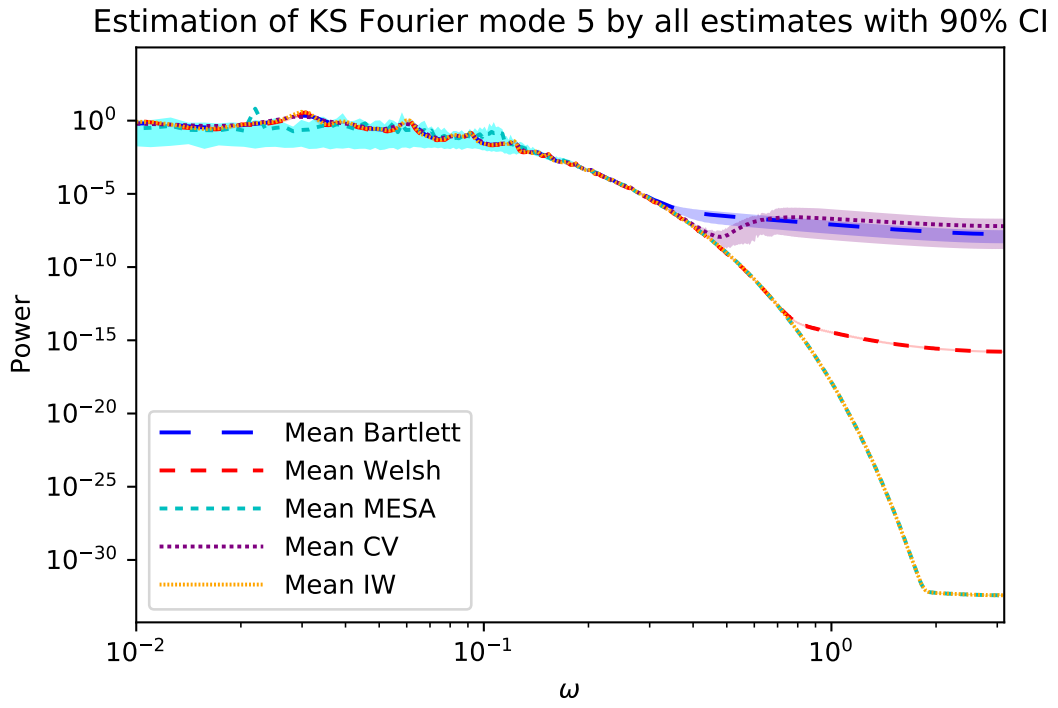


Figure 8.2: A numerical solution to the KS equation. Observe the stationary appearance of the data.



(a) Spectral estimates of KS modes 3. Here we bring together all 5 power spectrum estimates to observe their performance on the KS data. Observe that MESA follows IW very closely in high frequencies and Bartlett and Welsh in the low frequencies.



(b) Spectral estimates of KS modes 5.The true spectra (solid line) is plotted together with the mean of the spectrum estimated using the generalized Bartlett method with the Parzen lag window.

Figure 8.3: Spectral estimates of KS modes 3 and 5. As in Figure 8.3a all five estimates are displayed at once. Again, MESA is difficult to distinguish as it matches Bartlett at high powers and MESA at low powers.

Chapter 9

Discussion

I now conclude this dissertation with an overview of what has been discussed and highlighting key takeaways, a discussion of some questions about the IW method, and description of possible future work.

We began by noting an deficiency in certain commonly used spectral estimates. Namely, the accuracy of the power spectrum estimate in the low-power frequencies of high-contrast spectra is very poor. It was found that filtering a time series changes the power spectrum of the time series. in a straightforward way.

I then discussed high-contrast spectra and the estimator floor that appeared when their spectrum is being estimated by the so-called periodogram based estimators. We looked at two perspectives on the estimator floor, which can be summarized as spectral leakage and sampling properties (bias and variance).

Following that discussion the IW procedure was introduced as a solution to the problem of estimator floor. In order to discuss the IW method I introduced the concepts of spectral factorization, modeling filters and whitening filters. The idea of the IW method is to use the an inexpensive, method that is easy to implement, such as Bartlett, to fashion a whitening filter. The data is then whitened and another spectral estimate is produced. This is repeated until the whitened spectrum is sufficiently flat, suggesting all the powers in the spectrum were accurately reported and whitened out across all frequencies. The associated modeling filters where then used to modified the estimate of the final whitened spectrum to produce an accurate spectral estimate of the original process.

Two ways of producing the whitening filter was also presented causal and noncausal. I have found that in most cases the output of noncausal IW power spectrum estimation is indistinguishable from that of causal IW. And since it is significantly cheaper to implement, I recommend it as primary implementation of IW power spectrum estimation. However, if some numerical issues arise such as the spectral estimate becomes suspiciously large. Then, it would be wise to consider the causal implementation.

This new estimate was then analyzed numerically and it was found that over the estimation

floor. The bias and variance were significantly improved. However, off the estimation floor, where the Bartlett estimator already accurately estimated the power, there was no improvement in either the bias or variance. And, in fact, the empirical study suggested that it may be the case that the bias might not vanish as N and L get large. Meaning, there may not be away to control the bias to bring it to a desired tolerance.

Following that discussion the MESA estimator was discussed, by way of background. The estimate was later used to compare the performance of the IW estimate. It was also observed that MESA exhibited some numerical issues when estimating high-contrast spectra.

For an example, the Kuramoto-Sivashinsky PDE was introduced and the power spectrum of to of its Fourier modes where estimated using all of the forgoing methods. It was seen that iterated whitening matched the Bartlett and Welsh estimates at the high powers and method the MESA estimate in the low powers. Suggesting that the IW estimate was able to estimate the whole spectrum with an overall superior accuracy.

Spectral leakage gives an explanation of the estimator floor as the contributions of estimated power due to the side lobes of the spectral window. These contributions are negligible when the spectrum varies little over the frequencies because the side lobes are several decades smaller that the main lobe. However, when the spectrum exhibits 10 or more decades of contrast, especially over a short interval of frequencies, the amount of leakage can overwhelm the true power at some frequencies. Iterated whitening presents a way of iteratively flattening the spectrum (for it cannot be done all at once) eventually the leakage at any given frequency is mush smaller that the power at that frequency. It then adjust the estimate according to the successive stages of whitening in a way that does not introduce more leakage (it is by point-wise multiplication in frequency space rather than a convolution which was done in time domain).Thus the over all contribution of spectral leakage is reduced substantially.

Analytically, both bias and variance were the sum of a number of different terms that depended of different aspects of the true spectrum, L and N. I observed that the estimator floor seemed present only when the non-leading order terms for bias or variance dominated. For bias these terms depended on higher order derivatives of the true spectrum. When the process is whitened (even approximately) the true spectrum estimated in the next step is more flat than the step before and thus the bias decreases. Adjusting for the whitening steps (i.e. multiplying the estimated spectrum of the final step by the modeling filters of each step) seems to not affect the bias in any large degree.

The method of control variates for spectral estimation grew out of our study of whitening data

and was a way to reduce the variance of the estimators which was seen to reduce the estimator floor over some frequency. Particularly, the frequency bands for which the (once) whitened spectrum could be accurately estimated.

9.1 Open questions about IW

9.1.1 Controllability of bias and variance of IW estimate

The results of Section 5.1.2 suggest that off the floor, the variance of the IW estimator will vanish if, e.g., L is taken to be a positive power of N and $N \to \infty$. The same results also suggest that the bias of the IW estimate off the estimator floor cannot be made arbitrarily small by tuning L and N. As of this writing, it is unclear what features of the spectrum control the bias, and whether one can generally expect it to be small. It may also be possible to attenuate the bias by, e.g., stopping the IW process in a frequency-dependent way, or borrowing ideas from, e.g., the Monte Carlo literature.

Another related direction for future study is to construct a "self correcting" version of IW. Consider the KS third mode with the noncausal IW estimate on an even frequency grid. (See Figure 9.1.) The spectral estimate builds in power in the low frequencies for the first four or so iterations and then the spectral estimate remains high, so subsequent whitening filters approximate the unit impulse filter. This "memory" phenomenon may well be related to the bias off the floor, and a "self-correcting" version of IW that can compensate for over or underestimates in previous steps may also reduce the bias.

9.1.2 Relating the two perspectives of estimator floor

The spectral leakage and the bias and variance perspectives of the estimator floor both provide useful insight in understanding the estimator floor. However, *how* to relate these two points of view is not clear. A better understanding of how to capture leakage effects directly in the analysis may lead to a deeper understanding of IW and related estimators.

9.1.3 Analytic results on variance scaling

As can be seen in equation (A.57) there is a N^{-2} term in the expansion which originates from the error associated with approximating the Fejér kernel by a δ-function, this can be found under (A.47). This is, in fact, the only source of an error that could scale as N^{-2} and should be contained

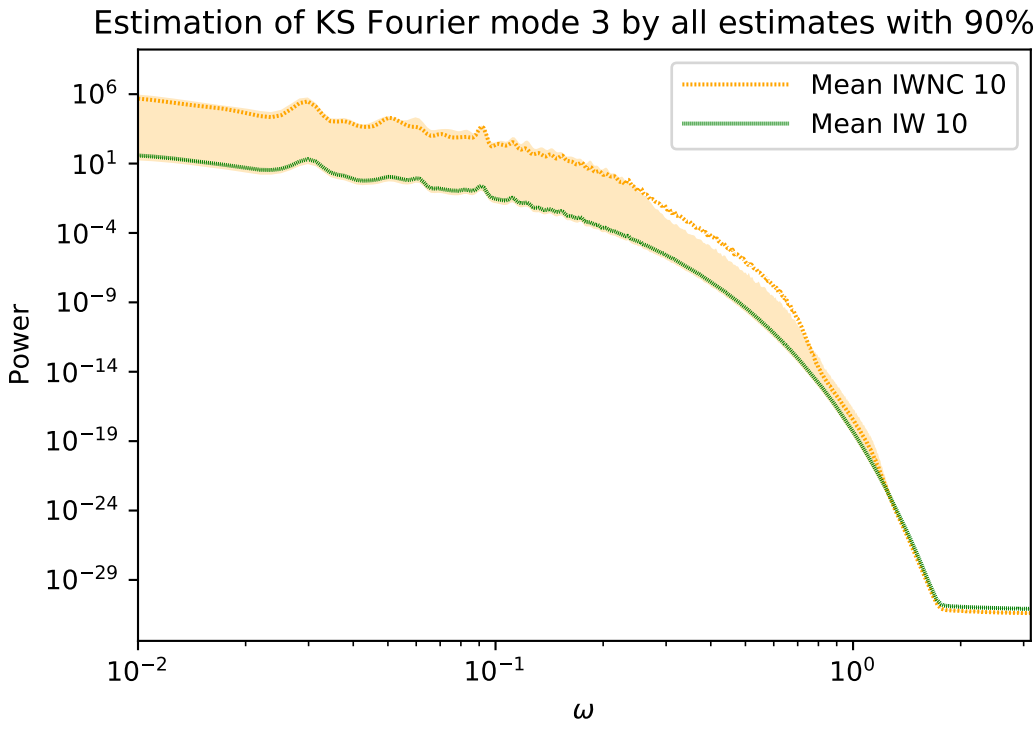


Figure 9.1: KS mode 3 spectrum estimated by noncausal IW and causal IW both on an odd frequency grid. Observe that noncausal IW exhibits some numerical issue, since the power at the low frequencies is significantly over estimated.

in the double integral under (A.27), which for Gaussian process, such as our example, constitutes the whole variance of the estimator sans the $O(1/N^{\alpha})$ term, which for the reason discussed above should decay very fast. However, when this double integral (thick, dashed black) is computed numerically there is yet a substantial gap between the empirical variance (thick dot-dashed red). Now the Bartlett estimate can be implemented as either the sum under (A.1) or the integral under (A.19). If the integral is to be used numerically one must account for aliasing by cause the equality between (A.1) and (A.19) relies on a convolution theorem for DFT, which will produce spurious terms (the aliases). Dealisaing is simply a matter of padding the in time domain [35]. So, I conducted an experiment where I took 100 realization of the Gaussian process and computed the Bartlett estimator using the integral with and without dealiasing. I then took the variance of these two collections of spectral estimates and found that one variance matched the double integral and the other matched the empirical variance of spectral computed the usual was, that is, by the sum under (A.1). The result plotted in Figure 9.2. The plot suggests that the gap between the double integral and the empirical variance has something to do with aliasing.

Through numerical experimentation surrounding this issue it was discovered that if the spectral

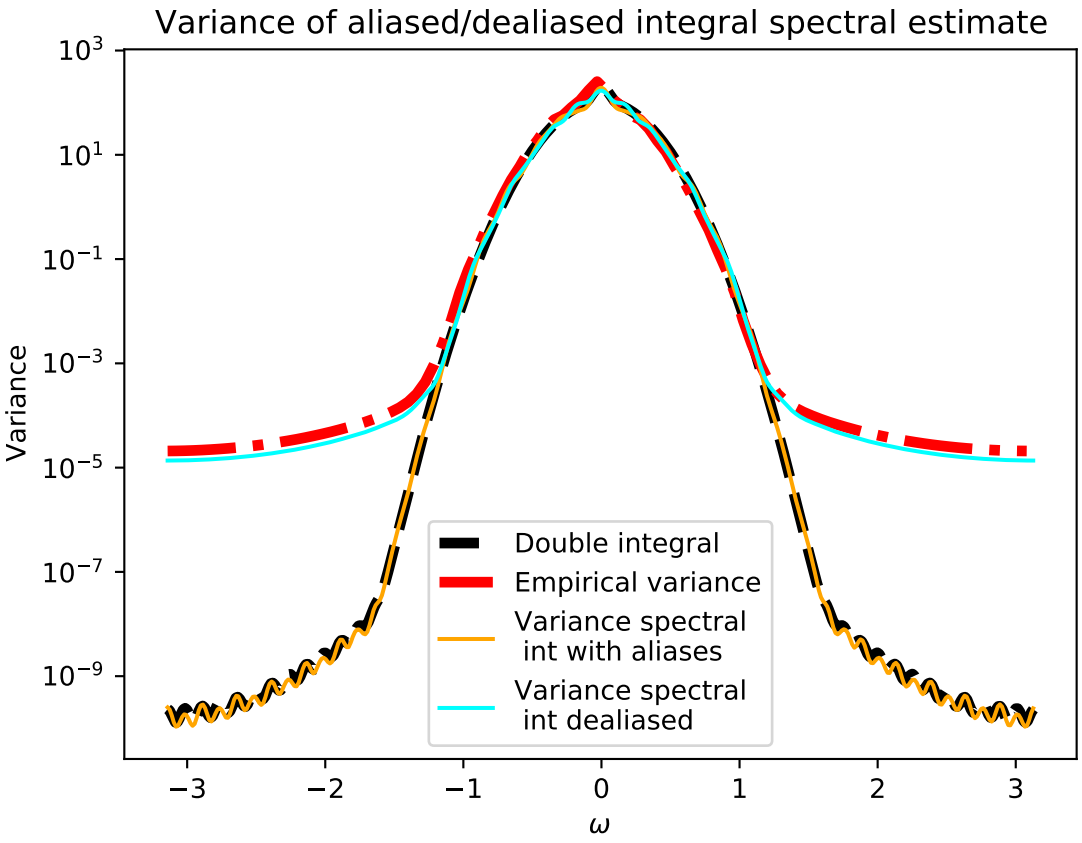


Figure 9.2: Variance of Bartlett estimate computed 3 ways and the double integral. See text.

estimate was computed as a numerical integral suggested by the form of the estimator found under (A.19) *without correcting for aliased terms. Aliaing* occurs when one seeks to apply the convolution theorem for the DFT. The result is spurious terms being added to the convolution. It is necessary to pad the sequences one wishes to convolve. However, when we failed to do that with the numerical periodogram empirical variance of those estimates appear to match the variance predicted by the double integral.

9.1.4 Parameter selection

I would like to study sensitivity L has on the final IW estimate.

More information with which to select L given N. Or perhaps even how to pick an N in situations that allow it. Also, with a little more work it seems likely that a the higher order expansion of the bias an variance of the Bartlett estimate can lead to a heuristic, or a back-of-an-envelope calculation to determine what size L and N need to be in order to reduce eliminate the estimator floor at a particular frequency.

9.1.5 Phase spectrum estimation

Similar to the power or magnitude spectrum is the phase spectrum. Some fields, such as astronomy, are interested in accurate methods of estimating these. It is currently unclear whether IW would improve the estimate of the phase spectrum.

9.2 Potential applications

Much of what I do in this section is to collect and elaborate on open questions already mention in the dissertation.

9.2.1 Applying the method of control variate to the welsh estimator

It was previously noted the method of control variate was basically able to lower the estimator floor as much as 1 iteration og iterated whiten, but unlike IW reduced the variance. It was also seen that one application of the welsh estimate accurately estimated power in a much larger spectral dynamic range than Bartlett. So, it seems reasonable to suggest that the control variate welsh procedure

could present a estimator that can accurately estimate perhaps as many as 20 decades doing so with a lower variance than IW while avoiding the numerical issues exhibited by MESA.

9.2.2 Noisy data

So, far the data on which the IW estimate has been tested is synthetic and does not exhibit the type of noisy that can be seen real-world measurement data. And so, as it stand this method cannot be recommended for use in industry until it has been test on "real" data.

As was mentioned in the introduction, many situations that are interested in frequency content deal with data that is not, strictly speaking, stationary, such as power systems engineering or audiology. It may be worthwhile to investigate the use of this technique in those fields.

9.2.3 Autocovariance estimation

The estimation of the autocovariances of a signal is an important problem in its own right and is of interest in many fields. Because of direct relation between power spectrum and autocovariances, I feel it is worth investigating the accuracy of autocovariances obtained from the IW power spectrum estimate.

9.2.4 Wiener filtering and Model reduction

One way to solve for the optimal, causal linear estimate of one signal given another signal (the Wiener filter) is to use the so-called Wiener-Hopf technique [6]. This technique requires the estimation of the power spectrum of the predictor signal and the cross spectrum of the two signals. I first became interested in power spectrum estimation when developing a numerical implementation of the Wiener-Hopf technique. It seems reasonable that the more accurate the spectral estimate the more accurate the solution should be. The mapping from data to the Wiener filter is a highly nonlinear map. I am hoping the analysis in this work and the development of this estimator may facilitate our understanding of that mapping.

One important area in which Wiener filtering gets used in data driven model reduction. Hear a Wiener filter is computed to fit a linear combination of user defined functions to to approximate the reduced order dynamics of the system.

Appendix A

Expansions of bias and variance formulas for estimators

The formulation below of the bias and variance of the spectral estimates is a modification of what can be found in ([23, 19, 3, 36]). The development found in those sources assume that the truncation parameter is much smaller than the length of the snapshot, $L \ll N$. It seems that, in practice, that assumption can be difficult to realize and as a result the estimator exhibits bias and variance that are not accurately modeled by the asymptotic formulas with some reasonable choice of L and N. Said differently, these classical formulas do not match the analogous empirical estimates. This not a critique on the asymptotic formulas found in the literature, just a statement that there does seem to be a place for formulas that do not assume $L \ll N$.

Recall the form of a Bartlett estimator. It is the discrete-time Fourier transform (DTFT, discrete in time but continuous in frequency) of a windowed version of the biased autocovariances estimate. This is in contrast with a Welsh estimator which is the modulus squared of the discrete Fourier transform (DFT, discrete in time and frequency) of a windowed (or tapered) version of the raw snapshot. For the analysis to follow we will write the Bartlett estimate, therefore, in the following form:

$$\tilde{S}(\omega) = \sum_{n=-N+1}^{N-1} \Lambda_L[n]\tilde{C}[n]e^{-in\omega}. \tag{A.1}$$

where $\tilde{C}[n]$ is the biased estimator, found under (2.30), of the autocovariances $C[n]$. Recall our assumption that the autocovariances decays exponentially as in (2.4). The development of the formulas below closely follows that of M. Priestly in [3, Sec. 6.2.4]. Now, if the lag window can be written in what Parzen calls "scale parameter form", that is, as $\Lambda_L(n) = \lambda(n/L)$ then $\lambda : [-1, 1] \to \mathbb{R}$ is called the *lag window generator*. Here, we assume the lag window is of this kind.

The remainder of this section is organized as follows. We first look at the bias of the estimator and supplement the discussion with an example. Then we consider the variance of the estimator and conclude with a discussion.

A.1 Bias

Recall the bias of an estimator is the expectation of the estimator subtracted by the (true) value being estimated.

$$\begin{aligned} b(\omega) &= \mathbb{E}\tilde{S}(\omega) - S(\omega) \\ &= \sum_{n=-N+1}^{N-1} \lambda(n/L)\mathbb{E}\tilde{C}[n]e^{-in\omega} - \sum_{n=-\infty}^{\infty} C[n]e^{-in\omega} \\ &= \sum_{n=-N+1}^{N-1} \left[\lambda(n/L)\left(1-\frac{|n|}{N}\right)-1\right] C[n]e^{-in\omega} - \sum_{|n|\geq N} C[n]e^{-in\omega} \\ &= \sum_{n=-N+1}^{N-1} \left[\lambda(n/L)-1\right] C[n]e^{-in\omega} - \frac{1}{N}\sum_{n=-N+1}^{N-1} \lambda(n/L)|n|C[n]e^{-in\omega} - \sum_{|n|\geq N} C[n]e^{-in\omega}. \end{aligned} \tag{A.2}$$

We seek to understand the scaling for the bias in L and N. Since the dependence of the bias on L is localized to the lag window generator λ, one way to analyze terms with higher order terms of L and N will be through an expansion of λ in L. It is usual to assume that λ is even and that $\lambda(0) = 1$, which we shall do here. Many useful choices of the lag window such as Bartlett and Parzen are not smooth at zero so that a Taylor expansion about zero may not exist. More generally though, if there is a number $r > 0$ such that

$$\lambda^{(r)} := \lim_{u\to 0} \frac{1-\lambda(u)}{|u|^r} \tag{A.3}$$

exists, is finite, and is nonzero, then we can write

$$\lambda(u) = 1 - \lambda^{(r)}|u|^r + o(|u|^r). \tag{A.4}$$

Further terms for the bias may be extracted by repeating this in a straightforward way. One derives the next term by finding $r' > r$ such that

$$\lambda_2^{(r')} := \lim_{u\to 0} \frac{1-\lambda_1^{(r)}|u|^r - \lambda(u)}{|u|^{r'}} \tag{A.5}$$

exists, is finite and is nonzero. (I have relabeled $\lambda^{(r)}$ as $\lambda_1^{(r)}$.) This would allow us then to write

$$\lambda(u) = 1 - \lambda_1^{(r)}|u|^r - \lambda_2^{(r')}|u|^{r'} + o(|u|^{r'}) \tag{A.6}$$

and so on.

Alternatively, we could also elect the forms

$$\lambda(u) = 1 - \lambda_1^{(r)}|u|^r + O(|u|^{r'}) \tag{A.7}$$

or

$$\lambda(u) = 1 - \lambda_1^{(r)}|u|^r - \lambda_2^{(r')}|u|^{r'} + O(|u|^{r''}) \tag{A.8}$$

where r'' is determined by continuing the process from (A.4) to (A.5) one more step. To be thorough let us continue the expansion in (A.2) by substituting (A.8) in for $\lambda(n/L)$. Doing so provides

$$b(\omega) = \sum_{n=-N+1}^{N-1} \left(-\frac{\lambda_1^{(r)}|n|^r}{L^r} - \frac{\lambda_2^{(r')}|n|^{r'}}{L^{r'}} + O\left(\frac{1}{L^{r''}}\right) \right) C[n]e^{-in\omega} \tag{A.9a}$$

$$- \frac{1}{N} \sum_{n=-N+1}^{N-1} \left(1 - \frac{\lambda_1^{(r)}|n|^r}{L^r} - \frac{\lambda_2^{(r')}|n|^{r'}}{L^{r'}} + O\left(\frac{1}{L^{r''}}\right) \right) |n| C[n]e^{-in\omega}$$

$$- \sum_{|n|\geq N} C[n]e^{-in\omega}. \tag{A.9b}$$

Now, we introduce what Parzen in his expansion of bias referred to as the "generalized r^{th} derivative," namely

$$S^{(r)}(\omega) = \sum_{n=-\infty}^{\infty} |n|^r C[n]e^{-in\omega} \tag{A.10}$$

which in truncate form, we label,

$$\tilde{S}^{(r)}(\omega) = \sum_{n=-N+1}^{N-1} |n|^r C[n]e^{-in\omega}. \tag{A.11}$$

Looking at (A.9) long enough one notices a truncated generalized derivative in every term but the last. That last term (A.9b) of the bias expansion actually shares a common form with the error $S^{(r)}(\omega) - \tilde{S}^{(r)}(\omega)$. To analyze both of these expressions, first recall our standing assumption that the autocovariances $C[n]$ decay exponentially as in (2.4), under this assumption we have

$$\left| \sum_{|n|\geq N} C[n]e^{-in\omega} \right| \leq 2K \sum_{n=N}^{\infty} \rho^{|n|} = \frac{2K\rho^N}{1-\rho} = O(\rho^N) \tag{A.12}$$

With this in hand it can be shown that

$$|S^{(r)}(\omega) - \tilde{S}^{(r)}(\omega)| = O(\rho^N). \tag{A.13}$$

And, with that in mind and using the new notation of the generalized derivative, we have the following theorem for the bias of the Bartlett estimator.

Theorem 3. *Let X be a WSS process with autocovariances that decay exponentially at rate ρ (as in (2.4)) and with snapshot x of length N. And suppose $\tilde{S}(\omega)$ is the Bartlett estimate with lag window in scaled parameter form with truncation parameter L, the lag window generator λ being even with $\lambda(0) = 1$, then the bias may be expanded as follows.*

$$\begin{aligned} b(\omega) = &- \frac{\lambda_1^{(r)}}{L^r} S^{(r)}(\omega) - \frac{\lambda_2^{(r')}}{L^{r'}} S^{(r')}(\omega) && \text{(A.14a)} \\ &- \frac{1}{N} S^{(1)}(\omega) + \frac{\lambda_1^{(r)}}{NL^r} S^{(r+1)}(\omega) + \frac{\lambda_2^{(r')}}{NL^{r'}} S^{(r'+1)}(\omega) \\ &+ O\left(\frac{1}{L^{r''}}\right) + O\left(\frac{1}{NL^{r''}}\right) + O(\rho^N) && \text{(A.14b)} \end{aligned}$$

where r and $\lambda_1^(r)$ are as in (A.3), r' and $\lambda_2^(r')$ are as in (A.6), r'' is as in (A.8) and $S^{(r)}$ is the generalized r^{th} derivative.

Using the fact that summation and differentiation commute for the power spectrum (2.5), which is valid as a consequence of the exponential decay of the autocovariances, we observe the relation between the generalized derivative and the ordinary derivative if r is even.

$$S^{(r)} = (-1)^{r/2} \frac{d^r S}{d\omega^r}. \tag{A.15}$$

So, if r is even (which is the case with the Parzen lag window where $r = 2$) in most books on the subject we find the following formula, which takes (A.14) only to leading order.

$$b(\omega) = \frac{\lambda^{(2)}}{L^2} S''(\omega) + o(L^{-2}). \tag{A.16}$$

I will now report the bias for the Bartlett window and the Parzen window. Indeed, with all this together Corollary 3.1 provides the bias for the Bartlett estimator with the lag windows above to demonstrate

Corollary 3.1. *Supposing all the conditions of Theorem 3 are met. If the Bartlett window is used the bias of this estimate is*

$$b(\omega) = -\left(\frac{1}{L} + \frac{1}{N}\right) S^{(1)}(\omega) + \frac{1}{NL} S^{(2)}(\omega) + O(\rho^N) \tag{A.17}$$

And for the Parzen window,

Corollary 3.2. *Supposing all the conditions of Theorem 3 are met. If the Parzen window is used the bias of this estimate is*

$$b(\omega) = -\frac{6}{L^2} S^{(2)}(\omega) + \frac{6}{L^3} S^{(3)}(\omega) - \frac{1}{N} S^{(1)}(\omega) + \frac{6}{NL^2} S^{(3)}(\omega) - \frac{6}{NL^3} S^{(4)}(\omega) + O(\rho^{L/2}) \tag{A.18}$$

Note that because the Parzen window expanded exactly in three terms, as seen under (2.32) the r'' term never materialized, as with the r' term in Bartlett above. The piece-wise nature of Parzen warranted a slight modification to the formulation above in that the point at which the sums in equation (A.9) are truncated, rather than being $\pm(N-1)$, is $\pm L/2$.

We now turn our attention to the variance formula.

A.2 Variance

For the variance, rather then using the estimate represented in series form as in (A.1) it is simpler to use the equivalent integral form under (3.7) for convenience I reproduce these form with its constituent parts here.

$$\tilde{S}(\omega) = \int_{-\pi}^{\pi} \hat{S}^{\text{per}}(\theta) W_L(\omega - \theta)\, d\theta \tag{A.19}$$

where

$$\hat{S}^{\text{per}}(\theta) = \frac{1}{N} \left| \sum_{n=1}^{N} x_n e^{i\theta(n-1)} \right|^2 \tag{A.20}$$

and

$$W_L(\theta) = \frac{1}{2\pi} \sum_{n=-L}^{L} \lambda(n/L) e^{-i\theta n}. \tag{A.21}$$

With this in hand, we start with the definition of the variance

$$\begin{aligned}\operatorname{var}\left(\hat{S}(\omega)\right) &= \mathbb{E}\left(\int_{-\pi}^{\pi} \hat{S}^{\mathrm{per}}(\theta) W_L(\omega-\theta)\, d\theta - \mathbb{E}\int_{-\pi}^{\pi} \hat{S}^{\mathrm{per}}(\theta) W_L(\omega-\theta)\, d\theta\right)^2 \\ &= \mathbb{E}\left(\int_{-\pi}^{\pi} \left[\hat{S}^{\mathrm{per}}(\theta) - \mathbb{E}\hat{S}^{\mathrm{per}}(\theta)\right] W_L(\omega-\theta)\, d\theta\right)^2 \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \operatorname{cov}\left(\hat{S}^{\mathrm{per}}(\theta_1), \hat{S}^{\mathrm{per}}(\theta_2)\right) W_L(\omega-\theta_1) W_L(\omega-\theta_1)\, d\theta_1 d\theta_2\end{aligned} \tag{A.22}$$

The covariance of the periodogram at different frequencies is given in the following Theorem from M. Priestly.

Theorem 4. Priestly [3, Sec. 6.2.2 p. 426] *Let X be a general linear process, i.e. it can be represented as*

$$X_n = \sum_{j=-\infty}^{\infty} g_j \varepsilon_{n-j} \tag{A.23}$$

in which (ε_n) are independent with $\mathbb{E}\varepsilon_n = 0$ and $\mathbb{E}(\varepsilon_n^4) < \infty$. Let $k_e = \mathbb{E}(\varepsilon_n)^4 - 3$ and $\alpha > 0$ be any number that satisfies

$$\sum_{j=-\infty}^{\infty} |g_j||j|^{\alpha} < \infty. \tag{A.24}$$

If $\hat{S}^{per}$ is the periodogram generated from a snapshot of N terms and S is the true power spectrum of X, then

$$\begin{aligned}&\operatorname{cov}\left(\hat{S}^{per}(\theta_1), \hat{S}^{per}(\theta_2)\right) \\ &\quad = \left(\frac{k_e}{N} + \frac{2\pi}{N}\left(F_N(\theta_1+\theta_2) + F_N(\theta_1-\theta_2)\right)\right) S(\theta_1)S(\theta_2) + O(1/N^{\alpha}).\end{aligned} \tag{A.25}$$

where F_N is the Fejér kernel of order N, given by

$$F_N(\theta) = \frac{1}{2\pi} \cdot \frac{\sin^2(N\theta/2)}{N\sin^2(\theta/2)}. \tag{A.26}$$

If it is the case that the coefficients (g_j) decay exponentially, then any $\alpha > 0$ will satisfy (A.24) and the error term $O(1/N^{\alpha})$ in (A.25) decays faster than any polynomial. Also it should be noted that is X is a Gaussian process than k_e the excess kurtosis vanishes.

Applying (A.25) to (A.22) provides

$$\begin{aligned}
\operatorname{var}\left(\hat{S}(\omega)\right) &= \frac{k_e}{N}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} S(\theta_1)S(\theta_2)W_L(\omega-\theta_1)W_L(\omega-\theta_2)\,d\theta_1 d\theta_2 \\
&\quad + \frac{2\pi}{N}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} S(\theta_1)S(\theta_2)W_L(\omega-\theta_1)W_L(\omega-\theta_2)\left(F_N(\theta_1+\theta_2)+F_N(\theta_1-\theta_2)\right)\,d\theta_1 d\theta_2 \\
&\quad + O(1/N^{\alpha}) \\
&= \frac{k_e}{N}\left(\int_{-\pi}^{\pi} S(\theta)W_L(\omega-\theta)\,d\theta\right)^2 \\
&\quad + \frac{2\pi}{N}\int_{-\pi}^{\pi} S(\theta_1)W_L(\omega-\theta_1)\int_{-\pi}^{\pi} S(\theta_2)W_L(\omega-\theta_2)\left(F_N(\theta_2+\theta_1)+F_N(\theta_2-\theta_1)\right)\,d\theta_2 d\theta_1 \\
&\quad + O(1/N^{\alpha})
\end{aligned} \tag{A.27}$$

It is very common at this point to observe that under the assumption $L \ll N$, the Fejér kernel F_N acts like a δ-function with respect to W_L and S. However, if L is not sufficiently small, approximating the Fejér kernel by the δ-function will generally incur an error of size $O(1/N)$. This fact is the point of the following lemma.

Lemma 1. *Let $\varphi(\theta_2)$ be smooth on the interval $[-\pi,\pi]$, then*

$$\int_{-\pi}^{\pi}\varphi(\theta_2)F_N(\theta_1-\theta_2)\,d\theta_2 - \varphi(\theta_1) = \frac{C(\theta_1)}{N} + O(N^{-3})$$

Proof. To see this, for the moment fix ω and consider the integral

$$\begin{aligned}
&\int_{-\pi}^{\pi}\varphi(\theta_2)F_N(\theta_2-\theta_1)\,d\theta_2 \\
&\qquad = \int_{-\pi}^{\pi}\varphi(\theta_2+\theta_1)F_N(\theta_2)\,d\theta_2 \qquad (2\pi\text{-periodicity of } \varphi \text{ and } F_N) \\
&\qquad = \varphi(\theta_1) + \varphi'(\theta_1)\int_{-\pi}^{\pi}\theta_2 F_N(\theta_2)\,d\theta_2 + \frac{\varphi''(\theta_1)}{2}\int_{-\pi}^{\pi}\theta_2^2 F_N(\theta_2)\,d\theta_2 + \ldots
\end{aligned} \tag{A.28}$$

Since F_N is even the integrands $\theta_2^n F_N(\theta_2)$ are odd if n is odd so that integrals vanish over a symmetric interval. Consequently, the terms with an odd derivative of φ in (A.28) vanish. To analyze integrals of the form

$$\int_{-\pi}^{\pi}\theta_2^n F_N(\theta_2)\,d\theta_2 \tag{A.29}$$

the following fact, which can be verified using elementary calculus, will be useful

$$\frac{1}{x^2}+\frac{1}{3}\leq\frac{1}{\sin^2(x)}\leq\frac{1}{x^2}+\left(1-\frac{4}{\pi^2}\right)\qquad \text{for}|x|\leq\frac{\pi}{2}. \tag{A.30}$$

Now, observe the following

$$\begin{aligned}\int_{-\pi}^{\pi}\theta^n F_N(\theta)\,d\theta &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\theta^n\frac{\sin^2(N\theta/2)}{N\sin^2(\theta/2)}\,d\theta &\text{(A.31)}\\ &\leq \frac{1}{N\pi}\int_0^{\pi}\theta^n\sin^2(N\theta/2)\left(\frac{1}{(\theta/2)^2}+\left(1-\frac{4}{\pi^2}\right)\right)\,d\theta &\text{(A.32)}\\ &\leq \frac{1}{N\pi}\int_0^{\pi}4\theta^{n-2}+\left(1-\frac{4}{\pi^2}\right)\theta^{n-2}\,d\theta &\text{(A.33)}\\ &= \frac{\pi^{n-2}}{N}\left(\frac{4}{n-1}+\frac{\pi^2-4}{n+1}\right). &\text{(A.34)}\end{aligned}$$

Similarly

$$\begin{aligned}\int_{-\pi}^{\pi}\theta^n F_N(\theta)\,d\theta &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\theta^n\frac{\sin^2(N\theta/2)}{N\sin^2(\theta/2)}\,d\theta &\text{(A.35)}\\ &\geq \frac{1}{N\pi}\int_0^{\pi}\theta^n\frac{\sin^2(N\theta/2)}{(\theta/2)^2}+\theta^n\frac{\sin^2(N\theta/2)}{3}\,d\theta &\text{(A.36)}\\ &= \frac{1}{N\pi}\int_0^{\pi}4\theta^{n-2}\sin^2(N\theta/2)+\frac{1}{3}\theta^n\sin^2(N\theta/2)\,d\theta &\text{(A.37)}\\ &= \int_0^{N\pi/2}\frac{2^{n+1}}{N^n\pi}\tau^{n-2}\sin^2(\tau)\,d\tau+\frac{2^{n+1}}{3N^{n+2}\pi}\tau^n\sin^2(\tau)\,d\tau &\text{(A.38)}\\ &\geq \sum_{j=1}^{N}\left(\frac{2^{n+1}}{N^n\pi}\left((j-1)\frac{\pi}{2}\right)^{n-2}+\frac{2^{n+1}}{3N^{n+2}\pi}\left((j-1)\frac{\pi}{2}\right)^{n}\right)\int_{(j-1)\frac{\pi}{2}}^{j\frac{\pi}{2}}\sin^2(\tau)\,d\tau &\text{(A.39)}\\ &= \frac{2\pi^{n-2}}{N^n}\sum_{j=1}^{N}(j-1)^{n-2}+\frac{\pi^n}{8N^{n+2}}\sum_{j=1}^{N}(j-1)^n &\text{(A.40)}\\ &= \frac{c}{N}+O(N^{-2}) &\text{(A.41)}\end{aligned}$$

where c is a positive constant. So that for and n even there exist a C_n such that

$$\pi^{n-2}\left(\frac{2}{n-1}+\frac{\pi^2}{6(n+1)}\right)\leq C_n\leq\pi^{n-2}\left(\frac{2}{n-1}+\frac{\pi^2+4}{2(n+1)}\right) \tag{A.42}$$

for which

$$\int_{-\pi}^{\pi}\theta_2^n F_N(\theta_2)\,d\theta_2=\frac{C_n}{N}+O(N^{-3}). \tag{A.43}$$

And so, the error

$$\int_{-\pi}^{\pi} \varphi(\theta_2) F_N(\theta_2 - \theta_1)\, d\theta_2 - \varphi(\theta_1) = \sum_{j=1}^{\infty} \frac{\varphi^{(2j)}(\theta_1)}{(2j)!} \int_{-\pi}^{\pi} \theta_2^{2j} F_N(\theta_2)\, d\theta_2 \tag{A.44}$$

$$= \frac{1}{N} \sum_{j=1}^{\infty} \frac{C_{2j}}{(2j)!} \varphi^{(2j)}(\theta_1) + O(N^{-3}) \tag{A.45}$$

$$= \frac{C_{S,W,L}(\theta_1)}{N} + O(N^{-3}) \tag{A.46}$$

□

Observe that if $\omega = 0$ and φ is even then $C_{S,W,L}(\theta_1)$ will be even in θ_1, since it is the sum of even order derivatives and the even order derivative of an even function is also even. However, if $\omega \neq 0$ then $C_{S,W,L}(\theta_1)$ will not be odd in general. I point this out be sure that the symmetry in the from the factor $F_N(\theta_1 - \theta_2) + F_N(\theta_1 + \theta_2)$ in (A.27) does not in general cancel the $1/N$ term in this error. Pressing forward, with this error term in hand, (A.27) can be written as

$$\begin{aligned}
\operatorname{var}\left(\hat{S}(\omega)\right) = {} & \frac{e}{N}\left(\int_{-\pi}^{\pi} S(\theta) W_L(\omega - \theta)\, d\theta\right)^2 \\
& + \frac{2\pi}{N} \int_{-\pi}^{\pi} S(\theta) W_L(\omega - \theta) \left(S(-\theta) W_L(\omega + \theta) + S(\theta) W_L(\omega - \theta)\right) d\theta \\
& + \frac{2\pi}{N} \int_{-\pi}^{\pi} S(\theta) W_L(\omega - \theta) \left(\frac{C_{S,W,L}(-\theta) + C_{S,W,L}(\theta)}{N} + O(1/N^3)\right) d\theta \\
& + O(1/N^{\alpha}) \\
= {} & \frac{e}{N}\left(\int_{-\pi}^{\pi} S(\theta) W_L(\omega - \theta)\, d\theta\right)^2 \\
& + \frac{2\pi}{N} \int_{-\pi}^{\pi} S^2(\theta) W_L(\omega - \theta) W_L(\omega + \theta)\, d\theta \\
& + \frac{2\pi}{N} \int_{-\pi}^{\pi} S^2(\theta) W_L^2(\omega - \theta)\, d\theta \\
& + \frac{4\pi}{N^2} \int_{-\pi}^{\pi} S(\theta) W_L(\omega - \theta) \big(C_{S,W,L}(-\theta) + C_{S,W,L}(\theta)\big) d\theta \\
& + O(1/N^{\alpha \wedge 3})
\end{aligned} \tag{A.47}$$

$$\begin{aligned}
= {} & \frac{e}{N}\left(\int_{-\pi}^{\pi} S(\theta) W_L(\omega - \theta)\, d\theta\right)^2 + (1 + \delta_{0,\pm\pi}(\omega)) \frac{2\pi}{N} \int_{-\pi}^{\pi} S^2(\theta) W_L^2(\omega - \theta)\, d\theta \\
& + \frac{4\pi}{N^2} \int_{-\pi}^{\pi} S(\theta) W_L(\omega - \theta) \big(C_{S,W,L}(-\theta) + C_{S,W,L}(\theta)\big) d\theta + O(1/N^{\alpha \wedge 3})
\end{aligned} \tag{A.48}$$

where

$$\delta_{0,\pm\pi}(\omega) = \begin{cases} 1, \text{ if } \omega = 0, \pm\pi \\ 0, \text{ otherwise} \end{cases} . \tag{A.49}$$

With regards the first term in (A.48), if the sequence ε_t is Gaussian, then $e = 0$. Otherwise, the term is still small since it essentially behave as $eS^2(\omega)/N$. However this can be expanded further if we as approximate W_L as a δ-function incurring this time an error of $C_{S,W}/L + o(1/L)$.

$$\frac{e}{N}\left(\int_{-\pi}^{\pi} S(\theta)W_L(\omega-\theta)\,d\theta\right)^2 = \frac{eS^2(\omega)}{N} + \frac{2eC_{S,W}S(\omega)}{NL} + \frac{eC_{S,W}^2}{NL^2} + o(1/NL) \tag{A.50}$$

$$= \frac{eS^2(\omega)}{N} + \frac{2eC_{S,W}S(\omega)}{NL} + o(1/NL). \tag{A.51}$$

In the third term we can again approximate W_L as a delta function and achieve

$$\frac{4\pi}{N^2}\int_{-\pi}^{\pi} S(\theta)W_L(\omega-\theta)\big(C_{S,W,L}(-\theta) + C_{S,W,L}(\theta)\big)d\theta$$

$$= \frac{4\pi S(\omega)C_{S,W,L}^{\pm}(\omega)}{N^2} + \frac{4\pi K_{S,W,L}(\omega)}{N^2 L} + o(1/N^2L) \tag{A.52}$$

where $C_{S,W,L}^{\pm}(\omega) = C_{S,W,L}(-\omega) + C_{S,W,L}(\omega)$ and $K_{S,W,L}(\omega)$ is the leading coefficient to the error in approximating W_L by a δ-function, this dependents on the shape of the functions S and W_L and also $C_{S,W,L}^{\pm}$.

The second term in (A.48) is much larger than the others. The function $W_L^2(\omega-\theta)$ is highly concentrated at ω. In fact

$$\int_{-\pi}^{\pi} W_L^2(\theta)\,d\theta \to \infty \qquad \text{as} \qquad L \to \infty. \tag{A.53}$$

So, then since S is continuous at zero we can write

$$\begin{aligned}\int_{-\pi}^{\pi} S^2(\theta)W_L^2(\omega-\theta)\,d\theta &= S^2(\omega)\int_{-\pi}^{\pi} W_L^2(\omega-\theta)\,d\theta + \mathrm{Err}(L,\omega) \\ &= S^2(\omega)\int_{-\pi}^{\pi} W_L^2(\theta)\,d\theta + \mathrm{Err}(L,\omega) \\ &= S^2(\omega)\sum_{n=-L}^{L}\lambda^2(n/L) + \mathrm{Err}(L,\omega)\end{aligned}$$

by Parseval's theorem. The form of $\mathrm{Err}(L,\omega)$ is difficult to determine in generality. Based on a

numerical study of this function with the Parzen window it was seen to obey a $1/L$ power law with a coefficient that depends of S and ω. by Parseval's theorem. Observe that

$$\frac{1}{L}\sum_{n=-L}^{L}\lambda^2(n/L) \to \int_{-1}^{1}\lambda^2(u)\,du \qquad \text{as } L\to\infty \tag{A.54}$$

where the sum is thought of as an approximating sum of the integral using the right-hand rule. The error can be found to be

$$\int_{-1}^{1}\lambda^2(u)\,du - \frac{1}{L}\sum_{n=-L}^{L}\lambda^2(n/L) = \frac{1}{L}\frac{d(\lambda^2)}{du}(u_0) = \frac{C_{\lambda,L}}{L} \tag{A.55}$$

Using this we have

$$\begin{aligned}(1+\delta_{0,\pm\pi}(\omega))\frac{2\pi}{N}\int_{-\pi}^{\pi}S^2(\theta)W_L^2(\omega-\theta)\,d\theta =&(1+\delta_{\omega,0,\pi})\frac{2\pi L}{N}S^2(\omega)\int_{-1}^{1}\lambda^2(u)\,du\\ &+\frac{L}{N}\frac{C_{\lambda,L}}{L}+\frac{\mathrm{Err}(L,\omega)}{N}\end{aligned}$$

Collecting all of these terms gives the following

$$\mathrm{var}\left(\hat{S}(\omega)\right) =(1+\delta_{\omega,0,\pi})\frac{2\pi L}{N}S^2(\omega)\int_{-1}^{1}\lambda^2(u)\,du+\frac{C_{\lambda,L}}{N}+\frac{\mathrm{Err}(L,\omega)}{N} \tag{A.56}$$

$$\frac{4\pi S(\omega)C^{\pm}_{S,W,L}(\omega)}{N^2}+\frac{4\pi K_{S,W,L}(\omega)}{N^2L}+o(1/N^2L) \tag{A.57}$$

$$+e\left(\frac{S^2(\omega)}{N}+\frac{2C_{S,W}S(\omega)}{NL}+o(1/NL)\right) \tag{A.58}$$

An empirical study was performed to verify the scaling of the variance of Bartlett estimator (this will be discussed in greater detail in Chapter 5). It was found that for frequencies above the estimator floor the variance of the estimator conformed to the leading order asymptotic scaling of L/N. However, for frequencies for which the power was consistently overestimated to the estimator floor, the scaling of the variance was independent of L and scaled in N like $1/N^2$. And so, in the case of the two Gaussian processes introduced earlier We see, as with bias, higher order terms dominating the variance in the region of the estimator floor.

‘

A.2.1 Discussion

This analysis also allows us to suggest the snapshot size N and truncation parameter L that would be needed for the Bartlett estimator to produce a reasonable low mean-square error for all frequencies.

In the demonstration, the Parzen window was used. However, I would like to point out that as the analysis below will show, altering the window will not change the performance of the Bartlett estimate in a way that will significantly reduce the estimator floor.

As was mentioned, it seems that for high-contrast spectra N and L may need to be prohibitively large before the above formulas accurately model the bias or variance of the estimator.

What I have learn from this analysis together with the numerical study associated with it is that when L and N are not sufficiently large the bias and variance are complex functions of the structure of spectrum being estimated and the lag window. I have provided an analysis that attempts to better expose the complexities of these relations. After we have discussed

Appendix B

Causal spectral factorization

B.1 Introduction

Given a wide-sense stationary (WSS) stochastic process $X = \big(X_n, n \in \mathbb{Z}\big)$ its z-spectrum $S_X(z)$ is defined by[1]

$$S_X(z) = \sum_{n=-\infty}^{\infty} C_X[n] z^{-n}, \tag{B.2}$$

where $C_X[n] = \mathbb{E} X_n X_0^*$ is the autocovariance sequence of the process X. This function is real-valued on the unit circle since, $C_X[-n] = C_X^*[n]$, and by the Wiener-Khinchin theorem, it is, in fact, nonnegative definite on the unit circle[2] Spectral factorization refers to a particular factorization in which

$$S_X(z) = L(z) L^*(z^{-*}) \tag{B.3}$$

where $L(z)$ is *minimum phase*, meaning both $L(z)$ and $L^{-1}(z) = 1/L(z)$ are analytic on and outside the unit circle. The function L will be referred to as a spectral factor of S_X. This factorization is not unique since for any spectral factor of S_X right-multiplication by a unitary matrix produces another spectral factor of S_X.

In this note I describe a method presented by Kailath et al. in [27] and [6, p. 336] for finding a spectral factor under certain conditions. The method employs a version of Kalman filtering, more specifically the Kalman recursion for finding the innovations process for a process with a finite

[1] A note on notation: the superscript $*$ denotes the complex conjugate transpose (Hermitian transpose) operator for matrices and just the complex conjugate for real or complex numbers. The superscripts $-*$ together is an abbreviation for the complex conjugate of the multiplicative inverse, i.e.

$$z^{-*} = \frac{1}{z^*}. \tag{B.1}$$

[2] Though what follows holds for scalar processes it also holds for vector valued processes. As such, I assume the vector case. Meaning, the covariance sequence is a sequence of square $d \times d$-matrices, where d is the dimension of X_0.

state-space model. The innovations for such a process will be defined below. Section B.2 provides a summary of relevant results from Kalman filtering, together with some intuition relevant to the present context. With those results in hand, Section B.3 presents the method for finding spectral factors. Then, in Section B.4 are examples.

B.2 Kalman filtering

The purpose of this section is to provide information from Kalman filtering theory that is pertinent to the development of the factorization algorithm to be discussed. To that end, the *summum bonum* of the section is the computation of what I refer to as a *weak* modeling filter for a given process. The associated *weak* whitening filter also bears significance in this work and will be briefly treated.

Suppose we have a (discrete-time) stochastic process[3] $X = \big(X_i,\ i = 0, 1, 2, \dots\big)$ with a finite state-space representation of the following form:

$$\begin{cases} \theta_{i+1} = F_i\theta_i + G_i u_i & \text{(B.4a)} \\ X_i = H_i\theta_i + v_i & \text{(B.4b)} \end{cases}$$

where $F_i \in C^{m\times m}$, $G_i \in C^{m\times p}$, and $H_i \in C^{d\times m}$ are known matrices, and $u = (u_i)$, $v = (v_i)$, and θ_0 are random variables with the following property[4]

$$\mathbb{E}\begin{pmatrix} \theta_0 \\ u_i \\ v_i \end{pmatrix}\begin{pmatrix} \theta_0 \\ u_j \\ v_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi_0 & 0 & 0 & 0 \\ 0 & Q_i\delta_{ij} & S_i\delta_{ij} & 0 \\ 0 & S_i^*\delta_{ij} & R_i\delta_{ij} & 0 \end{pmatrix} \tag{B.4d}$$

[3] From here till Section B.2.4 X need not be WSS.

[4] Throughout the note

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{B.4c}$$

In the context here, I think of X as a given series of *observations*, they are directly measurable, and $\theta = \big(\theta_i,\ i = 0, 1, 2, \dots\big)$ is the *state* of the process (usually not directly measurable). Both of these series are sequences of random variables, along with u and v. The matrices F_i, G_i, H_i, Π_0, Q_i, S_i and R_i are deterministic for all i.

B.2.1 A few words about the mathematical context

In the sequel there will be much discussion on linear least-mean-square estimators A nice thing about least-mean-square estimators is an orthogonality principle, namely the estimator error is orthogonal to the predictor. What follows also affords such a principle. However, there is no Hilbert space that comfortable houses this theory. So, I will build up the context here.

The context requires the iterations of vectors of more than one size, since the observations and the state need not be the same dimension. Since it is possible to write solutions to the problems I consider in terms of covariance information a natural "inner product" may be defined by

$$\langle x, y\rangle = \mathbb{E}xy^* \qquad \text{for all } x \in \mathbb{C}^d, y \in \mathbb{C}^m, \qquad d, m \in \mathbb{N}, \tag{B.5}$$

So, that $\langle x, x\rangle \in \mathbb{C}^{d\times d}$ is the covariance of x and $\langle x, y\rangle \in \mathbb{C}^{d\times m}$ is the cross covariance of x and y. We use the notation $x \perp y$ if x and y are uncorrelated, meaning $\langle x, x\rangle = 0 \in \mathbb{C}^{d\times m}$.

Because this inner product[5] is matrix-valued, rather than adding structure to a linear space with a scalar field the underlining space must be a module of vectors, with a ring of (square) matrix-valued scalars.The natural choice of scalars are matrices, of the appropriate size. This is an important point and leads me to another definition. Let

$$\operatorname{span}_d\{y_i,\ i = 0, 1, \dots, N\} := \left\{ \sum_{i=0}^{N} A_i y_i \;\middle|\; y_i \in \mathbb{C}^m, A_i \in \mathbb{C}^{d\times m}, \quad i = 0, 1, \dots, N \right\} \tag{B.6}$$

I would like us to think of the span as possibly set in a different space from the arguments.

There is a norm

$$\|x\| = \sqrt{\langle x, x\rangle} \tag{B.7}$$

associated with this inner product, defined in the usual way, using the unique (Hermitian) positive-semi-definite square root of the Hermitian positive-semidefinite matrix $\langle x, x\rangle$.

[5]If its arguments are restricted to the same space, It can be shown that $\langle\cdot,\cdot\rangle$ satisfies the axioms of an inner product, provided positivity is replaced with positive definiteness.

B.2.2 the Kalman filter

Now, the Kalman filter allows us to find, for some $i > 0$, the linear least-mean-squares estimator of the state θ_i given all preceding observations X_j for $j = 0, 1, 2, \ldots, i-1$. This quantity I will denote by $\hat{\theta}_{i|i-1}$. Since in this note I will only be concerned with linear least-mean-squares estimators given the X_i's the subscript $\cdot_{|i}$ will denote linear estimators with respect to and conditioned on all X_j for $j \leq i$. Consistent with this convention, when I write $\hat{X}_{i|i-1}$ I mean the linear least-mean-squares estimator of X_i given X_j for $j < i$.

For the stake of rigor and completeness I have provided lemma which describes the orthogonality principle in this context.

Lemma 2. *Given the inner product defined under (B.5) and the norm induced by it, under (B.7), let $x \in C^d$ and $y_i \in \mathbb{C}^m$ for $i = 0, \ldots, N$ be random variables. If $\hat{x}$ is the linear least-mean-squares estimator of x given y, meaning*

$$\mathbb{E}(x - \hat{x})(x - \hat{x})^* = \|x - \hat{x}\|^2 = \inf_{x' \in \text{span}_d\{y_i,\, \forall i\}} \|x - x'\|^2 \tag{B.8}$$

then $(x - \hat{x}) \perp y_i$ for $i = 0, 1, \ldots, N$.

We seek a least-mean-square estimator of θ_{i+1} redistricted to a linear combination of observations, $\{X_j,\ j < i\}$. In this context we seek $\hat{\theta}_{i+1|i}$ such that,

$$\hat{\theta}_{i+1|i} \in \mathcal{L}_i =: \text{span}\{X_i,\ i = 0, 1, \ldots, N\} \tag{B.9}$$

$$\hat{\theta}_{i+1|i} = \underset{\hat{\theta} \in \mathcal{L}_i}{\text{argmin}} \|\theta_{i+1} - \hat{\theta}\|^2 \tag{B.10}$$

Borrowing from Hilbert space theory, we consider building an orthogonal basis for $\mathcal{L}_i$, $e = \{e_j,\ j = 1, 2, \ldots, i\}$ and write

$$\hat{\theta}_{i+1|i} = \sum_{j=0}^{i} \langle \theta_{i+1}, e_j \rangle R_{e,j}^{-1} e_j \tag{B.11}$$

where $R_{e,i} = \langle e_i, e_i \rangle$. We are not in a Hilbert space, so the claim that the estimator under (B.11) satisfies (B.10) will need to be verified. But first, let us build this orthogonal set in a way similar to the Grahm-Schmidt process, but without normalizing. Put

$$e_i = X_i - \hat{X}_{i|i-1} \quad \text{for } i > 0 \qquad \text{and} \qquad e_0 = X_0 \tag{B.12}$$

The $\{e_i\}$ have the property $\mathcal{L}_i = \text{span}\{e_j,\ j \leq i\} = \mathcal{L}_i$ for any $i \geq 0$, and $\langle e_i, e_j \rangle = R_{e,i}\delta_{ij}$ by Lemma 2.

This stochastic process we will refer to as the *innovations* associated with X. Observe that because they are defined as the residual of least squares estimates over a progressively increasing space the innovations are an orthogonal set. And we having $\langle e_i, e_j \rangle = I\delta_{ij}$. There are other orthogonality relations I would generally like to highlight. First noting that $\theta_i \in \text{span}\{\theta_0, u_j,\ j < i\}$, then by use of (B.4b) and the definition of e_i one can show the following.

$$\langle \theta_i, u_j \rangle = 0 \quad \text{and} \quad \langle \theta_i, v_j \rangle = 0 \quad \text{for} \quad i \leq j; \tag{B.13a}$$

$$\langle X_i, u_j \rangle = 0 \quad \text{and} \quad \langle X_i, v_j \rangle = 0 \quad \text{for} \quad i < j; \tag{B.13b}$$

$$\langle X_i, u_j \rangle = S_i^* \quad \text{and} \quad \langle X_i, v_j \rangle = R_i \quad \text{for} \quad i = j; \tag{B.13c}$$

$$\langle e_i, u_j \rangle = 0 \quad \text{and} \quad \langle e_i, v_j \rangle = 0 \quad \text{for} \quad i < j; \tag{B.13d}$$

$$\langle e_i, u_j \rangle = S_i^* \quad \text{and} \quad \langle e_i, v_j \rangle = R_i \quad \text{for} \quad i = j. \tag{B.13e}$$

B.2.3 Kalman filter recursions for the innovations

Kalman developed a recursion to cheaply find the innovations process associated with the observations, which we will derive here. I like the way Kailath put it, "It turns out that the recursive construction of the innovations combines nicely with the recursive evolution of the state variables to give a recursion for the innovations..."[6, p. 312].

We seek a recursion for $\{e_i\}$ but start with finding a recursion for $\hat{\theta}_{i|i-1}$ using (B.11),

$$\begin{aligned}
\hat{\theta}_{i+1|i} &= \sum_{j=0}^{i} \langle \theta_{i+1}, e_j \rangle R_{e,j}^{-1} e_j \\
&= \left(\sum_{j=0}^{i-1} \langle \theta_{i+1}, e_j \rangle R_{e,j}^{-1} e_j \right) + \langle \theta_{i+1}, e_i \rangle R_{e,i}^{-1} e_i \\
&= \hat{\theta}_{i+1|i-1} + \langle \theta_{i+1}, e_i \rangle R_{e,i}^{-1} e_i
\end{aligned}$$

The first term of the last line undermines the recursion in $\hat{\theta}_{j|j-1}$; however, we can remedy this by conditioning equation (B.4a) on $\mathcal{E}_{i-1}$ to get

$$\hat{\theta}_{i+1|i-1} = F_i\hat{\theta}_{i|i-1} + G_i\hat{u}_{i|i-1} = F_i\hat{\theta}_{i|i-1} + 0 \tag{B.14}$$

by (B.13b). This provides

$$\hat{\theta}_{i+1|i} = F_i\hat{\theta}_{i|i-1} + \langle\theta_{i+1}, e_i\rangle R_{e,i}^{-1} e_i \tag{B.15}$$

The next thing to do is to make sure we can compute e_i from the $\hat{\theta}_{i|i-1}$'s. If we condition equation (B.4b) on $\mathcal{E}_{i-1}$ then

$$\hat{X}_{i|i-1} = H_i\hat{\theta}_{i|i-1} + \hat{v}_{i|i-1} = H_i\hat{\theta}_{i|i-1} + 0 \tag{B.16}$$

also by (B.13b). So, that

$$e_i = X_i - \hat{X}_{i|i-1} = X_i - H_i\hat{\theta}_{i|i-1}. \tag{B.17}$$

From this point I will use the abbreviated notation $\hat{\theta}_i = \hat{\theta}_{i|i-1}$.

Putting this together, we get a recursion for the innovations

$$\begin{cases} \hat{\theta}_{i+1} = F_i\hat{\theta}_i + K_i e_i, \quad \hat{\theta}_0 = 0 & \text{(B.18a)} \\ e_i = X_i - H_i\hat{\theta}_i. & \text{(B.18b)} \end{cases}$$

where $K_i = \langle\theta_{i+1}, e_i\rangle R_{e,i}^{-1}$.

Here then is the recursion we sought, which recursively generates the innovations associated with the process X. For our interests, and for many others, it would be useful if the K_i's and the $R_{e,i}$'s could be computed off-line, that is, independent of the observations. As it turns out this is possible and I will demonstrate how this may be accomplished.

Let $\tilde{\theta}_i = \theta_i - \hat{\theta}_i$ and $P_i = \langle\tilde{\theta}_i, \tilde{\theta}_i\rangle$. It turns out that this quantity is very useful in formulating an off-line recursion. It is easy to see by subtracting (B.16) from (B.4b) that $e_i = H_i\tilde{\theta}_i - v_i$, and this is enough to put $R_{e,i}$ in terms of P_i, because $\hat{\theta}_i$ is orthogonal to v_i by (B.13b), since $\hat{\theta}_i \in \mathfrak{E}_{i-1}$[6] , we get

[6]This means $\hat{\theta}$ is measurable with respect to $\mathfrak{E}_{i-1}$, which is independent of v_i.

$$R_{e,i} = H_i P_i H_i^* + R_i. \tag{B.19}$$

To get K_i in terms of only P_i and $R_{e,i}$, we start with $K_i = \langle \theta_{i+1}, e_i \rangle R_{e,i}^{-1}$ and observe that by (B.4b) $\langle X_{i+1}, e_i \rangle = F_i \langle X_i, e_i \rangle + G_i \langle u_i, e_i \rangle$. Reflecting on both inner products in turn provides

$$\begin{aligned} \langle X_i, e_i \rangle &= \langle X_i, \tilde{X}_i \rangle H_i^* + \langle X_i, v_i \rangle \\ &= \langle \hat{X}_i + \tilde{X}_i, \tilde{X}_i \rangle H_i^* + 0 \\ &= P_i H_i^* \end{aligned}$$

and $\langle u_i, e_i \rangle = S_i$, as we have seen. Substituting, we able to find that

$$K_i = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}, \tag{B.20}$$

as desired. Now, with $R_{e,i}$ and K_i in place a recursion for P_i can be established by combining equations (B.4a) and (B.18a) and taking the covariance of both sides, a somewhat long calculation results in the discrete Lyapunov recursion,

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_i R_{e,i} K_i^*, \qquad i \geq 0 \tag{B.21}$$

which we initialize by

$$P_0 = \langle \hat{\theta}_0, \hat{\theta}_0 \rangle = \langle \theta_0, \theta_0 \rangle = \Pi_0. \tag{B.22}$$

We have just have just found a recursion for the K_i and $R_{e,i}$, which are necessary in the computation of e_i, which is independent of observations. It also provides a recursion for the error covariance of the state estimate (which is useful in its own right). The discussion so far can be summarized into a single theorem

Theorem 5. *Consider the standard state-space model*

$$\begin{cases} \theta_{i+1} &= F_i \theta_i + G_i u_i \\ \\ X_i &= H_i \theta_i + v_i \end{cases} \tag{B.23}$$

for $i \geq 0$, and satisfying (B.4d). The innovations process of X can be recursively computed using the equations

$$\begin{cases} e_i = X_i - H_i\hat{\theta}_i, & i \geq 0, \quad \text{(B.24a)} \\ \hat{\theta}_{i+1} = F_i\hat{\theta}_i + K_{p,i}e_i, & \hat{\theta}_0 = 0, \quad \text{(B.24b)} \end{cases}$$

where

$$P_{i+1} = F_iP_iF_i^* + G_iQ_iG_i^* - K_iR_{e,i}K_i^*, \qquad P_0 = \Pi_0 \quad \text{(B.25a)}$$

$$R_{e,i} = H_iP_iH_i^* + R_i \quad \text{(B.25b)}$$

$$K_i = (F_iP_iH_i^* + G_iS_i)R_{e,i}^{-1} \quad \text{(B.25c)}$$

from (B.4d). Here, $P_i = \langle\tilde{\theta}_i, \tilde{\theta}_i\rangle$ where $\tilde{\theta}_i = \theta_i - \hat{\theta}_i$.

Observe that the discrete Lyapunov recursion (B.21) can now be written as a discrete Riccati recursion,

$$P_{i+1} = F_iP_iF_i^* + G_iQ_iG_i^* - (F_iP_iH_i^* + G_iS_i)(H_iP_iH_i^* + R_i)^{-1}(H_iP_iF_i^* + S_i^*G_i^*). \quad \text{(B.26)}$$

B.2.4 The stationary case and the modeling filter

So far in this section, all we have assumed about X is that it is a discrete-time stochastic process that has a finite state-space representation defined by (B.4). But now it is time to assume conditions native to the problem at hand, which will be delineated shortly. These assumptions are afforded by the stationarity of the processes I work with. The plan is to realize stationarity and adjust the innovations process so as to produce both stationary innovations and stationary state estimators.

Here are the further assumptions I impose on the process X,

- the state-space representation for X is time-invariant, meaning, all the parameters are constant in time. So,

$$F_i = F, \quad G_i = G, \quad H_i = H, \quad Q_i = Q, \quad S_i = S \quad \text{and} \quad R_i = R \qquad \text{for } i \geq 0 \quad \text{(B.27)}$$

  additionally I impose that F be stable (meaning its eigenvalues lie strict in the unit circle) and R be strictly positive-definite, $R > 0$, also

- θ_i is in steady state, meaning $\Pi_i = \Pi$ for $i \geq 0$ where is Π is the positive-semi-definite solution to the Lyapunov equation.

$$\Pi = F\Pi F^* + GQG^*. \tag{B.28}$$

  Since F is stable this has a unique solution.

A consequence of these assumptions in that X is wide-sense stationary, and its model can be written as

$$\begin{cases} \theta_{i+1} = F\theta_i + Gu_i & \text{(B.29a)} \\ X_i = H\theta_i + v_i & \text{(B.29b)} \end{cases}$$

where

$$\mathbb{E}\begin{pmatrix} \theta_0 \\ u_i \\ v_i \end{pmatrix}\begin{pmatrix} \theta_0 \\ u_j \\ v_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi & 0 & 0 & 0 \\ 0 & Q\delta_{ij} & S\delta_{ij} & 0 \\ 0 & S^*\delta_{ij} & R\delta_{ij} & 0 \end{pmatrix} \tag{B.29c}$$

It can be shown by direct computation that covariance function of X, $R_X(i-j)$, can be written as follows [6, p. 267]

$$R_X(i-j) = \begin{cases} HF^{i-j-1}N & i > j, \\ R + H\Pi H^* & i = j, \\ N^*(F^*)^{j-i-1}H^* & i < j, \end{cases} \tag{B.30}$$

where $N = F\Pi H^* + GS$.

Under these conditions we now apply Theorem 5 and find that the innovations sequence associated with X has the following state-space representation

$$e_i = X_i - H\hat{\theta}_i, \qquad i \geq 0, \tag{B.31a}$$

$$\hat{\theta}_{i+1} = F\hat{\theta}_i + K_{p,i}e_i, \qquad \hat{\theta}_0 = 0, \tag{B.31b}$$

where now

$$P_{i+1} = FP_iF^* + GQG^* - K_iR_{e,i}K_i^*, \qquad P_0 = \Pi \tag{B.32a}$$

$$R_{e,i} = HP_iH^* + R \tag{B.32b}$$

$$K_i = (FP_iH^* + GS)R_{e,i}^{-1} \tag{B.32c}$$

Rearranging (B.31) slightly we can write

$$\begin{cases} \hat{\theta}_{i+1} = F\hat{\theta}_i + K_{p,i}e_i, \qquad \hat{\theta}_0 = 0, & \text{(B.33a)} \\ X_i = H\hat{\theta}_i + e_i, \qquad i \geq 0. & \text{(B.33b)} \end{cases}$$

Which provides an alternative way of modeling the process X, this time driven by the innovations process e. Said differently (B.33) defines a model $\mathcal{L}$ which maps e to X.

Observe that $\mathcal{L}$, clearly causal, is also causally invertable since by algebra we can write,

$$\begin{cases} \hat{\theta}_{i+1} = F_p\hat{\theta}_i + K_{p,i}X_i, \qquad \hat{\theta}_0 = 0, & \text{(B.34a)} \\ e_i = -H\hat{\theta}_i + X_i, \qquad i \geq 0. & \text{(B.34b)} \end{cases}$$

This contrasts with the original model which took (θ_0, u, v) to X[7] , such is causally but not generally causally invertible.

Now, what can be said about the output $\mathcal{L}w$, for $w = (w_i,\ i = 0, 1, 2, \dots)$ a noise sequence with $\langle w_i, w_j \rangle = R_{e,i}\delta_{ij}$? Is it a stationary process with the same z-spectrum as X? The output in general is not stationary, since the model, found under (B.33), represents a time-variant system. This begs the question, does this model converge, in time, to a time-invariant model? It turns out that the system parameters P_i, $R_{e,i}$, and K_i all converge as $i \to \infty$. Which suggests a time-invariant causal system that, when fed in a stationary input, will certainly produce a stationary output. A *modeling filter* for X is causal, linear system which constructs the process X by passing a white-noise process through the filter. This is what we are seeking.

The convergence of P_i has been well studied [6, Sec. 14.5] and there are many conditions one can check to ensure that the limit of the recursion converges to the stabilizing solution of the appropriate Racatti equation necessary to provide the causal and casually invertible modeling filter, namely the

[7] One thing it has done was remove the additional input variable, the initial state

discrete algebraic Racatti equation (DARE), given by

$$P = FPF^* + GQG^* - (FPH^* + GS)(HPH^* + R)^{-1}(HPF^* + S^*G^*) \tag{B.35}$$

$$= FPF^* + GQG^* - KR_eK^* \tag{B.36}$$

Since this theory is more problem dependent I will defer the matter of convergence to when we have a specific problem.

So, assume that

$$\lim_{i\to\infty} P_i = P \tag{B.37}$$

So that P in the unique solution to

$$P = FPF^* + GQG^* - KR_eK^* \tag{B.38}$$

where $\lim R_{e,i} = R_e = HPH + R$ and $\lim K_i = K = (FPH^* + GS)R_e^{-1}$, and that P is such that $F_p = F - KH$ is stable. Let us, then, consider the system

$$L : w \mapsto Y \tag{B.39}$$

where $w = (w_i,\ i \geq -\infty)$ with $\langle w_i, w_j\rangle = R_e\delta_{ij}$ and $Y = (Y_i,\ i \geq -\infty)$[8] given by

$$\begin{cases} \eta_{i+1} = F\eta_i + Kw_i, \qquad \eta_{-\infty} = 0 & \text{(B.40a)} \\ Y_i = H\eta_i + w_i, \qquad i \geq -\infty. & \text{(B.40b)} \end{cases}$$

Since η_i was initialized in the remote past (independent of w_i) we have $\langle \eta_i, w_i\rangle = 0$ for $i > -\infty$ and $\langle \eta_i, \eta_i\rangle = \Sigma$ for $i > -\infty$ where

$$\Sigma = F\Sigma F^* + KR_eK^*. \tag{B.40c}$$

We want to check that the autocovariances of $\{Y_i\}$ is the same as for $\{X_i\}$ which is given in

[8]Notice the slight-of-hand, we now start our processes in the remote past and have them be bilaterally infinite. This is to obviate the need for the state variable to be initialized as an input in order for them to have a variance that obeys the Lyapunov equation below for each finite time

(B.30). This can be checked directly. For $i > 0$

$$\begin{aligned}
R_Y[i] &= \langle Y_i, Y_0 \rangle \\
&= \langle H\eta_i + w_i, H\eta_0 + w_0 \rangle \\
&= \langle HF\eta_{i-1} + HKw_{i-1} + w_i, H\eta_0 + w_0 \rangle \\
&= \langle HF^2\eta_{i-2} + HFKw_{i-2} + HKw_{i-1} + w_i, H\eta_0 + w_0 \rangle \\
&= \langle HF^i\eta_0 + HF^{i-1}Kw_0 + \cdots + HFKw_{i-2} + HKw_{i-1} + w_i, H\eta_0 + w_0 \rangle \\
&= \langle HF^i\eta_0, H\eta_0 \rangle + \langle HF^{i-1}Kw_0 + \cdots + HFKw_{i-2} + HKw_{i-1} + w_i, H\eta_0 \rangle \\
&\quad + \langle HF^i\eta_0, w_0 \rangle + \langle HF^{i-1}Kw_0 + \cdots + HFKw_{i-2} + HKw_{i-1} + w_i, w_0 \rangle \\
&= HF^i\Sigma H^* + 0 + 0 + HF^{i-1}KR_e \\
&= HF^i\Sigma H^* + HF^{i-1}(FPH^* + GS) \\
&= HF^{i-1}\big(F(\Sigma + P)H^* + GS)\big),
\end{aligned}$$

a similar calculation for $i < 0$ yields

$$R_Y[i] = \big(F(\Sigma + P)H^* + GS)\big)^*(F^*)^{i-1}H^* = R_Y^*[-i],$$

and for $i = 0$

$$\begin{aligned}
R_Y(0) &= \langle Y_0, Y_0 \rangle \\
&= \langle H\eta_0 + w_0, H\eta_0 + w_0 \rangle \\
&= H\Sigma H^* + R_e \\
&= H\Sigma H^* + HPH^* + R \\
&= H(\Sigma + P)H^* + R
\end{aligned}$$

Now what is $\Sigma + P$, by there definitions

$$\begin{aligned}
\Sigma + P &= F\Sigma F^* + KR_eK^* + FPF^* + GQG^* - KR_eK^* \\
&= F(\Sigma + P)F^* + GQG^*.
\end{aligned}$$

Meaning that $\Sigma + P$ solves the same discrete Lyapunov equation that Π solved in (B.28). We

therefore conclude

$$\Pi = \Sigma + P \tag{B.41}$$

and that

$$R_Y[i] = \begin{cases} HF^{i-1}N & i > 0, \\ R + H\Pi H^* & i = 0, \\ N^*(F^*)^{-i-1}H^* & i < 0, \end{cases} \qquad = R_X[i] \tag{B.42}$$

recalling that $N = F\Pi H^* + GS$, as in equation (B.30). This demonstrates that the causal LTI system L provides a model filter for produce process with the same autocovariance sequence as X given any white noise sequence with covariance $R_e\delta_{ij}$.

Furthermore, observe that L (given by (B.40)) is causal and causally invertible, since, by simple algebra again, L^{-1} can be written as

$$\begin{cases} \eta_{i+1} = F_p\eta_i + KY_i, & \eta_{-\infty} = 0, \\ w_i = -H\eta_i + Y_i, & i \geq -\infty. \end{cases} \tag{B.43a, B.43b}$$

with $\langle \eta_0, Y_i \rangle = 0$ for $i \geq 0$ and now η_0 is chosen such that $\langle \eta_0, \eta_0 \rangle = \Sigma_p$ where

$$\Sigma_p = F_p\Sigma_p F_p^* + KR_Y[0]K^* \tag{B.43c}$$

where $F_p = F - KH$ which is stable since P was a stabilizing solution to the Racatti equation given under (B.38). This then a whitening filter since.

I now summarize the above discussion into a theorem

Theorem 6. *Let $X = (X_i;\ i > 0)$ be a stationary process with a time-invariant state-space representation of the form given in (B.29). Then, the LTI system L given by (B.40) is causal and causally invertable, and constructs a stationary process with the same autocovariances as X by passing in any white-noise process with variance R_e.*

B.3 Spectral factorization by Kalman filtering

In this section I explain how the above results from Kalman filtering may be employed in finding an approximate spectral factor of a given power spectrum (z-spectrum).

Given a z-spectrum $S_X(z)$ of a WSS discrete-time process $X_n \in \mathbb{C}^d$ for $n > -\infty$, by definition,

$$S_X(z) = \sum_{n=-\infty}^{\infty} R_X[n] z^{-n}, \tag{B.44}$$

where

$$R_X[n] = \operatorname{cov}(X_{k+n}, X_k) = \mathbb{E} X_{k+n} X_k^* = \langle X_{k+n}, X_k \rangle \tag{B.45}$$

is the covariance. Now, if the decay of the covariance is sufficiently fast it is reasonable to truncate $S_X(z)$ to a Laurent polynomial

$$\tilde{S}_X(z) = \sum_{n=-m}^{m} R_X[n] z^{-n}. \tag{B.46}$$

However, this isn't always the best way to approximate the spectrum. It is usual to apply some windowing function to the autocovariances. The Parzen windowing function Λ_m, given below, for some $m > 0$ is an example.

$$\Lambda_m[n] = \begin{cases} 1 - 6\left(\frac{n}{m}\right)^2 + 6\left(\frac{|n|}{m}\right)^3, & |n| \le m/2 \\ 2\left(1 - \frac{|n|}{m}\right)^3, & m/2 < |n| \le m \\ 0, & |n| > m \end{cases}. \tag{B.47}$$

The Fourier transform $\widehat{\lambda}_m(\omega)$ of the Parzen windowing function is

$$\widehat{\Lambda}_m(\omega) = \frac{3}{4} m \left(\frac{\sin \pi\omega m/2}{\pi\omega m/2} \right)^4, \qquad -\infty \le \omega \le \infty \tag{B.48}$$

Let $R_{\text{win}}[n] = \Lambda_m[n] R_X[n]$ and

$$S_{\text{win}}(z) = \sum_{n=-m}^{m} R_{\text{win}}[n] z^{-n}. \tag{B.49}$$

The function $S_{\text{win}}(z)$ is generally a better approximation of $S_X(z)$ than $\tilde{S}_X(z)$.

Now that we have approximated the z-spectrum we seek to factor by a Laurent polynomial, we can construct a process $Y = \big(Y_i,\ i = 0, 1, 2, \dots\big)$ with a finite, time-invariant, state-space model whose z-spectrum is exactly $S_{\text{win}}(z)$, that is $S_Y(z) = S_{\text{win}}(z)$ [27, p. 488]. We then take the state-space model and using Kalman, compute the modeling filter.

The process Y is constructed using the following finite, time-invariant, state-space model:

$$\begin{cases} \theta_{i+1} = F\theta_i + Gv_i & \text{(B.50a)} \\ Y_i = H\theta_i + u_i & \text{(B.50b)} \end{cases}$$

Where[9]

$$\mathbb{E}\begin{pmatrix} \theta_0 \\ v_i \\ u_i \end{pmatrix}\begin{pmatrix} \theta_0 \\ v_j \\ u_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi & 0 & 0 & 0 \\ 0 & R\delta_{ij} & S\delta_{ij} & 0 \\ 0 & S^*\delta_{ij} & Q\delta_{ij} & 0 \end{pmatrix}. \tag{B.50c}$$

The parameters F, H, and N (think of the N from equation (B.30), are defined explicitly by

$$F = \begin{pmatrix} 0 & & & \\ I & 0 & & \\ & I & 0 & \\ & & \ddots & \ddots \\ & & & I & 0 \end{pmatrix} \in \mathbb{C}^{md\times md}, \tag{B.51}$$

$$H = \begin{pmatrix} 0 & \dots & 0 & I \end{pmatrix} \in \mathbb{C}^{d\times md}, \tag{B.52}$$

and

[9]Recall d is the dimension of the observation vectors and m is the degree of the Laurent polynomial approximation of the given z-spectrum

$$N = \begin{pmatrix} R_{\text{win}}[m] \\ R_{\text{win}}[m-1] \\ \vdots \\ R_{\text{win}}[1] \end{pmatrix} \in \mathbb{C}^{md \times d}. \tag{B.53}$$

The remaining quantities are less important almost auxiliary. With the exception of Π, they have no bearing on the covariance data of the process (recall equation (B.30)). The existence of Π is guaranteed by the fact that F is stable. It's value never needs to be determined, as we will see. For completeness if one were interested in actually realizing the process Y fro this state-space form the remaining variables must satisfy the following conditions

$$\Pi = F\Pi F^* + GQG^* \tag{B.54}$$

$$GS = N - F\Pi H^* \tag{B.55}$$

$$R = R_{\text{win}}[0] - H\Pi H^* \tag{B.56}$$

Now, provided all of the above can be satisfied it can be shown that Y is stationary and its autocovariances are

$$\begin{aligned} C_Y[i] &= \begin{cases} HF^{i-1}N & i > 0 \\ H\Pi H^* + R & i = 0 \\ NF^{*(-i-1)}H^* & i < 0 \end{cases} \\ &= \begin{cases} 0 & i > m \\ R_{\text{win}}[i] & m \geq i \geq -m \\ 0 & i < -m \end{cases} \end{aligned}$$

So, that $S_Y(z) = S_{\text{win}}(z) \approx S_X(z)$ as desired.

In Y, we have a stationary process with a time-invariant state-space representation, given under (B.50). This representation induces We now invoke Theorem 6, and assert that the system M : $w = (w_i,\ i > -\infty) \mapsto Y' = (Y'_i,\ i > -\infty)$ given by

$$\begin{cases} \eta_{i+1} = F\eta_i + Kw_i, & \eta_{-\infty} = 0 \qquad \text{(B.57a)} \\ Y'_i = H\eta_i + w_i, & i \geq -\infty, \qquad \text{(B.57b)} \end{cases}$$

is causal and causally invertible and constructs a stationary process Y' with the same autocovariances of Y by passing in any white-noise process with variance R_e. This system M constitutes a modeling filter for Y. Observe that from System (B.57) it can be shown that

$$Y'_i = \sum_{j=1}^{m} HF^{j-1}Kw_{i-j} + w_i = \sum_{j=0}^{m} K_i^{\#} w_{i-j} \qquad \text{(B.58)}$$

where

$$K_i^{\#} = \begin{cases} K_{m-i+1} & 1 \leq i \leq m \\ I & i = 0 \end{cases} \qquad \text{(B.59)}$$

The notation views $K \in C^{d \times md}$ as a column vector with matrix entries, $(d \times d)$-blocks. So in words, when $K \in C^{d \times md}$ is viewed as a block column vector with $d \times d$-blocks, $K^{\#}$ is the block-reverse of K appended on top by the $d \times d$ Identity matrix.

And so defines an $MA(m)$ (model) process.

B.4 An Example

Consider the MA(2) process given by $Y_n = \frac{1}{2}u_n + u_{n-1}$. This has a z-spectrum of

$$\begin{aligned} S_Y(z) &= \left(\frac{1}{2} + z^{-1}\right)\left(\frac{1}{2} + z\right) \\ &= \frac{5}{4} + \frac{1}{2}z^{-1} + \frac{1}{2}z \\ &= \sum_{n=-1}^{-1} a_n z^{-n} \qquad \text{for } a_0 = \frac{5}{4},\ a_1 = a_{-1} = \frac{1}{2} \end{aligned}$$

We seek a spectral factorization of $S_Y(z)$. Now this can be done by hand since

$$S_Y(z) = \left(\frac{1}{2} + z^{-1}\right)\left(\frac{1}{2} + z\right) \tag{B.60}$$

$$= \left(1 + \frac{1}{2}z^{-1}\right)\left(1 + \frac{1}{2}z\right) \tag{B.61}$$

$$\tag{B.62}$$

Observe that (B.60) is not a minimum phase factorization since $1/2 + z^{-1}$ has its root outside the unit disk, but (B.61) is. Guided by the discussion in Section B.3. We assign

$$F = 0, \qquad H = 1, \qquad G = Q = \Pi = 1, \qquad S = N = a_1 = \frac{1}{2}, \qquad \text{and} \tag{B.63}$$

$$R = a_o - 1 = \frac{1}{4}, \qquad K = a_1 = \frac{1}{2}, \qquad P_1 = \Pi = 1 \tag{B.64}$$

And check that the state-space model given by

$$\begin{cases} \theta_{i+1} = F\theta_i + Gu_i & \text{(B.65a)} \\ \tilde{Y}_i = H\theta_i + v_i, & \text{(B.65b)} \end{cases}$$

with

$$\mathbb{E}\begin{pmatrix} \theta_0 \\ v_i \\ u_i \end{pmatrix}\begin{pmatrix} \theta_0 \\ v_j \\ u_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi & 0 & 0 & 0 \\ 0 & R\delta_{ij} & S\delta_{ij} & 0 \\ 0 & S^*\delta_{ij} & Q\delta_{ij} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/4\delta_{ij} & 1/2\delta_{ij} & 0 \\ 0 & 1/2\delta_{ij} & \delta_{ij} & 0 \end{pmatrix}. \tag{B.66}$$

Satisfies

$$\Pi = F\Pi F^* + GQG^* = 0 + 1 \tag{B.67}$$

$$GS = N - F\Pi H^* = \frac{1}{2} - 0 \tag{B.68}$$

$$R = R_{\text{win}}[0] - H\Pi H^* = a_0 - 1 = \frac{1}{4}. \tag{B.69}$$

It can be shown that it does. And must therefore have the autocovariances of Y Now, these process have the same autocovariances but can strike one as being very different since no reference to we don't seem to ever regress of $v_i$in fact simplifying we obtain,

$$\tilde{Y}_i = v_i + u_{i-1}. \tag{B.70}$$

The key here is the correlation between u_i and v_i

$$|\langle u_i, v_i\rangle|^2 = a_1^2 = \frac{1}{4} = (a_0 - 1)(1) = \langle u_i, u_i\rangle\langle v_i, v_i\rangle \tag{B.71}$$

So that be Cauchy-Schwarz we see that these are completely correlated. So, that $v_i = \alpha u_i$ and

$$\frac{1}{2} = a_1 = \langle u_i, v_i\rangle = \langle u_i, \alpha u_i\rangle = \langle u_i, u_i\rangle\alpha^* = Q\alpha^* = \alpha^* \tag{B.72}$$

and so, $v_i = \frac{1}{2}u_i$. And therefor

$$\tilde{Y}_i = \frac{1}{2}u_i + u_{i-1}. \tag{B.73}$$

Now we run the Kalman recursion.

$$R_{e,i} = HP_iH^* + R = P_i + \frac{1}{4} \tag{B.74}$$

$$K_i = FP_iH^* + GS = 0 + a_1 = \frac{1}{2} \tag{B.75}$$

$$P_{i+1} = FP_iF^* + GQG^* - K_iR_{e,i}^{-1}K_i^* = 0 + 1 - \frac{1}{4}\frac{1}{P_i + \frac{1}{4}} = \frac{P_i}{P_i + 1/4} \tag{B.76}$$

The recursion in P_i can be solve analytically, with $P_1 = 1$ it can be shown $0 < P_i < 1$ for $i > 1$ and that P_i is increasing in if a limit does exist it is $\frac{3}{4}$ which solves the Racatti equation.

References

[1] Neil J Cornish and Tyson B Littenberg. Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches. *Classical and Quantum Gravity*, 32(13):135012, 2015.

[2] Kevin K. Lin and Fei Lu. Data-driven model reduction, wiener projections, and the koopman-mori-zwanzig formalism. *Journal of Computational Physics*, 424:109864, 2021.

[3] Maurice Bertram Priestley. *Spectral analysis and time series: Univariate series*, volume 1. Academic press, 1981.

[4] Athanasios Papoulis and S Unnikrishna Pillai. *Probability, random variables and stochastic processes*. McGraw-Hill Higher Education, 4th edition, 2002.

[5] James M Hyman and Basil Nicolaenko. The kuramoto-sivashinsky equation: a bridge between pde's and dynamical systems. *Physica D: Nonlinear Phenomena*, 18(1-3):113–126, 1986.

[6] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall information and system sciences series. Prentice Hall, 2000.

[7] Alan V Oppenheim, Alan S Willsky, and Syed Hamid Nawab. *Signals and systems*. Prentice hall Upper Saddle River, NJ, 2nd edition, 1997.

[8] Bernt Øksendal and Bernt Øksendal. *Stochastic differential equations*. Springer, 2003.

[9] P Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.

[10] Alessandro Martini, Stefano Schmidt, and Walter Del Pozzo. Maximum entropy spectral analysis: a case study, 2021.

[11] David J Thomson. Spectrum estimation techniques for characterization and development of wt4 waveguide—i. *Bell System Technical Journal*, 56(9):1769–1815, 1977.

[12] David J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

[13] Petr Vaníček. Approximate spectral analysis by least-squares fit: Successive spectral analysis. *Astrophysics and Space Science*, 4:387–391, 1969.

[14] Peter J Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[15] James B Elsner and Anastasios A Tsonis. *Singular spectrum analysis: a new tool in time series analysis*. Springer Science & Business Media, 1996.

[16] D S G Pollock, R C Green, and Truong Nguyen. *Handbook of time series analysis, signal processing, and dynamics*. Elsevier, 1999.

[17] Neil J Cornish and Tyson B Littenberg. Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches. *Classical and Quantum Gravity*, 32(13):135012, 2015.

[18] M S Bartlett. Periodogram analysis and continuous spectra. *Biometrika*, 37(1/2):1–16, 1950.

[19] E Parzen. Mathematical considerations in the estimation of spectra. *Technometrics*, 3(2):167–190, 1961.

[20] F.J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

[21] Silvia Maria Alessio. Digital signal processing and spectral analysis for scientists: concepts and applications. 2015.

[22] C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4):1088–1107, 1997.

[23] E Parzen. On consistent estimates of the spectrum of a stationary time series. *The Annals of Mathematical Statistics*, 28(2):329–348, 1957.

[24] Joseph L Doob. *Stochastic processes*. John Wiley & Sons, 1953.

[25] Ulf Grenander and Gabor Szegö. *Toeplitz forms and their applications*. Univ of California Press, 1958.

[26] Lasha Ephremidze, Gigla Janashia, and Edem Lagvilava. An analytic proof of the matrix spectral factorization theorem. 2008.

[27] Ali H Sayed and Thomas Kailath. A survey of spectral factorization methods. *Numerical linear algebra with applications*, 8(6-7):467–496, 2001.

[28] Harry Press and John W Tukey. Power spectral methods of analysis and their application to problems in airplane dynamics. In *Instrumentation Systems*, pages 1–IVC. Elsevier, 1956.

[29] J P Burg. Maximum entropy spectral analysis. Paper presented at the 37th Annual International Meeting, Society of Exploration Geophysicists, Oklahoma City, 1967.

[30] J P Burg. *Maximum Entropy Spectral Analysis*. PhD thesis, Stanford CA, 1975.

[31] E T Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, May 1957.

[32] E T Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, 1982.

[33] P F Fougere. Maximum entropy calculations on a discrete probability space. In *Maximum-Entropy and Bayesian Methods in Science and Engineering: Foundations*, pages 205–234. Springer, 1988.

[34] Aly-Khan Kassam and Lloyd N Trefethen. Fourth-order time-stepping for stiff pdes. *SIAM Journal on Scientific Computing*, 26(4):1214–1233, 2005.

[35] GS Patterson Jr and Steven A Orszag. Spectral calculations of isotropic turbulence: Efficient removal of aliasing interactions. *The Physics of Fluids*, 14(11):2538–2541, 1971.

[36] Ulf Grenander and Murray Rosenblatt. *Statistical analysis of stationary time series*. Chelsea Publishing Company, 2 edition, 1984.