Uniform FIR Approximation of Causal Wiener Filters with Applications to Causal Coherence

by Leighton P. Barnes

Submitted to the
Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

at the
Massachusetts Institute of Technology

June 2015

Copyright 2015 Leighton P. Barnes. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole and in part in any medium now known or hereafter created.

Author:

Department of Electrical Engineering and Computer Science
May 22, 2015

Certified by:
George Verghese, Henry Ellis Warren Prof. of Electrical Engineering,
Thesis Supervisor
May 22, 2015

Accepted by:
Prof. Albert R. Meyer, Chairman, Masters of Engineering Thesis Committee
Uniform FIR Approximation of Causal Wiener Filters with Applications to Causal Coherence

by Leighton P. Barnes

Submitted to the Department of Electrical Engineering and Computer Science

May 22, 2015

In Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science

Abstract

Leveraging the relationship between Wiener filtering and the coherence function, a version of coherence is defined that captures the causal relationship between WSS processes. This causal coherence is interpreted in a modeling context and used to demonstrate what a frequency dependent measure for causality both can and can’t represent. To understand how well frequency dependent coherence spectra can be estimated with finite order approximations, the convergence of the FIR causal Wiener filters to the full IIR causal Wiener filter is investigated as filter length goes to infinity. The main results prove $L^p$ convergence of the frequency responses for $p = 1, 2, \infty$ under certain Hölder continuity conditions on the power spectra, as well as give asymptotic upper bounds for the convergence error.

*Keywords:* Wiener filters, causality, coherence, FIR approximation
Acknowledgements

I would like to thank all of my professors and mentors at MIT. Special thanks to my advisor, George Verghese, for good ideas and great guidance. This thesis would not exist without his help. Additional thanks to Ehi Nosakhare and Peter Hagelstein for help with finding data and checking computations.

To my mother and brother, for teaching me to love numbers, puzzles, and games from an early age. To my father, for teaching me to be thoughtful and patient; that if I were to do something, I should take my time and do it right.

To my beautiful girlfriend Jenelle, for putting up with my obsessive nature. She’s been there for me through sleepless tooling and desperate punting.

Finally, to my friends on Second West and at the Muddy. Without you guys, the ’tvte would have eaten me long ago.
1. Introduction

When analyzing random signals it is useful to understand how they correlate and what information they share. One measure for determining how random signals correlate is the coherence function between two jointly wide-sense stationary (WSS) processes. Coherence is a frequency dependent analog of the usual determination coefficient (i.e., $r^2$) between random variables. One interpretation of the coherence function is that it represents how well one processes can be estimated from the other by minimum mean squared error (MSE) linear, time-invariant (LTI) filtering. The LTI filter used for this filtering is called the unconstrained Wiener filter. In this thesis, we define a new generalization of this interpretation of coherence which captures how well one WSS process can be estimated from another jointly WSS process using any arbitrary class of LTI estimation filters. We then consider this generalization with respect to specific classes of estimation filters. In particular, if the estimation filter is taken to be the causal Wiener filter, then this results in a coherence spectrum that captures causality. This idea of a frequency dependent measure for causal dependence also underlies other existing measures such as Granger causality [1] and Geweke’s formulation [2].

The classical coherence function (sometimes called the magnitude squared coherence function) between discrete-time jointly WSS processes $x[\cdot]$ and $y[\cdot]$ has the familiar form

$$K_{xy}(\Omega) = \frac{|S_{xy}(\Omega)|^2}{S_{xx}(\Omega)S_{yy}(\Omega)}.$$  \hspace{1cm} (1)
Given the power spectral densities (PSDs) $S_{xx}(\Omega), S_{yy}(\Omega)$ and cross spectral density $S_{xy}(\Omega)$ of $x[\cdot]$ and $y[\cdot]$, one can trivially compute (1). If instead we are given measured data from realizations of $x[\cdot]$ and $y[\cdot]$, then using standard spectral analysis methods such as Welch’s method ([3] Section 8.2), we can estimate the required power spectra and thus $K_{xy}(\Omega)$. In order to compute “causal” coherence spectra, however, we require an estimate of the causal Wiener filter.

The causal Wiener filter is the potentially infinitely long causal LTI filter for minimal MSE estimation of one WSS process from another jointly WSS process. It is often used in problems of denoising, estimation, and system identification. In general, this filter is difficult to calculate from measured data or known power spectra because it requires minimum phase spectral factorization and causal/anticausal decomposition. Additionally, in practice, only finitely many autocorrelation values can be used in any computation, and similarly, a resulting filter can only be computed at finitely many points in frequency or time. Because of these limitations, a causal, finite-duration impulse response (FIR) Wiener filter is often used as an approximation of the true causal Wiener filter.

Motivated by applications that could potentially benefit from the notion of “causal” coherence spectra between WSS processes, and the need to estimate these spectra, we investigate the convergence of the causal FIR Wiener filters to the true causal Wiener filter as filter length goes to infinity. It is found that if the power and cross spectral densities of the processes are sufficiently smooth, then the corresponding frequency responses converge uniformly (i.e., in $L^\infty$ norm) to that of the true causal Wiener filter. In par-
ticular, if they are \( k \geq 1 \) times differentiable with \( \alpha \)-Hölder continuous \( k \)th derivative then the convergence error will be \( O(N^{-r}) \) where \( N \) is the length of the FIR Wiener filter and

\[
r = \begin{cases} 
\frac{\alpha + k}{4}, & \text{if } \alpha + k \geq \frac{4}{3} \\
\alpha, & \text{if } \alpha + k < \frac{4}{3}.
\end{cases}
\]

Even in the case that \( k = 0 \), we show that uniform FIR approximations can be constructed by windowing sufficiently long FIR Wiener filters. The proof of uniform convergence is inspired by Fejér’s classical theorem regarding uniform convergence of Cesàro means of Fourier series of continuous functions.

In proving these results, we build on [4] which gave sufficient conditions for a causal Wiener filter to have Hölder continuous frequency response and have uniform FIR approximations. The uniform FIR approximations constructed there are those guaranteed by Jackson’s theorems (see e.g., [5], Chapter 3 Theorem 3.15). These are simply windowed versions of the true causal Wiener filter, and therefore they are no easier to compute. Our results show that the FIR Wiener filter, which can be computed easily with the autocorrelation normal equations, can be used instead. Other work in [6] investigated the limiting behavior of the inverse covariance matrix under similar spectral smoothness conditions, but did not consider uniform convergence of the full estimation filter. Aside from being independently interesting, uniform convergence is of particular importance when using a Wiener filter in the computation of a coherence spectrum, where we are interested in peaks that may show up anywhere in the entire frequency band.
1.1. Preliminaries

We will consider real-valued, discrete-time, jointly WSS processes \( x[\cdot] \) and \( y[\cdot] \); for simplicity, and with no real loss of generality, assume they are zero mean. Denote the associated autocorrelation and cross-correlation functions by \( R_{xx}[\cdot], R_{yy}[\cdot], \) and \( R_{xy}[\cdot] \). We will assume throughout that the process \( x[\cdot] \) has PSD \( S_{xx}(\Omega) \), and the processes \( x[\cdot] \) and \( y[\cdot] \) have cross spectral density \( S_{xy}(\Omega) \), that are both well-behaved enough for the autocorrelation and PSD to satisfy the discrete-time Fourier transform (DTFT) relationships

\[
S_{xx}(\Omega) = \sum_{n=-\infty}^{\infty} R_{xx}[n] e^{-j\Omega n} \quad (3)
\]

and

\[
R_{xx}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(\Omega) e^{j\Omega n} d\Omega \quad (4)
\]

with analogous relations for \( S_{xy}(\Omega) \). The technical conditions for this to be true will be made precise later.

Recall the useful identities

\[
S_{xy}(\Omega) = S_{yx}(-\Omega) = S_{yx}^*(\Omega) \quad (5)
\]

and

\[
S_{(h\ast x)y}(\Omega) = H(\Omega) S_{xy}(\Omega) , \quad (6)
\]

which we will use extensively. The process \((h \ast x)[\cdot]\) resulting from LTI filtering of \( x[\cdot] \) by a filter with impulse response \( h[\cdot] \) and associated frequency response \( H(\Omega) \) (which is the DTFT of \( h[\cdot] \)) is guaranteed to be jointly WSS with \( y[\cdot] \), and (6) is also valid, provided that \( h[\cdot] \) is square summable ([7]
Let $H_N(\Omega)$ be the frequency response of the FIR causal Wiener filter of length $N$ for minimum MSE estimation of $y[\cdot]$ from $x[\cdot]$, and let $H_c(\Omega)$ be the frequency response of the corresponding IIR causal Wiener filter. Formally, we have the following solutions [8]:

$$H_N(\Omega) = \sum_{n=0}^{N-1} h_N[n] e^{-j\Omega n}$$  \hspace{1cm} (7)

where

$$\begin{pmatrix} R_{xx}[0] & R_{xx}[1] & \ldots & R_{xx}[N-1] \\ R_{xx}[1] & R_{xx}[0] & \ldots & R_{xx}[N-2] \\ \vdots & \ddots & \vdots \\ R_{xx}[N-1] & R_{xx}[N-2] & \ldots & R_{xx}[0] \end{pmatrix} \begin{pmatrix} h_N[0] \\ h_N[1] \\ \vdots \\ h_N[N-1] \end{pmatrix} = \begin{pmatrix} R_{yx}[0] \\ R_{yx}[1] \\ \vdots \\ R_{yx}[N-1] \end{pmatrix},$$  \hspace{1cm} (8)

and

$$H_c(\Omega) = \frac{1}{F(\Omega)} \left[ \frac{S_{yx}(\Omega)}{F(-\Omega)} \right]_+.$$  \hspace{1cm} (9)

In (9), $S_{xx}(\Omega) = F(\Omega)F(-\Omega)$ is a minimum phase spectral factorization and $[\cdot]_+$ denotes the transform of the causal part. To be certain $H_N(\Omega)$ and $H_c(\Omega)$ are well-defined, we assume throughout that $S_{xx}(\Omega)$ is bounded away from zero, i.e., $S_{xx}(\Omega) \geq \delta > 0$. This condition ensures that (8) is invertible [9], that the minimum phase spectral factorization exists ([3] Section 3.5), and that $|H_c(\Omega)|$ is finite.

2. Coherence

The classical coherence function (1) between two WSS processes $x[\cdot]$ and $y[\cdot]$ is a normalized version of the magnitude squared cross spectral density.
It is a frequency dependent function that describes how \(x[\cdot]\) and \(y[\cdot]\) correlate at each frequency \(\Omega\). It is a real valued function that takes values \(0 \leq K_{xy}(\Omega) \leq 1\) and is symmetric with respect to the signals \(x[\cdot]\) and \(y[\cdot]\) (i.e., \(K_{xy}(\Omega) = K_{yx}(\Omega)\)). To see how the coherence function represents the relationship between \(x[\cdot]\) and \(y[\cdot]\), consider the following MSE calculation for any estimation filter \(H(\Omega)\) estimating \(y[\cdot]\) from \(x[\cdot]\). Let \(e[n] = y[n] - h * x[n]\) be the estimation error. In (10) the frequency arguments are suppressed for convenience.

\[
MSE(H) = E\left(e[n]^2\right) = R_{ee}[0] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{ee}(\Omega)d\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} (S_{yy} - H^* S_{yx} - HS_{xy} + HH^* S_{xx})d\Omega \tag{10}
\]

If we pick

\[
H(\Omega) = \frac{S_{yx}(\Omega)}{S_{xx}(\Omega)}, \tag{11}
\]

which is the unconstrained Wiener filter, then

\[
MSE(H) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( S_{yy}(\Omega) - \frac{|S_{yx}(\Omega)|^2}{S_{xx}(\Omega)} \right) d\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{yy}(\Omega) (1 - K_{xy}(\Omega))d\Omega. \tag{12}
\]

In other words, at frequencies where \(K_{xy}(\Omega) = 1\) the PSD of the error is zero, while at frequencies where \(K_{xy}(\Omega) = 0\) there is no predictive power and the PSD of the error is \(S_{yy}(\Omega)\). The coherence \(K_{xy}(\Omega)\) therefore represents how well the optimal filter estimates \(y[\cdot]\) from \(x[\cdot]\) at each frequency \(\Omega\).
There is actually a stronger statement we can make while interpreting $K_{xy}(\Omega)$. It does not just represent how well the optimal filter estimates $y[\cdot]$ from $x[\cdot]$ at each frequency $\Omega$, it represents how well any filter can estimate $y[\cdot]$ from $x[\cdot]$ at each frequency $\Omega$. This means that at any fixed frequency $\Omega$, the choice of $H(\Omega)$ in (11) is the one that minimizes $S_{ee}(\Omega)$. This can be seen by differentiating the expression for $S_{ee}(\Omega)$ from (10) with respect to the real and imaginary parts of the parameter $H(\Omega)$.

2.1. Causal Coherence

Now consider the following generalization of the classical coherence function. The definition of coherence (1) is equivalent to

$$K_{xy}(\Omega) = 1 - \frac{S_{ee}(\Omega)}{S_{yy}(\Omega)}$$

where $e[\cdot]$ is the error associated with the choice of $H(\Omega)$ in (11). However, the expression (13) makes sense even when $e[\cdot]$ is the error associated with any estimation filter $H(\Omega)$. So we can consider (13) to give a frequency dependent value representing how well $H(\Omega)$ estimates $y[\cdot]$ from $x[\cdot]$ at frequency $\Omega$. Regardless of the choice of $H(\Omega)$, we still have that (13) is real-valued, bounded by one, and will have value $K_{xy}(\Omega) = 1$ if and only if $S_{ee}(\Omega) = 0$. Note that in this more general case $K_{xy}(\Omega)$ will not necessarily be nonnegative, and in general $K_{xy}(\Omega) \neq K_{yx}(\Omega)$.

One particularly interesting choice of estimation filter is the causal Wiener filter $H_c(\Omega)$ from (9), and we restrict ourselves to this for the remainder of
the thesis. In this case we define

$$K_{x\rightarrow y}(\Omega) = 1 - \frac{S_{ee}(\Omega)}{S_{yy}(\Omega)}$$

$$= 1 - \frac{S_{yy} - H_c S_{yx} - H_c S_{xy} + H_c H_c^* S_{xx}}{S_{yy}}$$

$$= \frac{2\Re (H_c(\Omega) S_{xy}(\Omega)) - |H_c(\Omega)|^2 S_{xx}(\Omega)}{S_{yy}(\Omega)}$$

(14)

to be the causal coherence from \(x[\cdot]\) to \(y[\cdot]\). In (14), \(\Re(\cdot)\) denotes the real part and the quantity \(e[\cdot]\) is the estimation error of the causal filter. This causal coherence spectrum gives a sense of how well the two processes can be related by a causal LTI system with input \(x[\cdot]\) and output \(y[\cdot]\). While it is somewhat simple compared to the various causality spectra based on multivariate autoregressive (MVAR) models ([1],[2]), this definition of causal coherence has the nice property that it is immediately comparable to the classical coherence \(K_{xy}(\Omega)\). Since (11) minimizes \(S_{ee}(\Omega)\) at each \(\Omega\), the classical coherence \(K_{xy}(\Omega)\) is the maximum possible coherence value across all estimation filters. In particular we have

$$K_{x\rightarrow y}(\Omega) \leq K_{xy}(\Omega) .$$

(15)

The difference

$$\Delta K_{xy}(\Omega) = K_{xy}(\Omega) - K_{x\rightarrow y}(\Omega)$$

(16)

can be thought of as the price of causality. It is the amount of coherence that is lost when using the optimal causal estimation filter instead of the unconstrained optimum.
One limitation of this definition of causal coherence, or indeed any definition of causal coherence, is that it cannot represent the best coherence value achievable by a causal estimation filter at a given frequency. This can be demonstrated easily by fixing an \( \Omega_0 \) and considering the causal filter that is simply the constant \( H(\Omega) \equiv S_{yx}(\Omega_0)/S_{xx}(\Omega_0) \). This estimation filter minimizes \( S_{ee}(\Omega_0) \) and achieves the classical coherence value at \( \Omega_0 \). In this sense, causal coherence is not as natural a quantity as the classical value. It gives a frequency dependent metric for how well the causal Wiener filter models the relationship between \( x[\cdot] \) and \( y[\cdot] \), but it does not represent some sort of limit of how well a causal filter can estimate one process from another at each frequency.

2.2. Computation

The classical coherence (1) is relatively easy to compute by using existing spectral analysis techniques. In order to compute (14), however, one requires an estimate of \( H_c(\Omega) \). Furthermore, this estimate should be valid across all \( \Omega \in [-\pi, \pi) \) so that regardless of which frequency samples are being used, the estimate will be valid. Any erroneous peaks in a coherence spectrum could be misinterpreted as being significant. We will take the approach of estimating \( H_c(\Omega) \) with the FIR Wiener filter \( H_N(\Omega) \) for sufficiently long filter length \( N \). Section 3 is devoted to showing how exactly \( H_N(\Omega) \) converges to \( H_c(\Omega) \) as \( N \) goes to infinity. In Section 2.2.1, we compute \( K_{xy}(\Omega) \) for example processes \( x[\cdot] \) and \( y[\cdot] \), and we demonstrate how this approximation behaves for increasing \( N \).
2.2.1. An Analytical Example

Consider a unit variance white WSS process $w[\cdot]$ which we will put through the shaping filter

$$H(z) = 1 + \frac{1}{4}z^{-1}$$

(17)

to get $x[\cdot]$. We use $z = e^{j\Omega}$ as in the usual $z$-transform. Now suppose $x[\cdot]$ is filtered by

$$G(z) = \frac{1}{1 - \frac{1}{2}z^{-1}} + \frac{z}{1 + \frac{1}{2}z}$$

(18)

to get $y[\cdot]$. Note that $G(z)$ is the sum of a causal, stable lowpass filter and an anticausal, stable highpass filter. Figure 1 summarizes these relationships.

Figure 1

Now consider the causal coherence from $x[\cdot]$ to $y[\cdot]$. We expect that there will be high causal coherence at low frequencies and lower causal coherence at high frequencies. Analytically, the causal Wiener filter for minimum MSE estimation of $y[\cdot]$ from $x[\cdot]$ ends up being

$$H_c(\Omega) = \frac{1}{1 - \frac{1}{2}e^{-j\Omega}} + \frac{\frac{1}{4}}{1 + \frac{1}{4}e^{-j\Omega}} .$$

(19)

Figure 2 shows $|H_c(\Omega)|$ and its approximations by $|H_N(\Omega)|$ for increasing $N$. In Figure 2, the approximations $H_N(\Omega)$ were computed directly by solving the autocorrelation equations (8) with the analytically determined values for $R_{xx}[\cdot]$ and $R_{yx}[\cdot]$. Figure 3 shows the associated $K_{xy}(\Omega)$ that were computed using (14) and the estimates of $H_c(\Omega)$ from Figure 2. At low frequencies
the causal coherence is close to 1 which was expected – at those frequencies $x[\cdot]$ and $y[\cdot]$ can be related well by a causal filter. At frequencies where the noncausal filter is dominant, however, there is negative coherence.

In order to get a sense for how closely (14) can be estimated from measured realizations of $x[\cdot]$ and $y[\cdot]$, we generated a finite length realization of $w[\cdot]$. In this realization, each sample $w[n]$ was drawn independently from $\pm 1$ with equal probabilities. This is clearly a unit-variance white process. The corresponding realizations of $x[\cdot]$ and $y[\cdot]$ were then generated by filtering $w[\cdot]$ with $H(z)$ and $G(z)$ as in Figure 1. The resulting causal coherence spectrum can be seen in Figure 4, and compares very well with the analytical result,
Classical coherence is often used as a way of understanding if the components of two signals that occupy specific frequency bands can be related well by an LTI system. For example, suppose two signals \(x[\cdot]\) and \(y[\cdot]\) have high
coherence in a given frequency band. Then if those signals are bandpass filtered to isolate that frequency band, the resulting signals can be related well by the appropriate unconstrained Wiener filter. This relationship is depicted in Figure 5. In the diagram in Figure 5, if $H(z)$ is the unconstrained Wiener filter for estimating $y[\cdot]$ from $x[\cdot]$, then the estimate $\hat{y}_{bp}[\cdot]$ will closely resemble $y_{bp}[\cdot]$ and there will be low MSE between the two (relative to the variance of $y[n]$). In this situation, it is sometimes argued that the LTI system $H(z)$ can be used to model the relationship between the respective components of $x[\cdot]$ and $y[\cdot]$. One problem with this interpretation is that $H(z)$ is not guaranteed to be causal; and if it is not causal, then it may not make sense as a model for a physical system. This is one situation in which our definition of causal

Figure 4: The estimated causal coherence computed from the simulated realizations of $x[\cdot]$ and $y[\cdot]$ along with the true, analytically determined $K_{\Omega}^{xy}(\Omega)$.
coherence may lend additional insight, as discussed next.

Suppose that there is high causal coherence from $x[\cdot]$ to $y[\cdot]$ in a given frequency band, and that the filter $H(z)$ in Figure 5 is the causal Wiener filter for estimating $y[\cdot]$ from $x[\cdot]$. In this case we can still be sure that there is low MSE between $\hat{y}_{bp}[\cdot]$ and $\hat{y}_{bp}[\cdot]$, and now we can also be sure that $H(z)$ is a causal filter. There is therefore no conflict due to causality in interpreting the LTI system $H(z)$ as a model for the dynamics of that signal component.

Unfortunately, a strict converse of this interpretation for high causal coherence is not true. If $x[\cdot]$ and $y[\cdot]$ have high classical coherence and low causal coherence (in either direction), then it does not mean that there is no causal filter that relates the two signal components. In fact, if the frequency band does not have full measure, then such a filter must exist due to results regarding the approximation of elements of $L^p$ with elements of $H^p$ on subsets of the unit circle (e.g., [10]).

Take, for instance, the example from Section 2.2.1. Given a plot of $K_{xy}$ such as in Figure 4, we could infer that the causal Wiener filter (19) can causally estimate the low frequency component of $y[\cdot]$ with approximately zero MSE. At low frequencies, the causal Wiener filter (19) therefore makes more sense as a physical model than the unconstrained Wiener filter which is simply (18).

2.4. Remarks

The version of causal coherence (14) has some features worth noting. First, it shows an interesting generalization of the classical coherence value which could be useful in a modeling setting. Second, it allows us to see one of the limitations of a frequency dependent metric for causality, namely
that no frequency dependent metric for causality can represent a pointwise limit of how well a causal LTI filter can relate two time series. Finally, it gives us a simple framework in which to analyze the effects of a finite order approximation on a causality spectrum. Similar issues exist in the finite order MVAR models that are used in computing Geweke’s causality spectra [2], and techniques from this thesis could also be of use in that setting.

3. Convergence Theorems

3.1. Definitions

Before getting to some theorems regarding the convergence of the FIR Wiener filters, there are a few definitions to review. Let $\mathbb{T} \simeq \mathbb{R}/2\pi\mathbb{Z}$ be the unit circle, which we identify with the interval $[-\pi, \pi)$. We denote by $\|f\|_p$ the usual $L^p(\mathbb{T})$ norm of $f : \mathbb{T} \to \mathbb{C}$. The function $f$ is said to be (uniformly) $\alpha$-Hölder continuous if

$$\forall \Omega, \Omega' \in \mathbb{T}, \ |f(\Omega) - f(\Omega')| < C|\Omega - \Omega'|^\alpha$$
where $0 < \alpha \leq 1$ and $C$ is some constant that is independent of $\Omega, \Omega'$. We define $C_\alpha(\mathbb{T})$ to be the space of all such Hölder continuous $f$ on $\mathbb{T}$, and we denote by $C^k_\alpha(\mathbb{T})$ for $k \geq 0$ the space of all $k$-times differentiable functions on $\mathbb{T}$ with $f^{(k)} \in C_\alpha(\mathbb{T})$. Define $H^\infty(\mathbb{T})$ to be the Hardy space (on the unit circle) equipped with the supremum norm and let $\mathcal{A}^k_\alpha(\mathbb{T}) = H^\infty(\mathbb{T}) \cap C^k_\alpha(\mathbb{T})$. In other words, $H^\infty(\mathbb{T})$ is the space of bounded frequency responses of causal LTI filters, and $\mathcal{A}^k_\alpha(\mathbb{T})$ is the space of such frequency responses that have $\alpha$-Hölder continuous $k$th derivative.

We will make use of the smoothness classes $C^k_\alpha(\mathbb{T})$ because we are concerned with the smoothness of the power spectra $S_{xx}(\Omega)$ and $S_{xy}(\Omega)$. This, in turn, will affect the smoothness of the causal Wiener filter $H_c(\Omega)$. It turns out that we do not need similar restrictions on the power spectrum $S_{yy}(\Omega)$. Throughout we will assume that $S_{xx}, S_{xy} \in C_\alpha(\mathbb{T})$, and this guarantees the relations (3) and (4) by the Wiener-Khinchin Theorem ([11] Section 1.18).

We now turn to the issue of how $H_N(\Omega)$ converges to $H_c(\Omega)$ as $N \to \infty$. Consider the following reduction of this main convergence problem. If $\widehat{y}_c[\cdot] = x \ast h_c[\cdot]$ is the estimate of $y[\cdot]$ from $x[\cdot]$ using the IIR causal Wiener filter, then the FIR causal Wiener filter of order $N$ for estimating $\widehat{y}_c[\cdot]$ from $x[\cdot]$ is the same as the order-$N$ FIR filter for estimating $y[\cdot]$ from $x[\cdot]$. This follows directly from the orthogonality principle in that $R_{yx}[m] = R_{\widehat{y}_c,x}[m]$ for $m \geq 0$. Hence there is no difference between $R_{yx}[m]$ or $R_{\widehat{y}_c,x}[m]$ appearing in the autocorrelation normal equations (8). From now on we will view $H_N(\Omega)$ and $H_c(\Omega)$ as the filters for estimating $\widehat{y}_c[\cdot]$ rather than $y[\cdot]$. We proceed by first reviewing results from [4] in Lemmas 1 and 2 and then showing $L^1$ and $L^2$ convergence in Theorem 1.
Lemma 1. If $S_{xx}, S_{xy} \in C_{\alpha}^{k}(\mathbb{T})$ for $0 < \alpha < 1$, and $S_{xx}(\Omega) \geq \delta > 0$, then $H_{c} \in \mathcal{A}_{\alpha}^{k}(\mathbb{T})$.

Proof. The $k = 0$ case is treated explicitly in [4]. As mentioned there, the $k > 0$ case is similar. The main difference is that we need to check that the Hilbert transform of an $f \in C_{\alpha}^{k}(\mathbb{T})$ is also in $C_{\alpha}^{k}(\mathbb{T})$. This follows from Theorem 13.27 in [5] Chapter 3. \qed

Lemma 2. (Jackson) If $H_{c} \in \mathcal{A}_{\alpha}^{k}(\mathbb{T})$, then there exists a sequence $G_{N}(\Omega)$ of causal, length $N$ filters such that $\|G_{N} - H_{c}\|_{\infty} = O(N^{-(k+\alpha)})$.


The convergence rate of $G_{N}(\Omega)$ in Lemma 2 is achieved by the Fejér and Vallée-Poussin partial sums of $H_{c}(\Omega)$. Ideally, such partial sums would suffice for uniform FIR approximations of $H_{c}(\Omega)$, but computing them requires knowing the underlying IIR $H_{c}(\Omega)$. We will show that the FIR Wiener filters $H_{N}(\Omega)$ can be used instead, albeit with slower convergence.

3.2. $L^{1}$ and $L^{2}$ Convergence

Theorem 1. If $S_{xx}, S_{xy} \in C_{\alpha}^{k}(\mathbb{T})$ for $0 < \alpha < 1$, and $S_{xx}(\Omega) \geq \delta > 0$, then $H_{N} \to H_{c}$ in $L^{p}$ norm for $p = 1, 2$. The convergence rate is bounded by $\|H_{N} - H_{c}\|_{p} = O(N^{-(k+\alpha)})$.

Proof. Lemmas 1 and 2 ensure that $H_{c} \in \mathcal{A}_{\alpha}^{k}(\mathbb{T})$ and that there exists a sequence $G_{N}(\Omega)$ of causal, length $N$ filters such that

$$\sup_{\Omega}|G_{N}(\Omega) - H_{c}(\Omega)| \leq CN^{-(k+\alpha)}$$  \hspace{1cm} (20)
where the constant $C$ is independent of $N$. The core of the argument is that since $H_N(\Omega)$ is the causal, length $N$ filter that minimizes the MSE of estimation, we have

\[ \text{MSE}(H_N) \leq \text{MSE}(G_N). \tag{21} \]

Now note that because of the reduction mentioned at the beginning of Section 3,

\[
\text{MSE}(H_N) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(\Omega)d\Omega \\
= \frac{1}{2\pi} \int_{-\pi}^{\pi} (S_{y_c}\hat{y}_{c} - H_N S_{xx} - H_N^* S_{y_c} + H_N H_N^* S_{xx})d\Omega \\
= \frac{1}{2\pi} \int_{-\pi}^{\pi} (H_c H_c^* S_{xx} - H_N H_c^* S_{xx} - H_c H_N^* S_{xx} + H_N H_N^* S_{xx})d\Omega \\
= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx} \| H_c - H_N \|_2^2 d\Omega \tag{22}
\]

and similarly

\[
\text{MSE}(G_N) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx} \| H_c - G_N \|_2^2 d\Omega, \tag{23}
\]

which leads to

\[
\text{MSE}(G_N) \leq (CN^{-(k+\alpha)})^2 \max_\Omega S_{xx}(\Omega). \tag{24}
\]

Putting (21), (22), and (24) together gives

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx} \| H_c - H_N \|_2^2 d\Omega \leq (CN^{-(k+\alpha)})^2 \max_\Omega S_{xx}(\Omega), \tag{25}
\]

22
and finally
\[
\frac{\delta}{2\pi} \int_{-\pi}^{\pi} \|H_c - H_N\|_2^2 d\Omega \leq (CN^{-(k+\alpha)})^2 \max\Omega S_{xx}(\Omega).
\] (26)

Therefore
\[
\|H_c - H_N\|_2 \leq C' N^{-(k+\alpha)}
\] (27)

for some constant \(C'\) which is independent of \(N\). Because \(\mathbb{R}/2\pi\mathbb{Z}\) is a finite measure space, Hölder’s inequality gives \(L^1\) convergence with the same asymptotic rate.

\[
3.3. \text{Uniform Convergence}
\]

Now we consider the issue of uniform convergence. In Theorem 2 we demonstrate the uniform convergence of \(H_N(\Omega)\) to \(H_c(\Omega)\) if the smoothness of the power spectra is \(k \geq 1\).

**Theorem 2.** If \(S_{xx}, S_{xy} \in C^k_0(\mathbb{T})\) for \(0 < \alpha < 1\) and \(k \geq 1\), and \(S_{xx}(\Omega) \geq \delta > 0\), then \(H_N \to H_c\) in \(L^\infty\) norm. The convergence rate is bounded by \(\|H_N - H_c\|_\infty = O(N^{-r})\) where

\[
r = \begin{cases} 
\frac{\alpha+k}{4}, & \text{if } \alpha + k > \frac{4}{3} \\
\alpha, & \text{if } \alpha + k \leq \frac{4}{3}.
\end{cases}
\] (28)

**Proof.** Recall that the Fejér partial sums of \(H_c(\Omega)\) are

\[
P_N(\Omega) = \sum_{n=0}^{N-1} h_c[n] e^{-j\Omega n} (1 - n/N)
\] (29)
which is equivalent to multiplying \( h_c[\cdot] \) by a symmetric triangle of support \( 2N - 1 \) centered at the origin in the time domain. In other words,

\[
P_N(\Omega) = H_c \ast F_N(\Omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_N(\Omega')H_c(\Omega' - \Omega)d\Omega'
\]

(30)

where \( F_N \) is the Fejér kernel of length \( 2N - 1 \) (and \( \ast \) here denotes periodic convolution). In closed form,

\[
F_N(\Omega) = \frac{1}{N} \frac{\sin^2(N\Omega/2)}{\sin^2(\Omega/2)}.
\]

(31)

The Vallée-Poussin kernel is then defined as

\[
V_N(\Omega) = 2F_{2N}(\Omega) - F_N(\Omega).
\]

(32)

Convolving by the Vallée-Poussin kernel instead of the Fejér kernel represents multiplying by a trapezoid as seen in Figure 6 in the time domain. The Fejér kernel has the following two properties that we will use.

(i)

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} F_N(\Omega)d\Omega = 1.
\]

(33)

(ii) For \( K = 1, 2, \ldots, \)

\[
\sup_{1/K \leq |\Omega| \leq \pi} |F_N(\Omega)| = O\left( \frac{K^2}{N} \right).
\]

(34)

Both of these properties can be verified with (31) and carry over to \( V_N(\Omega) \) as well.
Figure 6: The time domain representation of $V_N(\Omega)$ for $N = 5$. 
For any \( M > N \), multiplying in the time domain by the trapezoid associated with \( V_M(\Omega) \) does not change the FIR filter \( H_N(\Omega) \), so

\[
H_N(\Omega) = H_N * V_M(\Omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_N(\Omega - \Omega') V_M(\Omega') d\Omega'.
\]

(35)

Then because of (33),

\[
H_N(\Omega) - H_c(\Omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (H_N(\Omega - \Omega') - H_c(\Omega)) V_M(\Omega') d\Omega'
\]

\[\implies |H_N(\Omega) - H_c(\Omega)| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_N(\Omega - \Omega') - H_c(\Omega)| |V_M(\Omega')| d\Omega'
\]

\[= \frac{1}{2\pi} \int_{\epsilon < |\Omega'| < \pi} |H_N(\Omega - \Omega') - H_c(\Omega)| |V_M(\Omega')| d\Omega' + \frac{1}{2\pi} \int_{|\Omega'| \leq \epsilon} |H_N(\Omega - \Omega') - H_c(\Omega)| |V_M(\Omega')| d\Omega'
\]

(36)

\[\leq \frac{1}{2\pi} \int_{\epsilon < |\Omega'| < \pi} |H_N(\Omega - \Omega') - H_c(\Omega)| |V_M(\Omega')| d\Omega'
\]

(37)

\[+ \frac{1}{2\pi} \int_{|\Omega'| \leq \epsilon} |H_c(\Omega - \Omega') - H_c(\Omega)| |V_M(\Omega')| d\Omega'
\]

(38)

The claim is that (36), (37), and (38) can be made arbitrarily small as \( \epsilon \) goes to zero and \( N, M \) get large. In order to understand the rates at which which these terms go to zero we will set \( M = \lfloor N^\beta \rfloor \) for some \( \beta \geq 1 \) and \( \epsilon = N^{-\gamma} \) for some \( \gamma > 0 \).
The expression (36) is bounded above by

\[ \frac{1}{2\pi} \sup_{\epsilon < |\Omega'| < \pi} \{ |V_M(\Omega')| \} \int_{-\pi}^{\pi} |H_N(\Omega - \Omega') - H_c(\Omega)|d\Omega' \]  

(39)

which is itself

\[ O\left( \sup_{\epsilon < |\Omega'| < \pi} |V_M(\Omega')| \right). \]  

(40)

This is because \( H_c(\Omega) \) and \( \|H_N(\Omega)\|_1 \) can be bounded independently of \( N \), the former because it is continuous on a compact set and the latter by Theorem 1 and the triangle inequality. Since we have chosen \( \epsilon = N^{-\gamma} \), the property (34) gives that (36) is \( O(\epsilon^{2\gamma-\beta}) \).

We can bound the next term (37) because \( H_c(\Omega) \) is continuously differentiable and thus \( |H_c(\Omega - \Omega') - H_c(\Omega)| \leq C\epsilon \) for \( |\Omega'| \leq \epsilon \) and a new constant \( C \). Therefore (37) is \( O(\epsilon^{-\gamma}) \).

For the last term (38) we use the \( L^1 \) convergence from Theorem 1. The expression (38) is less than or equal to

\[ \frac{M}{2\pi} \int_{-\pi}^{\pi} |H_N(\Omega - \Omega') - H_c(\Omega - \Omega')|d\Omega' \]  

(41)

which by Theorem 1 is \( O(N^{\beta-\alpha-k}) \).

Due to the fact that these estimates are true for all \( \Omega \), we have that

\[ \|H_c - H_N\|_\infty = O(N^{2\gamma-\beta}) + O(N^{-\gamma}) + O(N^{\beta-\alpha-k}). \]  

(42)

To optimize the bound on convergence rate that this estimate yields we can set

\[ \gamma = \beta - 2\gamma = \alpha + k - \beta \]  

(43)
which gives

\[ \gamma = \frac{\alpha + k}{4} \]  \hspace{1cm} (44)

and

\[ \beta = \frac{3(\alpha + k)}{4} . \]  \hspace{1cm} (45)

Under our assumption that \( \beta \geq 1 \), this optimal choice of \( \gamma, \beta \) can only be chosen when \( \alpha + k \geq \frac{4}{3} \). If \( \alpha + k < \frac{4}{3} \), the best choice is \( \beta = 1 \) and \( \gamma = \frac{1}{3} \).

We arrive at

\[ \|H_N - H_c\|_{\infty} = O(N^{-r}) \]  \hspace{1cm} (46)

where

\[ r = \begin{cases} \frac{\alpha + k}{4} & \text{if } \alpha + k \geq \frac{4}{3}, \\ \alpha & \text{if } \alpha + k < \frac{4}{3}. \end{cases} \]  \hspace{1cm} (47)

In Theorem 3 we conclude by showing that in the case \( k = 0 \), uniform approximations of \( H_c(\Omega) \) can still be constructed by windowing \( H_N(\Omega) \) for sufficiently large \( N \). In the case that the power spectra are continuous but not Hölder continuous, then there is no guarantee that uniform approximations of \( H_c(\Omega) \) exist. See [4] for an example where \( S_{xx}(\Omega) \) and \( S_{xy}(\Omega) \) are continuous, but \( H_c(\Omega) \) is not. If \( H_c(\Omega) \) is not continuous then it can not be uniformly approximated by FIR filters.

**Theorem 3.** Suppose \( V_M(\Omega) \) is the Vallée-Poussin kernel as in (32) for \( M = \lfloor N^\beta \rfloor \) and \( 0 < \beta < \alpha \). If \( S_{xx}, S_{xy} \in C^k_{\alpha}(\mathbb{T}) \) for \( 0 < \alpha < 1 \) and \( k = 0 \), and \( S_{xx}(\Omega) \geq \delta > 0 \), then \( V_M * H_N \to H_c \) in \( L^\infty \) norm.

**Proof.** The main difference here is that \( \beta < 1 \) so the Vallée-Poussin kernel
$V_M(\Omega)$ is not long enough for $V_M*H_N(\Omega)$ to equal $H_N(\Omega)$. The proof proceeds exactly as in Theorem 2 to get an estimate similar to (42). Then if $\gamma$ is chosen so that $\beta > 2\gamma$, all terms on the right hand side will go to zero as $N$ goes to infinity. □

4. Conclusion

We have given a more general interpretation of the notion of coherence by casting it in terms of the power spectral density of the error for any arbitrary LTI estimation filter. The classical coherence, then, becomes the frequency-wise maximum coherence achieved across all possible LTI estimation filters. In the case of the causal Wiener filter and the causal coherence (14), this could be used in a modeling setting to determine which frequency components of two signals can be related by the causal Wiener filter.

The convergence results proved in Section 3 are valid for a broad class of power spectra. The class of Hölder continuous power spectra that was considered is much larger than, for example, the class of rational power spectra. These are therefore very general theoretical results regarding the construction of FIR approximations for minimum MSE estimation filters. In the particular application of estimating causal coherence spectra between WSS processes, these results demonstrate that a sufficiently long FIR causal Wiener filter can be used in place of the true IIR causal Wiener filter.
Appendix A  Spectral Factorization

The problem of spectral factorization, also known as Wiener-Hopf factorization, is the problem of factorizing a power spectral density

\[ S_{xx} : T \to \mathbb{R} \quad (48) \]

as

\[ S_{xx}(\Omega) = F(e^{j\Omega})F(e^{-j\Omega}) \quad (49) \]

where \( F(z) \) is analytic for \(|z| > 1 \) and has no zeros in the region \(|z| > 1 \). In a more applied setting, particularly when using only rational power spectra, these conditions on \( F(z) \) are sometimes phrased as saying \( F(z) \) and its inverse \( 1/F(z) \) must be causal, stable systems. Any \( S_{xx}(\Omega) \) which is integrable and satisfies the Paley-Wiener condition

\[ \int_{-\pi}^{\pi} \log S_{xx}(\Omega) d\Omega > -\infty \quad (50) \]

can be factorized in this way and has the spectral factor [12]

\[ F(z) = \exp \left( \frac{1}{4\pi} \int_{-\pi}^{\pi} \log S_{xx}(e^{j\Omega}) \frac{e^{j\Omega} + z^{-1}}{e^{j\Omega} - z^{-1}} d\Omega \right) . \quad (51) \]

Note that throughout this thesis we have assumed \( S_{xx}(\Omega) \) is continuous, and thus integrable; and that \( S_{xx}(\Omega) \geq \delta > 0 \), which means (50) is satisfied.

The expression (51) can be rewritten as

\[ F(z) = \exp \left( \frac{1}{2} \left( \frac{1}{2} \psi_0 + \sum_{k=1}^{\infty} \psi_k z^{-k} \right) \right) \quad (52) \]
where the $\psi_k$ are the Fourier coefficients of $\log S_{xx}(\Omega)$. This forms the basis for a common practical algorithm for computing spectral factors $F(z)$ which is outlined in [12]. A finite number of Fourier coefficients of $\log S_{xx}(\Omega)$ are computed and then the frequency response (52) can be computed at a finite number of values $z = e^{j\Omega}$ (and then, if needed, a time-domain representation can be recovered). This Fourier analytic approach is based on the FFT algorithm and can be run in $O(n \log n)$ time with the number of Fourier coefficients $n$.

An alternative, iterative method for spectral factorization was developed by Wilson in [13]. A streamlined version of this method due to Burg is outlined in [14]. This Wilson-Burg method starts with a finite-length auto-correlation $R_{xx}[m]$ with associated PSD $S_{xx}(z)$, along with an initial guess of the minimum phase factor $F_0(z)$. Then, similarly to Newton’s iterative method for finding square roots, the estimate is updated by

$$S_{xx}(z) - F_t(z)F_t(z^{-1}) = F_t(z)(F_{t+1}(z^{-1}) - F_t(z^{-1})) + F_t(z^{-1})(F_{t+1}(z) - F_t(z)).$$

Dividing through by $F_t(z)F_t(z^{-1})$ leads to

$$\frac{S_{xx}(z)}{F_t(z)F_t(z^{-1})} + 1 = \frac{F_{t+1}(z)}{F_t(z)} + \frac{F_{t+1}(z^{-1})}{F_t(z^{-1})},$$

which directly gives the method for computing $F_{t+1}(z)$ from $F_t(z)$. Ignoring rounding errors, this method is guaranteed to converge to the correct minimum phase $F(z)$ [13]. The cost of factorization with Wilson-Burg is proportional to the number of autocorrelation coefficients times the length
of the minimum phase filter recovered times the number of iterations [14].

In both of the methods for spectral factorization mentioned above, finite order approximations are needed for application to general power spectra $S_{xx}(\Omega)$. In the Fourier analytic method only finitely many Fourier coefficients are used, whereas the Wilson-Burg method can only be applied to PSDs from finite-length autocorrelation functions. The question of how these finite order approximations affect the resulting minimum phase factor $F(\Omega)$ was addressed partially in [12]. It turns out that the spectral factorization mapping $S_{xx} \mapsto F$ is not a continuous mapping on the space of continuous power spectra equipped with the supremum norm. In other words, even if the power spectrum $S_{xx}(\Omega)$ is continuous, small perturbations in $S_{xx}(\Omega)$ (such as from a finite-order approximation) can cause large perturbations in the resulting $F(\Omega)$. This was part of the motivation behind investigating the FIR Wiener filter instead of attempting to approximate the IIR causal Wiener filter directly. A possible further research question would be to investigate the continuity of the spectral factorization mapping under the Hölder continuity conditions we required for the convergence of the FIR Wiener filters.
Appendix B  Proof of Lemma 1

Lemma 1 states that if $S_{xx}, S_{xy} \in C^k_\alpha(T)$ and $S_{xx} \geq \delta > 0$, then the causal Wiener filter $H_c \in A^k_\alpha(T)$. Here is a sketch of the proof. First recall that

$$H_c(\Omega) = \frac{1}{F(\Omega)} \left[ \frac{S_{yx}(\Omega)}{F(-\Omega)} \right]_+$$

where

$$S_{xx}(\Omega) = F(\Omega)F(-\Omega) = F(\Omega)F^*(\Omega)$$

is a minimum phase spectral factorization. The claim (which will be addressed shortly) is that both the operation of minimum phase spectral factorization and the operation of taking the causal part preserve which smoothness class $C^k_\alpha(T)$ a function is in. This means that $F \in C^k_\alpha(T)$. Furthermore, since $|F(\Omega)| \geq \sqrt{\delta}$, we also have $\frac{1}{F} \in C^k_\alpha(T)$. Because $C^k_\alpha(T)$ forms a Banach algebra with pointwise multiplication and the appropriate Hoelder norm, the product of two elements of $C^k_\alpha(T)$ is also in $C^k_\alpha(T)$. Hence $H_c \in C^k_\alpha(T)$. By construction $H_c(\Omega)$ is causal so $H_c \in A^k_\alpha(T)$.

Now we return to address the claim that $F \in C^k_\alpha(T)$. By construction, $F(\Omega)$ is equal almost everywhere to

$$\exp \left( \frac{1}{2} (\log S_{xx}(\Omega) - j\mathcal{H}(\log S_{xx})(\Omega)) \right)$$

where $\mathcal{H}$ is the Hilbert transform. It can be shown that under the condition that $S_{xx}(\Omega)$ is Hoelder continuous, this equality is valid on all of $T$. Because $S_{xx}(\Omega)$ is bounded away from zero, $\log S_{xx} \in C^k_\alpha(T)$. By Theorem 13.27 in Chapter 3 of [5], the Hilbert transform of an element of $C^k_\alpha(T)$ is also in $C^k_\alpha(T)$.
$C^k_\alpha(\mathbb{T})$. Finally, since everything is continuous and thus bounded, taking the exponential also results in an element of $C^k_\alpha(\mathbb{T})$.

In a similar way taking the causal time part of an $f \in C^k_\alpha(\mathbb{T})$ can be defined by

$$[f(\Omega)]_+ = \frac{1}{2} (f(\Omega) - j\mathcal{H}(f)(\Omega)) + \frac{1}{2} \int_{-\pi}^{\pi} \frac{f(\Omega)}{2\pi} d\Omega .$$

(57)

All operations involved preserve the smoothness class and we get $[f]_+ \in C^k_\alpha(\mathbb{T})$. 

References


