

©Copyright 2012
Charles Pascal Clark

Coherent Demodulation of Nonstationary Random Processes

Charles Pascal Clark

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Les Atlas, Chair

Ivars Kirsteins

Li Deng

Program Authorized to Offer Degree:

Electrical Engineering

University of Washington

Abstract

Coherent Demodulation of Nonstationary Random Processes

Charles Pascal Clark

Chair of the Supervisory Committee:
Professor Les Atlas
Electrical Engineering

Nonstationary processes have local properties which change over time. An example is speech, which can be represented as pitch harmonics multiplied by slower-varying syllabic modulations. Commonly-used power spectral analysis reveals the relative intensity of harmonics, but not their relative phase alignment. From an estimation standpoint, speech harmonics are random and thus not perfectly periodic. Syllabic modulations, at a longer time scale, are also ordered yet aperiodic. Instead, their relative alignment can be described as “rhythmic.” This thesis shows how a type of rhythm manifests as conjugate correlations in the frequency domain, a phenomenon called impropriety. A useful aspect of impropriety is its algebraic structure, which provides a new maximum-likelihood framework for synchronous estimation. Interestingly, impropriety can also appear in non-harmonic processes, such as underwater propeller noise. This suggests possible generalizations of rhythm and synchrony to a broader class of signals beyond speech. For this class, frequency-domain impropriety can be modeled in terms of linear systems with quickly time-varying subcomponents. Rather than impede analysis, these time variations provide a synchronization cue for coherent demodulation analysis of signals which are highly nonstationary.

TABLE OF CONTENTS

	Page
List of Figures	iv
Chapter 1: Introduction	1
1.1 Modulation Analysis of Natural Signals	1
1.2 Major Concepts	5
1.3 Organization	7
Chapter 2: Background	8
2.1 Introduction	8
2.2 Modulation Frequency Analysis in Speech	9
2.3 Stochastic Demodulation and Linear System Theory	18
2.4 Complex Random Processes and Statistics	31
2.5 Summary and Problem Statement	36
Chapter 3: Spectral Impropriety of Modulated Random Processes	38
3.1 Introduction	38
3.2 Definition of Spectral Impropriety	39
3.3 Modulation Frequency and Spectral Impropriety	41
3.4 The “Fourier Transform” of a Random Process	47
3.5 Linear Time-Varying System Model of Spectral Impropriety	60
3.6 Conclusion	67
Chapter 4: Coherent Subband Demodulation	69
4.1 Introduction	69
4.2 Coherent Modulation Signal Model	70
4.3 The Time-Varying “Fourier Transform” of a Random Process	73
4.4 Derivation of the Coherent System Equations	83
4.5 Complementary Square-Law Estimation	90
4.6 Comment on Underspread System Estimation	95

4.7	Conclusion	97
Chapter 5:	The Principal Components of Impropropriety	99
5.1	Introduction	99
5.2	Taxonomy of Improper and Proper Random Processes	100
5.3	Augmented Principal Components Analysis	112
5.4	Conclusion	123
Chapter 6:	Real-World Examples	125
6.1	Introduction	125
6.2	A Motivating Example	126
6.3	Hypothesis Testing for Subband Impropropriety	129
6.4	Impropropriety Detection in Speech	131
6.5	DEMONgram Analysis of Propeller Cavitation Noise	136
6.6	Coherent Enhancement of Propeller Cavitation Noise	145
6.7	Conclusion	155
Chapter 7:	Conclusion	156
7.1	Main Results	156
7.2	Future Work and Open Questions	157
	Bibliography	159
Appendix A:	Spectral Representation of Harmonizable Random Processes	173
A.1	Introduction	173
A.2	Harmonizable Processes	174
A.3	Time- and Frequency-Concentrated Processes	177
Appendix B:	Correlative Properties of Cyclic Demodulation	179
Appendix C:	Hypothesis Testing for Impropropriety	183
C.1	Scalar Noncircularity Coefficient	183
C.2	Multivariate Generalized Likelihood Ratio	184
C.3	A Special Case: Diagonal Covariance	185
C.4	Effect of SNR	187

Appendix D: Derivation of the Improper Likelihood Function	189
D.1 Derivation of the Complex Gaussian Log-Likelihood	189
D.2 Gradient of the Log-Likelihood	191
Appendix E: Augmented PCA Demodulation	193
E.1 Short-Time Augmented PCA	193
E.2 Time-Varying Covariance Estimation	194
E.3 Sum-of-Products Form	196
E.4 Generalized Coherent Demodulation	198
E.5 Summary	199

LIST OF FIGURES

Figure Number	Page
1.1 Annotated display of one subband envelope from a woman saying, “bird populations.” Blue indicates the subband signal, and black is the non-coherent envelope. The inset image is a zoomed-in portion showing the oscillations of the carrier.	2
1.2 One subband from a woman saying, “thing about.” Blue indicates the subband signal and black is the non-coherent envelope. Red and magenta are the real and imaginary parts of the complex modulator.	3
1.3 A synthetic example of a broadband coherent random process, with regular short-term structure which implies long-term variation.	4
2.1 Schematic for automatic synthesis of speech, reproduced from [1].	10
2.2 Subband time series taken from a single speech utterance (“bird populations”) overlaid with their upper envelopes in red. From top to bottom, the subband frequency boundaries are 625-875 Hz, 375-625 Hz, and 125-375 Hz.	11
2.3 Spectrogram of the same speech utterance as in Figure 2.2. The vertical axis indicates subband center frequency, and the colormap is a decibel scale.	11
2.4 Realizations from example AM-WSS processes, with the modulator overlaid in red. Left: WSS $x(t)$ with constant modulator. Right: periodically correlated $x(t)$ with sinusoidal modulator.	19
2.5 Comparison between power spectral estimates (top) and DEMON spectra (bottom), plotted on logarithmic frequency axes. The left panels are from merchant vessel propeller data, with a modulation rate of 2.8 Hz, while the right panels are from female speech, with syllabic rate around 3-4 Hz.	20
2.6 Multiband DEMON spectrum (left plot) for merchant propeller noise, showing periodic modulation with fundamental frequency of 2.8 Hz. Modulation is not uniform across acoustic (subband) frequency, as shown in the magnitude and phase plots for modulation frequency = 2.8 Hz.	23
2.7 Support regions of measureable spreading-functions, in modulation frequency ν and lag τ . In order of increasing generality from left to right: (a) Kailath’s bandwidth-lag product [2], (b) rotated-rectangular regions considered by Matz <i>et al.</i> [3, 4], and (c) Bello’s general underspread class [5]. In all cases, the area of the support region must be less than unity.	28

2.8	Time-frequency (spectrogram, left) and bi-frequency (Loève spectrum, right) displays for a synthetic, periodically correlated AM-WSS process. The colormap dynamic range is 20 dB for both plots.	30
2.9	Example scattergrams of scalar complex Gaussian random variables. Left: proper (second-order circular) with $\rho_{zz}^2 = 0$. Middle: improper (second-order noncircular) with real and positive ρ_{zz}^2 . Right: improper (second-order noncircular) with complex ρ_{zz}^2	33
3.1	Schematic for cyclic, or coherent, demodulation. The basebanded STFT is equivalent to quadrature branches with cosine and sine demodulators followed by lowpass filtering. After cyclic sampling, a subband centered on a harmonic of $\omega_0/2$ shows impropriety in the form of an elliptical histogram.	45
3.2	Scattergrams for subbands centered on $\omega_0/2$, ω_0 , and $3\omega_0/2$ for a modulator consisting of two cosines. Each black vector indicates the measured orientation of its respective distribution.	46
3.3	Magnitude (left) and phase (right) of the estimated complementary spectrum corresponding to a triangle-wave modulator. Gray curves indicate the known, theoretical value and circles indicate measured values. Dark circles indicate modulator frequencies at multiples of ω_0 and empty circles indicate in-between frequencies.	47
3.4	Generative signal model for a coherently modulated random process $y(t)$, as the output of an LTV system driven by PC white noise.	49
3.5	Synthesis and analysis chain for the observed process $y(t)$. The red vector corresponds to the reference angle determined by $m^2(t)$. The complex rotation to the measured, black vector is determined by the phase response of $H(\omega)$	50
3.6	Scatter plots for subbands centered on $\omega_0/2$, ω_0 , and $3\omega_0/2$ for Example 1. Each black vector is a measured orientation, and the red vectors are known a priori based on the defined modulator.	51
3.7	Complementary spectrum (top row) and LTI spectrum (bottom) for the speech-like signal in Example 2. Circles denote values measured from subbands. Dark and empty circles correspond to modulator frequencies and in-between frequencies. Curves are ground truth based on the signal synthesis. The light gray curve and circles show the magnitude of the noncircularity coefficient.	52
3.8	Same layout and key as in Figure 3.7, except for the modulator defined in Example 3. This time, the complementary spectrum (top row) must be divided by the squared-modulator spectrum to obtain the LTI estimate (bottom row).	53

3.9	Time-domain view of the speech-like Examples 2 (top) and 3 (bottom) in the text. The left-hand subplots show example segments of the PC process $x(t)$. The right-hand subplots show the corresponding segments of $y(t)$, with separate pulse responses shown as separate colors.	54
3.10	Schematic for Bedrosian over-modulation, where the dark shaded and pattern-shaded power spectra correspond to $w_L(t)$ and $w_H(t)$, respectively (top). The bottom plot shows the components of the analytic signal after the frequency shift.	62
3.11	Schematic showing A and B quadrants of two bifrequency LTV system functions: time-invariant (left) and modulated at 1000 Hz (right).	65
3.12	An example time-invariant linear system, in three representations: linear time operator (left), widely-linear frequency operator (middle), and modulation spectrum (right).	66
3.13	An example modulated linear system, in three representations: linear time operator (left), widely-linear frequency operator (middle), and modulation spectrum (right).	67
4.1	Revised signal model for a coherently modulated random process (see also Figure 3.4). The observed process, $y(t)$, is the output of an LTV system driven by white PC noise.	71
4.2	Example signals for a time-invariant system (top) and a time-varying system (bottom) driven by the same, periodically correlated excitation signal. . . .	72
4.3	Actual (left-hand side) and estimated (right-hand) squared system functions for a periodically-varying resonance, in magnitude (top) and phase (bottom).	75
4.4	Actual (left-hand side) and estimated (right-hand) squared system functions for a lowpass-varying resonance, in magnitude (top) and phase (bottom).	77
4.5	Phase-response estimates of a proper signal (left-hand side) and an improper signal (right-hand), for a periodic (top) and lowpass (bottom) resonance.	78
4.6	Actual (left-hand side) and estimated (right-hand) squared system functions for a 4th-order cyclic system, in magnitude (top) and phase (bottom).	79
4.7	Actual (left-hand side) and estimated (right-hand) squared system functions for a 4th-order lowpass system, in magnitude (top) and phase (bottom). Confusion between red and blue in the phase response is due to $\pm\pi/2$ ambiguity in the squared modulators.	80
4.8	Log-likelihood functions, plotted for a single moment in time over the complex plane of possible modulator values. The true modulator value is shown by a black dot. Left: proper signal, with no phase estimation. Right: improper signal, with only two possible solutions.	93

4.9	Illustration of square-law demodulation in the frequency domain, where the solid lines constitute the Fourier transform of $y_k(t)$ and the dashed lines constitute lowpass and bandpass components of $y_k^2(t)$	94
4.10	Hypothetical support region for a coherently-modulated linear system $H'(\nu, \tau)$, with underspread system subcomponent $H(\nu, \tau)$ and carrier rate ω_0 . The bounding rectangle is possibly overspread, while the true support region (filled) is still underspread.	96
5.1	Independent realizations of the sinusoidal process for the proper case (top, $\sigma_2^2 = \sigma_1^2$) and an improper case (bottom, $\sigma_2^2 \gg \sigma_1^2$), with corresponding Fourier coefficient distributions to the right. The phase delay, shown as a dotted line, is discernible only in the improper case.	102
5.2	Independent realizations of the sinusoidal process for the proper case (top, $\sigma_2^2 = \sigma_1^2$) and an improper case (bottom, $\sigma_2^2 \gg \sigma_1^2$), with corresponding Fourier coefficient distributions to the right. The phase delay, shown as a dotted line, is discernible only in the improper case.	105
5.3	Taxonomy for all second-order random processes. On the left are signals with balanced elements, versus unbalanced on the right. The relative size of boxes is meant to roughly reflect the cardinality of the sets.	108
5.4	Early understanding of Gaussian processes as either WSS or nonstationary (left), but later expanded to allow slowly-varying, or quasi-WSS, processes (right).	110
5.5	Modern, yet incomplete understanding of second-order processes, as either proper or improper, or alternatively, either balanced or imbalanced. However, propriety is only a sufficient condition for balance.	111
5.6	Samples of the modified DFT for $N = 8$, spaced uniformly around the unit circle by offset by an angle of $2\pi/16$	115
5.7	Time-domain (left) and dB-magnitude frequency-domain (right) covariance matrices for two sinusoidal random processes. Top: balanced, WSS. Bottom: unbalanced, improper polarized.	120
5.8	Time-domain (left) and dB-magnitude frequency-domain (right) covariance matrices for two quadrature random processes. Top: balanced, proper. Bottom: unbalanced, improper.	121
5.9	Time-domain (left) and dB-magnitude frequency-domain (right) covariance matrices for two non-quadrature random processes. Top: Haar basis, improper. Bottom: Daubechies-4 basis, also improper.	122
6.1	Time-domain view (top) of the random sinusoidal signal, which some would possibly call “quasi-periodic,” and average power spectral estimate (bottom) using a long-term periodogram.	127

6.2	Square-law spectral analysis of the random sinusoidal signal, showing the baseband spectrum (left) and sideband spectrum (right). These are equivalent to the Hermitian and complementary envelope spectra for the analytic signal.	128
6.3	Left: synchronous Hermitian power spectrum (dotted) overlaid with the synchronous complementary spectrum (solid), both found through averaging across T -length STFT frames. Right: complex-plane scattergram for the STFT bin corresponding to 1000 Hz.	128
6.4	Normalized histograms for 2000 independent trials of computing $ \hat{\gamma}_z $ for a proper (blue) and an improper (red) complex r.v. The dashed lines indicate the true NC for the improper sequence. Left: $N = 15$ samples, right: $N = 45$ samples.	130
6.5	Asymptotic distributions (dotted) overlaid with Monte Carlo histograms (solid) for five values of $ \gamma_z $. The H_0 , or proper, distribution is in blue. The 2% null-rejection threshold of 0.29 is indicated by the vertical dashed line. . .	131
6.6	Narrowband spectrogram of female speech, “bird populations,” on a dB color scale.	132
6.7	Real (red, solid) and imaginary (magenta, dashed) parts of three demodulated speech harmonics. The speech is the same as in Figure 6.6. Bottom: fundamental modulator. Middle: first harmonic. Top: second harmonic. . . .	133
6.8	Local noncircularity coefficients for the three basebanded subbands in Figure 6.7. The three averaging lengths are 40 (blue), 80 (green), and 160 (red) milliseconds. The dashed lines indicate the null-rejection threshold, with a p-value of 2%, for the weakest estimator.	134
6.9	Square-law modulation spectra, averaged over all harmonics. Left: Hermitian, showing syllabic-like peaks at 1.9 and 3.6 Hz. Right: Complementary, showing possible traces of syllabic-like modulations.	135
6.10	Impropriety measure as a function of subband center frequency, for three different ships and recording conditions. See text for details.	139
6.11	Sound pressure waveforms for the tanker (left) and self-noise (right) recordings used in Figure 6.10. Hard modulation bursts are prominent in the tanker data, suggesting a possible source of impropriety.	140
6.12	Hermitian (left) and complementary (right) DEMONgrams for merchant data, center-frequency = 450 Hz, bandwidth = 200 Hz, frame-length = 15 seconds. Impropriety is detected, as seen by complementary modulations. . .	142
6.13	Frame-by-frame GLR (red dots) for the subband observed in Figure 6.12, with null distribution in grayscale and the 2% p-value threshold shown as a dashed line.	142

6.14	Hermitian (left) and complementary (right) DEMONgrams for tanker data, center-frequency = 2250 Hz, bandwidth = 400 Hz, frame-length = 15 seconds. Impropropriety is detected, as seen by complementary modulations.	143
6.15	Frame-by-frame GLR (red dots) for the subband observed in Figure 6.14, with null distribution in grayscale and the 2% p-value threshold shown as a dashed line.	143
6.16	Hermitian (left) and complementary (right) DEMONgrams for self-noise data, center-frequency = 5000 Hz, bandwidth = 400 Hz, frame-length = 15 seconds. No impropropriety is detected.	144
6.17	Frame-by-frame GLR (red dots) for the subband observed in Figure 6.16, with null distribution in grayscale and the 2% p-value threshold shown as a dashed line.	144
6.18	Coherence chart for principal component pairs. A high Q-score corresponds to a quadrature pair, for which γ is meaningful as a measure of noncircularity or coherence.	148
6.19	Q-scores for tanker data, before (left) and after (right) thresholding above 0.6.	150
6.20	Thresholded Q-scores (left) compared to coherence scores (right) for tanker data.	150
6.21	Hilbert-envelope multiband modulation spectra for tanker data, before (top) and after (bottom) quadrature-subspace projection.	151
6.22	Narrowband spectrograms for tanker data, before (top) and after (bottom) quadrature-subspace projection.	152
6.23	Thresholded Q-scores (left) compared to coherence scores (right) for merchant data.	153
6.24	Hilbert-envelope multiband modulation spectra for merchant data, before (top) and after (bottom) quadrature-subspace projection.	153
6.25	Narrowband spectrograms for merchant data, before (top) and after (bottom) quadrature-subspace projection.	154
B.1	Magnitude (left) and phase (right) of the complementary spectrum for the same signal as in Figure 3.3, except with a colored noise source. The gray curves, derived analytically via (B.7), predict somewhat the measured subband complementary variances (circles). The correlation length of the signal is very short, around 0.5 millisecond, in order to make the predictions close.	181
B.2	A demonstration of synchronized filtering of a periodically correlated process, with $D = 3$. The modulator is shown in red, in this case a sinusoid. The process $x(t)$ is shown in light blue. Successive window placement is chosen such that samples in one frame are uncorrelated with samples in an adjacent frame, as dictated by the correlation length (illustrated by the triangle at the edge of the first frame).	182

C.1 Monte Carlo histograms for five values of L_z measured with additive proper noise. The H_0 , or proper, distribution is in blue. The 2% p-value threshold of 0.29 is indicated by the vertical dashed line. Top: Diagonal covariance GLR. Bottom: Full-covariance GLR. 186

C.2 Monte Carlo histograms (solid) for five values of $|\gamma_z|$ measured with additive proper noise. The H_0 , or proper, distribution is in blue. The 2% p-value threshold of 0.29 is indicated by the vertical dashed line. Top: SNR = 10 dB. Bottom: SNR = 0 dB. 187

ACKNOWLEDGMENTS

I wish to thank my advisor Les Atlas, who helped me realize a potential I didn't know I had. His persistence, creativity, and insight have been invaluable resources and an inspiration. I am also grateful to Ivars Kirsteins, whose guidance and expertise have been tremendous. My thanks go also to Bishnu Atal, who taught me what it means to seek the essence of a thing. Through their humility and openness, my mentors have shown by example that the highest aspiration is perhaps not to solve but to understand.

I am grateful to my dissertation committee members Li Deng and Richard Wright, whose advice helped guide this work to its current state. Over the course of my graduate years, I have had the opportunity to work with several exceptional individuals. I extend my thanks to Oded Ghitza and Steve Greenberg, for technical discussions as well as early and continued encouragement; to Don Percival, Louis Scharf, and Peter Schreier for mathematical discussions; to Geoff Zweig and Patrick Nguyen for their leadership in the 2010 CLSP summer workshop; to Andy Moorer for inspiration and technical discussions; to Shourov Chatterji for guidance at Lincoln Laboratory; to Jim Pitton, Kaibao Nie, and Shihab Shamma for many helpful discussions; and to my early role models Eric Klavins and Radha Poovendran.

This work was funded by the Office of Naval Research. I thank John Tague of the ONR for his suggestions and advice. I would also like to thank Willard Larkin of the AFOSR for his support over the years.

I wish to thank my colleagues over the years: Scott Philips, Steven Schimmel, and Jeff Cole, for showing me the ropes; Erik Edwards and KC Lee for theory and history discussions over beers; Greg Sell, Nima Mesgarani, and Brian King for collaboration and friendship; Rahul Vanam for musing on life, the universe, and everything; and all

of my lab-mates, past and present: David Slater, Josh Bishop, Fay Shaw, Eric Garcia, Cameron Colpitts, Danny Luong, Adam Greenhall, Patrick McVittie, Nicole Nichols, Xing Li, Elliot Saba, Laura Vertatschitsch, Po-Han Wu, Jason Silver, Jaehong Chon, Jessica Tran, Jongho Won, Dan Tidwell, Suresh Chandrasekan, Kai Wei, and Bill Kooiman.

Finally, I thank my family, friends, and girlfriend for their steadfast support and encouragement, and for making it all worthwhile.

DEDICATION

To my Mom and Dad.

Chapter 1

INTRODUCTION

1.1 Modulation Analysis of Natural Signals

Modulation is essential for transmitting message-bearing signals. Notable man-made examples are AM and FM radio, and subsequent evolutions for wireless networks and phones. Modulation representations are also a powerful tool for analyzing signals which occur in nature. Signals such as speech, machine noise, and brain waves contain informative “messages” which are useful to scientists and engineers. Therefore, analysis is a matter of finding the most appropriate decoding for technological applications. Possible applications are automatic speech recognition, marine vessel identification, brain-computer interfaces, and medical diagnosis, to name only a few.

Modulation is useful precisely because it solves the problem of mismatch between the message and the receiver. For example, the human ear is incapable of perceiving acoustic frequencies below 20 Hz. This is a biomechanical limitation due to the size of the ear and its transducers. Then how is it possible for humans to hear low-frequency messages, such as speech and music, when the tempo can be as low as 2-8 Hz? The answer is that speech and music are actually broadband signals, whose spectral content varies over time due to syllabic or rhythmic modulation. In other words, modulations are distinct from the realm of acoustic frequency. We experience this distinction every day, when we simultaneously perceive syllabic timing in speech as well as the talker’s pitch.

In its simplest form, a modulated signal $y(t)$ is the product

$$y(t) = m_R(t) \cos(\omega_0 t + \phi_0) - m_I(t) \sin(\omega_0 t + \phi_0) \quad (1.1)$$

where $m_R(t)$ and $m_I(t)$ are low-frequency modulators, and ω_0 is the frequency of the sinusoidal carriers. If $y(t)$ is an audio signal like speech, then ω_0 is in the range of acoustic

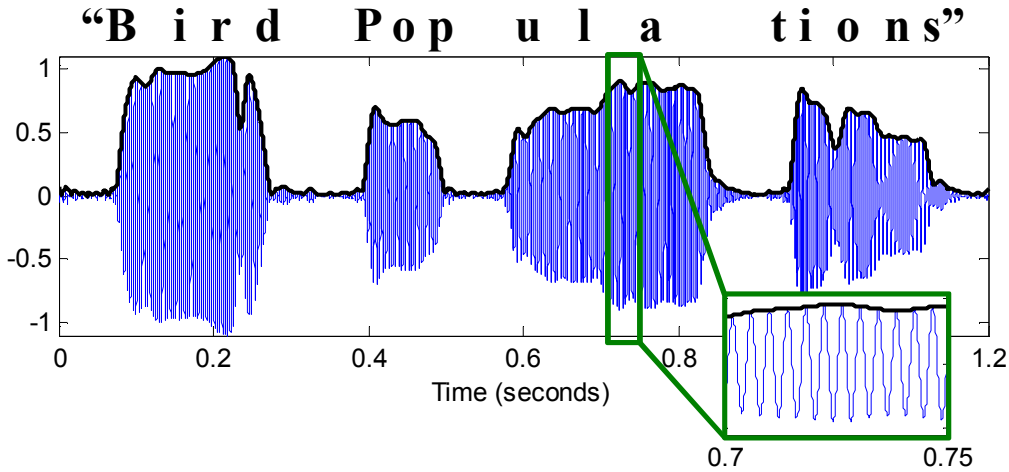


Figure 1.1: Annotated display of one subband envelope from a woman saying, “bird populations.” Blue indicates the subband signal, and black is the non-coherent envelope. The inset image is a zoomed-in portion showing the oscillations of the carrier.

frequency. The instantaneous power is the non-negative average

$$P(t) = m_R^2(t) + m_I^2(t) \quad (1.2)$$

which disregards the relative timing information between modulators. Although rarely acknowledged, $P(t)$ is the implicit estimator in spectrograms and many conventional subband methods. The subband envelope is often defined as $\sqrt{P(t)}$. To borrow a term from radar and radio, $\sqrt{P(t)}$ is the *non-coherent* envelope of $y(t)$, since it is agnostic to the sinusoidal timing of the carriers.

The non-coherent envelope is useful in its own right. It represents the power modulation of a signal, as shown for the speech example in Figure 1.1. What might a coherent envelope look like? As (1.1) might suggest, there are actually two envelopes, or, in a complex algebra, a single complex-valued “envelope” or modulator. The complex modulator is simply $m(t) = m_R(t) + jm_I(t)$, where $j^2 = -1$. Figure 1.2 demonstrates the interplay between the real and imaginary parts of the complex modulator for one speech subband. Complex representations are rooted in basic Fourier theory and have theoretical and practical advantages regarding analysis, modification, and synthesis [6][7][8]. Simply speaking, the complex modulator

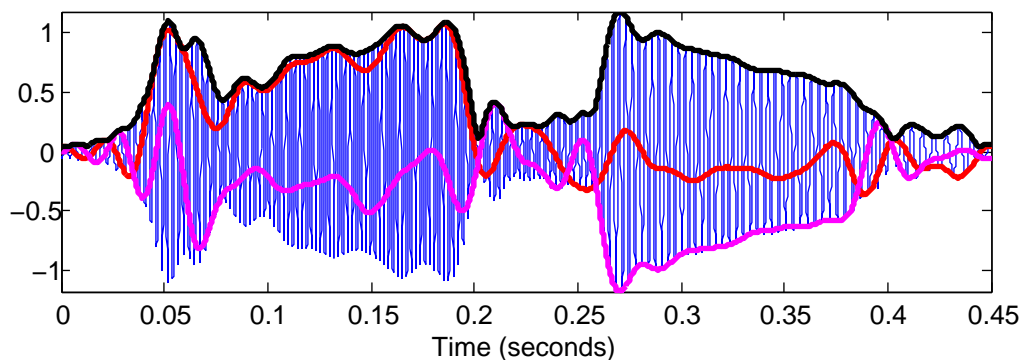


Figure 1.2: One subband from a woman saying, “thing about.” Blue indicates the subband signal and black is the non-coherent envelope. Red and magenta are the real and imaginary parts of the complex modulator.

encodes changes in local sinusoidal phasing or timing.

In speech, the harmonics of the fundamental frequency can be treated as subband carriers. In fact, Figure 1.2 was generated from a pitch-tracking subband. We should emphasize that the subject of this dissertation is not pitch-based processing, for speech or otherwise. Pitch detection is actually a notoriously difficult problem which often resorts to heuristics. Furthermore, pitch is speech-specific with limited application to non-harmonic signals.

We propose that harmonic pitch is an example of a more general concept. A signal is *coherently modulated* if it contains repeatable structure on a short time scale, which in turn provides a reference for detecting amplitude and phase modulations on a longer time scale. By repeatable, we do not mean periodic. Repeatable structure is potentially random to varying degree, but with an internal sense of timing. Therefore, periodicity is a subset of a more general and robust phenomenon of rhythm. For example, see the synthetic example in Figure 1.3. The signal is broadband, contains no harmonics, and yet appears periodic-like in its short-term carrier structure. At the same time, there is a long-term modulation period of about 60 milliseconds. This signal, like speech, is rhythmic on two time scales. This leads to our main question: By synchronizing with one time scale, what additional information can we learn about the other time scale?

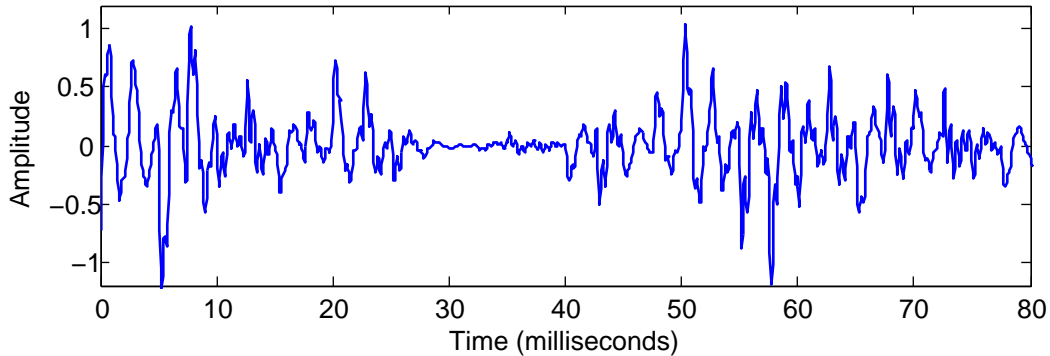


Figure 1.3: A synthetic example of a broadband coherent random process, with regular short-term structure which implies long-term variation.

Our intent is to mathematically define coherence and synchrony in a system-theoretic framework for Gaussian signals. We nominate random-phase sinusoids as a signal basis for coherent signals. Our main result is to show that sinusoidal coherence manifests in the frequency domain as a type of complex correlativity called *impropriety*. As a second-order property, impropriety is embedded in a system algebra with all the classical benefits of dimensional analysis, optimization metrics, and unambiguous definition of characteristic signal components. Such algebraic tools make it possible to generalize beyond pitch as a timing cue, and toward a general theory of synchronization which could apply to something as broadband as underwater propeller noise.

As mentioned, coherent signals are structured at both the carrier and modulator scale. Impropriety may apply to either scale, but as a starting point, we focus on impropriety of the carriers. Synchronizing with the carrier structure of a random process then resembles a system identification problem. Therefore, this thesis can be summarized by the following main concepts.

1.2 Major Concepts

1.2.1 Nonstationary Random Processes

A Gaussian random process is characterized by its second-order statistics. The literature on nonstationary processes usually assumes quasi-stationarity, in which the signal appears stationary on a short-enough time scale. Our contribution is to identify coherent modulation as *highly nonstationary*. Perhaps the easiest example is speech, which is sometimes mistakenly called quasi-stationary. It may be accurate to consider the vocal tract articulation as quasi-stationary, but the glottal pulses are highly structured on a short time scale.

The first key to estimation is separability of the short-term (carrier) and long-term (modulator) structure. The second is synchronization with the carriers. Both require a framework involving the following system-theoretic and algebraic principles.

1.2.2 Linear Time-Varying Systems

For a Gaussian random process $y(t)$, there exists a linear, time-varying system such that $y(t)$ is the output of the system driven by white, stationary noise [9]. Linking random processes to systems provides a deep reserve of analysis tools. For example, the eigenfunctions of a system translate to the principal component of a process. Also, the theory of underspread systems [2][5] has ramifications on the fundamental limits of estimating the covariance of a random process [3].

A key issue is whether a generative system can be factored into a cascade of sub-systems. Supposing

$$w(t) \longrightarrow H \longrightarrow y(t) \tag{1.3}$$

where $w(t)$ is white stationary noise and H is a linear system operator, the coherent modulation model assumes that H is separable in the sense that

$$w(t) \longrightarrow H_C \longrightarrow H_M \longrightarrow y(t). \tag{1.4}$$

The subsystems H_C and H_M are different in kind, and correspond respectively to short-term (carrier) and long-term (modulator) structure in the observed signal $y(t)$.

1.2.3 Estimation as an Inverse Problem

The above linear system model can be used to synthesize artificial signals on a computer. The real power of a synthesis model, though, is that it defines the maximum entropy estimator for a signal [10]. Given $y(t)$ and the model (1.3), the optimal estimator is the linear time-varying system H^{-1} such that

$$y(t) \longrightarrow H^{-1} \longrightarrow w(t). \quad (1.5)$$

The above criterion states that everything is known about $y(t)$ when the analysis residual is the maximum-entropy signal, white noise. This assumes, of course, that H is invertible. When H is a time-invariant system, then H^{-1} is the linear prediction filter which also minimizes squared prediction error.

From the standpoint of coherent modulation, we modify the above estimation criterion as follows. Demodulation is the systematic removal of time-variation from the statistics of $y(t)$. That is, $y(t)$ is reduced to stationary noise. This is reminiscent of Grenier's time-varying linear prediction algorithm [10], which estimates a parametric form of H . Coherent demodulation assumes the separable form in (1.4) with distinct properties on subsystems H_C and H_M . This in turn allows a more complete and numerically tractable estimation.

1.2.4 Complex Algebra and Statistics

Complex signals can be obtained by applying the Fourier transform, or equivalently, the Hilbert transform. This is advantageous because the algebra of complex numbers is compact for sinusoids. For example, the sine/cosine product in (1.1) becomes a scalar product

$$y_a(t) = m(t) e^{j\omega_0 t} = [m_R(t) + jm_I(t)] e^{j\omega_0 t} \quad (1.6)$$

where $y_a(t)$ is the analytic signal of $y(t)$. Similarly, the transfer function of a linear, time-invariant system is

$$H(\omega) = \frac{Y(\omega)}{W(\omega)}. \quad (1.7)$$

We seek analogous algebraic representations for stochastic signals with the separable subsystems H_C and H_M . Characterization in the frequency domain therefore requires complex

statistics, since the Fourier transform $Y(\omega)$ is a complex random process. Therefore, impropriety provides the link between the frequency domain and coherent, or synchronous, demodulation.

1.3 Organization

Chapter 2 gives the necessary background, beginning with the narrowband AM-FM theory of coherent modulation and modulation frequency, then proceeding to nonstationary random processes and system theory, and finally covering complementary statistics for complex random variables and processes.

Chapter 3 develops a basic theory of impropriety in the frequency domain of a random process. We find that amplitude modulation has a systematic relation to spectral impropriety, and under certain conditions, is separable from the convolutional element of a system. This leads to a frequency-domain transfer function for the phasing of improper carrier sinusoids.

Chapter 4 extends the signal model from Chapter 3 to account for a slowly time-varying convolutional element. We derive conditions for separability and estimation of the slowly-varying (long-term) modulations relative to the high-frequency (short-term) carrier. Coherent estimation is therefore based on subband filtering synchronized to the carrier.

Chapter 5 departs from filterbank analysis and investigates the principal components of impropriety. This leads to the most important theoretical result of this dissertation, a complete taxonomy of improper and proper random processes. An algebraic representation leads to several new contributions, one of which is coherent modulation defined with respect to impropriety in its most reduced form. We also define, for the first time, necessary *and* sufficient conditions for a signal to be coherent or non-coherent.

Chapter 6 applies the theories of previous chapters to real-world data. The primary contribution is a methodology for estimating impropriety and quantifying statistical significance. As examples, we demonstrate the existence of impropriety in speech and underwater propeller noise, by adapting a statistical tests from the literature.

Finally, Chapter 7 concludes the dissertation with a summary of main points and directions for future work.

Chapter 2

BACKGROUND

2.1 Introduction

Subband demodulation is the estimation of low-frequency, temporal envelopes from band-pass frequency regions of a signal. We refer to the subband frequencies as “acoustic” or carrier frequency, and the rate of envelope fluctuation as modulation frequency. A dual notion of frequency is intuitive in the sense of separating rhythm from pitch, but making this separation precise is a challenging mathematical problem. This is largely the objective of the family of coherent demodulation algorithms, with applications in characterizing the long-term structure of information in speech and machine noise, for example, corresponding to a time scale of hundreds of milliseconds or even seconds. This is our main focus, but questions of optimality require tools and concepts from other fields which may otherwise appear loosely related.

The first major connection is with the rich area of bifrequency, or generally bilinear, representations for random signals. We will see how modulation is related to classical time-frequency spectral representations which provide system-theoretic limits on measureability. This fundamentally imposes a sort of information limit on what a demodulation algorithm can estimate from a signal. Traditionally estimators assume a lowpass limit consistent with the notion of a slowly-varying system. In principle, however, the information limit can take many forms, where the lowpass assumption is one of many sufficient, but not necessary, communications schemes for demodulation. For random or noise-like signals, arising from physical systems, it is not immediately clear what other schemes are possible.

The approach taken by this dissertation is based on a relatively new type of second-order statistic. Conventional estimators do not take advantage of possible “complementary” statistics present in complex spectral decompositions of a signal. It remains to be seen how the correct use of these new, complex statistics may apply to optimal subband demodulation

of a signal. This leads to a new and rich problem, which is to characterize the so-called impropriety, or statistical phase preference, of a complex subband with a meaningful relation to the original real-valued signal.

The chapter is organized as follows. Section 2.2 gives the history of modulation analysis in speech, starting with the Hilbert envelope and leading to coherent demodulation theory. Section 2.3 then discusses stochastic demodulation, and how second-order statistics provide a basis for defining optimality and the limits of measurability. Section 2.4 expands upon bilinear, or second-order, estimation in the form of recently developed, complementary statistics connected to the impropriety of complex signals. Finally, Section 2.5 summarizes the chapter and defines the problem statement for the rest of the thesis.

2.2 Modulation Frequency Analysis in Speech

Speech has long been thought of as a modulated signal. In 1939, Dudley introduced two methods for artificially synthesizing speech, one manually operated in real time (known as the “Voder” [11]) and the other automatic [1]. In both systems speech is “remade” by multiplying modulator envelopes with narrowband carriers. A reproduction of Dudley’s schematic appears in Figure 2.1, showing the estimation of modulators by what is essentially the upper-envelope of each subband. Around the same time, the spectrogram gained popularity as the analysis dual to Dudley’s synthesis approach. Called “visual speech” by Potter [12], the spectrogram displays subband modulations over time and frequency dimensions to reveal structural variations such as formants and pitch harmonics. An example spectrogram and a few corresponding subband plots appear in Figures 2.2 and 2.3.

A widely adopted tool for subband demodulation is the so-called Hilbert envelope. We shall see that it is related to, yet more rigorously defined than, the rectified-and-smoothed envelopes in Dudley’s system. For a real bandpass signal $x(t)$, the Hilbert envelope is

$$m_H(t) = |x_a(t)| = |x(t) + j\hat{x}(t)| \quad (2.1)$$

where $x_a(t)$ is the complex analytic¹ signal [13] and $\hat{x}(t)$ is the Hilbert transform of $x(t)$. For

¹In this context, “analytic” is not in the usual mathematical sense of differentiability. An “analytic signal” is one for which the Fourier transform vanishes over the negative frequencies.

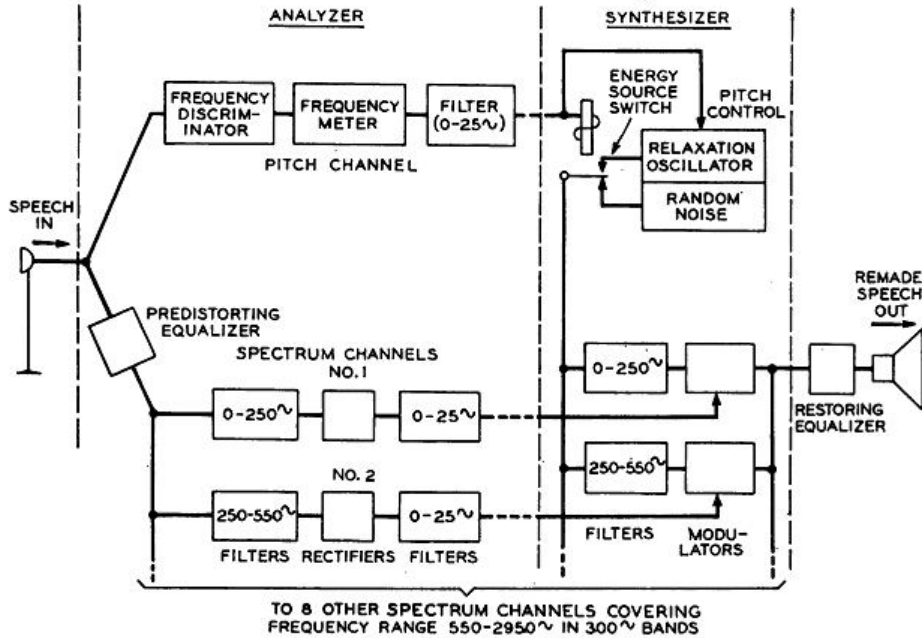


Figure 2.1: Schematic for automatic synthesis of speech, reproduced from [1].

now it suffices to understand that the Hilbert transform operator, denoted as \mathcal{H} , exchanges $\cos(\theta)$ for $\sin(\theta)$ and $\sin(\theta)$ for $-\cos(\theta)$. The corresponding Hilbert carrier is defined as

$$c_H(t) = \frac{x_a(t)}{m_H(t)} = \exp(j\angle x_a(t)) \quad (2.2)$$

and, since $x(t) = \text{Re } x_a(t)$, the real signal is given by

$$x(t) = m_H(t) \cos(\angle x_a(t)). \quad (2.3)$$

Introduced by Dugundji [14], the Hilbert envelope yields a real, non-negative, upper envelope of a bandpass signal (for examples, the red curves in Figure 2.2 are Hilbert envelopes). The envelope/carrier pair is also unique (ignoring points in time with zero amplitude, during which phase is indeterminate). The reason it works is due to Bedrosian's product theorem

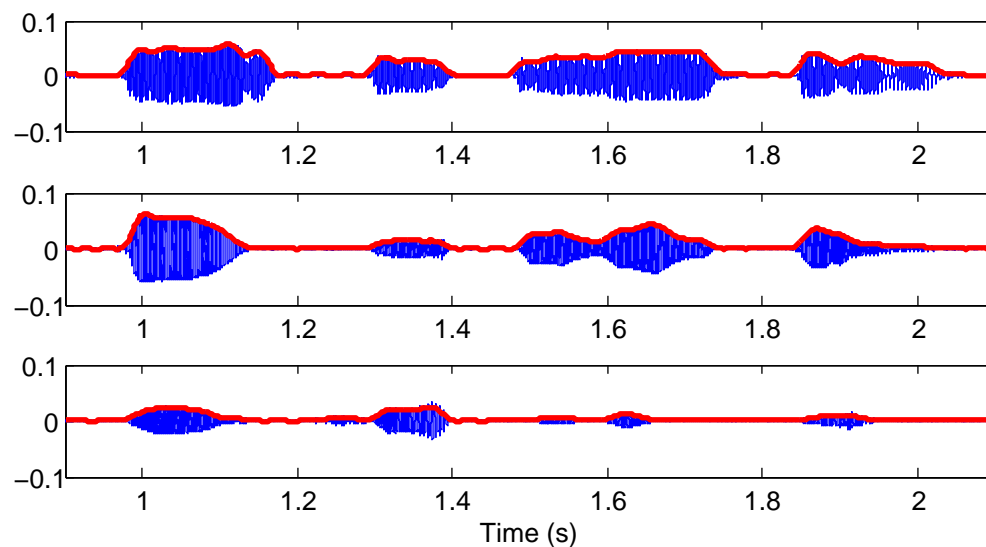


Figure 2.2: Subband time series taken from a single speech utterance (“bird populations”) overlaid with their upper envelopes in red. From top to bottom, the subband frequency boundaries are 625-875 Hz, 375-625 Hz, and 125-375 Hz.

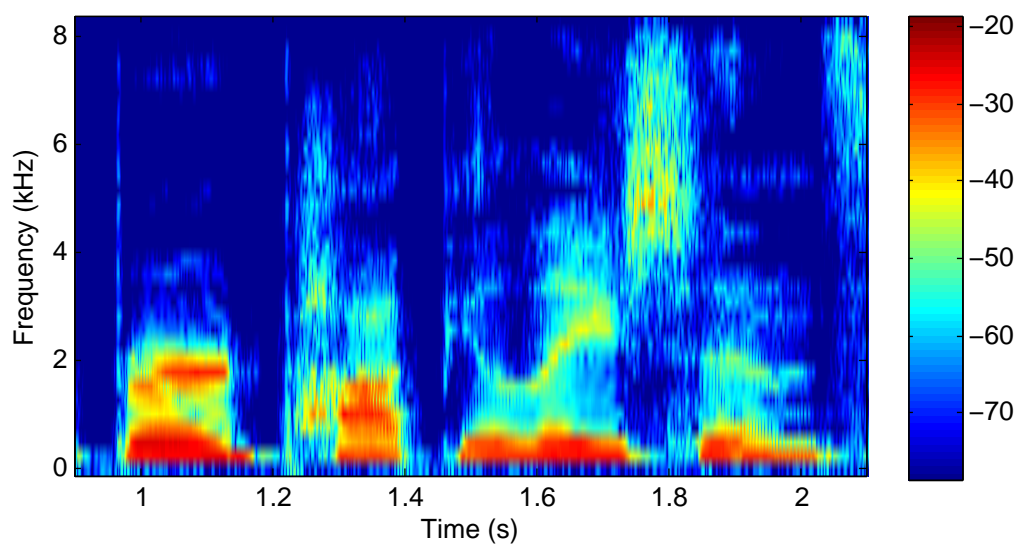


Figure 2.3: Spectrogram of the same speech utterance as in Figure 2.2. The vertical axis indicates subband center frequency, and the colormap is a decibel scale.

[15], which states the following. Given a signal in product-form, $x(t) = m(t)c(t)$, where the highest frequency in $m(t)$ is less than the lowest frequency in $c(t)$, the Hilbert transform is

$$\hat{x}(t) = \mathcal{H}\{x(t)\} = m(t)\mathcal{H}\{c(t)\} = m(t)\hat{c}(t). \quad (2.4)$$

It follows that $x_a(t) = m(t)c_a(t)$. Assuming that $|c_a(t)| = 1$ and $m(t)$ is non-negative, then $m_H(t) = m(t)$, which is lowpass by definition in the proposition (2.4).

The Hilbert envelope would be elegant if not for a subtle contradiction. Bedrosian’s theorem requires bandlimited components $m(t)$ and $c(t)$, such that they do not overlap in the frequency domain. Yet Dugundji himself observed that, while the *squared envelope* is bandlimited, this does “not allow one to draw the conclusion that the envelope itself is bandlimited; indeed, there seems to be no physical reason that it should be so” [14]. Therefore, the Hilbert envelope and its corresponding carrier may not belong to the Bedrosian class, but instead to another class of signals whose products are also analytic. The physical significance of the second class is doubtful if not simply distorted. Furthermore, it is unclear how one might systematically design or modify signals within this class. This is central to the argument in favor of coherent demodulation, to which we will return later in this section.

An early instance of non-Bedrosian signal decomposition occurred in Lerner’s 1960 treatment of the AM-FM signal $x(t) = m(t) \cos(\phi(t))$ with bandlimited $m(t)$ and $\phi(t)$. To some surprise (see Bedrosian’s contentious footnote in [15]), Lerner concluded that the Hilbert transform of $x(t)$ is only approximately equal to $m(t) \sin(\phi(t))$ with the approximation improving in the narrowband case. Dugundji made a similar observation in [14]. In response to Lerner, Rihaczek observed that, since $\cos(\phi(t))$ is not bandlimited², then it does not form a Bedrosian pair with $m(t)$. More to the point, the complex signal $m(t)e^{j\phi(t)}$ can be analytic only when $m(t)$ and $\phi(t)$ are carefully balanced against one another. In other words, $m(t)$ and $\phi(t)$ can not be independently specified, and are “merely mathematical functions with no relation to any physical modulation process” [17]. Although Rihaczek did not discuss envelope estimation, his analysis was an early indicator of the problem with non-bandlimited AM and FM components, such as the Hilbert envelope and its carrier.

²The non-bandlimited nature of FM signals, even when $\phi(t)$ is itself bandlimited, was first discussed by Carson [16].

Nevertheless, the Hilbert envelope has illuminated the functional relationship between speech modulation and the physiology of perception. Of particular interest is the concept of modulation frequency as a predictor for speech intelligibility³, pioneered by Houtgast and Steeneken starting in 1973 [20] and summarized in [21]. They made two key observations. First, the modulation spectrum of clean speech is concentrated below 16 Hz and is markedly peaked around 3-4 Hz. Second, they posited a Modulation Transfer Function (MTF) to describe the “filter characteristic” of a room acting upon the envelopes of speech, hence resulting in attenuation of the modulation spectrum [20]. From this they found a measure, or speech transmission index, for the intelligibility of speech depending on the MTF of the environment. Perhaps the earliest MTF, however, was due to Riesz [22] in 1928, who found that the human ear is most sensitive to low-frequency modulations (which he called “intensity fluctuations”) primarily around 4 Hz. Viemeister confirmed and extended Riesz’s work, employing the MTF for a “systems analysis” approach to measuring temporal processing in the human ear [23, 24]. “Given the transfer function based upon modulation,” he wrote in [23], “it should be possible to predict the output for arbitrary modulation waveforms. In this sense the modulation transfer function may provide a general and complete description of of temporal processing.”

The MTF concept led to two landmark studies in human perception by systematic manipulation of modulation frequency. In 1994 Drullman *et al.* [25] noted that Viemeister’s lowpass MTF, in conjunction with the acoustic analysis of Houtgast and Steeneken, suggested that “the ear’s capacity to detect temporal modulations is not a limiting factor in speech perception.” In order to learn the minimum modulation cutoff for the transmission of intelligible speech, they devised a modulation filtering experiment. Their algorithm 1) filtered speech into subbands, 2) found the Hilbert envelope for each subband, 3) lowpass filtered each envelope, and 4) recombined the smoothed envelopes with the original carriers

³Our discussion is biased toward psychophysical MTFs, based on subjective perception as studied by Riesz, Houtgast, Steeneken, and Viemeister. Another perspective is to directly measure the modulation-frequency response of individual cochlear neurons. See Møller [18] for a detailed discussion. Interestingly, Møller’s MTF curves are peaked between 100 and 200 Hz, which suggest selectivity for high-frequency *carrier* frequencies at the auditory periphery. Low-frequency modulations related to speech intelligibility might therefore be detected at higher stages of neural processing. For example, see the progression of single-unit MTFs – from the cochlear nucleus, to the auditory nerve, to the inferior colliculus – in Delgutte *et al.* [19].

to form a modified speech signal. Judging from listeners' responses they determined that the intelligibility of speech remained unchanged for lowpass cutoff frequencies above 16 Hz, regardless of the subband bandwidth.

A year later, Shannon *et al.* [26] demonstrated the importance of speech modulation by multiplying lowpass subband envelopes from speech with noise carriers matching the original subband bandwidths. With only three to four subbands, they found that speech remained intelligible with a modulation cutoff as little as 16 Hz. Both studies, Drullman *et al.* and Shannon *et al.*, have clear implications regarding the inherent compressibility of speech in the modulation-frequency domain. This encouraged applications to encoding for prosthetic cochlear implants (the original motivation for Shannon *et al.*) and audio compression [27].

Perhaps the most useful idea to emerge from these perceptual studies is that the human ear is a matched receiver for the principal modulations in speech. Furthermore, the compact MTF of speech suggests a class of modulation filters tuned to suppress “non-speech” modulations resulting from additive noise, reverberation, and microphone characteristics. This latter idea emerged in the field of automatic speech recognition (ASR), famously in the form of relative spectral (RASTA) processing developed by Hermansky and Morgan [28]. RASTA is essentially a bandpass modulation filter for the range 1-10 hz, and operates on the log-magnitude envelope of a subband. Other forms of temporal processing in ASR features are cepstral-mean subtraction [29, 30] and delta-cepstral coefficients [31], which comprise a simpler highpass subset of the more general RASTA method. Following RASTA, Greenberg and Kingsbury [32] specifically chose a 4-Hz bandpass modulation filter for the so-called modulation spectrogram, which was then used as an ASR front-end in [33]. More recently, Chiu and Stern [34] proposed an adaptive modulation filter for ASR features designed to jointly suppress noise modulations and preserve speech modulations.

This body of work, extending back to Houtgast and Steeneken and up to contemporary ASR front-end systems, demonstrates the general utility of analysis and modification based on modulation frequency. Nearly all of the aforementioned studies used Hilbert envelopes,

whereas a few used the closely related⁴ half-wave rectification method (e.g., [32] and [26]). Such modulation systems essentially yield the upper envelope of a subband as a measure of instantaneous power, which is known in the communications literature as “non-coherent” demodulation. The alternative, “coherent” demodulation, will be defined shortly. First, we must place the Hilbert envelope in a systems-theoretic framework so as to view its shortcomings, and hence motivate the recent interest in coherent modulation analysis.

In 2001, Ghitza [35] reproduced the modulation filter from Drullman *et al.* and observed an unexpected result. The modulation-filtered speech, when again filtered into subbands, actually *regenerated* high-frequency modulation content after re-estimating the Hilbert envelope. Moreover, this occurred for just one subband in isolation, which ruled out interference between subbands. If the smoothed envelope is not recoverable after being filtered and rectified by the human ear, Ghitza reasoned, then the findings of Drullman *et al.* may be flawed. The root of the problem is that, in the Hilbert product model,

$$x_k(t) = m_{k,H}(t)c_{k,H}(t) \tag{2.5}$$

both $m_{k,H}(t)$ and $c_{k,H}(t)$ are non-bandlimited. The subband signal $x_k(t)$ therefore remains bandlimited only when $m_{k,H}(t)$ and $c_{k,H}(t)$ are perfectly matched, so any modification of the envelope (e.g., modulation filtering) necessarily causes bandwidth expansion in the corresponding analytic signal. Ghitza’s observation is precisely the point made in Rihaczek’s assessment of non-Bedrosian factorization of an analytic signal [17] nearly 40 years prior, and foreshadowed by Dugundji’s statement [14] on the non-bandlimited nature of the Hilbert envelope in 1958.

Another alternative is to redefine the envelope altogether. Although rarely acknowledged, demodulation implicitly assumes a product model which is fundamentally under-determined. To illustrate, Hilbert demodulation proceeds as follows. From a bandpass signal $x(t)$, a) form the complex signal $x_a(t)$ via the Hilbert transform. Next, define the envelope and carrier such that b) the envelope is real and non-negative, and c) the carrier is a unimodular, or phase-only, signal. It becomes necessary at this point to reevaluate each of the steps (a)-(c). With regard to (a), Vakman [36] showed that the analytic signal

⁴We will return to this point in Section 2.3.

uniquely satisfies three axioms which preserve the scalar amplitude and phase relations of a simple test signal $x(t) = m \cos(\omega t + \phi)$. With regard to (b) and (c), however, Cohen *et al.* [37] illustrate a basic ambiguity in the definitions of “amplitude” $m(t)$ and “phase” $\phi(t)$ when they are functions of time. For example, suppose we synthesize a modulated signal $x(t) = m(t)c(t)$, where $m(t) = \cos(\omega_m t)$, $c(t) = \cos(\omega_c t)$, and $\omega_m \ll \omega_c$. Then the Hilbert envelope yields $m_H(t) = |m(t)|$, which leaves the sign-flips as discontinuities in the Hilbert carrier [37, 38]. Clearly, the continuous solution suggested by the synthesis is incompatible with the non-coherent Hilbert envelope.

How then does a coherent envelope behave? Loughlin and Tacer [39] defined a generally *complex-valued* modulator $m(t)$ satisfying the coherent-modulation signal model

$$z(t) = m(t)c(t) = m(t) \exp(j\phi(t)) \quad (2.6)$$

where the instantaneous frequency⁵ of the carrier is the time-dependent spectral center-of-gravity (COG) of $z(t)$. If the carrier is spectrally concentrated, we can say $z(t) \approx x_a(t)$, such that

$$x(t) \approx \operatorname{Re} z(t) = i(t) \cos(\phi(t)) - q(t) \sin(\phi(t)) \quad (2.7)$$

where $m(t) = i(t) + jq(t)$, so $i(t)$ and $q(t)$ are the real and imaginary parts of the complex modulator. Equation (2.7) is known as Rice’s representation.⁶ Furthermore, the complex modulator is related to the Hilbert envelope via

$$m_H(t) = |m(t)|. \quad (2.8)$$

As seen in (2.7), coherent demodulation is defined by a synthesis model for an observed $x(t)$. Each component of that model possesses qualities which satisfy explicit conditions and operational objectives. Loughlin and Tacer primarily used the model to define a bounded

⁵Here we mean “instantaneous frequency” as the time-derivative of $\phi(t)$ in the coherent signal model defined in (2.6). In the time-frequency literature, it is common to define the instantaneous frequency as $d/dt \angle x_a(t)$ which presupposes the Hilbert-envelope signal model.

⁶The reference is due to Papoulis [40]. Originally, Rice [41] defined the envelope of a subband as $m_R(t) = \sqrt{a^2(t) + b^2(t)}$, where $a(t)$ and $b(t)$ are such that $x(t) = a(t) \cos(\omega_0 t) - b(t) \sin(\omega_0 t)$ for some midband frequency ω_0 . Dugundji proposed the Hilbert envelope approach with the express intent of simplifying Rice’s envelope by showing $m_H(t) = m_R(t)$ without the need to estimate a midband frequency, or carrier. As we have labored to prove in this chapter, the rehabilitation of Rice’s original form has taken almost a half-century to regain standing alongside the Hilbert envelope in the analysis of natural sounds like speech.

instantaneous frequency, but Atlas and Janssen [42] proposed complex envelopes for more effective modulation filtering of speech and music. Here, “effective” refers to Ghitza’s recoverability criterion. Schimmel and Atlas [43] measured the effective modulation transfer function to evaluate filter performance, and showed nearly 25 dB improvement for coherent compared to Hilbert. Clark and Atlas [7, 44] proved the necessary and sufficient conditions for recoverability by way of an operator algebra. In coherent systems, recoverability was ultimately limited by the rate of change in the carrier’s instantaneous frequency (IF). Deliberate smoothing⁷ of the Hilbert-carrier IF is one solution [43], but the spectral center-of-gravity is physically more meaningful [39], while minimizing modulator bandwidth in a certain sense [46, 40] and satisfying idempotency for an operator theory of demodulation [7].

As a result, coherent demodulation aligns with the original meaning of the modulation transfer function. Presumably the 4-Hz modulation frequency has some perceptual significance by itself, which is similar in kind to the 4-Hz modulation at some other carrier frequency. In other words, the information connected to modulation should be frequency-shift (or carrier) invariant [6]. Coherent carriers realize this property by the fact that they are essentially time-varying frequency-shifts.

The difficulty with coherent demodulation is that *any* slowly-varying IF, provided it is bounded by the subband bandwidth, is consistent in the sense of preventing distortion in Ghitza’s recovery experiment. Therefore, carrier frequencies are entirely arbitrary without further specifications. A perceptually relevant possibility is to define carriers as pitch harmonics in tonal signals, as demonstrated for musical transposition [47], talker separation [48, 49], wind-noise reduction [50], and cochlear implant encoding [51]. Another notable example of harmonic-based demodulation of speech is by Kumaresan *et al.* [52, 53]. Brennan *et al.* [54] repeated the Drullman experiment from [25] using pitch-coherent modulation filtering and obtained more pronounced effects than were measured with the Hilbert envelope. These methods bear a resemblance to sinusoidal modeling [55, 56], except with an emphasis on decomposition rather than on parameterization.

⁷This method is related to the phase vocoder by Flanagan *et al.* [45], but quite different due to keeping complex modulators.

In each of the above examples, however, the modulation spectrum is highly sensitive to errors in the carrier frequencies [44]. This raises the issue of defining modulation optimally, to disambiguate one slowly-varying carrier from any other. Sell and Slaney proposed demodulation as a convex optimization problem in [57, 58], and Turner and Sahani [59, 60] used probabilistic inference to solve jointly for stochastic modulators and carriers. Each method constrains the modulator in some way, such as positivity, and bandwidth in the former or time-scale in the latter. Taken together, the methods of coherent, convex, and probabilistic demodulation may be classified as “constrained demodulation” [61]. They do not, however, specify optimal subband placement⁸, except for the limited case of pitch-coherent carriers in speech, particularly the voiced parts. Le Roux *et al.* [63] defined optimal subbands to maximize non-coherent modulations within 20 Hz, but the subbands were not time-varying.

In the interest of optimality, the next section approaches demodulation from a different point of view. What if the carrier is not sinusoidal, as presumed by the coherent methodology, but rather has some random characteristic? What are the sufficient statistics for demodulation? What are the elemental signals for modulation frequency, and when can they be considered invariant to, or independent from, carrier frequency? The answers to these questions require a foundation in classical random processes and system theory.

2.3 Stochastic Demodulation and Linear System Theory

In contrast to the previous section, we now turn to demodulation of a stochastic, or random, signal. We take the carrier to be a noise process, which in turn has fundamental connections to classical spectral representations of random processes. Section 2.3.1 begins with the square-law demodulator, which is essentially the squared Hilbert envelope. Section 2.3.2 generalizes the square-law to include second-order, or bilinear, time-frequency estimators with a system-theoretic underpinning. Finally, Section 2.3.3 compares subband demodulation to purely bilinear demodulation.

⁸This relates to Ville’s criticism of the “arbitrariness” [62] of Gabor’s time-frequency filterbank decomposition [13].

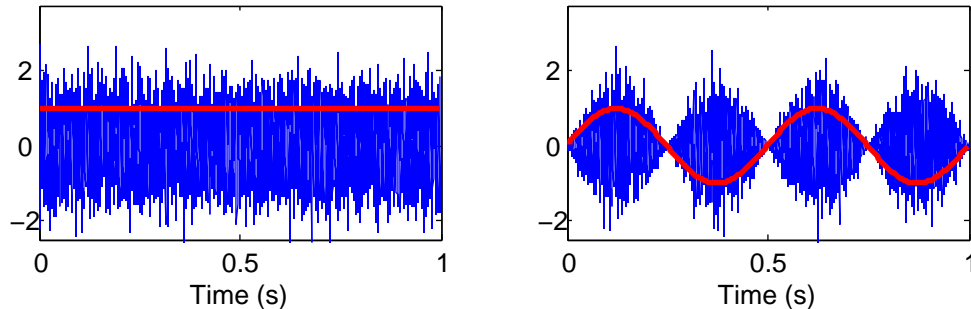


Figure 2.4: Realizations from example AM-WSS processes, with the modulator overlaid in red. Left: WSS $x(t)$ with constant modulator. Right: periodically correlated $x(t)$ with sinusoidal modulator.

2.3.1 Square-Law Demodulation Analysis

A theme for this thesis is prediction in signal-modeling. That is, a modulation signal model should be able to predict the temporal aspects of a random process $x(t)$ based on partial knowledge of the waveform. To illustrate, let $x(t)$ be a zero-mean random process given by the product

$$x(t) = m(t)w(t) \quad (2.9)$$

where the modulator $m(t)$ is a deterministic or random function independent of $w(t)$, and the signal $w(t)$ is a wide-sense stationary (WSS), Gaussian random process. We will refer to this as the AM-WSS model. Examples, with constant and sinusoidal $m(t)$, appear in Fig. 2.4. If $w(t)$ is a white process, then it might seem that there is no basis for predicting $x(0)$ based on neighboring points, such as $x(-1)$ or $x(1)$. Yet referring again to Fig. 2.4, the *intensity* of $x(t)$ in both examples is clearly predictable as a simple function of time. It is also audible, in Fig. 2.4(b), as a beat frequency (when less than 20 Hz). We view $m(t)$ as the model component which imparts predictable, temporal structure to $x(t)$.

The analysis of modulated noise is an important problem in passive sonar, since the noise radiated by underwater propellers is often noise-like yet rhythmically modulated. This phenomenon, known as cavitation, results from the collapse of air bubbles on propeller blades as they encounter nonuniformities in wake inflow and pressure [64]. Estimation of the

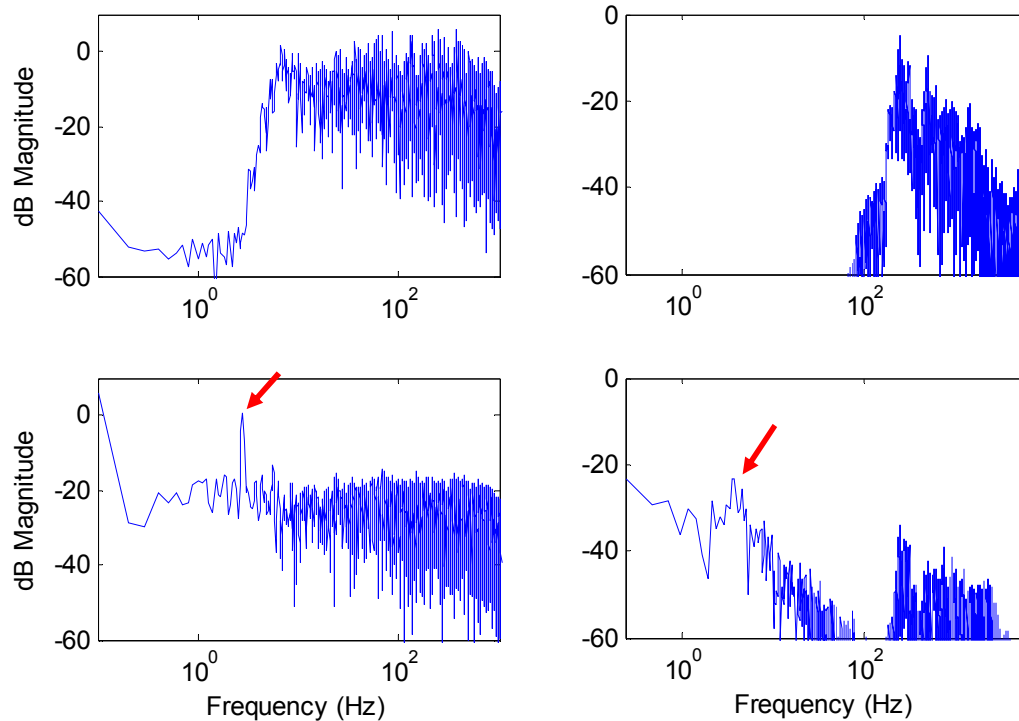


Figure 2.5: Comparison between power spectral estimates (top) and DEMON spectra (bottom), plotted on logarithmic frequency axes. The left panels are from merchant vessel propeller data, with a modulation rate of 2.8 Hz, while the right panels are from female speech, with syllabic rate around 3-4 Hz.

shaft rate, the number of blades, and the modulation waveform are important problems in passive sonar since the vessel type and speed can be inferred from these measurements [65]. Although cavitation noise is audibly rhythmic, traditional Fourier analysis of the waveform often does not reveal harmonics of the beating frequency as one might expect. Indeed, the modulation rate, sometimes as low as 2 Hz, may be below the highpass cutoff of the hydrophone receiver equipment. How then does the waveform support a 2-Hz percept? As in the speech case, rhythmic processes are better treated by a theory of “modulation frequency” separate from the usual acoustic-frequency.

In the naval sonar literature, modulation frequency analysis is known as the so-called⁹

⁹The term DEMON is variously attributed to be an acronym for “detection of envelope modulation on noise” [66] and “demodulated noise” [67]. It is commonly used yet appears to have no single origin in the

DEMON spectrum. In its simplest form the DEMON spectrum is the Fourier transform of $x^2(t)$, or $|z(t)|^2$ for a complex, possibly analytic, signal, and has many useful properties. Lourens and du Preez [65] showed that harmonic analysis of the DEMON spectrum is the maximum likelihood estimator for the propeller rate when $w(t)$ is white and Gaussian and $m(t)$ is periodic. The utility of the DEMON spectrum is immediately apparent in Fig. 2.5, since modulation features become visible in the frequency range associated with propeller rates (which happen to align with syllabic rate in speech). Kummert [66] further developed tracking algorithms for slowly time-varying DEMON spectra.

The DEMON spectrum is analogous to the modulation transfer function discussed in the previous section, owing to the close relationship between the square-law for random processes and the Hilbert envelope. Around the time of Rice and Dugundji, Parzen and Shiren [68] developed the square-law demodulator for modulated noise in 1954. With respect to (2.9), the square-law yields

$$m^2(t) = \sigma_x^2(t) = E\{x^2(t)\} \quad (2.10)$$

where we assume the variance $E\{w^2(t)\} = 1$ without loss of generality. Like the Hilbert envelope, the instantaneous variance $\sigma_x^2(t)$ is necessarily non-negative. To draw a tighter comparison, let us consider the case when $m(t)$ and $w(t)$ are a Bedrosian pair; that is, $m(t)$ is lowpass and the power spectrum of $w(t)$ vanishes below the highest modulation frequency. By (2.4), the complex analytic process $x_a(t) = m(t)w_a(t)$. Again assuming $E\{|w_a(t)|^2\} = 1$ without loss of generality, the square-law yields

$$|m(t)|^2 = \sigma_a^2(t) = E\{|x_a(t)|^2\}. \quad (2.11)$$

The squared Hilbert envelope is $m_H^2(t) = |m(t)|^2|w_a(t)|^2$, where $|w_a(t)|^2$ is unity on average by the preceding assumption. With the appropriate temporal averaging, the square-law is a modulation-filtered version of the Hilbert envelope. Note that, in the deterministic sinusoidal case where $|w_a(t)|^2 = 1$ by definition, the squared Hilbert is identically the square-law output.

Due to the rotary action of an underwater propeller, the sonar literature often assumes $m(t)$ is periodic with fundamental frequency F_0 equal to the propeller shaft rate. Assuming also $w(t)$ is WSS, the multiplicative periodicity of $x(t)$ induces periodicity in the covariance function of $x(t)$. Hence the AM-WSS model falls within the broader category of periodically-correlated (PC) processes, described by Gladyshev [69] in 1963, Dragan [70, 71], and popularized by Gardner (e.g., [72]) as “cyclostationary processes” for communications theory. Variants of the DEMON spectrum for PC processes also appear in cyclic statistical analysis and fourth-order cumulants [73, 74, 75] for noise-modulated signals (where $w(t)$ is usually called the “modulator” or random interference on the desired signal $m(t)$). In the context of maximum-likelihood, [76, 67] derived Crámer-Rao bounds for the estimation of modulation parameters in periodically-modulated noise. Voychishin and Dragan [77] discussed stochastic demodulation of PC processes (although not in the same terms) as the “detection of the period of correlativity” by means of a systematic removal of rhythm with the end result being a WSS process.

The square law and the AM-WSS signal model is insufficient for the class of PC processes. A full treatment requires bifrequency, or more generally bivariate representations [78]. Important examples are spectral coherence between acoustic frequencies [79], and spectral correlation density as a function of Doppler-frequency referenced to subband frequency [72]. Bivariate analysis of propeller noise has taken several forms, notably time-frequency in the model used by Kudryavtsev *et al.* [80], and subband modulation in the adaptive filterbank developed by Shi and Hu [81]. Clark *et al.* [82] proposed a signal model, based on a sum-of-products model, representing a PC process as a sum of narrowband AM-WSS processes. The resulting estimator, called the multiband DEMON spectrum, is defined as

$$\Gamma(\nu, \omega_k) = \frac{1}{\sigma_k^2} \int |x_{k,a}(t)|^2 e^{-j\nu t} dt \quad (2.12)$$

where $x_{k,a}(t)$ is the k th analytic subband signal from $x(t)$, with center frequency ω_k and long-term variance σ_k^2 . The variables ν and ω respectively pertain to modulation and acoustic frequency, and $\Gamma(\nu, \omega_k)$ is essentially a non-coherent modulation spectrum [27, 83]. An example multiband spectrum for merchant ship noise appears in Figure 2.6, showing prominent variation in both frequency dimensions which are lost in the simple AM-WSS

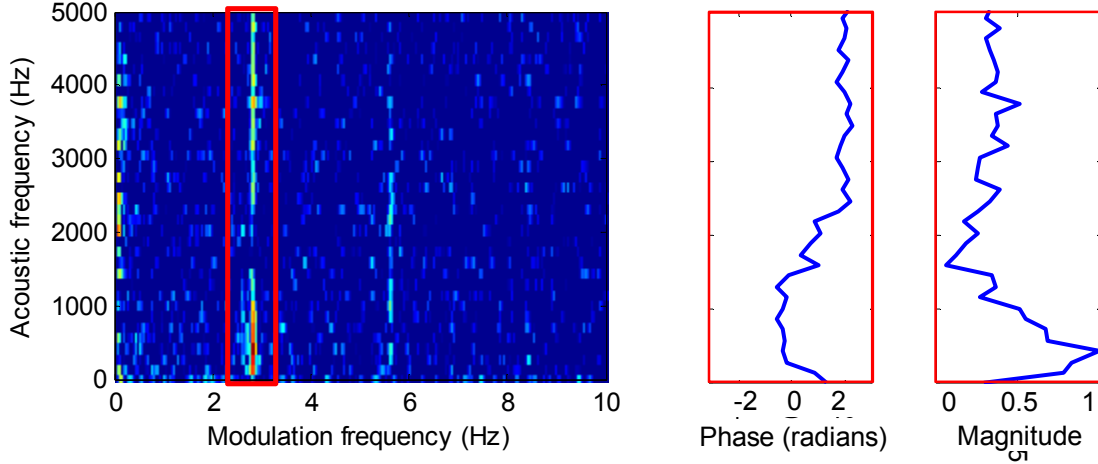


Figure 2.6: Multiband DEMON spectrum (left plot) for merchant propeller noise, showing periodic modulation with fundamental frequency of 2.8 Hz. Modulation is not uniform across acoustic (subband) frequency, as shown in the magnitude and phase plots for modulation frequency = 2.8 Hz.

model¹⁰.

By construction, the multiband DEMON model has some useful properties. It is the maximum-likelihood (ML) modulation-frequency estimator for $x(t)$, when it is AM-WSS and restricted by conditions of orthogonality between subbands and spectral flatness within-band. These conditions are met, to varying degree, by restricting the modulation bandwidth to be much less than the subband carrier bandwidth. At the same time, the subband bandwidth must be narrow enough to resolve the spectral detail of $x(t)$. The ratio of modulator bandwidth to carrier bandwidth is related to the notion of “underspread” system estimation, which we will discuss shortly. For now we note that orthogonality is also required for invertible modulation spectra [44], which provides an operational definition of acoustic and modulation frequency as discussed in Section 2.2.

Like modulation filtering, the multiband DEMON spectrum conceptually involves two

¹⁰In the present context, the simple AM-WSS model assumes that the frequency of amplitude modulation is far less than the bandwidth of the analysis subbands. Otherwise, it is conceivable that the modulation spectrum could have acoustic-frequency dependency, even in the AM-WSS case.

notions of frequency, one related to the subband analysis of $x(t)$, and the other pertaining to the modulation analysis after demodulation. Through Fourier transforms, the significance of both frequency variables derives from the joint nature of time and frequency, as well as two notions of “time.” The theory of two-dimensional analysis of a univariate signal was introduced by Gabor [13] and Ville [62] for time-frequency. In the next subsection, we review two-dimensional analysis of random processes and its connection to spectral representations.

2.3.2 Spectral Representations of Random Processes

We begin by relating the multiband DEMON model to a linear spectral decomposition of a random process. The sum-of-products model is equivalent to

$$x(t) = \sum_{k=1}^K z_k(t)e^{j\omega_k t} = \sum_{k=1}^K m_k(t)w_k(t)e^{j\omega_k t} \quad (2.13)$$

where $z_k(t)$ is the k th complex, lowpass AM-WSS subcomponent with lowpass, stationary noise carrier $w_k(t)$ and modulator $m_k(t)$ independent from $w_k(t)$. If we let the carrier bandwidth shrink to zero, we arrive at a Riemann-Stieltjes integral of the form (see Appendix A)

$$x(t) = \int M(t, \omega) dW(\omega) e^{j\omega t} \quad (2.14)$$

where $dW(\omega)$ is an orthogonal, complex increment process. Orthogonality of the increments guarantees that $E\{x^2(t)\} = \int |M(t, \omega)|^2 d\omega$, which earns $|M(t, \omega)|^2$ the role of a time-varying decomposition of variance or power over frequency.

Equation (2.14) was originally due to Priestley [84], who called $|M(t, \omega)|^2$ the evolutionary spectrum (ES), or time-varying spectrum, of $x(t)$. Defining a class of “semi-stationary” processes for which $M(t, \omega)$ varies slowly in t , Priestley prescribed a windowed Fourier estimator of the form

$$|M(t, \omega)|^2 \sim E \left| \int x(\tau) g(t - \tau) e^{-j\omega\tau} d\tau \right|^2 \quad (2.15)$$

where $g(t)$ is a window small enough to make $x(t)$ appear approximately stationary within its bounds. This is essentially the expected value of the spectrogram of $x(t)$, also called “physical spectrum” by Mark [85]. Taking the Fourier transform with respect to t on the

right side of (2.15) yields the multiband DEMON spectrum in (2.12). This suggests that the sum-of-products model is characteristically suited for semi-stationary (also called quasi-stationary) processes.

A related concept is Silverman's class of "locally stationary" (LS) random processes [86]. Given that $z(t)$ is zero-mean and LS, the autocovariance is of the form

$$E\{z(t)z^*(t + \tau)\} = R_{zz}(t, \tau) = E\{|z(t + \tau/2)|^2\}R_{ww}(\tau). \quad (2.16)$$

With a substitution of variables, Silverman showed that the two-dimensional Fourier transform of (2.16) is

$$\Psi(\nu, \omega) = S_{ww}(\nu + \omega/2)M(\omega) \quad (2.17)$$

which means that $z(t)$ is also LS in the frequency domain. Returning to the AM-WSS model, we note that $x_{k,a}(t) = m_k(t)w_k(t)$ with autocovariance $m_k(t)m_k^*(t + \tau)R_{ww,k}(\tau)$ by virtue of the carrier being WSS. Assuming $R_{ww,k}(\tau)$ vanishes beyond some $\pm\tau_{max}$, the modulator can be considered "slowly-varying" relative to the carrier if $m_k(t)$ is approximately constant in intervals of $2\tau_{max}$. The larger τ_{max} , the more stringent the constraint on $m(t)$. In the frequency domain, the converse is true. The broader the modulator bandwidth, the more stringent the flatness constraint on the carrier spectrum.

The LS criterion is therefore a formalism for the maximum-likelihood conditions on the multiband DEMON spectrum. In principle, the time-frequency LS duality suggests a quotient of the form $\nu_{max}\tau_{max} \simeq 1$, where the exact scale of the product depends on the units involved. This relationship concisely states the case where an AM-WSS process, and by extension a sum of narrowband AM-WSS processes, can also be considered slowly time-varying or semi-stationary. As evidenced by the physical spectrum and multiband DEMON spectrum, $\nu_{max}\tau_{max} \simeq 1$ appears to be a sufficient condition for demodulation of $x(t)$. But is it necessary? In other words, are there other classes of signals which demand new estimators?

A deeper treatment of this question lies within linear system theory. The reason for this is that the statistics of a random process, which are fixed, correspond to the deterministic properties of a generative filter. From Crámer [9], a second-order random process $x(t)$ can

be represented as a time-varying convolution

$$x(t) = \int h(t, t - \tau)w(\tau)d\tau \quad (2.18)$$

where $h(t, \tau)$ is the deterministic kernel of a linear time-varying (LTV) system, and $w(t)$ is a white, stationary noise process. Assuming zero-mean $x(t)$, the autocovariance is $E\{x(t_1)x(t_2)\} = \int h(t_1, \tau)h(t_2, \tau)d\tau$, by which we verify that the generating system determines the statistics of the observed random process.¹¹ When $h(t, \tau) = h(\tau)$ equation (2.18) reduces to the Wold theorem for a stationary process.

It is by this system definition that a spectral representation can exist for random processes, since the Fourier transform will not converge for a stationary process but will converge for $h(t, \tau)$ in τ , and possibly in t . Each of the time variables in $h(t, \tau)$ transforms to a different frequency variable in the “bifrequency system function” formulated by Zadeh [78], which is the generalized transfer function. Kailath [87] defined the various Fourier transforms for one or both time variables and their corresponding sampling theorems. With these tools, we may comprehend the spectrum of $x(t)$ in terms of its generating system. From an auditory point of view, Huggins presciently hypothesized that the ear performs “generalized frequency analysis upon the system function $h(t, \tau)$ rather than upon the sound wave itself, and that results of the analysis constitute an important group of invariants” [88].

A system-theoretic framework provides fundamental tools for estimating parameters of random signals. Notably, linear predictive coding [89, 90] and time-dependent autoregressive modeling [10] are least-squares optimal in terms of inverting (2.18) to obtain a white, stationary residual $w(t)$ from the observed $x(t)$. The system estimate is functional in the sense that its inverse removes temporal and spectral structure from $x(t)$, which can be considered a realization of Voychishin’s and Dragan’s [77] demodulation criterion mentioned in the previous subsection. Related to invertibility is minimum-phase, which plays a role in the STRAIGHT speech modification and synthesis system (see Kawahara *et al.* [91, 92]).

Arguably more fundamental than invertibility is the “underspread” criterion for LTV

¹¹The relationship is not unique, however. For example, given the WSS process $x(t)$, the phase response of the corresponding time-invariant filter is indeterminate. The nonstationary case is similarly undetermined, where the system, in matrix notation, is any of infinite solutions \mathbf{H} to the factorization of the covariance $\mathbf{R}_{xx} = \mathbf{H}\mathbf{H}^H$.

systems. Let us take the Fourier transform of $h(t, \tau)$ in t , to obtain $H(\nu, \tau)$. This is the spreading function, known in the radar literature as the Doppler-delay characteristic of a channel or system [93]. Conceptually, the modulation bandwidth in ν and the impulse width in τ are the extents to which the system will spread, respectively, an input sinusoid over frequency or an input impulse over time. Taking a sampling approach, Kailath [2] defined a measurable system as one for which $H(\nu, \tau)$ is concentrated over a rectangle of area less than unity in the ν, τ plane, or $\nu_{max}\tau_{max} < 1$, when ν and τ are in Hertz and seconds. This relation is the insight which allowed us to hypothesize the LS time-frequency relation earlier. Intuitively, Kailath’s condition requires that the system vary slowly relative to its impulse width. The system therefore appears “frozen” on this time scale, as described heuristically by Zadeh [78]. Conceptually, a compact spreading function means the impulse response of the system does not change appreciably until it dies out first; such a system can be sampled by reading off its impulse response at regular intervals [2].

Intriguingly, Kailath’s condition is not necessary for measureability. Bello [5] generalized the underspread class to include spreading functions concentrated over *any* area less than unity – including arbitrary non-rectangular and non-contiguous regions (see Fig. 2.7). Bello’s underspread class of LTV systems is a powerful concept because it is identical to the class of *measurable* systems. Given input and output pairs for a system, its kernel $h(t, \tau)$ is estimable only when it is underspread by Bello’s condition. For a modern reassessment of Bello’s proof, refer to Pfander and Walnut [94].

Underspread estimation links together the fields of stochastic demodulation and LTV system theory, as argued by Kozek *et al.* [95] and Matz *et al.* [3]. Matz *et al.* showed that Priestley’s semi-stationary assumption is actually Kailathian underspread. This is one of two “strictly underspread” variants, the other being quasi-white rather than quasi-stationary (see Fig. 2.7(b)). Their generalization is the rotated rectangle in Fig. 2.7(b), for which the Weyl spectrum is optimal [3]. In principle, this is sufficient but *not* necessary for measureability as derived by Bello. Generalization beyond the rectangular, origin-concentrated support regions is largely unconsidered in the literature, and it is not clear how to design corresponding estimators in practice [94].

The underspread concept arises also in the bilinear time-frequency literature, such as

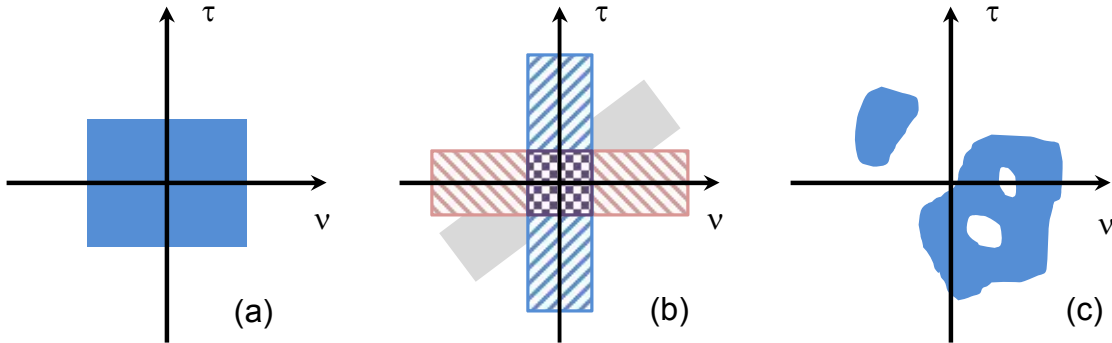


Figure 2.7: Support regions of measurable spreading-functions, in modulation frequency ν and lag τ . In order of increasing generality from left to right: (a) Kailath's bandwidth-lag product [2], (b) rotated-rectangular regions considered by Matz *et al.* [3, 4], and (c) Bello's general underspread class [5]. In all cases, the area of the support region must be less than unity.

Cohen's class of time-frequency estimators. Bilinear, also called quadratic, estimation has deep connections to the second-order nature of Gaussian processes. In the next subsection, we review the parallel development of bilinear estimation and subband-based estimation.

2.3.3 Linear versus Bilinear Time-Frequency Filtering

Given a real signal $x(t)$, its bilinear product is $x(t)x(t - \tau)$ or $x(t)x(t + \tau)$. In expectation, the bilinear product gives the autocovariance $R_{xx}(t, \tau)$ if $x(t)$ is zero-mean¹². The relationship between the autocovariance and the generating system $h(t, \tau)$ is given in the previous subsection. For a zero-mean, Gaussian process $x(t)$, its sufficient statistic is simply $x(t)x(t + \tau)$. We can therefore evaluate an estimator, be it a spectrogram or the multi-band DEMON spectrum, in terms of how faithfully it captures the structure of the signal covariance.

The Rihaczek spectrum [96] is a decomposition of the covariance over time and frequency.

¹²For a non-zero mean signal, the autocovariance is $E\{(x(t) - \mu(t))(x(t + \tau) - \mu(t + \tau))\}$, where $\mu(t) = E\{x(t)\}$.

It is the generally comple-valued, bilinear transform

$$S_{xx}(t, \omega) = \int_{-\infty}^{\infty} E\{x(t)x(t - \tau)\}e^{-j\omega\tau} d\tau. \quad (2.19)$$

Complex numbers¹³ arise in $S_{xx}(t, \omega)$ because $x(t)x(t - \tau)$ is generally non-symmetric in τ when $x(t)$ is nonstationary.

Permitting a Fourier transform on $x(t)$, expression (2.19) is equivalent to the product $E\{x(t)e^{j\omega t}X^*(\omega)\}$. The latter is Rihaczek's original form, which reveals $S_{xx}(t, \omega)$ as the weighted correlation between the signal at time t and frequency ω . Taking another Fourier transform over t results in the bilinear spectral correlation $X(\nu + \omega)X^*(\omega)$ [96, 97]. The related Loève spectrum is [98, 9]

$$\Gamma_{xx}(\omega, \omega') = E\{dX(\omega)dX^*(\omega')\} \quad (2.20)$$

assuming $x(t)$ is harmonizable with increments process $dX(\omega)$ (see Appendix A). For our purposes it suffices to say that $dX(\omega) = X(\omega)d\omega$ when $x(t)$ has finite energy. It can be shown¹⁴ that the multiband DEMON spectrum (2.12) is equal to $\Gamma_{xx}(k\Delta\omega + \nu, k\Delta\omega)$, where the K subbands uniformly sample the ω axis. Alternatively, $\Gamma_{xx}(\omega, \omega')$ is the two-dimensional Fourier transform of $E\{x(t)x(t')\}$. Therefore, variation in t induces off-diagonal structure in $\Gamma_{xx}(\omega, \omega')$. In particular, Hurd [79] showed that a periodically correlated process is spectrally correlated at intervals equal to the fundamental frequency. Refer to Fig. 2.8 to see the Loève spectrum for a periodically modulated process.

In practice, the expected value of a bilinear product can only be estimated by means of averaging, especially when only one realization of the process is available. Cohen's class, reviewed in detail in [99], is a general time-frequency estimator which uses a bilinear kernel for smoothing in both dimensions. Lowpass kernel design usually aims for time-frequency localization, as in the cone kernel [100], which reflects a methodology of Kailathian underspread estimation [4]. The Wigner-ville distribution [62], a prominent member of Cohen's class, is often smoothed in practice (see [101]). Also, Kirsteins *et al.* [102] derived

¹³The notion of complex energy may seem contradictory. Another view is to treat (2.19) as a distribution of correlation rather than energy, as discussed at length by Schreier and Scharf [97].

¹⁴We thank Prof. Louis Scharf for pointing out this connection to the Loève spectrum.

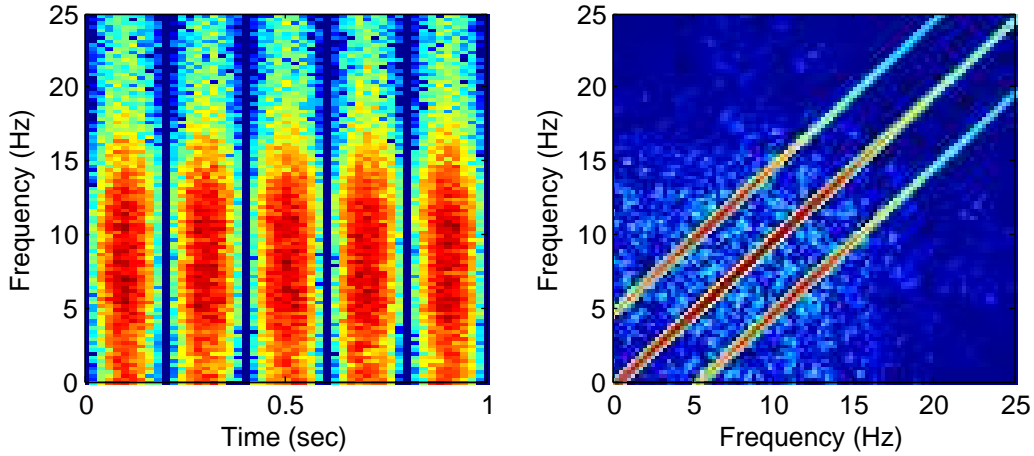


Figure 2.8: Time-frequency (spectrogram, left) and bi-frequency (Loève spectrum, right) displays for a synthetic, periodically correlated AM-WSS process. The colormap dynamic range is 20 dB for both plots.

a maximum-likelihood detector for periodic modulation through cyclic smoothing of the Rihaczek distribution.

In contrast to Cohen’s bilinear class, subband filtering is linear in $x(t)$ while also affording a time-frequency representation. We owe this to Gabor, who in 1946 defined an “information quantum” as a component signal maximally localized¹⁵ in time and frequency [13]. Projecting a real signal $x(t)$ onto a time-frequency lattice yields the short-time Fourier transform (STFT)

$$X(t, \omega) = \int x(\tau)g(t - \tau)e^{-j\omega\tau} d\tau \quad (2.21)$$

where $g(t)$ is the analysis window, and $g(t - \tau)e^{j\omega\tau}$ is one time-frequency basis function. Expression (2.21) is also the convolution with a halfband (approximately analytic) bandpass filter, consistent with subband methods such as Dudley’s vocoders [11, 1] and spectrographic displays [12]. Indeed, the spectrogram is simply $|X(t, \omega)|^2$. If $x(t)$ is a random process, then $X(t, \omega)$ is a complex random process.

One advantage to the subband approach is that it maintains a direct relation to the

¹⁵Kailath’s rectangular underspread criterion is strikingly relevant in light of Gabor’s famous time-frequency uncertainty relation, $\Delta t \Delta f \geq 1/2$.

original signal, which allows modification and synthesis. The subbands can have arbitrary placement and bandwidth, such as on the mel-frequency scale [103], or can be signal-adaptive [63]. Also, linearity allows bandlimits for effective coherent modulation filtering (Section 2.2). Linearity is important when, for example, separating a mixture of audio signals. This is the motivation behind source separation based on complex matrix factorization [8, 104] and Wiener filtering [105, 106].

The STFT has several other useful properties. Under certain constraints, the discrete STFT can be invertible, also known as satisfying “perfect reconstruction” (PR) [107]. This is possible in discrete-time implementation, even when the subbands are partially aliased by downsampling, and simply requires a synthesis window to complement the analysis. See Portnoff [108] and Crochiere [109] for details. Therefore, a discrete, PR STFT $X[n, k]$ completely describes the sampled random process $x[n]$. It follows that the covariance of $X[n, k]$ is a sufficient statistic for $x[n]$ when $x[n]$ is Gaussian. Under certain conditions, the STFT can also be a decorrelating transform akin to a Karhunen-Loève expansion. Matz and Hlawatsch [4] showed this relationship for the case where $x(t)$ is Kailathian underspread in the locally- or semi-stationary sense.

Consistent with the above observations, we treat the discrete STFT as an invertible mapping from a univariate time series $x[n]$ to a multivariate, complex process $X[n, k]$. Stochastic demodulation of the STFT therefore requires tools for analyzing complex random processes. This includes the non-traditional concepts of impropriety and complementary statistics, which we discuss in the next section.

2.4 Complex Random Processes and Statistics

Many of the following results are compactly stated in vector and matrix notation. We will consider a complex random process in discrete time $z[n]$, also represented as an N -dimensional vector \mathbf{z} when of finite length. As always, we assume every signal has zero mean. Traditionally, the second-order statistic of \mathbf{z} is the Hermitian autocovariance matrix $\mathbf{R}_{zz} = E\{\mathbf{z}\mathbf{z}^H\}$, where H denotes Hermitian (conjugate) transpose. A less common, “complementary” second-order statistic has recently been introduced, $\mathbf{C}_{zz} = E\{\mathbf{z}\mathbf{z}^T\}$, where T denotes non-conjugated transpose. The distinction may appear trivial, but \mathbf{C}_{zz} contains

information that is unobtainable from the usual \mathbf{R}_{zz} matrix, and vice versa. Therefore, both \mathbf{C}_{zz} and \mathbf{R}_{zz} are necessary for a complete second-order description of a random process, which is important for maximum-likelihood estimation problems involving complex random variables and processes.

We first require some basic terminology, which is unfortunately not standardized. Picinbono, who is often credited with initiating the recent wave of complex processing, defined a random process \mathbf{z} “circular” if the multidimensional probability distribution of $e^{j\phi}\mathbf{z}$ is invariant to the phase angles ϕ [110]. A “noncircular” variable or process is one whose distribution is variant to phase rotations. For a second-order process, circularity is equivalent to $\mathbf{C}_{zz} = \mathbf{0}$, which Picinbono and Bondon called second-order circular [110, 111]. Neeser and Massey [112] called this case “proper,” referring to the historical assumption that Hermitian correlations alone are sufficient for complex noise in communications systems. Naturally, an “improper” or second-order noncircular process is one for which $\mathbf{C}_{zz} \neq \mathbf{0}$. Since our present interest is in Gaussian processes, we will adopt the proper/improper terminology in the following discussions. We also adopt the term “complementary” in reference to non-Hermitian correlations, which is due to Schreier and Scharf [97].

An improper random process is correlated with its own complex conjugate. We can interpret this phenomenon in terms of the real and imaginary parts of the process, where $\mathbf{z} = \mathbf{i} + j\mathbf{q}$, and \mathbf{i} and \mathbf{q} are both real-valued random processes. The complementary covariance is therefore

$$\mathbf{C}_{zz} = E\{\mathbf{i}\mathbf{i}^T - \mathbf{q}\mathbf{q}^T\} + jE\{\mathbf{q}\mathbf{i}^T + \mathbf{i}\mathbf{q}^T\}. \quad (2.22)$$

The Hermitian covariance is

$$\mathbf{R}_{zz} = E\{\mathbf{i}\mathbf{i}^T + \mathbf{q}\mathbf{q}^T\} + jE\{\mathbf{q}\mathbf{i}^T - \mathbf{i}\mathbf{q}^T\} \quad (2.23)$$

where T and H are equivalent for real processes. The real part of the Hermitian covariance acts as an average power descriptor, whereas \mathbf{C}_{zz} describes structural differences and dependency between the real and imaginary parts. The clearest example of this is when \mathbf{z} is a univariate random variable Z ($N = 1$). Its Hermitian and complementary variances are, respectively

$$\sigma_{zz}^2 = E\{I^2 + Q^2\}, \quad \rho_{zz}^2 = E\{I^2 - Q^2\} + j2E\{IQ\}. \quad (2.24)$$

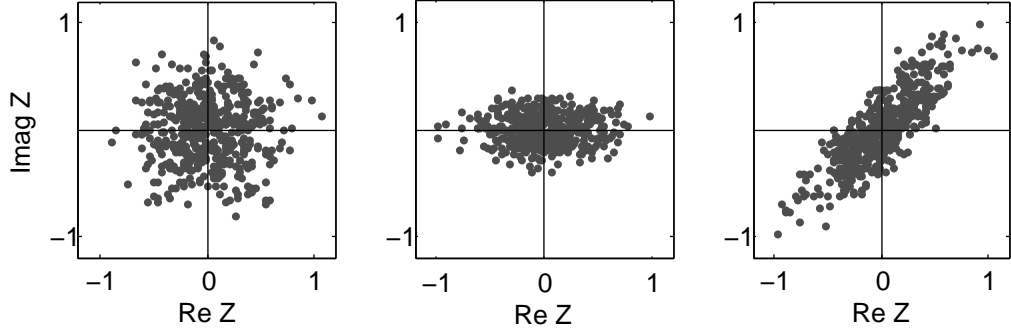


Figure 2.9: Example scattergrams of scalar complex Gaussian random variables. Left: proper (second-order circular) with $\rho_{zz}^2 = 0$. Middle: improper (second-order noncircular) with real and positive ρ_{zz}^2 . Right: improper (second-order noncircular) with complex ρ_{zz}^2 .

It is natural to attribute the Hermitian variance to the “power” of the variable. What about the complementary variance?

Returning to Picinbono’s definition of phase-rotation invariance, the univariate distribution of a proper variable Z is literally circular when viewed from above the complex plane. This is the case for a complex Gaussian random variable as usually defined, with $\rho_{zz}^2 = 0$. An improper Gaussian distribution, as defined by van den Bos [113], is elliptical in the complex plane (refer to Fig. 2.9 for simulated examples). Ollila [114] gave an in-depth analysis of elliptical improper distributions.

It is often useful to deal with complex statistics in a concatenated or “augmented” [97] linear algebra. Given \mathbf{z} , the conjugate-augmented vector is

$$\underline{\mathbf{z}} = \begin{pmatrix} \mathbf{z} \\ \mathbf{z}^* \end{pmatrix} \quad (2.25)$$

where $*$ denotes complex conjugation. The resulting augmented covariance matrix is

$$\mathbf{\Gamma}_{zz} = E\{\underline{\mathbf{z}}\underline{\mathbf{z}}^H\} = \begin{pmatrix} \mathbf{R}_{zz} & \mathbf{C}_{zz} \\ \mathbf{C}_{zz}^* & \mathbf{R}_{zz}^* \end{pmatrix} \quad (2.26)$$

which completely captures the second-order statistics of \mathbf{z} . Van den Bos [113] introduced the above form for the probability distribution of an improper Gaussian vector process.

It is sometimes desirable to estimate the statistics of a complex signal with an assumption of wide-sense stationarity. Returning to signal notation, let $z[n]$ be a random process with Hermitian autocovariance $R_{zz}[n, \tau] = E\{z[n]z^*[n + \tau]\}$ and complementary autocovariance $C_{zz}[n, \tau] = E\{z[n]z[n + \tau]\}$. Schreier and Scharf [97] defined a WSS process¹⁶ as one where $R_{zz}[n, \tau] = R_{zz}[\tau]$ and $C_{zz}[n, \tau] = C_{zz}[\tau]$. A complex WSS random process $z(t)$ therefore possesses two spectral transforms. The Hermitian autocovariance corresponds to the usual power spectral density, which we denote $S_{zz}^{(H)}(\omega)$, whereas the complementary spectrum is $C_{zz}[\tau] \leftrightarrow S_{zz}^{(C)}(\omega)$. For what they called the “modeling problem,” Picinbono and Bondon [111] derived necessary and sufficient conditions guaranteeing $S_{zz}^{(H)}(\omega)$ and $S_{zz}^{(C)}(\omega)$ constitute a valid model for a second-order process. This means the spectra are not arbitrary pairs of functions, since $S_{zz}^{(H)}$ must be non-negative and $S_{zz}^{(C)}(\omega)$ is upper bounded by $S_{zz}^{(H)}(\omega)$.

More interesting is the fact that in general, $z[n]$ can not be represented as the output of a linear filter driven by white noise [111]. The complex analogue of Wold’s theorem instead requires a model of the form

$$z[n] = \sum_p h[n-p]\varepsilon[p] + \sum_p g[n-p]\varepsilon^*[p] \quad (2.27)$$

where $z[n]$ is complex WSS, $h[n]$ and $g[n]$ are shift-invariant deterministic filters, and $\varepsilon[n]$ is white, proper noise. Expression (2.27) was introduced as “conjugate-linear” by Brown and Crane [115] in 1969, but has since come to be known as “widely linear” [116, 117] due to the fact that the system satisfies superposition but not the scaling property of a linear system. We note the special role played by $\varepsilon[n]$. From the perspective of maximum entropy, $\varepsilon[n]$ should be maximally unpredictable. For the Gaussian case, this is equivalent to the least-squares solution, and further requires that $\varepsilon[n]$ be *proper* complex noise [112].

At this point we should remind ourselves that complex numbers are an abstraction fundamentally rooted in observables from the world of real numbers¹⁷. Likewise, impropriety is meaningful only in terms of how it relates back to real signals. The analytic signal

¹⁶We adopt this definition as a clearer alternative to the one put forward by Picinbono and Bondon [111], who only required $R_{zz}[n, \tau] = R_{zz}[\tau]$ for $z[n]$ to be WSS.

¹⁷It was a long time before mathematicians accepted the notion of a negative number, let alone a complex one. We now know that negativity merely indicates direction with respect to a reference point. Complex

is ubiquitous in time-frequency processing, introduced by Gabor [13] in order to define measures of temporal and spectral concentration. In time-frequency estimation it is common to use the analytic signal $x_a(t)$ for Cohen’s class, yet only with the Hermitian bilinear product $x_a(t)x_a^*(t + \tau)$. This would be sufficient if $x_a(t)$ were proper. However, Picinbono and Bondon [111] showed that a proper analytic signal is guaranteed only when $x(t)$ is real WSS¹⁸. In general, a nonstationary $x(t)$ might be improper in $x_a(t)$, in which case Cohen’s class is not sufficient. Schreier and Scharf [97] remedied this shortcoming by introducing the complementary Rihaczek distribution, obtained from the non-conjugated bilinear product $x_a(t)x_a(t + \tau)$.

The analytic signal will be our main concern throughout the rest of this thesis. Empirical studies, however, are surprisingly scant throughout the literature. Rivet *et al.* [119] found evidence of impropriety in STFT subbands from speech, but it is still not clear how such impropriety arises. By contrast, theoretical results and properties are rather well developed, which are usually motivated by communication systems employing man-made, complex modulated signals. In general, widely-linear estimation of a signal embedded in proper noise has up to 3-dB gain over purely linear (i.e., proper) processing [120]. This result was first discovered by Brown and Crane [115], who remarked that the improvement requires “phase-lock” with the complex signal carrier, which is why Schreier *et al.* [120] later referred to widely linear estimation as “coherent.” Further theoretical development of signal estimation in noise appears in Navarro-Moreno *et al.* [121]. For examples of widely linear detection in the context of communications signals, refer also to [122, 123].

It is worth mentioning that complementary statistics are used in other complex processing problems not involving the analytic signal. Namely, directional data such as wind velocity can be treated in a complex statistical framework [124]. Complex blind source ex-

numbers are the planar generalization of a number with magnitude and direction. Although j is widely called “imaginary,” the term was originally a pejorative. A better term is perhaps the “lateral” part of a number, to indicate the inherently geometric nature of the complex plane. It is this lateral generality which makes complex numbers essential for factoring real, constant coefficient polynomials, as well as trigonometric polynomials via the Fourier transform. In the end the lateral component collapses back onto the real line due to conjugate symmetry, as if the real world is the shadow of a symmetrically-expanded superspace. For an illuminating history of complex numbers, refer to Nahin [118].

¹⁸However, Dugundji was perhaps the first to show that the real and imaginary parts of the analytic signal are uncorrelated when the original process is WSS [14].

traction has also been used to remove eye-movement artifact from electroencephalographic (EEG) recordings, where the real and imaginary parts of complex signals correspond to readings from the left and right brain hemispheres [125].

2.5 Summary and Problem Statement

We are now ready to summarize the main ideas from previous sections and formulate a problem statement for the remainder of this thesis.

In Section 2.2 we introduced coherent subband demodulation as a new alternative to the conventional Hilbert envelope for speech analysis. The coherent methodology uses a functional modulator definition to allow modulation-frequency analysis, modification, and synthesis. A key observation is that the resulting “envelope” is generally complex-valued, which dispels the notion of the envelope as a curve sitting on top of a signal. Instead, the envelope satisfies operational criteria such as preservation of bandwidth in modulation filters. This is essential to the precise definition of modulation frequency as separate from acoustic frequency.

The coherent framework, however, is still underdetermined. Bandwidth is preserved by any carrier with sufficiently bounded instantaneous frequency. How then might we formulate the optimal signal modulator? Considering the second-order statistics of a signal, Section 2.3 framed optimal estimation in terms of the underspread criterion for LTV system. An unresolved issue is the estimation of Bello’s generalized class of underspread signals, which by exclusion of the slowly-varying subtype, necessitates new demodulation estimators.

It is not immediately clear that Bello’s class may be accessible through subband demodulation. However, the expanding field of complementary statistics, briefly surveyed in Section 2.4, suggests a new class of complex subband estimators which embody coherent estimation of carrier phase. Is complex demodulation possible in a maximum-likelihood sense? What is the corresponding underspread interpretation?

The aim of this dissertation is to develop a theory of maximum-likelihood complex demodulation for subbands, where the subbands derive from short-time Fourier decomposition of a real-valued random process. Pertinent to this goal are the following technical questions:

- How does subband propriety correspond to the physical characteristics and spectral properties of real-valued signals?
- Does propriety provide a phase reference for optimal detection of complex modulators and carriers?
- What are the necessary and sufficient conditions for a real process to have proper/improper subbands?

Chapter 3

**SPECTRAL IMPROPRIETY OF
MODULATED RANDOM PROCESSES**

3.1 Introduction

Many signals are usefully characterized as a sum of interacting frequencies. Examples are overtones in voiced speech, rhythms of brains waves, or subbands in machine noise. Such signals are non-stationary in the sense that temporal information, like syllable rate in speech or modulation in propeller noise, arises from correlativity in the frequency domain. In other words, nonstationarity is not always evident in the frequency amplitudes themselves, but rather in the temporal co-variation between different frequencies.

To second order, temporal nonstationarity of a real process manifests as correlation in the frequency domain. At frequency ω , the Fourier amplitude $X(\omega)$ is a complex random variable with two distinct second-order statistics: the real-valued Hermitian variance $E\{|X(\omega)|^2\}$, and the lesser-known, complex complementary variance $E\{X^2(\omega)\}$. If the latter is equal to zero, then $X(\omega)$ is proper [112], or second-order circular [110]. Otherwise, it is improper or non-circular. In the communications literature, it has been shown that optimal detection requires the use of all second-order statistics, Hermitian and complementary (e.g., [123]). It is also increasingly common to use complementary statistics for data which is bivariate and naturally represented by complex numbers, such as fMRI analysis [126].

Impropriety in the spectrum and analytical signal calculated from real, univariate signals, like speech or machine noise, has comparatively little coverage in the literature. Speech can be empirically improper in the frequency domain [119], but impropriety is poorly understood in terms of its relation to physical properties of real-world signals. A necessary, but not sufficient condition for impropriety is that the real-valued random process be at least nonstationary [97].

It would seem that impropriety is related to the phase characteristics of a signal. Under

circumstances described later in this chapter, an improper Fourier coefficient is elliptically distributed in the complex plane. The angle of the ellipse is the principal phase of the coefficient. This suggests that spectral impropriety corresponds to deterministic-like, relative timing of sinusoidal components. Phase information becomes important in the analysis of signals which have regular timing, but are not precisely periodic. This concept receives more attention in future chapters.

In this chapter, we systematically relate impropriety to physically meaningful signal model components. Section 3.2 defines essential concepts, and Section 3.3 develops estimation tools which connect impropriety to modulation frequency. Our main result appears in Section 3.4. Section 3.5 gives a broader view of the main result in the context of linear time-varying system theory. Finally, we conclude in Section 3.6.

3.2 Definition of Spectral Impropriety

Let $x(t)$ be a real-valued, zero-mean, univariate random process. We assume $x(t)$ is Gaussian, so wide-sense stationary (WSS) is equivalent to stationary. Our definition of impropriety occurs primarily in the frequency domain. The frequency spectrum¹ of $x(t)$ is defined by

$$X(\omega) = \int x(t)e^{-j\omega t} dt \quad (3.1)$$

where $X(\omega)$ is a complex random process with zero mean. The spectrum is conjugate symmetric, or $X(\omega) = X^*(-\omega)$, which means $x(t)$ is characterized by its positive frequencies only. The analytic signal [13, 14, 97], defined as

$$x_a(t) = \frac{1}{\pi} \int_0^\infty X(\omega)e^{j\omega t} d\omega \quad (3.2)$$

is a complex, time-domain random process in one-to-one correspondence with $x(t)$.

Definition 3.2.1. The process $x(t)$ is *spectrally improper* if and only if the complementary spectral correlation $E\{X(\omega)X(\omega')\} \neq 0$ for some $0 < \omega, \omega' < \infty$.

¹We use the Fourier transform with the understanding that $x(t)$ has finite energy. Refer to Appendix A for a more detailed treatment of the theory of harmonizable processes.

When $\omega = \omega'$, the covariance simply reduces to the complementary variance $E\{X^2(\omega)\}$. We will refer to this as the complementary spectrum

$$S_{xx}^{(C)}(\omega) = E\{X^2(\omega)\} \quad (3.3)$$

which is the counterpart of the more familiar Hermitian variance, or power spectrum

$$S_{xx}^{(H)}(\omega) = E\{|X(\omega)|^2\}. \quad (3.4)$$

Unlike the complementary spectrum, the power spectrum always exists, and is simply the Fourier transform of the autocovariance $R_{xx}(t, \tau)$ in τ and averaged over t .

The analytic signal itself is a time-domain indicator of spectral impropriety. Its complementary autocovariance is

$$E\{x_a(t)x_a(t')\} = 4 \int_0^\infty \int_0^\infty E\{X(\omega)X(\omega')\}e^{j(\omega t - \omega' t')}d\omega d\omega' \quad (3.5)$$

which is the two-dimensional Fourier transform of the spectral correlation quadrant defined over $\omega, \omega' \geq 0$. An improper analytic signal is therefore a sufficient and necessary condition for spectral impropriety [97].

The purpose of this chapter is to understand the properties of $x(t)$ which correspond to spectral impropriety. That is, we wish to use spectral impropriety as an indicator for a certain class of signals. First, we can rule out all WSS processes. If $x(t)$ is WSS, then it is proper in both its analytic signal $x_a(t)$ [111, 14] and its spectrum $X(\omega)$ [97]. For clarity, we state this established result as the following theorem.

Theorem 3.2.2. *Let $x(t)$ be a wide-sense stationary random process. Then $E\{X(\omega)X(\omega')\} = 0$ for all $0 < \omega, \omega' < \infty$.*

As discussed in Chapter 2, this provides the one-way implication of “WSS \rightarrow spectrally proper.” The contrapositive is simply “spectrally improper \rightarrow nonstationary.” It remains to be seen exactly what *kind* of nonstationarity induces impropriety. As seen next, modulation frequency has a clear and systematic connection to spectral impropriety.

3.3 Modulation Frequency and Spectral Impropriety

Modulation frequency, and the associated modulation transfer function, was discussed thoroughly in Chapter 2. Briefly, modulation frequency is a useful predictor for the intelligibility of speech [21, 25] and for classifying underwater propeller noise [82], to name a few examples. In the present context, we formalize modulation frequency in terms of the AM-WSS signal product model (introduced in Section 2.3.1). The model is a type of periodically correlated (PC) [79], or cyclostationary [72], random process. We show that spectral impropriety has a clear and instructive relation with the modulator of a PC process. A similar point was made obliquely in [127], but we take a more deliberate and generative approach here.

3.3.1 Amplitude Modulation Signal Model

Suppose $x(t)$ is a continuous-time AM-WSS process defined by the point-by-point multiplication

$$x(t) = m(t) \cdot w(t) \quad (3.6)$$

where all components are real-valued, $w(t)$ is stationary white noise normalized to unit variance, and $m(t)$ is a periodic modulator function. We verify that $x(t)$ is PC by inspecting its autocovariance,

$$E\{x(t)x(t + \tau)\} = R_{xx}(t, \tau) = m^2(t)\delta(\tau) \quad (3.7)$$

which is itself periodic in t with the same period as $m(t)$. We assume the excitation noise is normalized so that its autocovariance $R_{ww}(\tau) = \delta(\tau)$. Let us denote the period T and the fundamental frequency $1/T = f_0 = \omega_0/2\pi$.

Recall that $w(t)$ is stationary and hence spectrally proper, which means impropriety in $x(t)$ results from $m(t)$ alone. We find $X(\omega)$ as

$$X(\omega) = \frac{1}{2\pi} \int M(\omega - u)W(u)du \quad (3.8)$$

which is simply the time-frequency dual to (3.6). Substituting this into (3.3), the complementary spectrum is

$$S_{xx}^{(C)}(\omega) = \frac{1}{4\pi^2} \iint M(\omega - u)M(\omega - v)E\{W(u)W(v)\}du dv. \quad (3.9)$$

Since $w(t)$ is real and stationary, $W(\omega)$ is proper and conjugate-symmetric. Therefore,

$$E\{W(u)W(v)\} = 2\pi\delta(u+v) \quad (3.10)$$

where $\delta(u)$ is the Dirac delta function, which follows from the fact that $E\{W(u)W(v)\}$ is nonzero only when $W(v) = W^*(u)$, or when $v = -u$. Integrating out the $\delta(u+v)$ in (3.9) yields

$$\begin{aligned} S_{xx}^{(C)}(\omega) &= \frac{1}{2\pi} \int M(\omega-u)M(\omega+u) du \\ &= \frac{1}{2\pi} \int M(2\omega-u)M(u) du \end{aligned} \quad (3.11)$$

or equivalently,

$$S_{xx}^{(C)}(\omega) = \int m^2(t)e^{-j2\omega t} dt. \quad (3.12)$$

This is an important result, which ties together spectral impropriety and the modulation spectrum for the AM-WSS signal model. From (3.12) we observe that $x(t)$ is improper at frequencies corresponding to the halved harmonic frequencies of $m^2(t)$. The fundamental frequency of impropriety is thus $\omega_0/2$. The reader can verify this fact from the simulation results in the next subsection, wherein we propose a practical estimator for $S_{xx}^{(C)}(\omega)$.

3.3.2 Cyclic Demodulation Estimator

Although $x(t)$ conceptually represents an ensemble, we assume that only one realization of the process is available for analysis. That is, we have a single time-domain waveform, and a single corresponding frequency-domain waveform $X(\omega)$. How then can we compute expected values for spectral impropriety as required by Definition 3.2.1? Unlike WSS processes, ergodicity does not immediately apply. However, the periodicity of our AM-WSS process creates a virtual ensemble of T -length realizations arrayed sequentially over time.

Our estimator is therefore based on synchronous time-frequency filtering. The complex, short-time Fourier transform (STFT) [13, 108, 109] is the multivariate random process given by

$$X[n, \omega] = \int x(\tau)g(nT - \tau)e^{-j\omega\tau} d\tau \quad (3.13)$$

where n is an integer and $g(t)$ is a lowpass window function of finite duration. We recognize (3.13) as an array of subband signals frequency-shifted to baseband and sampled discretely in time. Our estimators for the Hermitian and complementary spectra are the *cyclic* time averages

$$\hat{S}_{xx}^{(C)}(\omega) = \frac{1}{2N+1} \sum_{n=-N}^N X^2[n, \omega], \quad \hat{S}_{xx}^{(H)}(\omega) = \frac{1}{2N+1} \sum_{n=-N}^N |X[n, \omega]|^2. \quad (3.14)$$

The Hermitian estimator is essentially a Welch power spectral estimate [128], but the complementary estimator is our new contribution.

Under mild conditions, $\hat{S}_{xx}^{(C)}(\omega)$ is an unbiased and consistent estimator for the complementary spectrum $S_{xx}^{(C)}(\omega)$. This follows from two observations.

1) *Convergence of the time averages:*

Holding ω fixed, $X[n, \omega]$ is a stationary subband signal shifted to baseband. We prove this in Appendix B for any PC white process and for a Kailathian underspread process. The result holds as long as the STFT sampling period T is the true modulator period. Thus, the true subband variances $E\{X^2[n, \omega]\} = \rho_X^2(\omega)$ and $E\{|X[n, \omega]|^2\} = \sigma_X^2(\omega)$ are constants for each subband. We need only assume the subbands are second-order ergodic for the time averages in (3.14) converge, as $N \rightarrow \infty$, to their respective expected values of $\rho_X^2(\omega)$ and $\sigma_X^2(\omega)$.

2) *The subband complementary variance equals the process complementary variance:*

Substituting (3.13) into $E\{X^2[n, \omega]\}$, we find

$$\begin{aligned} \rho_X^2(\omega) = E\{X^2[n, \omega]\} &= \int m^2(\tau) g^2(nT - \tau) e^{-j2\omega\tau} d\tau \\ &= \int m^2(\tau) g^2(-\tau) e^{-j2\omega\tau} d\tau \end{aligned} \quad (3.15)$$

which follows from the whiteness of $w(t)$, and is indeed constant in n due to the periodicity relation $m(\tau + nT) = m(\tau)$. This is a windowed version of (3.12), which means

$$\rho_X^2(\omega) = \frac{1}{2\pi} \int S_{xx}^{(C)}(\omega + u) \tilde{G}(u) du \quad (3.16)$$

where $\tilde{G}(\omega)$ is the Fourier transform of $g^2(t)$. Since $S_{xx}^{(C)}(\omega)$ is a Fourier series with spacing of $\omega_0/2$, we have

$$\rho_X^2(k\omega_0/2) = S_{xx}^{(C)}(k\omega_0/2) \quad (3.17)$$

for integer k , when $g(t)$ is rectangular² of length $2T$ and unit energy (i.e., $\int g^2(t)dt = \tilde{G}(0) = 1$).

By transitivity, observations 1) and 2) combine such that

$$\lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N X^2[n, k\omega_0/2] = S_{xx}^{(C)}(k\omega_0/2) \quad (3.18)$$

which is the main result of this section. In words, the cyclic estimate of subband complementary variance converges to the complementary spectrum of the AM-WSS process, for subbands located at integer multiples of half the modulator fundamental frequency.

We prove the stationarity of $X[n, \omega]$ in Appendix B. Here, it is more important to examine $X[n, \omega]$ as a coherent demodulation of the PC process $x(t)$. In principle, T is the value which systematically removes nonstationarity from $x(t)$ through synchronized sampling in $X[n, \omega]$. In a different context, Voychishin and Dragan [77] first proposed the removal of periodicity as an estimation criterion, and we called this “stochastic demodulation” in Section 2.3.1. Taking this idea further, we suggest that the ideal demodulator produces stationary improper subbands which are also white. This eliminates extraneous correlations and guarantees that $\sigma_X^2(\omega)$ and $\rho_X^2(\omega)$ are sufficient statistics for $x(t)$. (We discuss correlative properties of $X[n, \omega]$ in the next subsection.)

If the subbands are white, then we can plot a histogram for each subband and observe impropriety in terms of elliptical distributions. Each sample corresponds to a point in the complex plane, as seen in the schematic in Figure 3.1. Over time the samples fill in an ellipse defined by the noncircularity quotient

$$\gamma_X(\omega) = \frac{\rho_X^2(\omega)}{\sigma_X^2(\omega)} \quad (3.19)$$

²One might wish to use a tapered window for $g(t)$, for which the equivalence becomes approximate as long as the energy condition is still maintained.

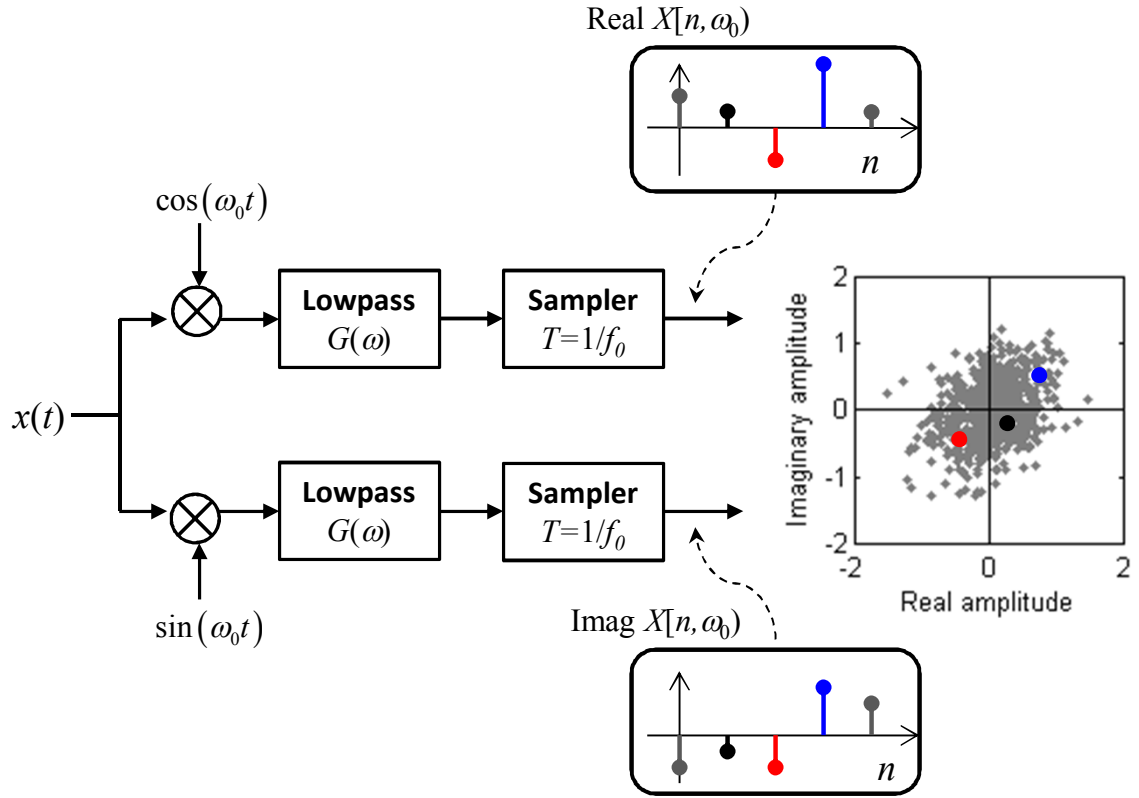


Figure 3.1: Schematic for cyclic, or coherent, demodulation. The basebanded STFT is equivalent to quadrature branches with cosine and sine demodulators followed by lowpass filtering. After cyclic sampling, a subband centered on a harmonic of $\omega_0/2$ shows impropriety in the form of an elliptical histogram.

where the eccentricity $\epsilon_X(\omega) = \sqrt{\gamma_X(\omega)}$ and the angle of orientation $\theta_X(\omega) = 1/2\angle\gamma_X(\omega)$ [114]. Thus, each ellipse is characterized by the respective subband variances alone. Naturally, these functions have special significance for subband frequencies $\omega = k\omega_0/2$ for integer k .

3.3.3 Synthetic Example of a Spectrally Improper Signal

Let us consider the following example. Suppose $w(t)$ is white Gaussian (actually a sequence of iid variables simulated by computer), and the modulator $m(t) = \cos(\omega_0 t + \pi/4) + \cos(2\omega_0 t + 2\pi/3)$. The value of ω_0 is not very important, but it is meaningful for acoustic

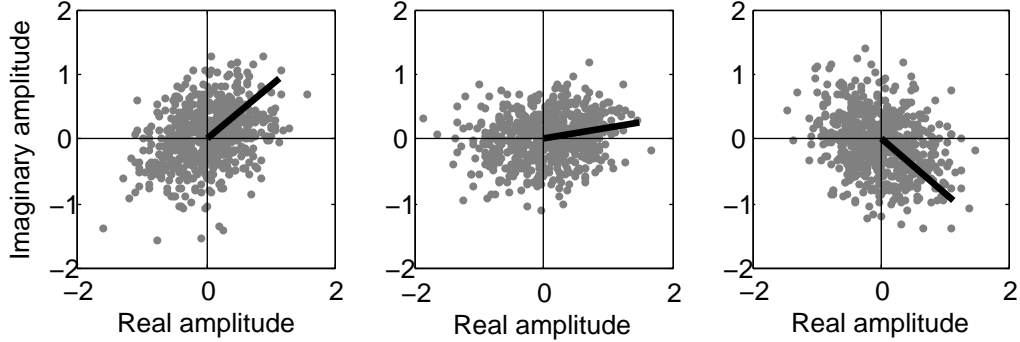


Figure 3.2: Scattergrams for subbands centered on $\omega_0/2$, ω_0 , and $3\omega_0/2$ for a modulator consisting of two cosines. Each black vector indicates the measured orientation of its respective distribution.

subband analysis when it is within the range of human hearing. For the simulation, we chose $\omega_0 = 2\pi 160$. In Figure 3.2, impropriety is apparent in subbands centered on $\omega = \omega_0$, $\omega = 3\omega_0/2$, and $\omega = 2\omega_0$. This confirms our prediction of the fundamental frequency of impropriety being $\omega_0/2$. The black vectors show the estimated orientation θ for each subband, computed from cyclic estimates of complementary and Hermitian variance. As seen, the angles are approximately $\pi/4$, $\pi/24$, and $-\pi/6$ radians, or 45° , 7.5° , and -30° . Since $\rho_X^2(k\omega_0/2) = S_{xx}^{(C)}(k\omega_0/2)$, the angle of a subband distribution for $\omega = k\omega_0/2$ should correspond to the phase of the Fourier coefficients of $m^2(t)$ via (3.12). As predicted, the angles shown match the phase of the first, second, and third harmonics of $m^2(t)$ as defined.

The number of improper subbands depends on ω_0 and the frequency content of $m(t)$. For this second example, $m(t)$ is a periodic triangle wave. In discrete-time implementation, each period of the modulator consists of sample values $[0.5, 1.0, 0.5]$ followed by zeros. With a fundamental frequency of $\omega_0 = 2\pi 1600$, we expect impropriety at subbands spaced every 800 Hz. The complementary spectrum estimate appears in Figure 3.3, with magnitude and phase displayed separately. Since $m(t)$ is a causal, symmetric triangle, the phase plot is simply a linear trend (modulo- 2π) corresponding to the group delay of one sample. The phase at each frequency is two times the orientation angle of the subband distribution at

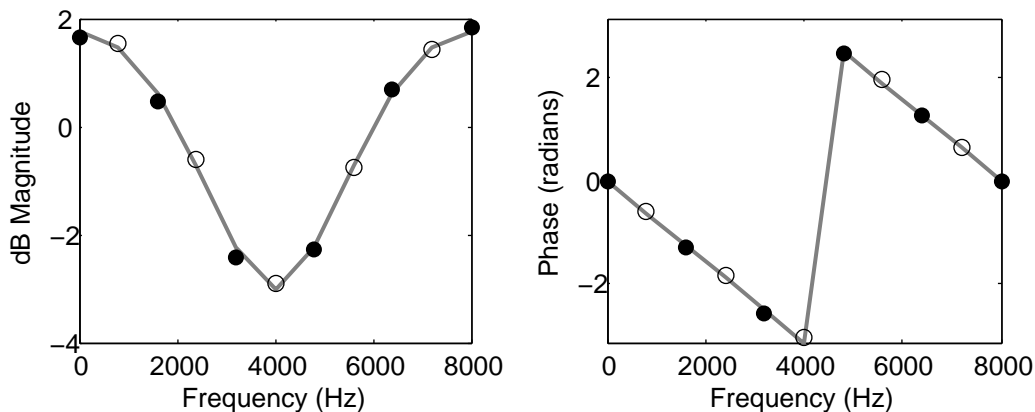


Figure 3.3: Magnitude (left) and phase (right) of the estimated complementary spectrum corresponding to a triangle-wave modulator. Gray curves indicate the known, theoretical value and circles indicate measured values. Dark circles indicate modulator frequencies at multiples of ω_0 and empty circles indicate in-between frequencies.

that frequency.

3.4 The “Fourier Transform” of a Random Process

The previous section demonstrated how to build a simple, improper signal using amplitude modulation of stationary noise. Our present aim is to define a richer signal model where impropriety provides a phase estimate of the generating system. This is like taking the “Fourier transform” of a random process, or more accurately, the Fourier transform of a model for the process. We will find that the method requires the process to be rapidly modulated and hence highly nonstationary, even within an analysis window. This violates the assumption of “quasi-stationarity” which is commonly used but often a convenient fiction in time-frequency processing. For example, speech is highly nonstationary during voiced sounds generated by glottal pulses.

In the course of writing this chapter, the author discovered similar derivations published as early as 1977. In retrospect, this is not surprising considering the fundamental nature of the main result. The previous literature deals with the equalization problem, or equivalently system identification, but makes no explicit mention of complementary statistics. Therefore

our contribution is the connection between spectral impropriety, an empirical observation, and an existing estimation framework from communication theory. With this comes the realization that even a random signal, such as speech, machine noise, or brain waves, may contain recoverable phase information. This section provides a novel perspective for the analysis of improper signals.

We will begin with the signal model, followed by a statement of the main theoretical result and several examples. Then we will derive the main result, and finally discuss connections to the aforementioned existing literature.

3.4.1 Separable PC Signal Model

Let $y(t)$ be the observed random process. We model it as the output of a generative linear system,

$$\begin{aligned} y(t) &= \int h(t, t - \tau)x(\tau)d\tau \\ &= \int h(t, t - \tau)m(\tau)w(\tau) d\tau \end{aligned} \tag{3.20}$$

where, as before, $w(t)$ is white Gaussian noise, $m(t)$ is a periodic modulator with period T , and $x(t) = m(t)w(t)$. This model was briefly discussed as a “Type II separable system” in [87], but is also a subtype of the Crámer-Wold decomposition (See Section 3.5). Our motivation in proposing this model is the source-filter representation for speech, originally proposed by Huggins [88]. A signal synthesis diagram appears in Figure 3.4.

When we define $y(t)$ as the “observed” signal, this means the remaining components – $h(t, \tau)$, $m(t)$, and $w(t)$ – are behind the curtain, so to speak, as strictly theoretical entities. The purpose of (3.20) is to attribute measured spectral impropriety in $y(t)$ to the signal model and thereby infer $h(t, \tau)$ and $m(t)$. Later, we relate this estimation problem to coherent demodulation in Chapter 4. We therefore refer to $y(t)$ as a coherently modulated random process.

Let us restrict our focus to the time-invariant case, or when $h(t, t - \tau) = h(t - \tau)$. We will revisit the time-varying case in Section 3.5 and in detail in Chapter 4. For now, the

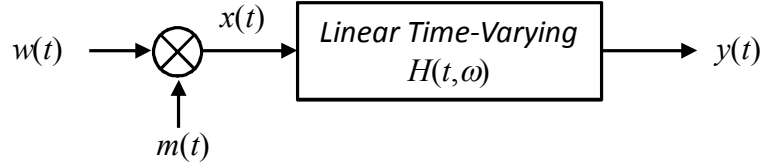


Figure 3.4: Generative signal model for a coherently modulated random process $y(t)$, as the output of an LTV system driven by PC white noise.

signal model reduces to

$$\begin{aligned} y(t) &= \int h(t - \tau)x(\tau)d\tau \\ &= \int h(t - \tau)m(\tau)w(\tau) d\tau \end{aligned} \quad (3.21)$$

which is sufficient for demonstrating the main principles of phase estimation via spectral impropriety. Note that $y(t)$ is PC [69] with period T , but is *not* AM-WSS, because the convolution occurs after the product term. We refer to this model instead as separable PC, which is the intersection of PC processes and coherently modulated processes. The key attribute of $y(t)$ is that the pre-modulator $m(\tau)$ is a reference for detecting both the magnitude and phase of the frequency response of $h(\tau)$.

3.4.2 System Estimation of a Separable PC Process

Given that $y(t)$ is a separable PC process defined by (3.21), we claim the following. The magnitude *and phase* of the LTI system $H(\omega)$ are recoverable from a new transfer function algebra given by

$$S_{yy}^{(C)}(k\omega_0/2) = H^2(k\omega_0/2) \cdot S_{xx}^{(C)}(k\omega_0/2) \quad (3.22)$$

for integer k , indicating the k th harmonic of $\omega_0/2$. Therefore, the samples of $H(\omega)$ are simply the complex square roots of the ratio of complementary spectra. The signal $y(t)$ is “observed,” so we may estimate the left-hand side of the equation as the complementary variances of the synchronous, or coherent, STFT $Y[n, k\omega_0/2]$. This requires prior knowledge of the modulator period T . Since $x(t)$ is a theoretical quantity, its complementary spectrum

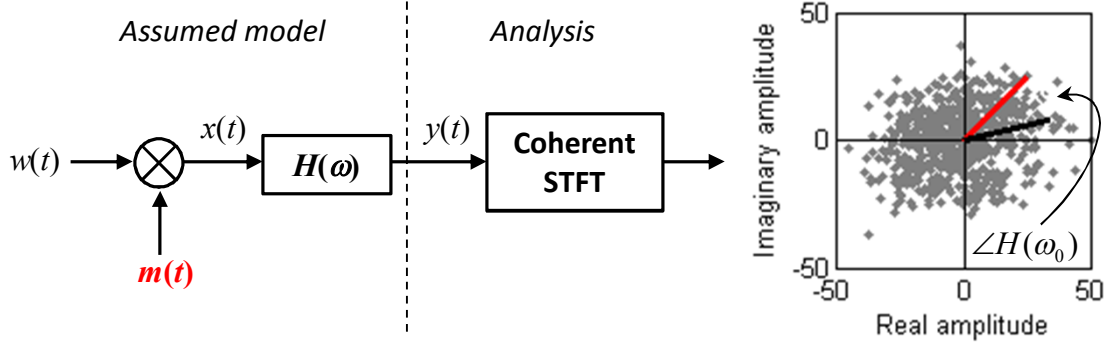


Figure 3.5: Synthesis and analysis chain for the observed process $y(t)$. The red vector corresponds to the reference angle determined by $m^2(t)$. The complex rotation to the measured, black vector is determined by the phase response of $H(\omega)$.

must be jointly estimated or known a priori. From (3.12) we know that, when $x(t)$ is white, $S_{xx}^{(C)}(\omega)$ depends on the modulator only. To be explicit,

$$S_{yy}^{(C)}(k\omega_0/2) = H^2(k\omega_0/2) \cdot \tilde{M}(k\omega_0) \quad (3.23)$$

where $\tilde{M}(\omega)$ is the Fourier transform of $m^2(t)$, which follows from (3.12). Throughout the rest of this section, estimates of $H(\omega)$ are based on prior knowledge of $m^2(t)$. As such, the following results are meant as proofs of concept rather than practical estimators.

The following examples help explain the connection between impropriety and system estimation. In all examples, $w(t)$ is white Gaussian with unit variance.

Example 1

Let $m(t) = \cos(\omega_0 t + \pi/4) + \cos(2\omega_0 t + 2\pi/3)$. The fundamental frequency is $\omega_0 = 2\pi 160$, but the particular value is of minor importance. Note that $x(t) = m(t) \cdot w(t)$ is identical to the earlier example in Figure 3.2. By introducing the LTI component $h(\tau)$, we may observe the effects of $H^2(\omega)$ on the subband distributions. Substituting (3.19) into (3.19), the noncircularity coefficient for $Y[n, \omega]$ is

$$\gamma_Y(k\omega_0/2) = \frac{|S_{yy}^{(C)}(k\omega_0/2)|}{S_{yy}^{(H)}(k\omega_0/2)} = \gamma_X(k\omega_0/2) \cdot e^{j2\angle H(k\omega_0/2)} \quad (3.24)$$

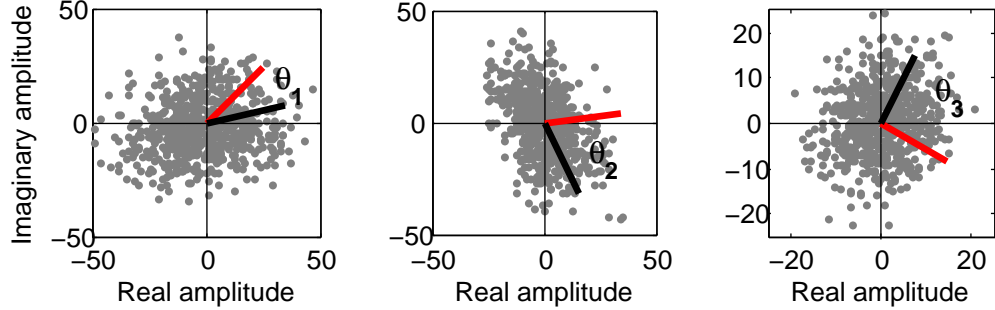


Figure 3.6: Scatter plots for subbands centered on $\omega_0/2$, ω_0 , and $3\omega_0/2$ for Example 1. Each black vector is a measured orientation, and the red vectors are known a priori based on the defined modulator.

where we have used the power spectrum $S_{yy}^{(H)}(k\omega_0/2)$ as derived in (3.38) in the next subsection. Therefore, the subband distributions of $Y[n, k\omega_0/2]$ are rotated by an angle equal to the phase response of $H(k\omega_0/2)$, with no change in eccentricity. This relationship is illustrated conceptually in Figure 3.5.

For the present example, the angular displacement of subband distribution is apparent in Figure 3.6. As the reader may verify, the scatter plots are rotated and scaled versions of those appearing in Figure 3.2, without the system function $H(\omega)$. Thus, Figures 3.2 and 3.6 show the input-output relationship between $X[n, k\omega_0]$ and $Y[n, k\omega_0]$, respectively. Red vectors in the new scatter plots indicate the original orientations of $X[n, k\omega_0]$, and the displacement angles, measured as the difference between the red (known a priori) and the black (measured) vectors, correspond to $\pm\angle H(k\omega_0/2)$ for $k = 2, 3, 4$. The system function $h(\tau)$ is the impulse response of a 20th-order, all-pole filter estimated from merchant ship cavitation noise. This means our synthesized signal $y(t)$ is somewhat realistic in terms of its spectral coloring. The true phase angles are -30.8° , -73.1° , -91.0° which are close to the observed displacements θ_1 , θ_2 , and θ_3 to within a \pm ambiguity.

Example 2

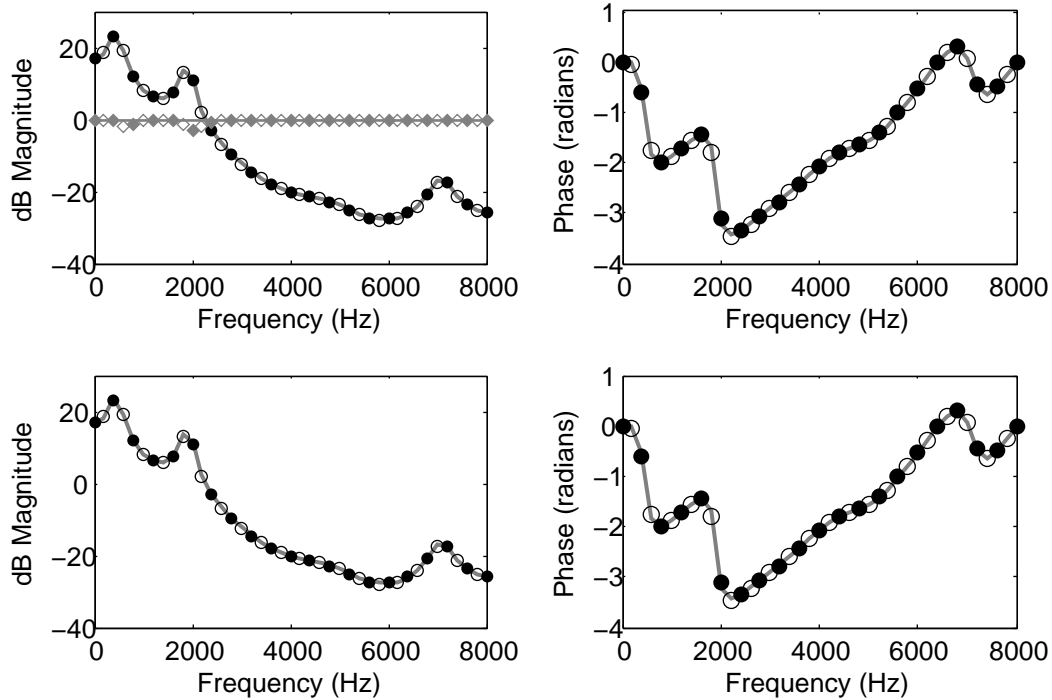


Figure 3.7: Complementary spectrum (top row) and LTI spectrum (bottom) for the speech-like signal in Example 2. Circles denote values measured from subbands. Dark and empty circles correspond to modulator frequencies and in-between frequencies. Curves are ground truth based on the signal synthesis. The light gray curve and circles show the magnitude of the noncircularity coefficient.

In the previous example, only subbands up to $\omega = 2\omega_0$ are improper due to the fact that the modulator contained only two sinusoids. In this example we define a broad-spectrum modulator $m(t) = \sum_i \delta(t - iT)$, the Fourier transform of which is itself a periodic pulse train with intervals of ω_0 . One way to interpret $x(t) = m(t)w(t)$ is as a crude simulation of the vibrations of the human glottis during voiced speech. Accordingly, $h(\tau)$ for this example is a 12th-order all-pole filter estimated from a short utterance of female speech. We choose $\omega_0 = 2\pi 400$, which is somewhat high for human speech (except perhaps for a child), but lower frequencies make for over-crowded plots.

Spectral estimates appear in Figure 3.7. These are based on a synchronous STFT

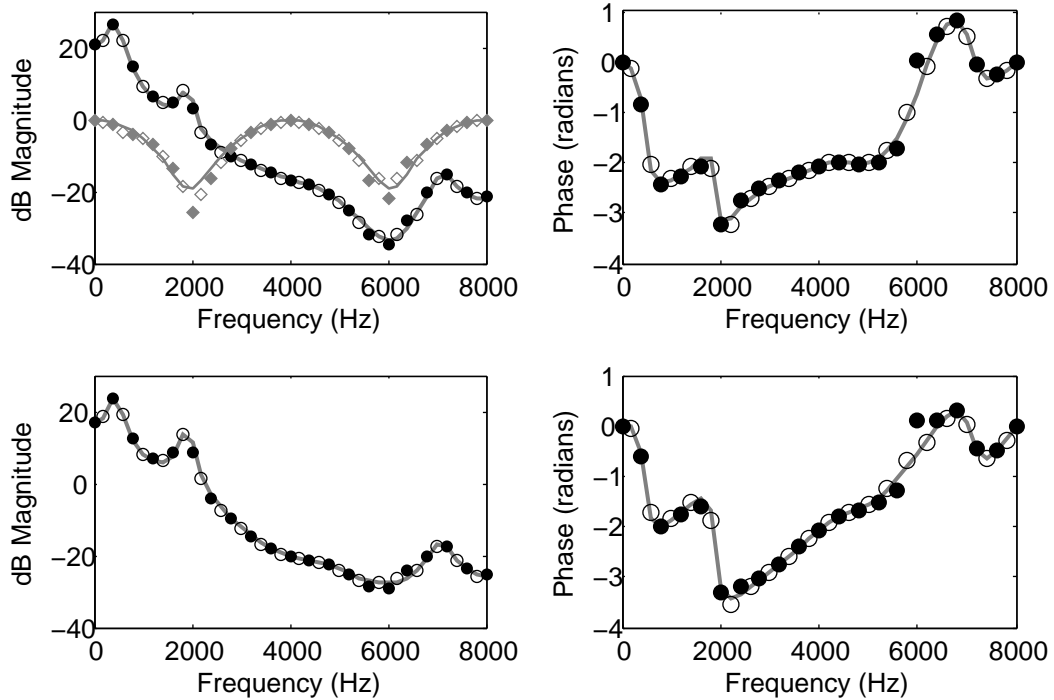


Figure 3.8: Same layout and key as in Figure 3.7, except for the modulator defined in Example 3. This time, the complementary spectrum (top row) must be divided by the squared-modulator spectrum to obtain the LTI estimate (bottom row).

$Y[n, k\omega_0/2]$ using a non-rectangular window with better mitigation of spectral leakage³. In all sub-plots, gray curves denote known, analytic values. Circles indicate values measured from $Y[n, k\omega_0/2]$, either on modulator frequencies (dark) or in-between frequencies (empty), that is, odd multiples of $\omega_0/2$. The top plots show the magnitude and phase of the complementary spectrum, which happens to align exactly with the frequency response of $H(\omega)$, seen in the bottom plots. This is because the modulator spectrum simply has unit amplitude at each harmonic frequency. The light gray plot in the top-left of Figure 3.7 displays the magnitude of noncircularity coefficient for each subband, which one can see is uniformly 1.

³Specifically, a windowed and truncated sinc function with -3 dB bandwidth equal to $\omega_0/2$.

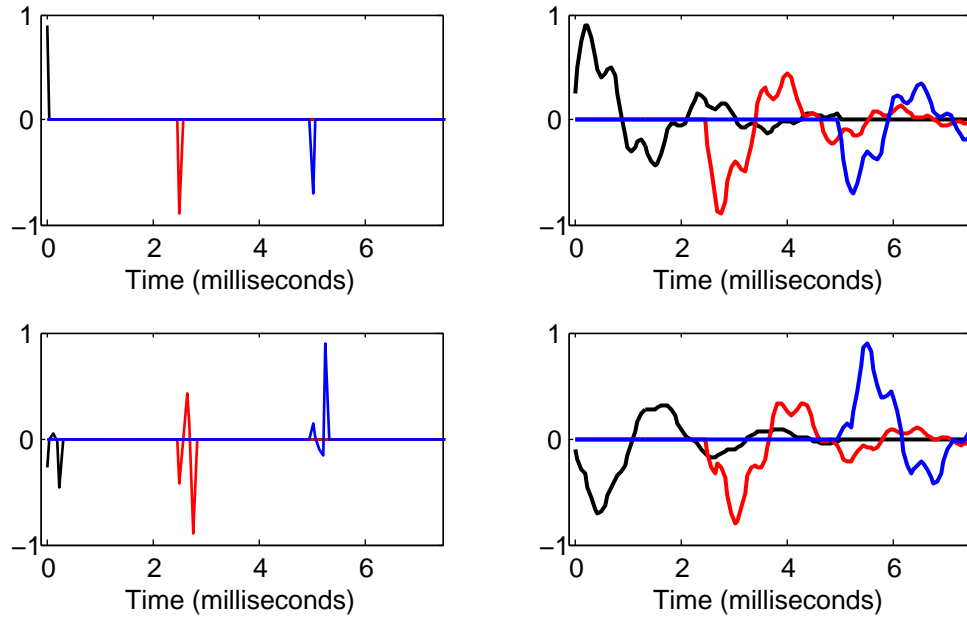


Figure 3.9: Time-domain view of the speech-like Examples 2 (top) and 3 (bottom) in the text. The left-hand subplots show example segments of the PC process $x(t)$. The right-hand subplots show the corresponding segments of $y(t)$, with separate pulse responses shown as separate colors.

Example 3

This example is identical to Example 2 except for a more interesting modulator sequence. With a sampling rate of 16000 Hz, each period of the modulator consists of the sequence $[1, 0, -1, 0.2, 0.5]$ followed by zeros. The more elaborate pulse pattern is evident in Figure 3.8, where the light gray curve traces out the squared-modulator Fourier transform magnitude. Note that the complementary spectrum in the top row is not equivalent to the LTI estimate of the bottom row, which is the quotient of the top row and $\tilde{M}(k\omega_0)$. It is also interesting to note errors occurring at 2000 Hz and 6000 Hz, which correspond to low noncircularity coefficients. This implies that improper subband distributions with low eccentricity are more difficult to estimate than those with high eccentricity. We return to this point in a later subsection.

Examples 2 and 3 also have a time-domain interpretation. When $m(t)$ is a T -periodic impulse train, we can view $y(t)$ as a sequence of overlapping impulse responses with random amplitude. If the impulse response is less than T in duration, then one could simply read off one of the impulse responses and find its Fourier transform to within a multiplicative gain [2]. The situation is more difficult in Example 3, where the input pulse shape has structure which is randomized by the product $m(t) \cdot w(t)$. Therefore, deconvolution is required to estimate the impulse response $h(\tau)$, which must be done stochastically due to the Gaussian driving noise. This is essentially what we have accomplished here, except the deconvolution occurs as a demultiplication of subband signals. Figure 3.9 shows the input $x(t)$ and the system output $y(t)$ for both of the speech-like examples, color-coded to identify the separate pulse responses. Coherent averaging in the STFT is equivalent to synchronous framing of $y(t)$.

From the preceding results, we know that spectral impropriety results from high-frequency modulation in the signal. These modulations provide a coherent phase reference for determining a complex system function, as long as the estimator is synchronized with the modulations. Interestingly, our methodology not only acknowledges but makes constructive use of high-frequency nonstationarity, rather than treating $y(t)$ as “quasi-stationary.” The assumption of stationarity is flatly wrong in speech, for example, which is highly modulated during voiced sounds as a result of glottal pulses.

The transfer function relation (3.22) holds under rather mild conditions. Clearly, the estimate of $H(\omega)$ is limited by the frequency content of $m^2(t)$. Also, we prove in the next subsection that the impulse response $h(\tau)$ must be “short” in the sense of having energy concentrated in an interval of length $2T$. This is what allows $H(\omega)$ to be sampled accurately at intervals of $\omega_0/2$. Finally, a caveat is that the square-root is plus/minus ambiguous, which means the estimated system phase will be $\pm\angle H(\omega)$. The math presented here is incapable of resolving this ambiguity. In all preceding examples, phase unwrapping happens to find the correct solution, but this will not be the case in general.

3.4.3 Derivation of the PC System Estimator

The frequency-domain version of (3.21) is simply the multiplication

$$y(t) = \frac{1}{2\pi} \int H(\omega)X(\omega)e^{j\omega t}d\omega \quad (3.25)$$

where $H(\omega)$ is the frequency response of the system component, or the Fourier transform of $h(\tau)$. Substituting (3.13) into the STFT definition from (3.25) results in

$$Y[n, \omega] = \frac{1}{2\pi} \int G(-u)H(u + \omega)X(u + \omega)e^{junT}du \quad (3.26)$$

which is again synchronized to T , the period of correlation of $x(t)$. The complementary variance of the subband at frequency ω is

$$E\{Y^2[n, \omega]\} = \frac{1}{2\pi} \iint G(-u)G(-v)H(u + \omega)H(v + \omega)\dots \\ E\{X(u + \omega)X(v + \omega)\}e^{j(u+v)nT}du dv. \quad (3.27)$$

The above statement takes some effort to reduce. We require the full complementary covariance of $X(\omega)$, which is, with slight adjustment to (3.11),

$$E\{X(u + \omega)X(v + \omega)\} = \frac{1}{2\pi} \int M(u + \omega - s)M(v + \omega + s)ds. \quad (3.28)$$

One can imagine the integral as two $M(\omega)$'s sliding past each other on the s -axis. Knowing that $M(\omega)$ consists of impulses separated at multiples of ω_0 , it is clear that $E\{X(u + \omega)X(v + \omega)\} = 0$ for $u + v + 2\omega \neq k\omega_0$, where k is an integer. Returning to (3.27), recall that $G(\omega)$ is a lowpass function. Assuming it is bandlimited to *at most* $|\omega| < \omega_0/4$, we simplify the complementary correlation term as

$$E\{X(u + \omega)X(v + \omega)\} = \delta(u + v)S_{xx}^{(C)}(\omega) \quad \text{for } |u|, |v| < \omega_0/4 \quad (3.29)$$

where $S_{xx}^{(C)}(\omega)$ is itself a series of harmonics $\omega_0/2$ apart. Perfect bandlimitation idealizes the filter $g(t)$, but it expedites our derivation. In general, $G(\omega)$ can be concentrated within the required bandwidth, such as when $g(t)$ is rectangular as suggested in Section 3.3.2.

Substituting (3.29) into (3.27) gives an expression which can be integrated in the variable v . This yields

$$E\{Y^2[n, \omega]\} = \frac{1}{2\pi} S_{xx}^{(C)}(\omega) \int G(-u)G(u)H(\omega + u)H(\omega - u)du \quad (3.30)$$

Assuming $g(t)$ is real so $G(-u)G(u) = |G(u)|^2$, we reduce further to

$$E\{Y^2[n, \omega]\} = \frac{1}{2\pi} S_{xx}^{(C)}(\omega) \int |G(u)|^2 H(u + \omega) H(\omega - u) du. \quad (3.31)$$

Due to the periodic nature of $m(t)$, which is vital to our derivation, the only subbands of consequence are those where $\omega = k\omega_0/2$ for integer k . By inspection, the subband complementary variance is also constant over time, so we use the notation $\rho_Y^2[k] = E\{Y^2[n, k\omega_0/2]\}$.

If $H(\omega)$ is relatively flat within a bandwidth of $\omega_0/2$, then we have

$$\begin{aligned} \rho_Y^2[k] &\approx \frac{1}{2\pi} S_{xx}^{(C)}(k\omega_0/2) H^2(k\omega_0/2) \int |G(u)|^2 du \\ &= S_{xx}^{(C)}(k\omega_0/2) H^2(k\omega_0/2) \end{aligned} \quad (3.32)$$

where the integral equals 2π as a result of Parseval's theorem and $g(t)$ having unit energy as prescribed in Section 3.3.2. If $H(\omega)$ is not sufficiently smooth, then $\rho_Y^2[k]$ will instead contain a blurred estimate of $H(k\omega_0/2)$.

From Section 3.3.2 we know that the subband complementary variance equals the process complementary spectrum, or $\rho_Y^2[k] = S_{yy}^{(C)}(k\omega_0/2)$. Therefore,

$$S_{yy}^{(C)}(k\omega_0/2) = H^2(k\omega_0/2) \cdot S_{xx}^{(C)}(k\omega_0/2) \quad (3.33)$$

which is what we set out to prove. In practice, we solve for the system function via

$$H(k\omega_0/2) = \pm \sqrt{\frac{\hat{S}_{yy}^{(C)}(k\omega_0/2)}{\hat{S}_{xx}^{(C)}(k\omega_0/2)}} \quad (3.34)$$

where the hat notation indicates the cyclic STFT estimator for the complementary spectrum. The examples from the previous subsection, however, assume that only $y(t)$ is available for observation. The process $x(t)$ is theoretical, but we assumed that $m(t)$ was known. Denoting the modulator harmonic amplitudes as $M_k = M(k\omega_0)$, it is apparent from (3.11) that the harmonics of $S_{xx}^{(C)}(\omega)$ result from the discrete convolution

$$S_{xx}^{(C)}(k\omega_0/2) = \frac{1}{2\pi} \sum_i M_{k-i} M_i \quad (3.35)$$

from which we derive the system function

$$H(k\omega_0/2) \approx \pm \sqrt{\frac{2\pi \hat{S}_{yy}^{(C)}(k\omega_0/2)}{\sum_i M_{k-i} M_i}}. \quad (3.36)$$

Surprisingly, this result suggests that we can estimate the transfer function of the system on a grid twice as fine as the spacing of harmonics in $m(t)$. This is consistent with the STFT window length of $2T$ used earlier.

It is also interesting to note that $\omega_0/2$ is the smallest frequency resolution possible with complementary subband analysis. Making the window arbitrarily long, e.g., $4T$ or $100T$, affords no extra benefit in estimating the phase of $H(\omega)$. Instead, the fundamental frequency of the modulator defines the ultimate limit of frequency resolution. Conversely, Hermitian power spectral estimation achieves arbitrary frequency resolution provided a long enough window, but will only estimate $|H(\omega)|$. We find the Hermitian subband variance to be

$$E\{|Y[n, \omega]|^2\} = \iint G(-u)G^*(-v)H(u+\omega)H^*(v+\omega) \dots \\ E\{X(u+\omega)X^*(v+\omega)\}e^{j(u-v)nT} du dv. \quad (3.37)$$

After some reduction it becomes

$$E\{|Y[n, \omega]|^2\} = \sigma_Y^2(\omega) = S_{xx}^{(H)}(\omega) \int |G(u)|^2 |H(u+\omega)|^2 du \quad (3.38)$$

which follows from the fact that $E\{X(u+\omega)X^*(v+\omega)\}$ is nonzero for $u-v$ equal to multiples of ω_0 , or 0 within the $u-v$ bounds imposed by $G(u)G^*(-v)$. As seen, the spectral resolution of the estimator depends on the window of analysis, and the phase of $H(\omega)$ is impossible to determine.

We should review the assumptions that led us to this point. First, the STFT is sampled synchronously with the period of correlation T . Second, the driving noise is white. And finally, $H(\omega)$ is smooth such that it appears nearly flat within a bandwidth of $\omega_0/2 = \pi/T$. This is the same as saying the energy of $h(\tau)$ is roughly concentrated over an interval of length $2T$.

As a final observation, it is plausible that the derivations of this section extend also to the general time-varying case, with $h(t, \tau)$. If the system varies slowly enough in t , then the preceding results apply on time intervals for which $h(t, \tau)$ appears “frozen,” to use Zadeh’s term [78]. This is essentially Kailath’s underspread condition on $h(t, \tau)$, but considering the high-frequency modulator $m(t)$, the overall system is decidedly non-Kailathian. Is it

underspread by Bello's generalized criterion? We discuss the limits of complex estimation in the time-varying case in Chapter 4

3.4.4 Connection to System Identification Literature

The results of this section have appeared in prior publications, although in a considerably different context and language. In fact, the signal model given by (3.21) was considered as early as 1977, when Cerrato and Eisenstein [129] investigated the problem of identifying $H(\omega)$ given its output $y(t)$ and knowledge of the statistics of a PC, or cyclostationary, input signal $x(t)$. Others have followed suit, usually motivated by channel equalization in communications or deconvolution in geophysics. In the following we contrast the results of the present dissertation with related works, in chronological order.

Perhaps the closest match is [129] by Cerrato and Eisenstein. Similar to our transfer function relation, their general formula is also a cyclic average,

$$H[k]H^*[l] = \frac{\frac{1}{N} \sum_{n=0}^{N-1} Y[n, k]Y^*[n, l]}{E\{X[k]X^*[l]\}} \quad (3.39)$$

where $Y[n, k]$ and $X[n, k]$ are discrete STFTs of output and input, and $H[k]$ and $X[k]$ are discrete Fourier transforms. Noting that the system of equations are over-determined, they developed a complicated least-squares fit to estimate $H[k]$ separately in its magnitude and phase. Interestingly, the above equation involves conjugate (i.e., complementary) correlations as well as Hermitian. Thus, impropriety is implicit. However, their paper makes no mention of stationary subbands, which is what allows us to associate elliptical distributions to magnitude and phase of the system function.

Unlike Cerrato and Eisenstien, Kormylo and Mendel [130] discussed the excitation product model $x(t) = m(t)w(t)$. They remark on the necessity of synchronizing with major features in $m^2(t)$, which need not be periodic. Broadly related to our approach, their consideration for non-periodic modulators suggests a way to extend the theory of spectral impropriety.

In 1991, Gardner [131] derived the transfer function relation

$$H(\nu) = \sqrt{\frac{S_y^{2\nu}(0)}{S_x^{2\nu}(0)}} \quad (3.40)$$

where $S_y^\nu(\omega)$ is the cyclic spectral correlation density of $y(t)$ [72]. Conceptually, $S_y^\nu(\omega)$ is the correlation between subbands centered on $\omega - \nu/2$ and $\omega + \nu/2$. Gardner's $S_y^{2\nu}(0) = S_{yy}^{(C)}(\nu)$, so the above expression is identical to our (3.22). Gardner's commentary is relevant: "Although it is often said... that second-order autostatistics contain no phase information, this is not true for cyclostationary signals as demonstrated with the new approach proposed in this paper."

Also relevant is the work of Tong, Xu, and Kailath [132, 133], who developed a subspace method for system identification based on oversampling the process $y(t)$. By "oversampling" they refer to higher than the baud rate, or in our terminology, at a rate greater than $1/T$. Our approach is also oversampled in the sense that, for each cycle, we obtain multiple Fourier coefficients corresponding to discrete subbands.

As stated at the introduction of this section, the purpose of the present work is to connect spectral impropriety with the signal model previously studied in different contexts. When researchers and engineers work in the frequency domain, their first guess is perhaps not to assume a model of the form (3.21). Accepting the possibility of spectral impropriety, however, introduces an unusual modeling challenge. In this section we have shown that impropriety can be modeled constructively. Our analysis tools are similar to those in the cited literature, but we buttress them with fundamental concepts such as demodulation and the removal of nonstationarity. Furthermore, subband distribution analysis, as in Figures 3.2 and 3.6, is to our knowledge a novel interpretation of high-frequency modulation.

Nor is spectral impropriety limited to the PC-separable model discussed in here and in [129][130][131][132, 133]. In the following, we discuss more general system properties which can generate impropriety in the frequency domain.

3.5 Linear Time-Varying System Model of Spectral Impropriety

The signal models and estimators of the previous section are formalisms for analyzing spectral impropriety. A more general view arises from correlative analysis of linear time-varying systems. We begin with an alternative analysis of the AM-WSS signal model, which in turn defines the notion of "widely-linear mixing" in the frequency domain as the origin of spectral impropriety.

3.5.1 Impropropriety Caused by Sinusoidal Over-Modulation

Section 3.3 showed that the AM-WSS signal is improper. It was also clear that the impropriety resulted directly from the action of the modulator alone. But why is that? We turn now to the frequency domain for an explanation.

Let $x(t) = m(t)w(t)$, where $w(t)$ is stationary noise. The simplest case is a sinusoidal modulator, $m(t) = r_0 \cos(\omega_0 t + \theta_0)$, where $\omega_0 = 2\pi f_0$ is the modulation frequency. Now, let $w_L(t)$ and $w_H(t)$ be stationary random processes with non-overlapping power spectra, such that $w(t) = w_L(t) + w_H(t)$. Further suppose that the upper and lower cutoff frequency of $w_L(t)$ and $w_H(t)$ is equal to the modulation frequency ω_0 , as shown in the top plot of Figure 3.10.

Given the product of a lowpass signal and a highpass signal with no spectral overlap, the Bedrosian product theorem (Section 2.2, from [15]) states that the analytic signal is equal to the lowpass component times the analytic signal of the highpass component. With respect to $x(t)$, we have

$$x_a(t) = m_a(t) \cdot w_L(t) + m(t) \cdot w_{H,a}(t) \quad (3.41)$$

where subscript- a in every case denotes an analytic signal. The complementary variance of the analytic signal is

$$E\{x_a^2(t)\} = \sigma_L^2 m_a^2(t) \quad (3.42)$$

where σ_L^2 is the variance of $w_L(t)$. The above result follows from the fact that $w_H(t)$ is stationary and its analytic signal $w_{H,a}(t)$ is proper (Theorem 3.2.2). Cross-terms similarly vanish due to orthogonality of $w_L(t)$ and $w_H(t)$.

We conclude that the analytic signal is improper for any modulation frequency $\omega_0 > 0$. Equivalently, $x(t)$ is spectrally improper (see (3.5)). From (3.42), however, we finally see that this impropriety results from the interaction between the modulator and the lowpass component $w_L(t)$. This links impropriety to what we call over-modulation, which is when the frequency of modulation overlaps the spectral content of a multiplicative carrier signal. In the frequency domain, modulation becomes a convolution:

$$X(\omega) = \frac{1}{2\pi} \int [W_L(\omega - s) + W_H(\omega - s)] M(s) ds. \quad (3.43)$$

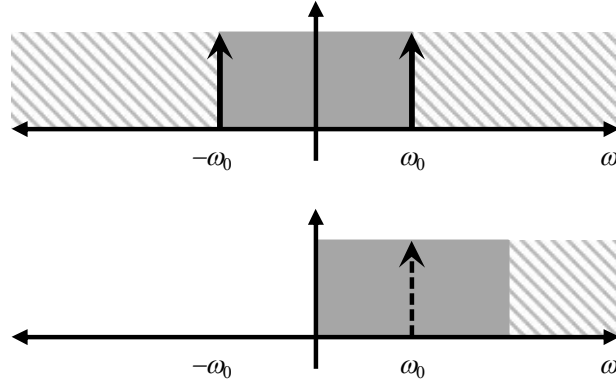


Figure 3.10: Schematic for Bedrosian over-modulation, where the dark shaded and pattern-shaded power spectra correspond to $w_L(t)$ and $w_H(t)$, respectively (top). The bottom plot shows the components of the analytic signal after the frequency shift.

Sinusoidal $m(t)$ corresponds to a simple frequency shift, such that the analytic spectrum is

$$X_a(\omega) = m_0 W_L(\omega - \omega_0) + m_0 W_{H,a}(\omega - \omega_0) + m_0^* W_{H,a}(\omega + \omega_0) \quad (3.44)$$

where $m_0 = 1/2r_0 e^{j\theta_0}$. As seen in the bottom plot of Figure 3.10, the entire spectrum $W_L(\omega)$ is present in $X_a(\omega)$, and since $w_L(t)$ is real, $W_L(\omega - \omega_0)$ is conjugate-symmetric around ω_0 . In other words, the frequency-shift of the amplitude modulation has moved conjugate amplitudes from $W_L(\omega)$ into the positive frequencies of $X_a(\omega)$.

Although $W(\omega)$ is entirely proper, its conjugate redundancy contributes to impropriety through the frequency-shift action resulting from amplitude modulation. In other words, impropriety arises from the mixing of negative- and positive-frequency components. This is just one example of the general class of linear, time-varying systems, discussed next.

3.5.2 Linear in Time is Widely-Linear in Frequency

Here we derive a general characterization of spectral impropriety in a fashion parallel to Schreier and Scharf [97]. Focusing on the analytic signal for bilinear time-frequency estimation, Schreier and Scharf demonstrated the basic need for complementary statistics when the signal of interest is nonstationary. This follows from the simple, yet historically overlooked, observation that complementary statistics are naturally embedded in the Loève

spectrum $\Gamma_{yy}(\omega, \omega') = E\{Y(\omega)Y^*(\omega')\}$. Due to the conjugate symmetry of $Y(\omega)$, we have, for example, $\Gamma_{yy}(\omega, -\omega) = E\{Y^2(\omega)\} = S_{yy}^{(C)}(\omega)$. In the following we show that a linear time-varying (LTV) generative system is in fact widely linear (WL) in the frequency domain. This is a new, system-theoretic perspective with implications discussed in detail in Chapter 5.

In Section 3.4, we defined $y(t)$ as the linear system output

$$y(t) = \int h(t, t - \tau)m(\tau)w(\tau) d\tau. \quad (3.45)$$

This random process also has the Crámer-Wold decomposition

$$y(t) = \int h'_2(t, t - \tau)w(\tau) d\tau \quad (3.46)$$

where $h'_2(t, \tau)$ is the linear system which completely defines the autocovariance of $y(t)$ because the input signal $w(t)$ is white, Gaussian, and stationary. Therefore, (3.45) is the Crámer-Wold subtype where $h'_2(t, \tau)$ is separable in the sense that

$$h'_2(t, \tau) = h(t, \tau)m(t - \tau). \quad (3.47)$$

In general, $h'_2(t, \tau)$ need not satisfy (3.47), but is generally a square-integrable function. The reason we use subscript-2 for this function is to distinguish it from the alternate form [87]

$$h'_1(t_1, t_2) = h'_2(t_2, t_2 - t_1) = h(t_2, t_2 - t_1)m(t_1). \quad (3.48)$$

This is the system function referenced to output time t_2 and input time t_1 , rather than global time t and lag τ . Therefore,

$$y(t_2) = \int h'_1(t_1, t_2)w(t_1)dt_1. \quad (3.49)$$

In the terminology of linear algebra, $h'_1(t_1, t_2)$ is the “matrix” one would left-multiply the “vector” $w(t_1)$ to obtain $y(t_2)$.

Following Zadeh [78], the bifrequency spectrum is

$$H'_1(\omega_1, \omega_2) = \iint h'_1(t_1, t_2)e^{j\omega_1 t_1 - j\omega_2 t_2} dt_1 dt_2 \quad (3.50)$$

which substitutes into (3.49) to yield the frequency-domain expression

$$Y(\omega_2) = \frac{1}{2\pi} \int H'_1(\omega_1, \omega_2) W(\omega_1) d\omega_1. \quad (3.51)$$

Since $w(t)$ is real, it is conjugate symmetric in the frequency domain, or $W(\omega_1) = W^*(-\omega_1)$.

Expression (3.51) is hence equivalent to

$$\begin{aligned} 2\pi Y(\omega_2) &= \int_0^\infty H'_1(\omega_1, \omega_2) W(\omega_1) d\omega_1 + \int_{0^+}^\infty H'_1(-\omega_1, \omega_2) W(-\omega_1) d\omega_1 \\ &= \int_0^\infty H'_1(\omega_1, \omega_2) W(\omega_1) d\omega_1 + \int_{0^+}^\infty H'_1(-\omega_1, \omega_2) W^*(\omega_1) d\omega_1 \end{aligned} \quad (3.52)$$

If we define, for $0 \leq \omega_2 < \infty$,

$$\begin{aligned} A(\omega_1, \omega_2) &= U(\omega_1) H'_1(\omega_1, \omega_2), \\ B(\omega_1, \omega_2) &= [1 - U(-\omega_1)] H(-\omega_1, \omega_2) \end{aligned} \quad (3.53)$$

where $U(\omega)$ is the unit step function, then (3.52) becomes

$$2\pi Y(\omega_2) = \begin{cases} \int A(\omega_1, \omega_2) W_a(\omega_1) d\omega_1 + \int B(\omega_1, \omega_2) W_a^*(\omega_1) d\omega_1 & \omega_2 \geq 0 \\ \int A^*(\omega_1, \omega_2) W_a^*(\omega_1) d\omega_1 + \int B^*(\omega_1, \omega_2) W_a(\omega_1) d\omega_1 & \omega_2 < 0 \end{cases} \quad (3.54)$$

where $W_a(\omega) = U(\omega)W(\omega)$ is the analytic excitation spectrum. The second line is simply the complex-conjugate of the first since $y(t)$ is real and $Y(\omega)$ is also conjugate-symmetric.

Expression (3.54) is the main result of this section, since it reveals that the linear system $h'_1(t, \tau)$ is *widely linear* (WL) in the frequency domain (Refer to Section 2.4 for basic properties of WL systems). Since $W(\omega)$ is entirely proper (Theorem 3.2.2), the impropriety of $Y(\omega)$ results from the widely-linear mixing of complex-conjugate amplitudes from $W(\omega)$. Or equivalently, impropriety results from mixing positive- and negative-frequency spectral elements of a stationary excitation process.

Since $W_a(\omega)$ is white and proper, we have, respectively,

$$\begin{aligned} E\{W_a(\omega)W_a^*(\omega')\} &= \delta(\omega - \omega') \\ E\{W_a(\omega)W_a(\omega')\} &= 0 \end{aligned} \quad (3.55)$$

It follows that the output complementary covariance is, for $\omega_2, \omega'_2 \geq 0$

$$4\pi^2 E\{Y(\omega_2)Y(\omega'_2)\} = \int_0^\infty A(\omega_1, \omega_2)B(\omega_1, \omega'_2)d\omega_1 + \int_0^\infty B(\omega_1, \omega_2)A(\omega_1, \omega'_2)d\omega_1 \quad (3.56)$$

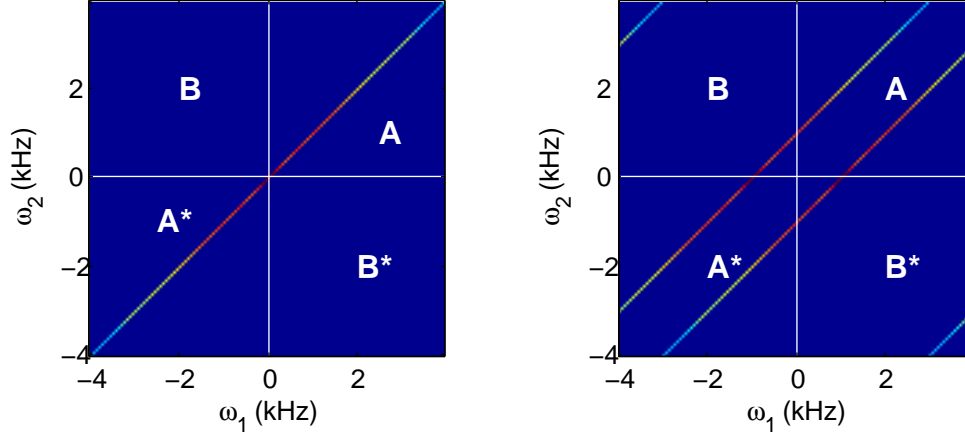


Figure 3.11: Schematic showing A and B quadrants of two bifrequency LTV system functions: time-invariant (left) and modulated at 1000 Hz (right).

and the Hermitian covariance is

$$4\pi^2 E\{Y(\omega_2)Y^*(\omega'_2)\} = \int_0^\infty A(\omega_1, \omega_2)A^*(\omega_1, \omega'_2)d\omega_1 + \int_0^\infty B(\omega_1, \omega_2)B^*(\omega_1, \omega'_2)d\omega_1. \quad (3.57)$$

Together, the above expectations define the Loève spectrum of $y(t)$. Specifically, $\Gamma_{yy}(\omega_2, \omega'_2) = E\{Y(\omega_2)Y^*(\omega'_2)\}$ and $\Gamma_{yy}(\omega_2, -\omega'_2) = E\{Y(\omega_2)Y(\omega'_2)\}$ for $\omega_2, \omega'_2 \geq 0$. In other words, $\Gamma_{yy}(\omega_2, \omega'_2)$ is the *outer product* of $H'_1(\omega_1, \omega_2)$.

As shown in [97], $\Gamma_{yy}(\omega_1, \omega'_1)$ is an augmented covariance matrix consisting of Hermitian (3.57) and complementary (3.56) quadrants. Our contribution here is the generative view that $H'_1(\omega_1, \omega_2)$ also consists of quadrants $A(\omega_1, \omega_2)$ and $B(\omega_1, \omega_2)$ which determine correlative structure in the frequency domain. A schematic example appears in Figure 3.11. In the next subsection, we provide illustrative examples and connect the WL-in-frequency result to the formalism of modulation frequency.

3.5.3 Examples of WL Systems in the Frequency Domain

Our first example is a synthetic time-invariant system, where $h'_2(t, \tau) = h(\tau)$. Consequently, $y(t)$ is stationary, an example of which appears in Figure 3.12. Plot (a) is the function

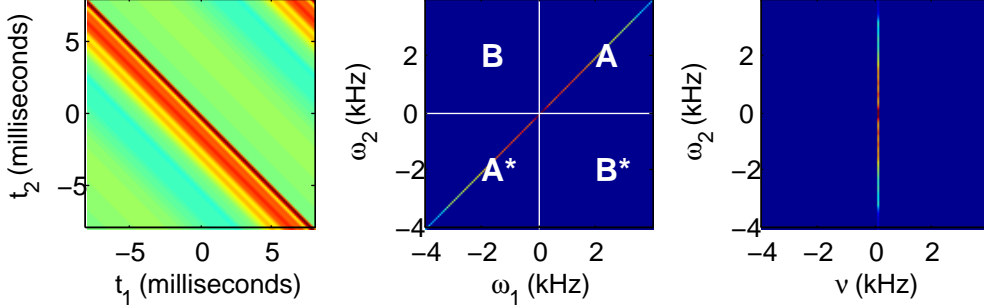


Figure 3.12: An example time-invariant linear system, in three representations: linear time operator (left), widely-linear frequency operator (middle), and modulation spectrum (right).

$h'_1(t_1, t_2)$, or “matrix view” of the system operator as a mapping from input time to output time. One can see that each row is a reversed and shifted copy of a constant impulse response $h(\tau)$. Plot (b) is the magnitude of $H'_1(\omega_1, \omega_2)$, with A and B quadrants indicated. As for any LTI system, the frequency operator is diagonal. From (3.53), $A(\omega_1, \omega_2) = H(\omega_1)\delta(\omega_2 - \omega_1)$ and $B(\omega_1, \omega_2) = 0$ and the integrals for $E\{Y(\omega_1)Y(\omega'_1)\}$ vanish in (3.56). Simply put, $B(\omega_1, \omega_2) = 0$ is a sufficient condition for proper $Y(\omega)$. As a result,

$$Y(\omega_2) = \int H(\omega_2)\delta(\omega_2 - \omega_1)W(\omega_1)d\omega_1 = H(\omega_2)W(\omega_2) \quad (3.58)$$

for $H'_1(\omega_1, \omega_2) = H(\omega_2)\delta(\omega_2 - \omega_1)$, which is strictly linear with respect to $W_a(\omega_2)$.

Plot (c) is the joint-frequency, linear modulation spectrum $H'_2(\nu, \omega)$. As the image would suggest, it is $H'_1(\omega_1, \omega_2)$ sheered such that the diagonal $\omega_1 = \omega_2$ aligns with the ordinate ω and the frequency-shift $\omega_2 - \omega_1 = \nu$ is the abscissa. The modulation spectrum is defined as

$$H'_2(\nu, \omega) = \iint h'_2(t, \tau)e^{-j\nu t - j(\omega - \nu)\tau} dt d\tau \quad (3.59)$$

which, through substitution of variables, also satisfies $H'_2(\nu, \omega) = H'_1(\omega - \nu, \omega)$. In words, diagonal lines in $H'_1(\omega_1, \omega_2)$ appear as vertical lines in $H'_2(\nu, \omega)$. Hence, Plot (c) shows the LTI system as having zero modulation bandwidth.

Our second example is a system with time-variation due to modulation. That is, $h'_2(t, \tau) = h(\tau)m(t - \tau)$. The three plots in Figure 3.13 depict the input-output time

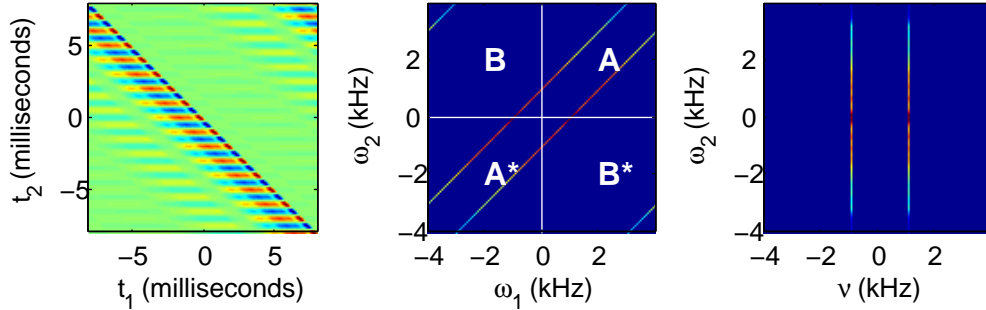


Figure 3.13: An example modulated linear system, in three representations: linear time operator (left), widely-linear frequency operator (middle), and modulation spectrum (right).

function $h'_1(t_1, t_2)$ in (a), the frequency operator $H'_1(\omega_1, \omega_2)$ in (b), and the modulation spectrum $H'_2(\nu, \omega)$ in (c). High-frequency temporal modulations appear as ripples along the diagonals in (a) and lines parallel to the main diagonal in (b). True to its name, the modulation spectrum in (c) represents the signal as uniformly modulated at 1000 Hz for all acoustic frequencies. Thus, subbands up to 1000 Hz are “over-modulated” in the sense of (3.42), and hence will be improper. This explains the corners of the B quadrants in subplot (b) as generators for impropriety.

3.6 Conclusion

Impropriety is an often-overlooked, second-order property of complex random variables. For example, signal analysis based on the power spectrum is inherently proper. Through the Fourier transform, impropriety in the frequency domain has a special relation to the corresponding real-valued time-domain signal. We characterized this relationship by means of a signal model, wherein modulation frequency systematically determines the complementary variance of complex subbands.

The signal model is generative, meaning it represents the observed signal as the output of a system, which in this case is a linear system driven by white, periodically correlated noise. The corresponding estimator requires (1) coherence, meaning the subband frequencies align

with modulation frequencies; and (2) synchronization, meaning the time-average must align with the fundamental period of the driving modulation. The resulting cyclic STFT is thus frequency- and phase-locked to the modulation. Since subbands usually correspond to high frequencies, such as the human range of hearing for acoustic signals, we attributed spectral impropriety primarily to high-frequency modulations. That is, an improper subband at 1000 Hz necessitates modulation frequencies at least as high as 1000 Hz. These modulations have a pulse-shape in the time-domain, which is modeled by the frequency response of a linear system component. This property is embodied by the transfer function relation presented and derived in Section 3.4.

Additional insight derives from a general theory of linear time-varying systems, which we showed are equivalent to widely linear systems in the frequency domain. As a result, impropriety can be seen as a result of mixing positive-frequency amplitudes with their negative-frequency conjugates. This is tightly linked to the concept of high-frequency modulation. In fact, over-modulation, where the modulating frequency overlaps the spectrum of a stationary excitation signal, is precisely the phenomenon underlying the PC-driven system model and transfer function for improper signals. Having laid this foundation, we proceed in the next chapter to discuss coherent demodulation of more general, time-varying systems driven by PC noise.

Chapter 4

COHERENT SUBBAND DEMODULATION

4.1 Introduction

Modulation is the systematic perturbation of a baseline carrier signal over time. The receiver in turn detects deviations from the carrier as the transmitted message. The detection stage is called demodulation, which, in the algebra of complex numbers, amounts to demultiplying the signal from its carrier. The spectrogram, common in signal processing, is a simple example of demodulation because it makes minimal assumptions on the carrier. In the terminology of radio theory, it is “non-coherent.” In this chapter, we describe a signal model and estimation framework for coherent demodulation of signals which arise from physical processes. Such signals are often random and broadband. We will see how subband impropriety, as defined in the previous chapter, leads to analytic tools for detecting modulations with respect to a carrier reference frequency. This view provides a more complete representation of a signal beyond the usual spectrographic approaches. In fact, the resulting representation is complex-valued. This is precisely analogous to the distinction between a complex transfer function and a real power spectrum.

The proposed coherent signal model assumes the signal is modulated at two principal rates. The carrier is itself a high-frequency nonstationarity which provides a synchronous reference for the slower, system modulations which are complex. Coherent signals are thus decidedly not quasi-stationary. We do, however, assume they are *separable* into a high-frequency carrier and a modulation system component which is itself quasi-stationary. Separability is what distinguishes the coherent approach from Grenier’s related, time-varying ARMA model [10] (see also [134]).

An instructive example of a coherent-like signal is voiced speech, since glottal pulses destroy the notion of stationarity even within a short window. This chapter discusses a new, stochastic framework for what is essentially pitch-synchronous processing. Similar

methods exist in the literature (e.g., [135, 136, 137, 138, 92]), but the principles here are different. We treat pitch as an entry point for understanding deeper theoretical constructs relating impropriety, coherence, demodulation, and the removal of nonstationarity from a signal.

Starting with Section 4.2, we define the coherent modulation signal model. Then, Section 4.3 presents the main result of this chapter, a time-varying system estimator for coherent signals, accompanied by examples. We derive essential formulas in Section 4.4. In Section 4.5, we connect the main result with square-law estimation and introduce the complementary envelope of a bandpass signal. Finally, we comment on generalizations of underspread systems in Section 4.6 and conclude in Section 4.7.

4.2 Coherent Modulation Signal Model

In Chapter 3, we introduced a linear signal model to explain impropriety in the frequency domain. The model is repeated below as

$$\begin{aligned} y(t) &= \int h(t, t - \tau) c_0(\tau) w(\tau) d\tau \\ &= \int h(t, t - \tau) x(\tau) d\tau \end{aligned} \quad (4.1)$$

where $w(t)$ is white, Gaussian noise and $c_0(t)$ is a periodic modulator function with fundamental frequency $\omega_0 = 2\pi f_0 = 2\pi/T$. Everything to the right of the equation sign is, again, behind a curtain. The task is to infer $h(t, \tau)$ based on spectral impropriety measured in $y(t)$. In practice, one might have to jointly estimate $c_0(t)$ with $h(t, \tau)$.

The renaming of $m(t)$ to $c_0(t)$ is deliberate. We wish to distinguish $c_0(t)$ from the *complex modulators* defined as

$$M(t, k] = \int h(t, \tau) e^{-jk(\omega_0/2)\tau} d\tau \quad (4.2)$$

which reflect the time-varying spectral amplitudes of the system component¹. This is an important semantic shift, since we will soon see how $c_0(t)$, although technically a modulation, behaves like a carrier for the process. To prevent misunderstanding, we refer to

¹Our reference to time-invariance or lack thereof pertains to the model component $h(t, \tau)$. Due to $c_0(t)$, $y(t)$ is nonstationary in either case.

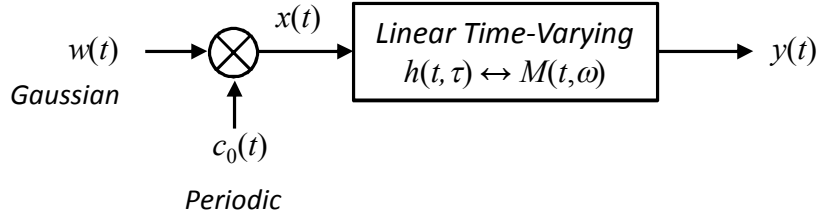


Figure 4.1: Revised signal model for a coherently modulated random process (see also Figure 3.4). The observed process, $y(t)$, is the output of an LTV system driven by white PC noise.

$c_0(t)$ as the “pre-modulator” or “carrier” with fundamental frequency ω_0 , distinct from the time-frequency, complex modulators $M(t, k]$.

Is a subband estimator still possible for the time-varying case? The answer is affirmative under certain assumptions. The LTV synthesis equation is equivalent to

$$y(t) = \sum_k M(t, k] \cdot X(t, k] \cdot e^{jk(\omega_0/2)t}. \quad (4.3)$$

where $X(t, k]$ is the improper, continuous-time STFT of $x(t)$ from Section 3.3.2. Therefore it would seem that subbands of $y(t)$ could extract information from each k element. Let $Y(t, k]$ be the STFT of $y(t)$ with $2T$ -length analysis window of $g(t)$. It is possible to show it has the form

$$Y(t, k] \approx M(t, k] \cdot C(t, k] \quad (4.4)$$

where $C(t, k]$ is improper and stationary in t for each k . Both $X(t, k]$ and $C(t, k]$ are theoretical STFTs induced by the analysis of $y(t)$, and they are statistically equivalent. Therefore, we may revise the system transfer function relation from (3.22) as

$$\begin{aligned} S_{yy}^{(C)}(t, k\omega_0/2) &\approx M^2(t, k] \cdot S_{xx}^{(C)}(t, k\omega_0/2) \\ &= M^2(t, k] \cdot \tilde{C}_0(k\omega_0) \end{aligned} \quad (4.5)$$

where $M(t, k]$ is the time-varying version of $H(k\omega_0/2)$, and $\tilde{C}_0(k\omega_0)$ is the Fourier transform of $c_0^2(t)$. In this model, the subbands are nonstationary due to modulators $M(t, k]$ and improper due to carriers $C(t, k]$. We derive (4.3) and (4.4) in Section 4.4, and these equations govern what we refer to as “coherent demodulation” of a random process.

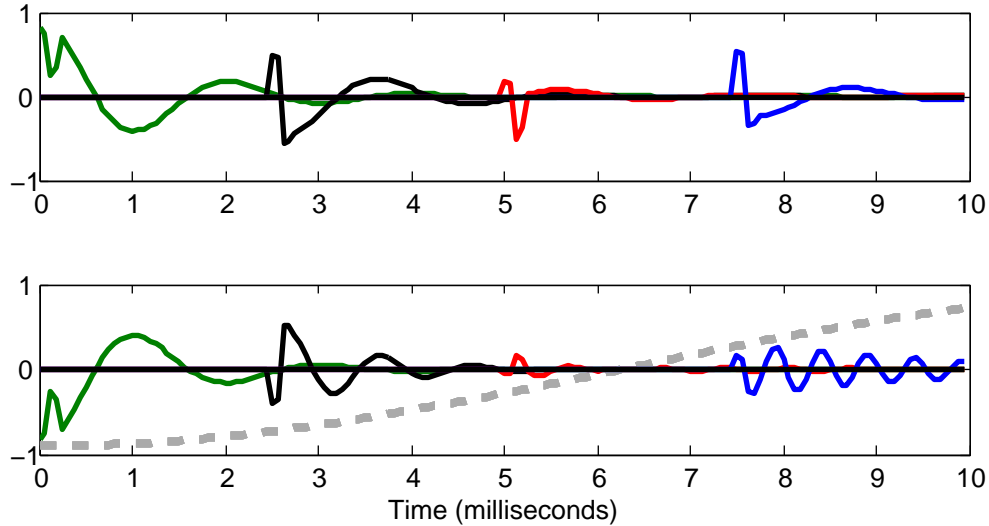


Figure 4.2: Example signals for a time-invariant system (top) and a time-varying system (bottom) driven by the same, periodically correlated excitation signal.

Let us take a moment to understand what is happening in the signal model. A key assumption is that $M(t, k]$ is smooth in t , with maximum modulation frequency ν_{max} limited by $2\nu_{max} < \omega_0$. The observed signal $y(t)$ is therefore modulated at two rates, ω_0 due to $c_0(t)$, and in the range $[0, \nu_{max}]$ due to $h(t, \tau)$. These modulations are in cascade, such that the slow rates appear as modulations on top of the fast rate. We liken this separation to Bedrosian’s product theorem [15]. In Section 2.2, Bedrosian’s theorem defines the frequency separation between modulator and carrier. In the present context, the carrier is itself a type of modulation, but specifically one preceding a slowly-varying system element. Thus, the Bedrosian-like boundary of $\omega_0/2$ separates “carrier modulations” from “slow” or “system modulations.”

A visualization appears in Figure 4.2. In the top plot, the system is time-invariant so that the variation in the pulse responses (individually colored) is due to the intrinsic randomness of $x(t)$. The spacing of the pulses is due to the carrier period T , which in this example is 2.5 milliseconds. Also, the retention of phase in $M^2(t, k]$ means that Fourier components

have a preferred relative phasing, as seen by the consistent oscillations in the pulse shapes. In the bottom plot, the system varies slowly in its gross amplitude (dashed gray line) as well as with a resonance migrating from low to high frequency. The result is that the pulse shapes become increasingly oscillatory over time. The slowly-varying effects – such as amplitude changes and shifting resonances – are what we call the system modulations. These modulations impinge upon the underlying, high-frequency periodicity induced by the carrier.

System estimation requires that the STFT subbands align with, or be coherent with, the harmonics of the fast modulator $c_0(t)$. There are clear connections to the pitch-synchronous STFT used in speech analysis [135, 136, 138], and the STRAIGHT algorithm [91, 92]. The benefit to our proposed approach is that we now have a statistic – namely, impropriety – for defining synchrony. Portnoff [139, 140] also discussed, in considerable detail, speech vocal tract identification via STFT. What he called amplitude- and phase-modulation correspond in principle to the magnitude and phase of $M(t, k)$. By developing a theory of impropriety, we hope to simplify Portnoff’s dense calculations while remaining flexible enough to extend the results to signals other than speech.

4.3 The Time-Varying “Fourier Transform” of a Random Process

In this section, we validate the theory with synthetic examples which are nevertheless based on real-world signals. In each case, the time-varying “Fourier transform” of a process $y(t)$ corresponds to the transfer function of the system component $h(t, \tau)$, referred to here as the complex modulator function $M(t, k)$. This is measured exclusively from the second-order statistics of improper STFT subbands.

We must first address one issue. Using hat notation to indicate an estimated value, we rearrange (4.5) to obtain

$$\widehat{M}^2(t, k) = \frac{\widehat{S}_{yy}^{(C)}(t, k\omega_0/2)}{\widetilde{C}_0(k\omega_0)}. \quad (4.6)$$

This raises the problem of estimating the carrier spectrum $\widetilde{C}_0(k\omega_0)$ and its fundamental frequency ω_0 . We briefly discuss the estimation of ω_0 in a later subsection. As for the carrier amplitudes, we treat $\widetilde{C}_0(k\omega_0)$ as an unknown scalar gain for each subband. Thus,

the long-term temporal modulations in each band are detectable “to within a multiplicative factor,” which effectively sets $\tilde{C}_0(k\omega_0) = 1$ in our analysis. True deconvolution of $h(t, \tau)$ from $x(t)$, however, requires knowledge of $\tilde{C}_0(k\omega_0)$ but we save this for future work.

4.3.1 Synthetic Time-Varying Examples

Let us begin with a time-varying, resonant system. In discrete-time, the LTV system is

$$h[n, p] = (0.8)^p \cos(\phi_M[n]p) \quad (4.7)$$

where n is global time, p is lag time, and $\phi_M[n]$ is the slowly time-varying angle of the system resonance. The corresponding transfer function is

$$M[n, \omega] = \frac{1 - 0.8 \cos(\phi_M[n])e^{-j\omega}}{1 - 1.6 \cos(\phi_M[n])e^{-j\omega} + 0.64e^{-j2\omega}} \quad (4.8)$$

from which it is clear that the system is second-order. Over time, the conjugate poles migrate with a fixed radius of 0.8 while the zero moves along the real axis. On a computer, we synthesize $y[n]$ via the difference equation

$$y[n] + a_1[n]y[n-1] + a_2[n]y[n-2] = x[n] + b_1[n]x[n-1] \quad (4.9)$$

where $x[n] = c_0[n] \cdot w[n]$, and the time-varying b and a coefficients read left-to-right from the numerator and denominator of (4.8).

In the following, we demonstrate two estimators for $M^2(t, k\omega_0/2)$, where ω_0 is the carrier, or “fast,” frequency introduced by the periodically-correlated excitation $x[n]$. The first estimator is cyclic and synchronized to the slow system modulation induced by periodic $\phi_M[n]$. The second is a local average matched to the bandwidth of a lowpass $\phi_M[n]$. In both cases, neither estimator makes no use of the fact that the system function $M(t, \omega)$ is technically minimum-phase.

Cyclic Example

Let $\phi_M[n]$ be a periodic function of time with Fourier series

$$\phi_M[n] = \pi/2 + \pi/16 [\cos(\omega_M n) + \cos(2\omega_M n + \pi/3)] \quad (4.10)$$

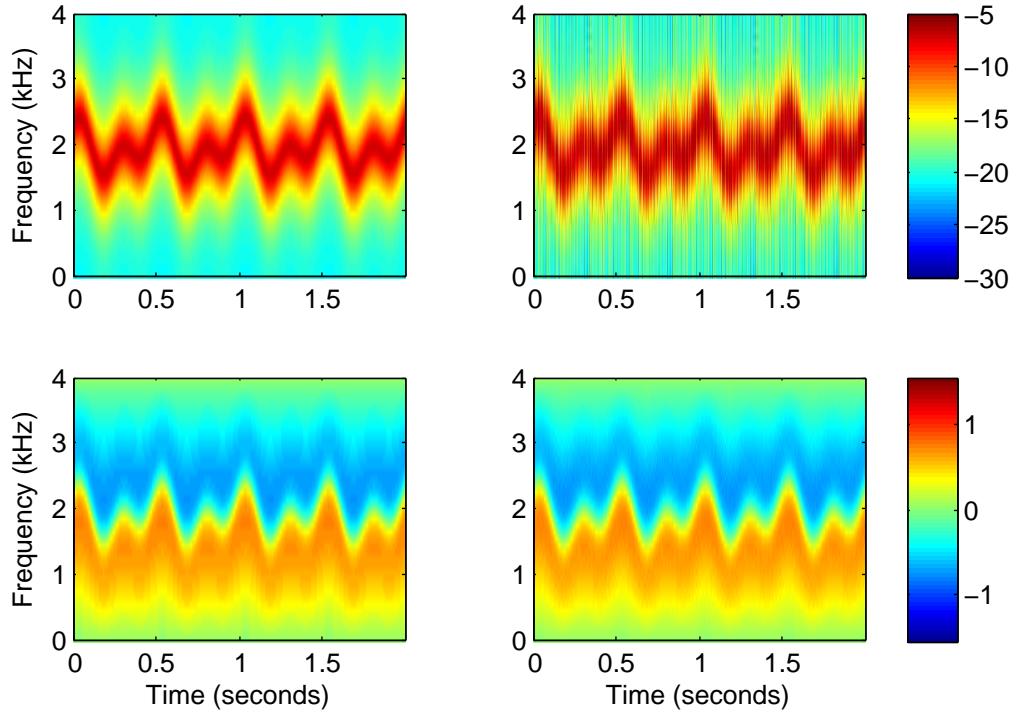


Figure 4.3: Actual (left-hand side) and estimated (right-hand) squared system functions for a periodically-varying resonance, in magnitude (top) and phase (bottom).

where ω_M is the slow modulation rate corresponding to 2 Hertz. We set ω_0 to correspond to 200 Hertz, which safely satisfies the Bedrosian-like constraint of $\omega_M \ll \omega_0$. To obviate the need for carrier estimation, $c_0[n]$ is an impulse train with unit amplitude.

Figure 4.3 displays the results of cyclic estimation of the system function, compared with the true result. As seen, the system resonance migrates periodically over time in both magnitude and phase of $M^2[n, k]$. The frequency axes are discrete and spaced every 100 Hz, which is precisely due to the $\omega_0/2$ rule for improper subband spacing. As seen, the impropriety of the subband signals allows accurate estimation of the phase response of the system.

The estimate is based on a cyclic average of 20 cycles, or 10 seconds of data. Keep in mind that this cyclic average is different from the cyclic demodulator discussed in Section

3.3.2, for which the cyclic interval was $T = 2\pi/\omega_0$. Here, the interval is much longer, 500 milliseconds to be exact, in order to synchronize with the modulators $M[n, k]$ rather than the carrier. Coherence with the carrier is already implicit in the fact that subbands are located at integer multiples of $\omega_0/2$.

Lowpass Example

For this example, $\phi_M[n]$ is a lowpass function with maximum modulation frequency ω_M corresponding to 2 Hertz. All other parameters are equivalent to those of the previous example. Estimation results appear in Figure 4.4, where the right-hand plots were obtained by filtering the squared subbands with a Hamming window 255 milliseconds long. This window has a -3 dB cutoff of almost 3 Hz, where the extra bandwidth accommodates the doubling of bandwidth in the square modulators $M^2[n, k]$.

Improper Versus Proper Estimation

In both Figures 4.3 and 4.4, the top magnitude plots can be obtained by short-time power spectral estimation. The contribution made by spectral impropriety is the ability to estimate the phase response in the bottom plots. To verify this claim, we refer to Figure 4.5 which compares system phase estimates for each case, periodic and lowpass, for proper and improper versions of the previous examples. The improper versions have a periodically-correlated excitation modulated at 200 Hz. By contrast, the proper versions have stationary excitation, or $c_0[n] = 1$.

4.3.2 Propeller-Like and Speech-Like Examples

The previous examples were entirely synthetic, yet the time-varying resonance is reminiscent of real-world signals with similarly slow modulation rates. For instance, propeller cavitation noise can be nearly periodic with rates around 2 Hz for merchant traffic (see Section 2.3.1). Also, the syllabic rate of speech is typically between 3 and 4 Hz [21, 25]. In the following examples, we synthesize $y[n]$ from a fourth-order difference equation

$$y[n] + a_1[n]y[n-1] + a_2[n]y[n-2] + a_3[n]y[n-3] + a_4[n]y[n-4] = x[n] \quad (4.11)$$

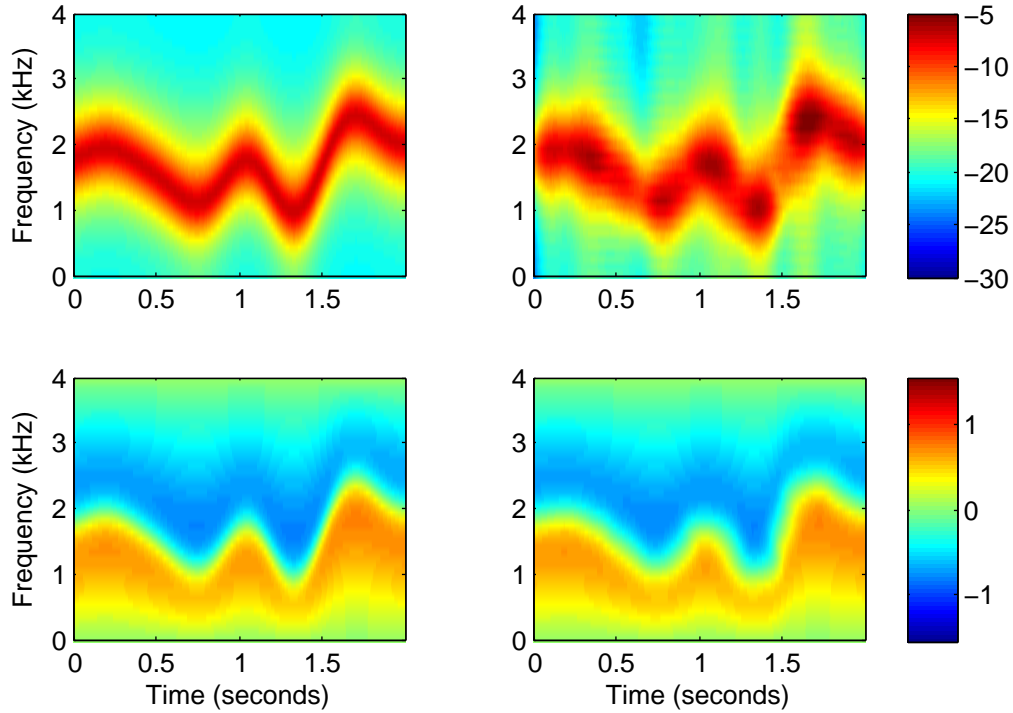


Figure 4.4: Actual (left-hand side) and estimated (right-hand) squared system functions for a lowpass-varying resonance, in magnitude (top) and phase (bottom).

with coefficients determined by feeding actual audio data of cavitation and speech into Grenier’s least-squares LTV estimator [10]. As such, the time-varying coefficients, and corresponding $h[n, p]$, are a minimum-phase approximation to the cavitation or speech signal.

By exciting the systems with PC noise $x[n]$, we may demonstrate the ability of cyclic and lowpass impropriety to recover the system phase. We again choose ω_0 to correspond to 200 Hz. The carrier is thus entirely synthetic, but we claim the systems are cavitation-like in the cyclic case (2 Hz), and speech-like in the lowpass case (8 Hz cutoff). In a realistic situation, the carrier would have to be separately estimated for to provide the coherent subband frequencies. Here, the carrier rate is a known control parameter. Figures 4.6 and 4.7 illustrate phase estimation via improper estimation for both signals.

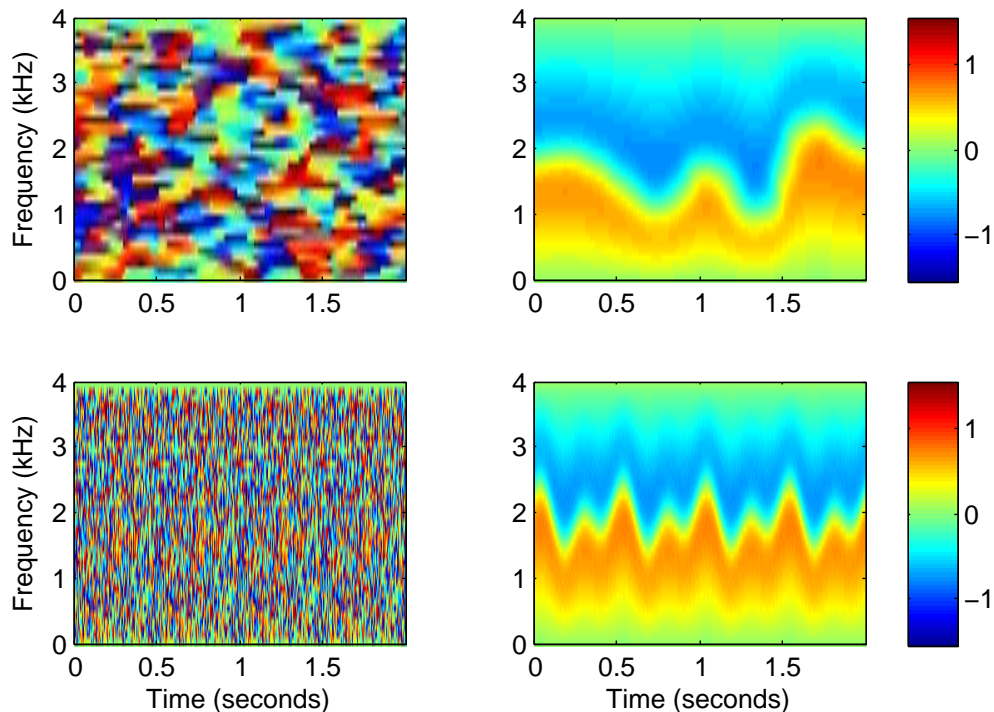


Figure 4.5: Phase-response estimates of a proper signal (left-hand side) and an improper signal (right-hand), for a periodic (top) and lowpass (bottom) resonance.

4.3.3 Modulation Subspace Filtering for Time-Varying Estimation

The preceding examples demonstrate how filtering the squared subbands $Y^2(t, k]$ can yield a reasonable estimate of the squared modulators $M^2(t, k]$. For the k th subband,

$$M^2(t, k] \approx \widehat{M}^2(t, k] = \int Q(t - \tau) Y^2(\tau, k] d\tau \quad (4.12)$$

as long as the filter $Q(t)$ is matched to the frequency content of $M^2(t, k]$. As we will show next, the approximation improves with increasing integration time and is inversely proportional to the bandwidth of $M^2(t, k]$. Furthermore, this result extends to non-Fourier signal bases, where the sparsest or most compact basis is, in fact, preferable.

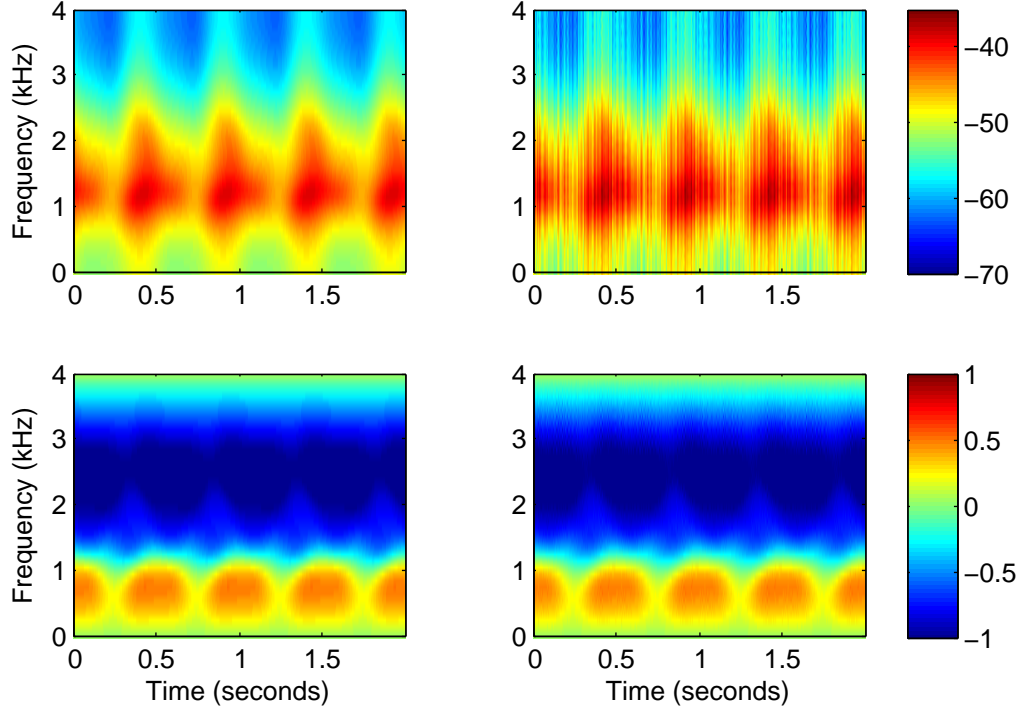


Figure 4.6: Actual (left-hand side) and estimated (right-hand) squared system functions for a 4th-order cyclic system, in magnitude (top) and phase (bottom).

Starting with the transfer function relation (4.6), repeated here as

$$\widehat{M}^2(t, k] = \frac{\widehat{S}_{yy}^{(C)}(t, k\omega_0/2)}{\widetilde{C}_0(k\omega_0)} \quad (4.13)$$

demodulation is equivalent to estimating the time-varying statistic $S_{yy}^{(C)}(t, k\omega_0/2) = E\{Y(t, k]^2\}$. Since the subbands $Y(t, k]$ are no longer stationary in t , the cyclic estimator in Chapter 3 is inadequate. How can one implement a time-varying average when ergodicity no longer holds?

In the DEMON literature, cavitation noise is assumed to have periodic instantaneous variance. The maximum-likelihood estimator is the Fourier transform of the squared signal [65] (see also Section 2.3.1). The principle is identical for the instantaneous complementary

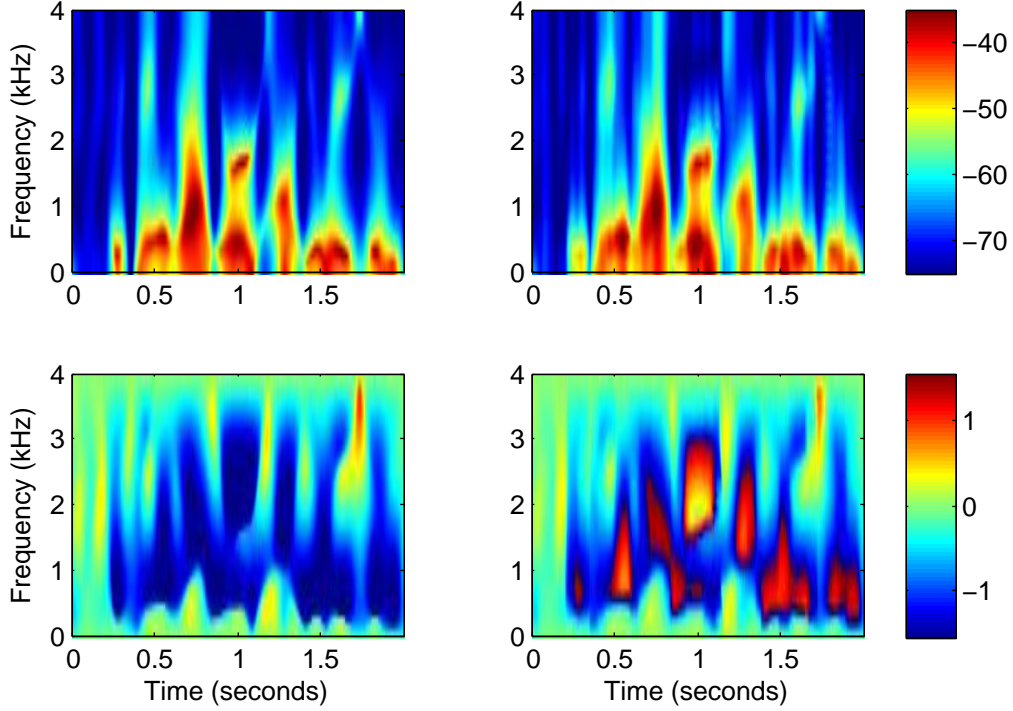


Figure 4.7: Actual (left-hand side) and estimated (right-hand) squared system functions for a 4th-order lowpass system, in magnitude (top) and phase (bottom). Confusion between red and blue in the phase response is due to $\pm\pi/2$ ambiguity in the squared modulators.

variance of a complex random process. From (4.4), the k th STFT band is

$$Y(t) = M(t) \cdot C(t) \quad (4.14)$$

where $C(t)$ is stationary in t . (We refrain from k indexing for simplicity.) For a long enough integration time, ergodicity implies

$$\frac{1}{T_{ave}} \int_0^{T_{ave}} C^2(t) \approx 1 \quad (4.15)$$

where 1 is the assumed complementary variance in the subband, as justified at the beginning of Section 4.3. Therefore,

$$C^2(t) = 1 + \epsilon(t) \quad (4.16)$$

where $\epsilon(t)$ is a zero-mean perturbation which is negligible for sufficiently large T_{ave} .

Supposing $M^2(t)$ is periodic in t , we have

$$M^2(t) = \sum_p m_p e^{jp\omega_M t} \quad (4.17)$$

where m_p are complex coefficients and ω_M is the fundamental frequency of the *slow* modulation, distinct from ω_0 . Therefore, the Fourier series of $Y^2(t)$ is

$$\begin{aligned} \frac{1}{T_{ave}} \int_0^{T_{ave}} Y^2(t) e^{-jq\omega_M t} dt &= \frac{1}{T_{ave}} \sum_p m_p \int_0^{T_{ave}} [1 + \epsilon(t)] e^{j(p-q)\omega_M t} dt \\ &= m_q + \sum_p m_p \varepsilon [(p-q)\omega_M] \end{aligned} \quad (4.18)$$

assuming T_{ave} is a multiple of $2\pi/\omega_M$, and $\varepsilon(\omega)$ is the Fourier transform of $\epsilon(t)$. We refer to the left-hand side of the equation as the DEMON spectrum of $Y(t)$, given knowledge of ω_M .

From the above expression, the Fourier transform of the squared subband will reliably estimate *salient* modulation coefficients. The signal-to-noise ratio, for a P -component modulator is

$$\frac{|m_q|^2}{\left| \sum_p m_p \varepsilon [(p-q)\omega_M] \right|^2} \geq \frac{|m_q|^2}{\sigma_\varepsilon^2 \sum_p |m_p|^2} \approx \frac{T_{ave}}{P} \quad (4.19)$$

assuming the noise $\epsilon(t)$ is stationary with power decaying at a rate of T_{ave}^{-1} . The approximation is crude yet reflects the principal relationship behind the DEMON spectrum. Estimation is more accurate for a given integration time when the spectrum of $M^2(t)$ is compact or sparse.

In general, basis expansion of the squared subband signal requires that $M^2(t)$ be accurately spanned by a low-rank subspace of zero-mean basis elements. Let us assume the expansion

$$M^2(t) = \sum_{p=0}^{P-1} m_p v_p(t) \quad (4.20)$$

where $\int v_p^*(t) v_q(t) dt = \delta[p-q]$ and $\int v_p(t) dt = 0$. Then, for small P ,

$$\int Y^2(t) v_q^*(t) dt = \sum_{p=0}^{P-1} m_p \int v_p(t) v_q^*(t) \cdot [1 + \epsilon(t)] dt \approx m_q. \quad (4.21)$$

A corresponding de-noising filter is

$$\widehat{M}^2(t_2) = \int Y^2(t_1) Q(t_1, t_2) dt_1 \quad (4.22)$$

where $Q(t_1, t_2)$ is the linear projection onto the P -dimensional subspace, defined by

$$Q(t_1, t_2) = \sum_{p=0}^{P-1} v_p^*(t_1)v_p(t_2). \quad (4.23)$$

When $M^2(t)$ is periodic, as in the DEMON spectrum example, the signal basis consists of harmonic sinusoids and $Q(t_1, t_2)$ is a convolution with a comb-filter. If $M^2(t)$ is generally lowpass, then the basis covers frequencies up to the maximum modulation frequency and $Q(t_1, t_2)$ is a sinc convolution kernel. In both cases, it may be desirable to taper the higher frequencies so as to avoid ringing artifacts. We note that $Q(t_1, t_2)$ is linear in $Y^2(t)$ and quadratic in $Y(t)$. See Fano [141] for a fundamental treatment, as well as a recent resurgence due to [142, 143].

It is important to note that the range of modulation frequencies in $M^2(t)$ must be upper-bounded by $\omega_0/2$. This is a consequence of the Bedrosian-like principle mentioned in Section 4.2. In the periodic case, estimation requires synchronization with the carrier rate ω_0 as well as the maximum modulation rate $\omega_M < \omega_0/2$. When the modulation is low-frequency, the corresponding lowpass projection described in this section is essentially a local quadratic average of coherently basebanded subbands. Although $y(t)$ is not quasi-stationary, its coherent subbands are.

4.3.4 Estimation of Carrier Frequency ω_0

Until now, we have assumed that the carrier rate ω_0 is known a priori. This is because our chief interest is to illustrate a simple point: that spectral impropriety in a signal $y(t)$ corresponds to the phase response of a linear time-varying system element in a generative signal model. However, we should give some consideration to the estimation of ω_0 , which is essential for obtaining coherent subbands.

The conventional DEMON spectrum is sufficient for estimating ω_0 , at least as a preliminary suggestion. Given real-valued $y(t)$, the DEMON spectrum is

$$D_y(\nu) = \int y^2(t)e^{-j\nu t} dt \quad (4.24)$$

where we use ν to indicate modulation frequency. From the LTV synthesis equation (4.1),

we have

$$D_y(\nu) = \iiint h(t, t - \tau)h(t, t - \tau')x(\tau)x(\tau')e^{-j\nu t} dt d\tau d\tau'. \quad (4.25)$$

Since the excitation is white, the expected value of the DEMON spectrum is

$$\begin{aligned} E\{D_y(\nu)\} &= \iiint h(t, t - \tau)h(t, t - \tau')c_0^2(\tau)\delta(\tau - \tau')e^{-j\nu t} dt d\tau d\tau' \\ &= \iint h^2(t, t - \tau)c_0^2(\tau)e^{-j\nu t} dt d\tau \\ &= \frac{1}{2\pi} \iint \tilde{H}(\nu - \omega, \tau)\tilde{C}_0(\omega)e^{-j\omega\tau} d\omega d\tau \end{aligned} \quad (4.26)$$

where the last line is the frequency-domain convolution corresponding to the product of $h^2(t, \tau)$ and $c_0^2(t - \tau)$. Evaluating the Fourier transform in τ yields

$$E\{D_y(\nu)\} = \frac{1}{2\pi} \int \tilde{H}(\nu - \omega, \omega)\tilde{C}_0(\omega) d\omega. \quad (4.27)$$

Recall that $\tilde{C}_0(\omega)$ is a series of impulses spaced ω_0 apart. Furthermore, the assumption that $\nu_M < \omega_0/4$ requires that $H(\nu, \omega)$ is bandlimited to $-\omega_0/2 < \nu < \omega_0/2$. Thus, $D_y(\nu)$ will, in expectation, contain spectra concentrations shifted to the harmonic frequencies of $\tilde{C}(\omega)$. The shape of the spectral mass will vary from harmonic to harmonic, but the underlying spacing of ω_0 can be estimated by, for example, least-squares harmonic fitting [144, 145].

For a maximum-likelihood estimate of ω_0 , we refer the reader to the PC-MLE algorithm described in [102]. Unlike the DEMON spectrum, the PC-MLE takes into account the lagged bilinear product $y(t)y(t + \tau)$ for multiple lags τ .

4.4 Derivation of the Coherent System Equations

In this section we derive the equations (4.4) and (4.3) based on definitions (4.2) and (4.1). These proofs underpin the system estimation method of Section 4.2.

4.4.1 Proof of the Sum-of-Products Model (4.3)

The following is a derivation of the sum-of-products model (4.3) for $y(t)$ as defined in the second line of (4.1), or

$$y(t) = \int h(t, t - \tau)x(\tau)d\tau. \quad (4.28)$$

In words, we must show that $y(t)$ is also equivalent to a sum of modulated subband signals. First, we can substitute the inverse Fourier transform relation for $x(t)$ to obtain

$$y(t) = \int h(t, t - \tau) \frac{1}{2\pi} \int X(\omega) e^{j\omega\tau} d\omega d\tau. \quad (4.29)$$

Integrating in τ gives

$$y(t) = \frac{1}{2\pi} \int M(t, \omega) X(\omega) e^{j\omega t} d\omega \quad (4.30)$$

where $M(t, \omega)$ is the Fourier transform of $h(t, \tau)$ in τ [78].

To simplify further, let us hypothesize a basis expansion for $M(t, \omega)$ such that

$$M(t, \omega) = \sum_k M(t, k] F_k(\omega) \quad (4.31)$$

with $F_k(\omega)$ being the k th frequency-domain basis function, and $M(t, k]$ is the set of coefficients at time t . Substituting this expansion into (4.30) yields

$$y(t) = \frac{1}{2\pi} \sum_k M(t, k] \int F_k(\omega) X(\omega) e^{j\omega t} d\omega \quad (4.32)$$

which is equivalent to the time-domain convolution

$$y(t) = \sum_k M(t, k] \int f_k(t - \tau) x(\tau) d\tau \quad (4.33)$$

where $f_k(t)$ is the inverse Fourier transform of $F_k(\omega)$.

The next step is crucial to the subband formulation used in this chapter. Let $f_k(t) = g(t) e^{jk(\omega_0/2)t}$, such that $f_k(t)$ is a complex bandpass filter for $k \neq 0$ when $g(t)$ is a lowpass window. Recall that $g(t)$ is our STFT analysis window. Substitution into (4.33) yields

$$y(t) = \sum_k M(t, k] e^{jk(\omega_0/2)t} \int g(t - \tau) x(\tau) e^{-k(\omega_0/2)\tau} d\tau \quad (4.34)$$

where we recognize the integral as the STFT of $x(\tau)$. That is,

$$y(t) = \sum_k M(t, k] \cdot X(t, k] \cdot e^{jk(\omega_0/2)t} \quad (4.35)$$

which is what we set out to prove. This is a sum-of-products signal model [44, 61], which we reviewed in Sections 2.2 and 2.3.1. The difference here is that $y(t)$ is explicitly the sum of *improper* AM-WSS, or modulated-stationary, random processes.

What types of signals are amenable to the preceding analysis? The STFT is a frequency analyzer for which it only makes sense to use a lowpass window shape for $g(t)$. Therefore, the basis elements $F_k(\omega) = G(\omega - k\omega_0/2)$ are bandpass with fixed bandwidth. This suggests that $M(t, \omega)$ must be locally smooth within the window bandwidth, or equivalently, that $h(t, \tau)$ be concentrated in τ within the window length. A special case is when $g(t)$ is rectangular and the basis functions $G(\omega - k\omega_0/2)$ are orthogonal, translated sinc functions. In this case, $h(t, \tau)$ is exactly $2T$ -long or shorter in the τ -axis.

In the time-invariant case of Chapter 3, $M(t, k) = M[k] = H(k\omega_0/2)$, and the smoothness constraint on $H(\omega)$ is the same as stated here for $M(t, \omega)$.

4.4.2 Proof of Modulated STFT Subbands (4.4)

Having shown that $y(t)$ is equal to a sum of modulated subbands, we require some means of extracting the subband modulations from $y(t)$. As in Chapter 3, we use linear filtering in the form of the short-time Fourier transform (STFT). The following shows that STFT subbands are in turn equivalent to modulated, random processes of the form (4.4).

From the definition of the STFT, we have

$$Y(t, k) = \int g(t - t')y(t')e^{-jk(\omega_0/2)t'} dt'. \quad (4.36)$$

Substituting the LTV synthesis equation (4.30) for $y(t')$ yields

$$Y(t, k) = \int g(t - t') \int h(t', \tau)x(t' - \tau)e^{-jk(\omega_0/2)t'} d\tau dt' \quad (4.37)$$

which follows from substitution of variables in the original form, $h(t', t' - \tau)x(\tau)$.

Further reduction requires a simplifying assumption. Suppose $h(t', \tau)$ varies slowly in t' such that it appears almost constant within the window $g(t - t')$ for all time shifts. (Remember that $x(t)$ is quickly-varying so as to embody the impropriety of $y(t)$.) Then we have $Y(t, k) \approx \tilde{Y}(t, k)$, where

$$\tilde{Y}(t, k) = \int h(t, \tau) d\tau \int g(t - t')x(t' - \tau)e^{-jk(\omega_0/2)t'} dt'. \quad (4.38)$$

The integral in t' is an STFT with delayed input. Therefore,

$$\tilde{Y}(t, k) = \int h(t, \tau)X(t - \tau, k)e^{-jk(\omega_0/2)\tau} d\tau. \quad (4.39)$$

At this point, we must constrain $h(t, \tau)$ to have certain properties amenable to our chosen method of STFT analysis. Let us assume² a basis expansion in τ of the form

$$h(t, \tau) = \sum_i M(t, i] f_i(\tau) \quad (4.40)$$

where the sum is possibly infinite. This is the τ -domain version of (4.31). As before, we wish to retain a time-frequency interpretation for the modulator array $M(t, i]$, which means each $f_i(\tau)$ function occupies a distinct bandpass region in frequency. We can also remain consistent with the STFT by defining $f_i(\tau) = f(\tau)e^{j(i\omega_0/2)\tau}$, for some lowpass function $f(\tau)$ analogous to $g(t)$. With such a basis expansion, we have

$$\tilde{Y}(t, k] = \sum_i M(t, i] \int f(\tau)e^{j(i-k)(\omega_0/2)\tau} X(t - \tau, k] d\tau. \quad (4.41)$$

By the associative property of convolution, the above integral is the same as a new STFT with implicit analysis window $r(t) = f(t) * g(t)$. Let us define $C(t, k]$ as the STFT of $x(t)$ with this implicit window. Therefore,

$$\begin{aligned} \tilde{Y}(t, k] &= M(t, k] \cdot C(t, k] + \sum_{i \neq k} M(t, i] \int f(\tau)e^{j(i-k)(\omega_0/2)\tau} X(t - \tau, k] d\tau. \\ &= M(t, k] \cdot C(t, k] + \epsilon_k(t). \end{aligned} \quad (4.42)$$

If $\epsilon_k(t)$ is comparatively small, and if the assumption underlying (4.38) is not too severe, then we have

$$Y(t, k] \approx M(t, k] \cdot C(t, k] \quad (4.43)$$

which is what we set out to prove.

Since $C(t, k]$ has the same impropriety as $X(t, k]$ (see Remark 2 below), we find the instantaneous complementary variance to be

$$\begin{aligned} E\{Y^2(t, k]\} &\approx M^2(t, k] \cdot E\{C^2(t, k]\} \\ &\approx M^2(t, k] \cdot S_{xx}^{(C)}(k\omega_0/2). \end{aligned} \quad (4.44)$$

²This assumption differs from the one taken in Section 3.4.3, specifically equation (3.32), for a similar problem. Both arrive at the same idea, but the present discussion is more rigorous.

From Chapter 3, we know that $S_{xx}^{(C)}(k\omega_0/2)$ is equivalent to the Fourier transform of $c_0^2(t)$. Thus,

$$S_{yy}^{(C)}(t, k\omega_0/2) \approx M^2(t, k] \cdot \tilde{C}_0(k\omega_0) \quad (4.45)$$

as was stated in (4.5).

We will attempt to clarify some points in the derivation with the following remarks.

Remark 1: Implications of the analysis window $r(t)$

We emphasize that $Y(t, k]$ is an STFT computed with analysis window $g(t)$. The mathematical relationship between $Y(t, k]$ and the hypothetical excitation signal $x(t)$ induces an *effective* STFT window, $r(t)$, on $x(t)$. In practice, this allows us to use $g(t)$ and then model the impropriety of $Y(t, k]$ in terms of $C(t, k]$ and hence $x(t)$.

In the previous subsection, we assumed that $f(t) = g(t)$. The purpose was to obtain a convenient relation between $y(t)$ and $X(t, k]$, which the reader may recall is the STFT of $x(t)$ computed with $g(t)$. This would then mean $r(t)$ is the auto-convolution of $g(t)$. In general, $f(t)$ is arbitrary except for the lowpass requirement, which maintains the time-frequency understanding of the modulators $M(t, k]$. This in turn implies a smoothness constraint on $M(t, \omega)$ in ω , the samples of which are represented in $M(t, k]$.

Remark 2: Properties of $C(t, k]$

Both $X(t, k]$ and $C(t, k]$ are induced STFTs of the theoretical signal $x(t)$. In Chapter 3, we found that $X(t, k]$ is complex and stationary improper in t for each k . Furthermore, the complementary variance in the k th band, $\rho_X^2[k]$, is equal to the k th harmonic amplitude of the Fourier series of $c_0^2(t)$. These properties hold as long as $g(t)$, the analysis window used on $y(t)$, is lowpass and frequency-concentrated between $-\omega_0/4$ and $+\omega_0/4$. Therefore, $C(t, k]$ is also stationary improper with the same correspondence to $m_0^2(t)$ provided that $r(t)$ is similarly bandlimited.

Remark 3: Constraints on $h(t, \tau)$

The preceding derivation made two assumptions on the system function $h(t, \tau)$. Namely,

$h(t, \tau)$ is slowly-varying in t as well as slowly-varying in its τ -axis Fourier transform. This is really one assumption, a property we have referred to as Kailathian underspread. That is, its Doppler-delay spreading function is concentrated over a rectangle in the $\nu - \tau$ plane. We address a more general notion of underspread later in Section 4.6, by also taking into account the pre-modulator $c_0(t)$.

Remark 4: Analysis of the estimation noise

The noise $\epsilon_k(t)$ in the k th band is intrinsic to the STFT and should be regarded as a type of estimation variance. It is defined as (with substitution of variables)

$$\epsilon_k(t) = \sum_{i \neq 0} M(t, k - i] \int f(\tau) e^{-ji(\omega_0/2)\tau} X(t - \tau, k] d\tau. \quad (4.46)$$

This is the summation of modulated convolutions of $X(t, k]$ with frequency-shifted versions of $f(t)$. Loosely speaking, the frequency-shift moves $f(t)$ beyond the bandlimits of $X(t, k]$, which is a lowpass function of t , and hence attenuates the output. If $f(t)$ has sufficient stopband suppression, then $\epsilon_k(t)$ will be small in comparison with the desired $M(t, k] \cdot C(t, k]$.

The summation over i , however, may pose a problem. Could a large summation possibly accumulate to a point where $\epsilon_k(t)$ overcomes the signal of interest? The following is a quick refutation of this concern. Suppose that $M(t, \omega)$ is bandlimited in the sense that almost all of its energy, for all t , is contained within $-B\omega_0/4 < \omega < B\omega_0/4$. For integer B , we require up to B subbands to faithfully represent the system spectrum. This is a reasonable assumption due to the fact that most, if not all, natural systems attenuate high frequencies. Further suppose the basis is orthonormal, such that

$$\int f_i^*(\tau) f_k(\tau) d\tau = \frac{1}{2T} \delta[i - k] \quad (4.47)$$

where $\delta[\cdot]$ is the Kronecker delta function. Now we can say that the basis preserves the power of the transformed signal.

Since $x(t)$ and $g(t)$ are both normalized (see Chapter 3), the k th subband signal $X(t, k]$ has unit power. Thus the signal-to-noise ratio, ignoring the effect of $M(t, k]$, is

$$R = \frac{E\{|C(t, k]|^2\}}{E\{|\epsilon_k(t)|^2\}} \approx \frac{E\{|C(t, k]|^2\}}{1 - E\{|C(t, k]|^2\}}. \quad (4.48)$$

where $C(t, k]$ is the cascade of $x(t)e^{-jk(\omega_0/2)\tau}$ convolved with $g(t)$ and then with $f(t)$. The power of $C(t, k]$, which is stationary in t , follows from Parseval's theorem and we have

$$E\{|C(t, k]|^2\} = \frac{1}{2\pi} \int |F(\omega)G(\omega)|^2 d\omega = \int |r(t)|^2 dt \quad (4.49)$$

where $\frac{1}{2\pi} \int |G(\omega)|^2 d\omega = 1$ and $\frac{1}{2\pi} \int |F(\omega)|^2 d\omega = 1/2T$ by definition. The closer (4.49) is to 1, the greater the signal-to-noise ratio in (4.42).

An example satisfying the above conditions is when $g(t)$ and $f(t)$ are both rectangular windows of length $2T$. The resulting $r(t)$ is a triangle with width $4T$ and peak amplitude $\sqrt{1/2T}$. It is a basic calculus problem to evaluate (4.49) and find $E\{|C(t, k]|^2\} = 2/3$ for an effective signal-to-noise ratio of approximately 2, or +3 dB. Rectangles are not ideal in terms of bandlimiting, so this example provides an analytic lower bound on STFT performance.

4.4.3 An Open Theoretical Question

There was a point in the previous derivation, at equation (4.38), where we made the following assumption:

$$\int g(t-t') \int h(t', \tau) x(t'-\tau) dt' d\tau \approx \int h(t, \tau) \int g(t-t') x(t'-\tau) dt' d\tau. \quad (4.50)$$

The ideal operator, if one exists, commutes with $h(t, \tau)$ such that

$$g(t) *_t [h(t, \tau) *_\tau x(\tau)] = h(t, \tau) *_\tau [g(\tau) *_\tau x(\tau)]. \quad (4.51)$$

When $h(t, \tau)$ is time-invariant, any LTI filter $g(t)$ obviously satisfies this relation through the commutativity of convolution.

For the general time-varying case, Tjøstheim [146] defined a spectral analyzer \mathcal{G} with frequency response $G(\omega)$ defined by

$$\mathcal{G}\{y(t)\} = \frac{1}{2\pi} \int G(\omega) H_2'(t, \omega) W(\omega) e^{j\omega t} d\omega \quad (4.52)$$

where $W(\omega)$ corresponds to stationary white noise $w(t)$, and $y(t)$ is given by the LTV synthesis equation

$$y(t) = \frac{1}{2\pi} \int H_2'(t, \omega) W(\omega) e^{j\omega t} d\omega. \quad (4.53)$$

The reader may recall $H_2'(t, \omega)$ as the overall system function, incorporating both $c_0(t)$ and $h(t, \tau)$, from Section 3.5.2. For example, a bandpass \mathcal{G} perfectly isolates a range of frequencies in ω without disturbing the temporal modulations in $H_2'(t, \omega)$. Tjøstheim proved that \mathcal{G} exists and is generally linear and time-varying. However, the construction of \mathcal{G} , given $y(t)$, is left unresolved.

For our problem, Tjøstheim's operator is a linear bandpass filter which isolates carriers, essentially distinct harmonics of $m_0^2(t)$, while preserving complex amplitude modulations within the carrier frequency band. That is, the operator satisfies

$$\mathcal{G}\{y(t)\} = \frac{1}{2\pi} \int G(\omega) M(t, \omega) X(\omega) e^{j\omega t} d\omega \quad (4.54)$$

where $X(\omega)$ corresponds to a PC process $x(t)$. The construction of \mathcal{G} in this context is a possible direction for future work.

4.5 Complementary Square-Law Estimation

The previous sections were largely concerned with furnishing a plausible signal model and methodology for a random process which is improper in the frequency domain. We have focused on LTV system theory and a justification for a subband approach which reduces a broadband process to a collection of modulated complex signals. The purpose of this section is to point the reader's attention to the behavior of just one subband. It has a complex representation as well as an equivalent, real-valued one. In either representation, we show that the signal has *two* envelopes. This fact is a consequence of assuming Gaussianity, for which the square law envelope detector has a traditional form as well as a new, complementary form.

4.5.1 Complex Representation

Given a product model as in (4.4), the square law for the k th complex subband is simply

$$E\{|Y(t)|^2\} = |M(t)|^2 \cdot E\{|C(t)|^2\} = \sigma_C^2 |M(t)|^2 \quad (4.55)$$

where $Y(t) = M(t) \cdot C(t)$, and $C(t)$ is stationary Gaussian noise with variance σ_C^2 . The only other assumption is that the modulator $M(t)$ is statistically independent from $C(t)$.

As discussed in Section 2.3.1, the square law is related to the Hilbert envelope and provides the foundation for DEMON analysis.

An important tool in previous sections is the complementary square law, which states, for the k th subband,

$$E\{Y^2(t)\} = M^2(t) \cdot E\{C^2(t)\} = \rho_C^2 M^2(t) \quad (4.56)$$

where ρ_C^2 is the complementary variance of $C(t)$. The reader may recall from Section 2.4 that a complex Gaussian process is characterized entirely by its Hermitian and complementary second-order statistics. Since $Y(t)$ is also Gaussian, and assuming it is principally white³, there is nothing more to be gained beyond the two square laws. Thus, the detection of a complex modulator, and hence a linear system function $M(t, k]$, is possible in the Gaussian case only when statistical impropriety is present in $C(t)$.

It is possible to solidify this point even further by considering the optimal detector for a complex modulator. Here, we choose to define the optimal estimate as the one which maximizes the probability of observing the signal $Y(t)$. This is the maximum-likelihood approach common to Gaussian signal estimation [147]. However, the introduction of impropriety requires a new formulation, given briefly below.

Suppose we sample $Y(t)$ to obtain a white, complex time series $Y[n]$, with length N . We presume the signal model

$$Y[n] = M[n] \cdot C[n] + \varepsilon[n] \quad (4.57)$$

where all signals are complex, $C[n]$ is Gaussian noise, and $\varepsilon[n]$ is strictly proper, additive Gaussian noise independent from $C[n]$. Again, we understand the above as occurring within a subband for some k . With no loss of generality, let $C[n]$ be normalized such that $\sigma_C^2 = 1$. The multivariate probability distribution of \mathbf{Y} , treated as an N -dimensional vector, is given by van den Bos [113]. When $Y[n]$ is white, the probability distribution is a function of the Hermitian variance

$$\sigma_Y^2[n] = E\{|Y[n]|^2\} = |M[n]|^2 + \sigma_\varepsilon^2 \quad (4.58)$$

³We assume that the subband signal is white in the sense that coloration is principally a function of the subband filter. If we sampled $Y(t)$ at its Nyquist rate, this means the discrete samples would be uncorrelated. A generalization of the square law for multiple time lags appears in Kirsteins *et al.* [102]

and complementary variance

$$\rho_Y^2[n] = E\{Y^2[n]\} = \rho_C^2 M^2[n]. \quad (4.59)$$

It is possible to show (see Appendix D) that the log-likelihood function is

$$J(\mathbf{Y}|\mathbf{M}) = \sum_{n=0}^{N-1} \ln(\sigma_Y^4[n] - |\rho_Y^2[n]|^2) + \frac{2\sigma_Y^2[n]\hat{\sigma}_Y^2[n] - \rho_Y^{2*}[n]\hat{\rho}_Y^2[n] - \rho_Y^2[n]\hat{\rho}_Y^{2*}[n]}{\sigma_Y^4[n] - |\rho_Y^2[n]|^2}. \quad (4.60)$$

where hat-notation indicates an estimated value, such as obtained by cyclic or lowpass averaging (see Section 4.3). We can also express the log-likelihood as individual time-sample likelihoods, or

$$J(\mathbf{Y}|\mathbf{M}) = \sum_{n=0}^{N-1} J_n(\mathbf{Y}|\mathbf{M}). \quad (4.61)$$

The role of impropriety becomes strikingly apparent when we plot the log-likelihood over a grid of modulator values. Only one point in the grid is the true modulator, which should have a high likelihood. Figure 4.8 plots $J_n(\mathbf{Y}|\mathbf{M})$ over the real and imaginary parts of $M[n]$ at one moment in time. There are two cases in Figure 4.8, one where $C[n]$ is assumed proper ($\hat{\rho}_C^2 = 0$) and the other where $C[n]$ is improper. The proper case shows a ring of equiprobable modulator values and is thus unable to detect modulator phase. On the other hand, the improper case resolves the phase ambiguity to two options separated by π radians.

The log-likelihood function for an improper process is an example of how complementary statistics can aid estimation problems. One example is the solution for $M[n]$, not just $M^2[n]$, by optimizing over the log-likelihood function with a bandwidth constraint on $M[n]$.

4.5.2 Real Representation

Complex numbers are abstract, and have discouraged many researchers who are otherwise interested in analysis tools offered by signal processing theory. This section shows that complex demodulation is equivalent to a parallel operation in the domain of real-valued signals.

Let $y_k(t)$ be a real, bandpass signal. Its center frequency is ω_k with bandlimits $\omega_k \pm B/2$. It could be from speech or some other natural process. As such, there is no pre-defined

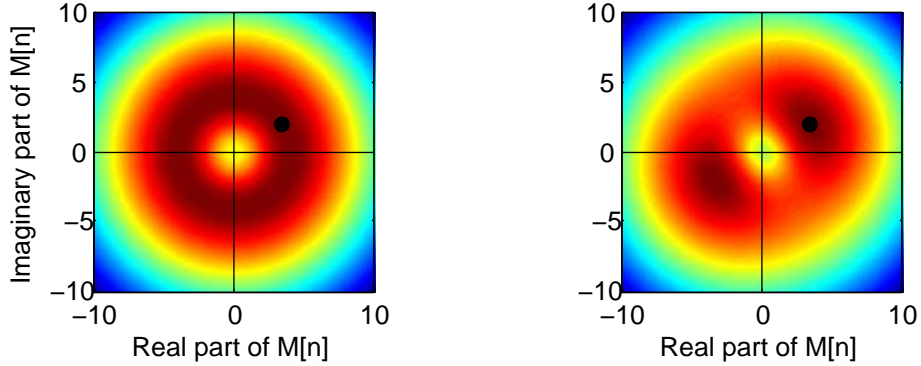


Figure 4.8: Log-likelihood functions, plotted for a single moment in time over the complex plane of possible modulator values. The true modulator value is shown by a black dot. Left: proper signal, with no phase estimation. Right: improper signal, with only two possible solutions.

method of estimating its envelope, unlike man-made communication signals. A common envelope-detection method is to impose a nonlinearity followed by lowpass filtering (see, e.g., [26]). The square law [68] is attractive since $y_k^2(t)$ is still bandlimited at baseband. Therefore, let us define the envelope as

$$m_k^{(H)}(t) = \int q(\tau) y_k^2(t - \tau) d\tau \quad (4.62)$$

where $q(t)$ is a lowpass filter with cutoff equal to B . Squaring in time also leads to double-frequency sidebands with center frequency $2\omega_k$. These can be extracted by a bandpass operation, such as

$$y_k^{(SB)}(t) = \int \cos(2\omega_k \tau) q(\tau) y_k^2(t - \tau) d\tau. \quad (4.63)$$

Both signals comprise all of the information in the squared signal, or

$$y_k^2(t) = m_k^{(H)}(t) + y_k^{(SB)}(t). \quad (4.64)$$

The Fourier transforms of $y_k(t)$ and $y_k^2(t)$ appear in Figure 4.9 as solid and dotted lines, respectively.

Sidebands are often thrown away in demodulation algorithms, since the desired envelope is lowpass. However, we intend to show that spectral impropriety in $y_k(t)$ manifests uniquely

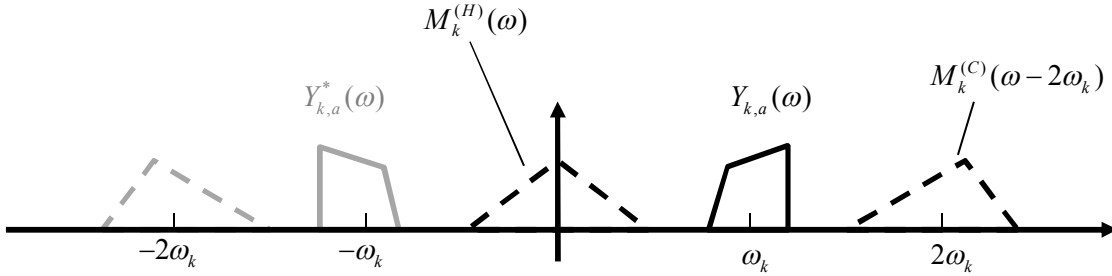


Figure 4.9: Illustration of square-law demodulation in the frequency domain, where the solid lines constitute the Fourier transform of $y_k(t)$ and the dashed lines constitute lowpass and bandpass components of $y_k^2(t)$.

in the the high-frequency sidebands. Let $y_{k,a}(t)$ denote the analytic signal of $y_k(t)$, with a Fourier transform vanishing over negative frequencies. Therefore,

$$y_k(t) = \frac{1}{2} [y_{k,a}(t) + y_{k,a}^*(t)]. \quad (4.65)$$

Due to the conjugate-symmetry of the Fourier transform of $y_k(t)$, we have simply defined analytic and anti-analytic components to represent the signal over positive and negative frequencies separately. These appear as the dark solid line and gray solid line in Figure 4.9. Squaring the subband as in (4.62), we have

$$y_k^2(t) = \frac{1}{2} |y_{k,a}(t)|^2 + \frac{1}{4} [y_{k,a}^2(t) + y_{k,a}^{*2}(t)] \quad (4.66)$$

which is clearly demarcated into lowpass and bandpass elements. The real, lowpass *Hermitian envelope* is

$$m_k^{(H)}(t) = \frac{1}{2} |y_{k,a}(t)|^2 \quad (4.67)$$

while the real, bandpass sidebands are

$$y_k^{(SB)}(t) = \frac{1}{4} [y_{k,a}^2(t) + y_{k,a}^{*2}(t)]. \quad (4.68)$$

It is clear that the sidebands contain the complementary variance term $y_{k,a}^2(t)$. Therefore, the complex, lowpass *complementary envelope* is given by

$$m_k^{(C)}(t) = \frac{1}{4} y_{k,a}^2(t) \cdot e^{-j2\omega_k t} \quad (4.69)$$

which is nontrivial when $y_{k,a}(t)$ is improper.

We can think of the real sidebands $y_k^{(SB)}(t)$ as containing frequency-shifted versions of the complementary envelope. In other words, the Hermitian and complementary envelopes pertain to the lowpass and double-frequency terms in the real-valued square-law detector. This is most clear in the frequency domain, shown by Figure 4.9. Our point in this discussion is to argue against the usual practice of ignoring sideband information in envelope analysis.

With respect to the LTV generative model proposed in Section 4.2, this k th subband signal $y_k(t)$ is the non-basebanded version of $Y(t, k]$ for $\omega_k = k\omega_0/2$. If the signal is modulated at a rate of $k\omega_0$, then the Hermitian envelope $m_k^{(H)}(t) = |Y(t, k]|^2$ and the complementary envelope $m_k^{(C)}(t) = Y^2(t, k]$. In expectation, these quantities reduce to $m_k^{(H)}(t) = |M(t, k]|^2$ and $m_k^{(C)}(t) = M^2(t, k]$. The LTV phase response is therefore embedded within the sideband information of the real-valued square-law detector.

4.6 Comment on Underspread System Estimation

Priestley [84] and Matz *et al.* [3] both argue that the concept of “frequency” has meaning when the time-frequency amplitude function $M(t, \omega)$ varies slowly in t . This is what motivated Matz *et al.* to derive a time-frequency estimation framework upon Kailath’s underspread criterion [87], which is part of a long tradition spanning concepts of “quasi-stationary” and “locally stationary” [86] processes (see Section 2.3.2). They all essentially mean the same thing, which is the assumption that signals of interest are nearly stationary within a window of analysis.

From the results presented in this chapter, it is apparent that other options are available. The process $y(t)$ from (4.1) is highly nonstationary due to carrier modulations from $c_0(t)$. Indeed, the more broadband $c_0(t)$ is, the more benefit we derive via subband impropriety. That is evident from the complementary transfer function relation (4.5). Seen through the STFT window $g(t)$, the process $y(t)$ can not be considered “quasi-stationary” since $g(t)$ covers at least one whole period of $c_0(t)$. We do, however, require precise knowledge of the modulation period, in order to synchronize with $c_0(t)$ in the form of coherent STFT subband frequencies. All further estimation hinges upon this fact.

As envisioned here, $h(t, \tau)$ is, however, underspread or quasi-stationary in the classi-

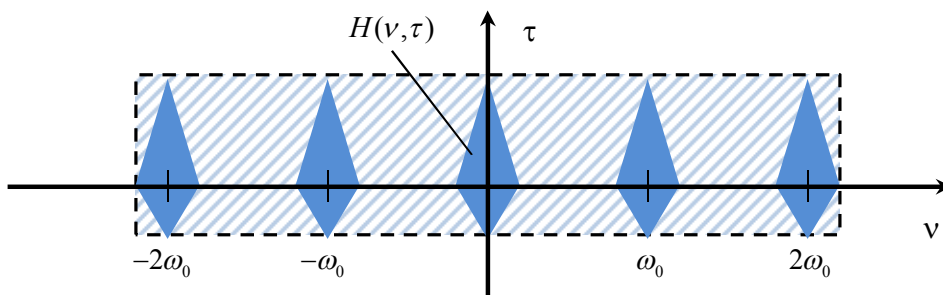


Figure 4.10: Hypothetical support region for a coherently-modulated linear system $H'(\nu, \tau)$, with underspread system subcomponent $H(\nu, \tau)$ and carrier rate ω_0 . The bounding rectangle is possibly overspread, while the true support region (filled) is still underspread.

cal Kailathian sense. Assuming we know the harmonic Fourier coefficients of $c_0(t)$, the estimation of $M(t, k]$ through improper subband demodulation only confirms Kailath's conjecture. For we must remember the strict definition of an underspread system: one which is identifiable given knowledge of its output *and* input.

Inferring a generative system is necessarily more difficult, since we only have access to the observed output $y(t)$. Indeed, the system input $x(t)$ is merely a theoretical abstraction. It would seem, then, that estimating the system function of a process is probably less forgiving than the underspread criterion. Nevertheless, Kailath's and Bello's classes are useful constructs in our understanding of observability, as ably demonstrated by Matz *et al.* [3, 4].

The point we wish to leave with the reader is that subband impropriety is a new signal feature with new implications in estimation. The literature is overwhelmingly, and unnecessarily, confined by the Kailathian concept of underspread systems and processes. By exploiting high-frequency modulation in the signal, we can interpret $h(t, t - \tau)c_0(\tau)$ as akin to Bello's class of generalized underspread systems [5, 94]. Defining the overall system as

$$h'(t, \tau) = h(t, \tau)c_0(t - \tau) \quad (4.70)$$

we find the Doppler-range spreading function as

$$H'(\nu, \tau) = \frac{1}{2\pi} \int H(\nu - s, \tau)C_0(s)e^{-js\tau} ds \quad (4.71)$$

which is a convolution in modulation frequency. Further expanding the carrier in terms of its harmonic amplitudes,

$$H'(\nu, \tau) = \sum_k H(\nu - k\omega_0, \tau) C_k e^{jk\omega_0\tau} ds \quad (4.72)$$

where ω_0 is the fundamental frequency. Therefore, the effective spreading function appears as periodic copies of $H(\nu, \tau)$ across the modulation-frequency axis. Figure 4.10 gives a conceptual example reminiscent of Bello's generalized class shown in Figure 2.7.

$H'(\nu, \tau)$ is easily overspread by Kailath's criterion, as long as the carrier has enough harmonics and a large-enough fundamental frequency ω_0 . That is, the bounding rectangle for the support region has an area in the (ν, τ) plane exceeding 1. However, the true support region is disjoint with area far less than the bounding rectangle. Although we are not able to perfectly estimate $M(t, k]$, obtaining at best $M^2(t, k]$, the methods in this chapter point toward a realization of Bello's criterion.

4.7 Conclusion

The contribution of this chapter was to prove the necessity of impropriety for detecting a complex modulator within a subband. We motivated this problem in terms of a signal modulated at two principal rates. Namely, slow subband modulations appear on top of high-frequency, carrier modulations. The signal is therefore highly nonstationary, even within a short analysis window. This would seem to preclude conventional methods which assume quasi-stationarity. A coherent estimator, however, can synchronize with the carrier modulations in order to produce slowly-modulated, complex subbands.

In this way, we transformed the problem of system identification to a problem of complex demodulation. Underlying our derivations is the key assumption of separability between carrier and system modulations, which took on a Bedrosian-like form. By applying these concepts, we were able to estimate a time-varying spectral representation for synthetic, stochastic signals. The representation is complex, which preserves the relative phasing of carrier components in a signal.

On the surface, the methods discussed here are similar to common pitch-synchronous spectral analysis. Our contribution is a new statistical definition for synchrony. The gener-

ality of impropriety suggests an extension beyond the concept of pitch for a broader class of signals. We introduced the complementary envelope, related to the sidebands of a real-valued square law, as a feature related to impropriety and not necessarily to pitch.

The pitch-synchronous model is simply a logical consequence of certain axiomatic conditions. First, we chose an STFT analyzer with uniformly spaced subbands. Second, we assumed the stochastic demodulation model, which treats each subband as a modulated, stationary process. Therefore, the Hermitian and complementary envelopes are sufficient statistics for an LTV system excited by periodically correlated noise. Modifying one of these assumptions, by allowing non-uniform subbands or non-stationary subband carriers, will necessarily lead to a more flexible signal model. The next chapter addresses a relaxation of the uniform subband assumption, by taking the approach of principal components.

Chapter 5

THE PRINCIPAL COMPONENTS OF IMPROPRIETY

5.1 Introduction

This chapter marks a distinct shift from previous chapters. As usual, we are motivated by the question of defining the properties of improper and proper signals. Here, we reduce impropriety to its principal components. Gaussian signals are completely characterized by their covariances, and every covariance is uniquely determined by its eigenvalues. For proper signals, it is known that the eigenvalues appear in pairs [148]. Until now, the literature has paid little attention to the eigenfunctions, or principal components. As shown in this chapter, the principal components are like the “atoms” of impropriety, dictated fundamentally by quadrature, or Hilbert-transform, pairing. From studying the interaction of just one pair of principal components, we can define an algebra and a taxonomy for all second-order random processes.

This leads to a number of new insights, chief among them the unambiguous definition of the necessary and sufficient conditions for propriety. Other important issues are the separation of mixed improper and proper signals, and the removal of impropriety from a signal. An important realization is that removal of impropriety is not generally possible in a least-square sense. In other words, proper signals are not defined by a unique signal subspace. Impropriety is instead a feature of eigenvalue *balance* between paired dimensions in the signal subspace. This is akin to how different phase responses of a filter do not define new filtering subspaces. All of these important results are embodied by the taxonomy of random processes and its associated algebra.

It is perhaps useful to contrast our present method with the approaches taken in earlier chapters. Both Chapter 3 and Chapter 4 developed generative models whose sole function was to convert a proper signal, specifically random white noise, into an improper signal $y(t)$. Both chapters relied on the derivation of systematic relationships between complementary

statistics and functional components of the generative model. The models themselves, although rooted in fundamental LTV theory for Gaussian processes, assume a particular separability assumption which is sufficient but not necessary for generating impropriety. Instead of generative models, this chapter defines *both* necessary and sufficient conditions. We accomplish this by simply representing a signal $y(t)$ as a sum of non-interacting principal component pairs. Nevertheless, the elementary results derived in this chapter apply readily to the framework of coherent subband demodulation in Chapter 4.

The Chapter proceeds as follows. In Section 5.2, we construct the taxonomy of proper and improper processes from elemental examples. Then in Section 5.3, we formally derive the taxonomic results using a widely linear algebra. Finally, we conclude in Section 5.4.

5.2 Taxonomy of Improper and Proper Random Processes

Chapters 3 and 4 explained spectral impropriety in terms of a generative linear model. Here, we take a slightly different approach in terms of the principal components of a random process. Keeping in line with a time-frequency perspective, we treat the simple case of a signal comprising a sine and cosine of the same frequency. The random process, although real, is improper in the frequency domain when the cosine and sine are “unbalanced” in a way that forces a preferred time alignment across the ensemble. Progressively generalizing from the sinusoidal case, we will arrive at a full taxonomy for proper and improper signals.

It is first necessary to review some concepts before proceeding with the sinusoidal example. We have mentioned these concepts earlier, but it helps to collect the prerequisites in one place.

5.2.1 Review of Impropriety and the Hilbert Transform

We are concerned with a real random process $y(t)$ with Fourier transform $Y(\omega)$, which is itself a complex random process. From Definition 3.2.1, the process is spectrally improper, or simply improper, if $E\{Y(\omega)Y(\omega')\} \neq 0$ for any $0 < \omega, \omega' < \infty$.

Spectral impropriety is equivalent also to analytic impropriety. That is, $y(t)$ is spectrally improper if and only if its analytic signal is improper. We will use the analytic signal often

in this section. It is the complex signal defined as

$$y_a(t) = \frac{1}{\pi} \int_0^{\infty} Y(\omega) e^{j\omega t} d\omega. \quad (5.1)$$

It follows that $y(t) = \text{Re}\{y_a(t)\}$. Furthermore, the imaginary part of the analytic signal is an invertible function of $y(t)$. Namely, $y_a(t) = y(t) + j\hat{y}(t)$, where $\hat{y}(t)$ denotes the Hilbert transform of $y(t)$. The Hilbert transform is an all-pass filter with the transfer function

$$\hat{Y}(\omega) = \begin{cases} jY(\omega), & \omega < 0 \\ -jY(\omega), & \omega > 0 \end{cases} \quad (5.2)$$

Finally, a real, deterministic signal $\phi(t)$ is orthogonal to its own Hilbert transform. From Parseval's theorem,

$$\int \phi(t)\hat{\phi}(t)dt = \frac{1}{2\pi} \int \Phi(\omega)\hat{\Phi}^*(\omega)d\omega \quad (5.3)$$

which evaluates to

$$\begin{aligned} 2\pi \int \phi(t)\hat{\phi}(t)dt &= -j \int_{-\infty}^0 |\Phi(\omega)|^2 d\omega + j \int_0^{\infty} |\Phi(\omega)|^2 d\omega \\ &= 0 \end{aligned} \quad (5.4)$$

since $|\Phi(\omega)|^2$ is a symmetric function.

We are now ready to begin the construction of improper signals based on the theory of Fourier and Hilbert transforms.

5.2.2 A Motivating Example: Sinusoidal Basis

Suppose $y(t)$ is a sinusoidal random process defined by

$$y(t) = s_1 \cos(\omega_0 t + \theta_0) - s_2 \sin(\omega_0 t + \theta_0) \quad (5.5)$$

where s_1 and s_2 are independent, zero-mean, real, and Gaussian random variables. When is $y(t)$ spectrally improper? The analytic signal is

$$\begin{aligned} y_a(t) &= (s_1 + js_2)e^{j\theta_0} e^{j\omega_0 t} \\ &= Y e^{j\omega_0 t} \end{aligned} \quad (5.6)$$

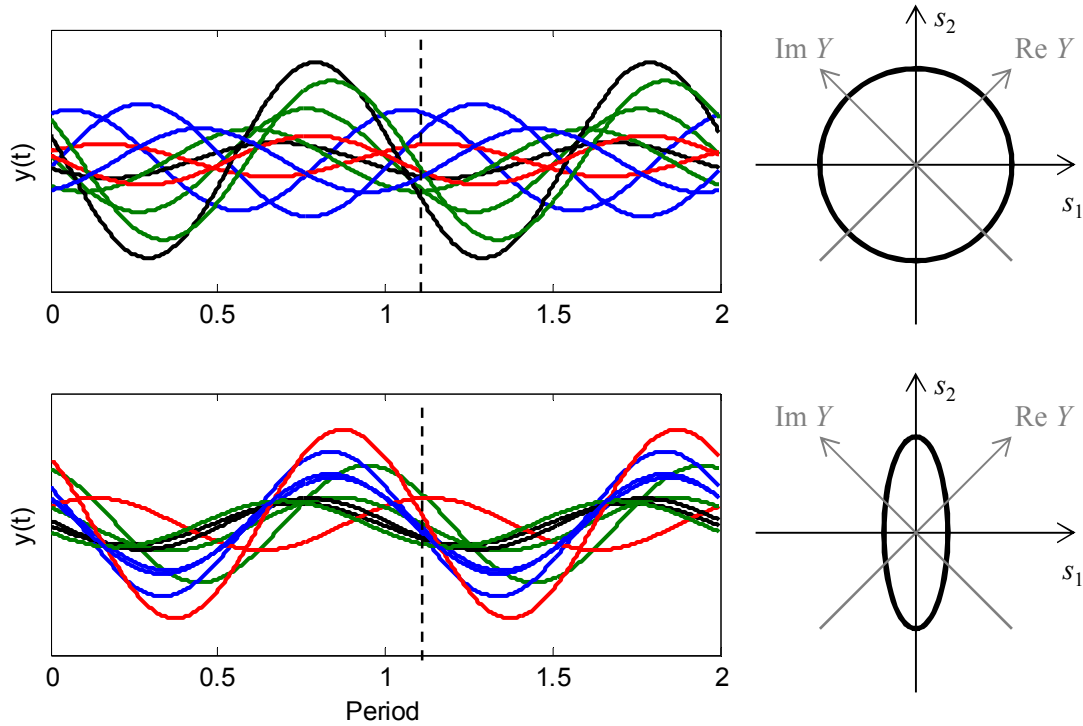


Figure 5.1: Independent realizations of the sinusoidal process for the proper case (top, $\sigma_2^2 = \sigma_1^2$) and an improper case (bottom, $\sigma_2^2 \gg \sigma_1^2$), with corresponding Fourier coefficient distributions to the right. The phase delay, shown as a dotted line, is discernible only in the improper case.

where Y is a zero-mean, complex Gaussian random variable. Thus, $y_a(t)$ is improper if and only if Y is improper. Since $s_1 \perp s_2$, the complementary variance of m is

$$E\{Y^2\} = e^{j2\theta_0} (\sigma_1^2 - \sigma_2^2) \quad (5.7)$$

which means impropriety results if and only if $\sigma_1^2 \neq \sigma_2^2$. Impropriety is therefore a function of the power imbalance between coefficients s_1 and s_2 . However, we note that the real and imaginary parts of Y are correlated, but is decorrelated by the phase counter-rotation $e^{-j\theta_0}$.

Let us take one extreme, where $\sigma_1^2 = \sigma_2^2$. In this case, Y is proper and $y(t)$ is wide-sense stationary. The autocovariance is

$$R_{yy}(t, \tau) = \sigma_1^2 \cos(\omega_0 \tau) \quad (5.8)$$

which follows from the trigonometric formula, $\cos \alpha \cos \beta + \sin \alpha \sin \beta = \cos(\beta - \alpha)$. Another way of interpreting wide-sense stationarity is the uniform phase distribution. For proper Y , it is well known that the magnitude is Rayleigh distributed and the angle is uniformly distributed in the interval $[0, 2\pi]$. In other words, $y(t)$ has the equivalent form

$$y(t) = |Y| \cos(\omega_0 t + \theta_0 + \angle Y) \quad (5.9)$$

where $\angle Y \sim U[0, 2\pi]$. Therefore, $y(t)$ has no determinate phase, and takes on all time offsets with equal probability. This explains the phase-agnosticism of WSS processes¹, which are determined by the power spectrum $E\{|Y|^2\}$. Figure 5.1 shows several independent realizations of $y(t)$ with proper M . The angle is $\theta_0 = -\pi/4$, but is indiscernible from the plots.

Impropriety, however, induces sinusoidal phasing across the ensemble. One maximally improper case is where $\sigma_1^2 = 0$ and $y(t) = s_2 \sin(\omega_0 t + \theta_0)$. The instantaneous variance is

$$E\{y^2(t)\} = \sigma_2^2 \sin^2(\omega_0 t + \theta_0) \quad (5.10)$$

which is a function of time and means $y(t)$ is nonstationary. Regardless of the amplitude, every realization of $y(t)$ has periodic zero-crossings for $t = (k\pi - \theta_0)/\omega_0$ and integer k . The average phase offset θ_0 is constant across the ensemble. As impropriety decreases, the time offsets of the realizations of $y(t)$ will appear increasingly “jittered.” This is akin to losing synchronization with an AC voltage on an oscilloscope. For $\sigma_1^2 \neq \sigma_2^2$, the distribution of the sinusoidal jitter is given by the Hoyt phase distribution [149]

$$p(\angle Y) = \frac{\sqrt{1 - b^2}}{2\pi [1 - \gamma \cos(2\angle Y)]} \quad (5.11)$$

where γ is the noncircularity coefficient. When either σ_1^2 or σ_2^2 is zero, this distribution approaches two spikes at $\angle Y = 0$ and $\angle Y = \pi$. Jitter vanishes, but the sign of the signal is a Bernoulli coin flip between positive and negative amplitude.

To summarize, a random sinusoidal signal is spectrally proper if and only if it is stationary in the time-domain. Furthermore, impropriety pertains to nonstationary sinusoidal

¹The orthogonality of Y and Y^* can also be derived directly from the assumption of uniformly distributed phase, which is essential to Fourier theory for WSS processes [128, pp. 127-128].

phasing. An increase in impropriety manifests as an increase in temporal alignment across the process ensemble. This relationship is just one special case of what we call a quadrature process, discussed next.

5.2.3 Quadrature Signal Basis

Noting that $\sin(t)$ is the Hilbert transform of $\cos(t)$, let us generalize the previous example to

$$y(t) = s_1\phi(t) - s_2\widehat{\phi}(t) \quad (5.12)$$

where $\widehat{\phi}(t)$ is the Hilbert transform of $\phi(t)$. By definition, $\widehat{\phi}(t)$ and $\phi(t)$ are an orthogonal pair. Again, s_1 and s_2 are independent, zero-mean, real, and Gaussian random variables. We find the analytic signal to be

$$\begin{aligned} y_a(t) &= (s_1 + js_2)\phi_a(t) \\ &= Y\phi_a(t) \end{aligned} \quad (5.13)$$

where $\phi_a(t) = \phi(t) + j\widehat{\phi}(t)$ is an analytic signal, and $Y = s_1 + js_2$ is a complex Gaussian random variable.

As in the previous subsection, $y_a(t)$ is improper if and only if Y is improper. Which is to say, if there is a power imbalance² between the real random variables s_1 and s_2 . Again, impropriety dictates the phase jitter over the ensemble. A simple example is the Gabor function

$$\phi(t) = g(t)\cos(\omega_0 t + \theta_0) \quad (5.14)$$

where $g(t)$ is a lowpass function with cutoff less than ω_0 . This is essentially one time-frequency “atom” [13] used, for example, in STFT decomposition. By Bedrosian’s product theorem [15], we have $\widehat{\phi}(t) = g(t)\sin(\omega_0 t + \theta_0)$. Figure 5.2 shows realizations of $y(t)$ for proper as well as improper cases. In the latter, the random “fine structure” jitters around the underlying periodicity of the basis function.

We refer to the Hilbert pairing of $\phi(t)$ and $\widehat{\phi}(t)$ as a *quadrature* basis. The term “quadrature” comes from communications theory, and refers to the fourth-period (i.e., $\pi/2$) phase

²Equivalently, Y has power imbalance and/or correlation between its real and imaginary parts.

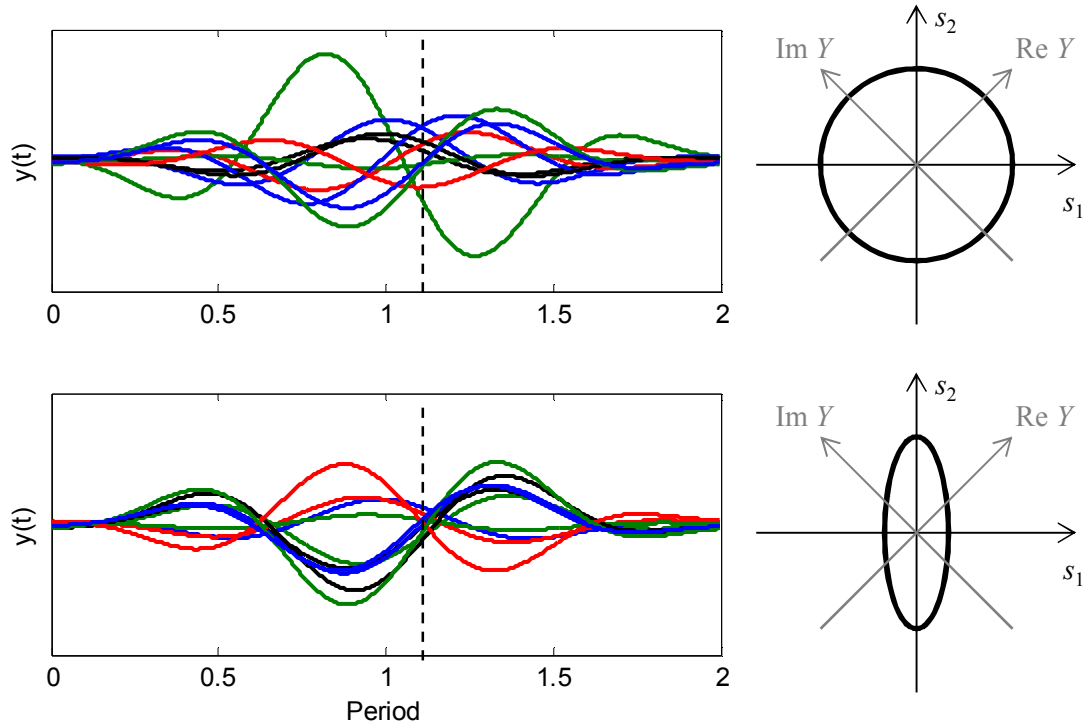


Figure 5.2: Independent realizations of the sinusoidal process for the proper case (top, $\sigma_2^2 = \sigma_1^2$) and an improper case (bottom, $\sigma_2^2 \gg \sigma_1^2$), with corresponding Fourier coefficient distributions to the right. The phase delay, shown as a dotted line, is discernible only in the improper case.

shift caused by the Hilbert transform. We can express $y(t)$ in terms of real components in quadrature, as in (5.12), which is essentially Rice's representation (2.7). Equivalently, $y(t)$ is the complex modulation

$$y(t) = \text{Re}\{y_a(t)\} = \text{Re}\{Y\phi_a(t)\}. \quad (5.15)$$

When Y is improper, we call $y(t)$ a coherent process because $y_a(t)$ has a meaningful complex envelope. The coherent detection operation is

$$\int y_a(t)\phi_a^*(t)dt = Y \quad (5.16)$$

which yields an improper random variable whose complementary variance is real. Of course, this assumes $\phi_a(t)$ is known. A more useful detector is $\phi_a(t, \theta)$, where θ is the parameterized

phase delay of the component. Let us assume

$$\int \phi_a(t, \theta_0) \phi_a^*(t, 0) dt = e^{j\theta_0} \quad (5.17)$$

which is satisfied by narrowband signals such as sinusoids and Gabor functions, each of the form $\phi_a(t, \theta) = g(t)e^{j(\omega_0 t + \theta)}$. The general detector is therefore

$$\int y_a(t) \phi_a^*(t, 0) dt = Y e^{j\theta_0} \quad (5.18)$$

where $Y e^{j\theta_0}$ is the complex envelope with respect to $\phi_a(t, 0)$. Improperity in Y , or power imbalance between s_1 and s_2 , is crucial for coherent detection of θ_0 . We find the phase by

$$\theta_0 = \frac{1}{2} \angle \rho_a^2 + (3 \pm 1) \frac{\pi}{2} \quad (5.19)$$

where ρ_a^2 is the complementary variance of $Y e^{j\theta_0}$, and the \pm term accounts for π -phase ambiguity in the square-root of a complex number.

To conclude this section, we show that a quadrature signal is WSS only when Y is proper and sinusoidal. Since the Hilbert transform is a linear time-invariant filter, it follows that $y(t)$ is WSS if and only if $y_a(t)$ is also WSS. Furthermore, it is known that improperity implies $y(t)$ is nonstationary [97]. We therefore must show that proper wide-sense stationarity is equivalent with the Fourier basis. When $y_a(t)$ is proper, the autocovariance is

$$R_{aa^*}(t, \tau) = \sigma_Y^2 \phi_a^*(t) \phi_a(t + \tau) \quad (5.20)$$

where $\sigma_Y^2 = E\{|Y|^2\}$. It is easy to show that $R_{aa^*}(t, \tau) = R_{aa^*}(\tau)$ when $\phi_a(t) = e^{j(\omega_0 t + \theta_0)}$.

To prove the converse, we first transform to the bi-frequency domain via

$$\begin{aligned} S_{aa^*}(\nu, \omega) &= \int R_{aa^*}(t, \tau) e^{-j\nu t} e^{-j\omega \tau} dt d\tau \\ &= \sigma_Y^2 \Phi_a^*(\omega - \nu) \Phi_a(\omega). \end{aligned} \quad (5.21)$$

Assuming stationarity, $S_{aa^*}(\nu, \omega) = \delta(\nu) S_{aa^*}(\omega)$, which requires $\Phi_a(\omega) = e^{j\theta_0} \delta(\omega - \omega_0)$ and hence $\phi_a(t) = e^{j(\omega_0 t + \theta_0)}$. This completes our proof.

To summarize, the quadrature basis is a superset of the sinusoidal, or Fourier, basis. A quadrature process is stationary if and only if it is proper *and* sinusoidal. This is notable because we have discovered a type of process which is nonstationary yet still proper, namely, a quadrature process with balanced non-sinusoidal components. In the next subsection, we discuss the final category of non-quadrature processes.

5.2.4 Non-Quadrature Signal Basis

The most general case for our two-component signal model is

$$y(t) = s_1\phi_1(t) - s_2\phi_2(t) \quad (5.22)$$

where the only constraint on $\phi_1(t)$ and $\phi_2(t)$ is that they are real, orthogonal functions. That is, $\int \phi_1(t)\phi_2(t)dt = 0$. As usual, s_1 and s_2 are independent, zero-mean, real, and Gaussian random variables. We find the analytic signal to be

$$y_a(t) = s_1\phi_{1,a}(t) + s_2\phi_{2,a}(t). \quad (5.23)$$

Since s_1 and s_2 are real, $y_a(t)$ is always improper, regardless of the power ratio between s_1 and s_2 . Invoking the contrapositive of Theorem 3.2.2, we know that this $y(t)$ is WSS under no circumstances. Furthermore, the lack of quadrature prevents an association of phase alignment with this model.

The category of non-quadrature processes is obviously a large one. With respect to impropriety and eigenvalue pairing, however, there are no obvious internal distinctions. The main discovery here is that only a quadrature process can be proper. In the next subsection, we combine the three classes into a taxonomy of second-order random processes.

5.2.5 Taxonomy of Paired Basis Elements

As in the previous subsections, we may categorize a basis pair as either sinusoidal, quadrature, or non-quadrature. Impropriety means something different for each class, especially the latter compared to the first two. We therefore suggest *balance/imbalance*, not propriety/impropriety, as the organizing dichotomy for second-order processes. A basis pair is balanced when $\sigma_1^2 = \sigma_2^2$. The resulting signal taxonomy appears in Figure 5.3, and is the primary result of this dissertation.

Figure 5.3 introduces a new term, “polarized,” for the class of unbalanced sinusoidal processes. We use polarization as an analogy to describe how an improper, sinusoidal processes is associated with systems which react differently to sine and cosine of the same frequency. The other categories should be familiar from the previous subsections. We emphasize the following important relationships:

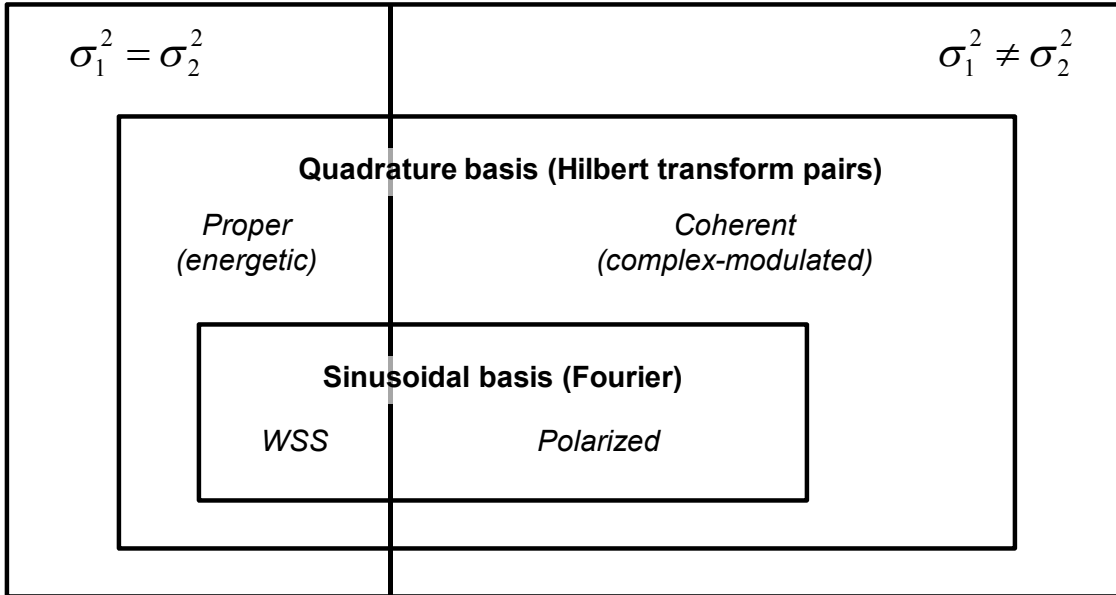


Figure 5.3: Taxonomy for all second-order random processes. On the left are signals with balanced elements, versus unbalanced on the right. The relative size of boxes is meant to roughly reflect the cardinality of the sets.

- The entire class of proper signals is subsumed by the quadrature class. As stated earlier, balanced quadrature is a necessary and sufficient condition for propriety.
- Every signal outside of the WSS category is nonstationary. Some nonstationary signals are still proper, notably the class of balanced quadrature signals.

Of course, some signals comprise more than just two basis elements. The taxonomy extends easily to any real, Gaussian $y(t)$ in terms of its principal components, or Karhunen-Loève (KL) series expansion. Signals consisting of more than just two orthogonal basis elements. We have

$$y(t) = \sum_k s_{k,1} \phi_{k,1}(t) - s_{k,2} \phi_{k,2}(t), \quad 0 \leq t \leq T \quad (5.24)$$

where the sum is possibly infinite and the principal components $\phi_k(t)$ are arranged into

pairs. By Mercer's theorem, we know that the basis functions are orthonormal, or

$$\int_0^T \phi_{k,i}(t)\phi_{p,q}(t)dt = \delta[k-p]\delta[i-q], \quad i = 1, 2 \quad (5.25)$$

and the random coefficients are uncorrelated and defined by

$$s_{k,i} \sim \mathcal{N}(0, \sigma_{k,i}^2), \quad E\{s_{k,i}s_{p,q}\} = \sigma_{k,i}^2 \delta[k-p]\delta[i-q], \quad i = 1, 2. \quad (5.26)$$

Due to Gaussianity, uncorrelated implies independence.

Equation (5.24) treats $y(t)$ as a sum of independently-weighted, non-interacting basis pairs. If the total number of principal components is a finite, odd number K , then the last basis pair simply has $s_{K,2} = 0$. Generally, $y(t)$ may comprise any number of sinusoidal, quadrature, and/or non-quadrature elements. These occupy three orthogonal subspaces,

$$y(t) = y_s(t) + y_q(t) + y_n(t) \quad (5.27)$$

where the subscripted signals contain sinusoidal, non-sinusoidal quadrature, and non-quadrature elements, respectively. In the same order, these subspaces correspond to the inner to outer concentric boxes in Figure 5.3. Of course, the principal components of $y(t)$ can be paired in any arbitrary fashion. Our point here is that sine/cosine and Hilbert-transform pairing are meaningful in terms of defining coherent time-frequency signal models. For $y_s(t)$ and $y_q(t)$, impropriety relates specifically to sinusoidal coherence.

To show the importance of the new signal paradigm shown in Figure 5.3, we will take a moment now to place it into historical context. Pioneering work by Wiener and Khintchine related Fourier analysis to wide-sense stationary processes in the early 20th century. This effectively demarcated the world of second-order processes into WSS and nonstationary, shown in Figure 5.4 on the left. Time-frequency analysis then evolved as the study of the "slowly-varying" superset of WSS. This is shown in Figure 5.4 on the right, and is exemplified by Priestley's evolutionary spectrum [84] and the spectrogram, to name a few.

Recent understanding of impropriety, due to Picinbono [111] and Schreier and Scharf [97], suggests a new taxonomy shown in Figure 5.5. This view is lacking in concrete statements about impropriety and different types of nonstationarity. A further complication arises from the paradigm of balance, initially due to [148] and used extensively in this section. It is not

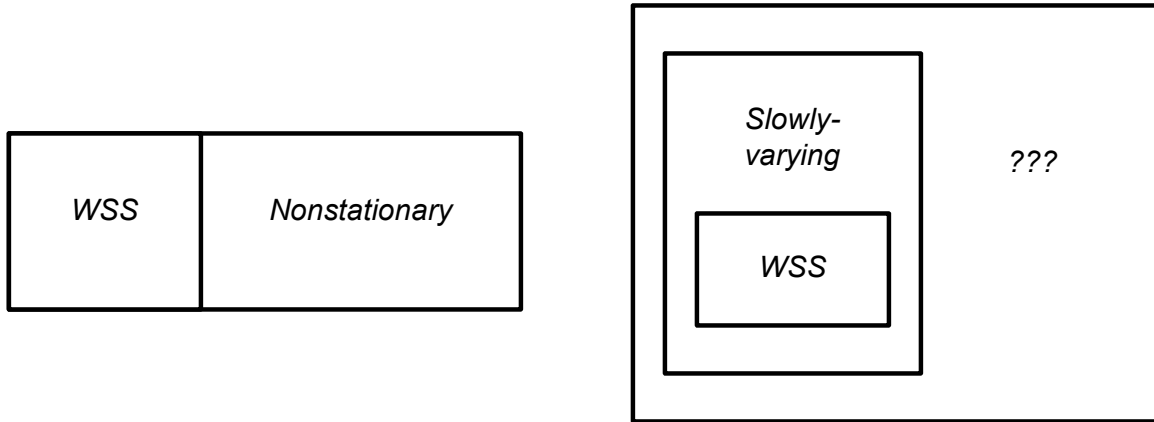


Figure 5.4: Early understanding of Gaussian processes as either WSS or nonstationary (left), but later expanded to allow slowly-varying, or quasi-WSS, processes (right).

clear how the case $\sigma_1^2 = \sigma_2^2$ relates to propriety other than the fact that all proper processes are balanced, although so too are some improper processes.

Our taxonomy in Figure 5.3 remedies the vagueness of impropriety, nonstationarity, and eigenvalue balance. It does so by simply grouping a signal's principal components into pairs based on their attributes. As a result, the dichotomy of power balance gives an appealing structure where every subclass for $\sigma_1^2 = \sigma_2^2$ has a fully defined counterpart in which $\sigma_1^2 \neq \sigma_2^2$. We see how coherent processing differs from conventional time-frequency analysis, the former being sensitive to local phase alignment.

There is a small matter remaining – that of the time limit T in Equation (5.24). A signal $y(t)$ defined over a finite time is by definition not WSS. We use the term to refer to a sinusoidal basis which is analogous to a true WSS process. If T is long enough, one may wish to consider $y(t)$ “effectively stationary” inside the timespan of interest. Either way, an orthogonal sinusoidal basis over T permits only frequencies for which $\omega = k2\pi/T$ for integer k . The KL expansion is therefore a Fourier-series expansion. This is a natural extension considering that we could also define $y(t)$ over all t but periodically correlated with period T .

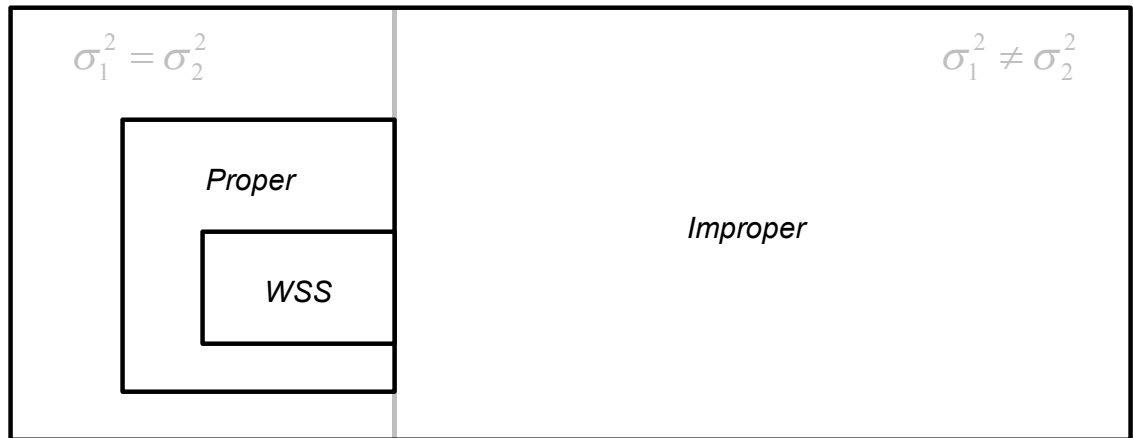


Figure 5.5: Modern, yet incomplete understanding of second-order processes, as either proper or improper, or alternatively, either balanced or imbalanced. However, propriety is only a sufficient condition for balance.

5.2.6 Removing Impropropriety from a Signal

The signal theory depicted in Figure 5.3 clarifies the difference between proper and improper signals. Even though the broader dichotomy is between balanced and unbalanced signals, propriety is interesting in its own right as the realm of noncoherent time-frequency estimation methods. A compelling question is, what do we lose when using a purely proper method of signal analysis? Conversely, what is to be gained from coherent analysis?

One way to address these questions is to systematically remove impropriety from a signal. We refer to this as “properization.” A possible application is for testing the perceptual importance of impropriety in images and audio. From the Karhunen-Loève framework of the previous subsection, one might be tempted to find a properizing filter, which would ideally project an improper signal to its nearest proper subspace. Closer inspection, however, reveals this is not possible in general.

Within the class of quadrature signals, propriety has only to do with the relative power balance of Hilbert-pair coefficients. An analogy is how the phase response of an LTI filter

does not determine new subspaces for the filter. Given the quadrature signal

$$y(t) = s_1\phi(t) - s_2\widehat{\phi}(t) \quad (5.28)$$

the only way to “properize” $y(t)$ is by adjusting the gain on s_1 with respect to s_2 . That is,

$$y_{proper}(t) = As_1\phi(t) - Bs_2\widehat{\phi}(t) \quad (5.29)$$

where A and B are constants satisfying $A^2\sigma_1^2 = B^2\sigma_2^2$. It is a simple exercise to show that the least-squares solution is the trivial $A = B = 1$, and that $y_{proper}(t)$ is by no means orthogonal to $y(t)$ since they occupy the same subspace.

Regardless, properization can proceed by finding A and B in (5.29) for each basis pair in $y_s(t)$ and $y_q(t)$, which are defined in (5.27). A possible constraint on A and B is to preserve overall energy by defining $A = \sqrt{(\sigma_1^2 + \sigma_2^2)/2\sigma_1^2}$ and $B = \sqrt{(\sigma_1^2 + \sigma_2^2)/2\sigma_2^2}$.

If $y(t)$ also has a non-quadrature component $y_n(t)$, then this is one case where subspace projection might properize $y(t)$. Assuming $y_s(t)$ and $y_q(t)$ are both proper, then the only improper component is $y_n(t)$, which is improper by necessity. The properized signal is therefore

$$y_{proper}(t) = y_s(t) + y_q(t). \quad (5.30)$$

The removal of $y_n(t)$ is, in fact, an orthogonal projection. That is, $y_{proper}(t)$ is orthogonal to $y_n(t) = y(t) - y_{proper}(t)$. It is also possible to convert $y_n(t)$ to a quadrature subspace via coordinate rotation and balancing of coefficients. Such an approach is rather ad hoc, however, with many possible configurations.

5.3 Augmented Principal Components Analysis

Up until now, we have discussed impropriety in terms of real-valued signal components in the time domain. Since impropriety is technically a property of the Fourier transform $Y(\omega)$, it is appropriate to develop a parallel theory for the frequency domain. This allows algebraic proofs for the taxonomic structure of 5.3, while emphasizing the observation that a “linear system in time is a widely linear system in frequency” from Section 3.5. The following is essentially a system algebra for proper and improper random processes.

5.3.1 Discrete Formulation of Improper Principal Components

The following discussion on principal component analysis (PCA) is easiest in vector notation. Supposing that $y(t)$ is effectively bandlimited, its discrete samples form the N -dimensional vector \mathbf{y} . We will also use discrete-time signal notation $y[n]$, for $0 \leq n < N$, when it is appropriate.

The previous section can be summarized succinctly as the eigenvalue decomposition

$$E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{R}_{yy} = \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}^T \quad (5.31)$$

where $\mathbf{\Phi}$ is a real, orthonormal matrix, and $\mathbf{\Sigma}$ is diagonal with real, non-negative entries. The k th column of $\mathbf{\Phi}$ is one time-domain function, $\phi_k[n]$. Our modification to PCA is defined as

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 \end{pmatrix} \quad (5.32)$$

where $\mathbf{\Phi}_i$ is an $N \times N/2$ matrix. The k th column of $\mathbf{\Phi}_1$ is paired with the k th column of $\mathbf{\Phi}_2$. The eigenvalue matrix is correspondingly

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{pmatrix} \quad (5.33)$$

where $\mathbf{\Sigma}_i$ has diagonal entries $\sigma_{k,i}^2$.

Based on the taxonomy of Figure 5.3, we give preference to quadrature pairing. That is, if the k th column of $\mathbf{\Phi}_1$ has a Hilbert-transform counterpart in $\mathbf{\Phi}$, then that counterpart is in the k th column of $\mathbf{\Phi}_2$. The remaining components without quadrature counterparts can then be paired arbitrarily, or by some other rule beyond the scope of this work. Finally, the eigenvalues are sorted accordingly, so that $\sigma_{k,1}^2$ is the power of the k th component in $\mathbf{\Phi}_1$ and so on.

In Section 5.2, we showed that spectral impropriety arises from imbalances between $\sigma_{k,1}^2$ and $\sigma_{k,2}^2$, but also for any non-quadrature principle components. In the Fourier transform of \mathbf{y} , it is possible to express (5.31) in terms of the widely-linear algebra devised by Schreier and Scharf [148]. That is the topic of the next subsection.

5.3.2 Widely-Linear Frequency-Domain View

To analyze propriety in the frequency domain, let \mathbf{F} denote the orthogonal, $N \times N$ matrix operator for the discrete Fourier transform (DFT). Since \mathbf{y} is real, we have conjugate-symmetry in the frequency domain. Therefore,

$$\underline{\mathbf{z}} = \begin{pmatrix} \mathbf{z} \\ \mathbf{z}^* \end{pmatrix} = \mathbf{F}^H \mathbf{y} \quad (5.34)$$

where $*$ indicates conjugation and H indicates conjugate (i.e., Hermitian) transpose. We assume that N is even so that \mathbf{z} is $N/2$ -dimensional.

Note that the DFT of \mathbf{y} is in augmented form (see Sec 2.4), which will allow convenient manipulation of Hermitian and complementary statistics later. As it is usually defined, the DFT is not compatible with augmented form, because it orders DFT coefficients circularly, and includes real coefficients for DC and the Nyquist rate. For our purposes, the modified DFT is

$$z[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y[n] e^{-j\pi n/N} e^{-j2\pi kn/N}, \quad 0 \leq k < N/2 - 1 \quad (5.35)$$

for N even. The modulation by $e^{-j\pi n/N}$ is such that the effective discrete frequencies are offset by half the DFT bandwidth. Also, $z[k]$ is defined for only the upper half of the unit circle in the complex-frequency plane. Refer to Figure 5.6 for an explanatory diagram. It is easy to show that the DFT defined here is unitary and hence, invertible³.

Since \mathbf{z} is Gaussian, the second-order statistics are completely surmised by the “augmented covariance” matrix,

$$\mathbf{\Gamma} = E\{\underline{\mathbf{z}}\underline{\mathbf{z}}^H\} = \begin{pmatrix} \mathbf{z}\mathbf{z}^H & \mathbf{z}\mathbf{z}^T \\ \mathbf{z}^*\mathbf{z}^H & \mathbf{z}^*\mathbf{z}^T \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{zz} & \mathbf{C}_{zz} \\ \mathbf{C}_{zz}^* & \mathbf{R}_{zz}^* \end{pmatrix} \quad (5.36)$$

where T is non-conjugated transpose. The augmented covariance has the special property of being widely-linear (WL) with respect to $N/2$ -dimensional vectors \mathbf{z} . We reviewed the algebra of WL matrices in Section 2.4 but cover the most essential relationships here.

³A non-unitary solution would be to discard DC and Nyquist coefficients from the usual DFT, since those frequencies are often not important.

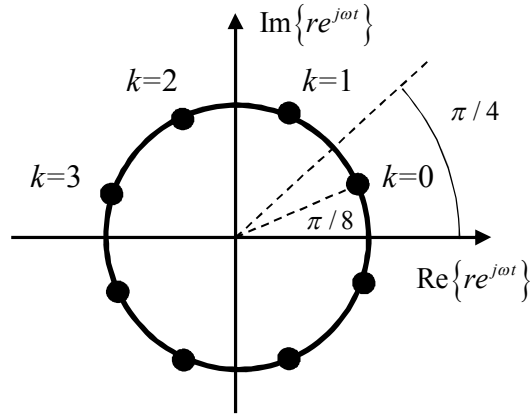


Figure 5.6: Samples of the modified DFT for $N = 8$, spaced uniformly around the unit circle by offset by an angle of $2\pi/16$.

Schreier and Scharf [148] discovered the WL eigen-factorization

$$\mathbf{\Gamma} = \mathbf{V}\mathbf{T}\mathbf{\Sigma}\mathbf{T}^H\mathbf{V}^H \quad (5.37)$$

where

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_2 \\ \mathbf{V}_2^* & \mathbf{V}_1^* \end{pmatrix}, \quad \mathbf{T} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{pmatrix} \quad (5.38)$$

where \mathbf{I} is the $N/2$ -dimensional identity matrix. The \mathbf{V} and \mathbf{T} matrices are orthonormal, meaning $\mathbf{V}\mathbf{V}^H = \mathbf{V}^H\mathbf{V} = \mathbf{I}$, and $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are diagonal matrices with real, non-negative eigenvalues $\sigma_{1,k}^2$ and $\sigma_{2,k}^2$. This decomposition is appealing in that every matrix in (5.37) is a WL operator. Hence, (5.37) is closed under the WL algebra. Another feature is that the eigenvectors of $\mathbf{\Gamma}$ are guaranteed to be conjugate-augmented. That is,

$$\mathbf{\Gamma} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H, \quad \text{where} \quad \mathbf{U} = \mathbf{V}\mathbf{T} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \\ \mathbf{U}_1^* & \mathbf{U}_2^* \end{pmatrix}. \quad (5.39)$$

It is easy to verify that \mathbf{U} is also orthonormal.

Our first, new contribution is the observation that the DFT of a real signal \mathbf{y} naturally obeys augmented form. Substituting (5.34) into (5.36), the bifrequency covariance is related to \mathbf{y} via

$$\mathbf{\Gamma} = E\{\mathbf{F}^H\mathbf{y}\mathbf{y}^T\mathbf{F}\} = \mathbf{F}^HE\{\mathbf{y}\mathbf{y}^T\}\mathbf{F}. \quad (5.40)$$

Substituting (5.31) into the above yields

$$\mathbf{\Gamma} = \mathbf{F}^H \mathbf{\Phi} \mathbf{\Sigma} \mathbf{\Phi}^T \mathbf{F}. \quad (5.41)$$

Since the eigenvalues of a matrix are unique, it follows that $\mathbf{\Sigma}$ is the same in (5.41) as in (5.37), and

$$\mathbf{U} = \mathbf{F}^H \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \\ \mathbf{U}_1^* & \mathbf{U}_2^* \end{pmatrix}. \quad (5.42)$$

The frequency-domain eigenvectors \mathbf{U} are simply the Fourier transforms of real time-domain principal components $\mathbf{\Phi}$. Through this relationship, augmentation in the frequency domain naturally formalizes the conjugate-symmetry of the DFT for real signals.

So far, we have connected a linear theory of time-domain signals to a widely-linear theory in the frequency domain. There is also a corresponding system theory, which we first introduced in Section 3.5. The following is a discrete formulation of the same principles. Let \mathbf{L} be an $N \times N$ matrix such that

$$\mathbf{\Gamma} = \mathbf{L} \mathbf{L}^H. \quad (5.43)$$

The complex signal \mathbf{z} is always expressible as

$$\begin{pmatrix} \mathbf{z} \\ \mathbf{z}^* \end{pmatrix} = \mathbf{L} \begin{pmatrix} \mathbf{w} \\ \mathbf{w}^* \end{pmatrix} \quad (5.44)$$

where \mathbf{w} is a proper, white, Gaussian random vector, and \mathbf{L} is a generative system. Therefore, \mathbf{L} is a WL operator in the frequency domain. In the time-domain, we have the equivalent linear relation

$$\mathbf{y} = \mathbf{F} \mathbf{L} (\mathbf{F}^H \mathbf{r}) = \mathbf{H} \mathbf{r} \quad (5.45)$$

where $\mathbf{H} = \mathbf{F} \mathbf{L} \mathbf{F}^H$ is a real time-domain linear operator, and \mathbf{r} is a real, white, Gaussian noise process.

As established in Section 3.5, a linear system in the (real) time domain is widely linear in the frequency domain. Additionally, the eigenfunctions $\mathbf{\Phi}$ and \mathbf{U} are related via Fourier transform. This summarizes an algebra of spectral impropriety, which prepares us for deriving several important theoretical results in the next subsection.

5.3.3 Algebraic Results for Spectral Impropriety

Our first main result is the following theorem.

Theorem 5.3.1. *A complex random vector \mathbf{z} is improper if and only if $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ and $\mathbf{U}_2 = \pm j\mathbf{U}_1$, where \pm is a column-wise.*

Proof. From the previous subsection, the principal components of \mathbf{z} satisfy

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{R}_{zz} & \mathbf{C}_{zz} \\ \mathbf{C}_{zz}^* & \mathbf{R}_{zz}^* \end{pmatrix} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H. \quad (5.46)$$

From (5.39), the complementary covariance is

$$\mathbf{C}_{zz} = \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{U}_1^T + \mathbf{U}_2\boldsymbol{\Sigma}_2\mathbf{U}_2^T. \quad (5.47)$$

If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ and $\mathbf{U}_2 = \pm j\mathbf{U}_1$, then

$$\mathbf{C}_{zz} = \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{U}_1^T - \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{U}_1^T = \mathbf{0} \quad (5.48)$$

which is the definition of proper. Thus, the stated conditions are sufficient for propriety.

Now, assume $\mathbf{C}_{zz} = \mathbf{0}$. The augmented covariance $\boldsymbol{\Gamma}$ therefore contains only \mathbf{R}_{zz} and its conjugate. It therefore seems that $\boldsymbol{\Gamma}$ should retain a twofold redundancy in its eigenvalues. Since \mathbf{R}_{zz} is always positive semi-definite, there exist a unitary matrix \mathbf{V}_0 and a real diagonal matrix $\boldsymbol{\Sigma}_0$ such that

$$\mathbf{V}_0^H \mathbf{R}_{zz} \mathbf{V}_0 = \boldsymbol{\Sigma}_0. \quad (5.49)$$

The eigenvectors of \mathbf{R}_{zz} also diagonalize $\boldsymbol{\Gamma}$, as seen by

$$\begin{pmatrix} \mathbf{V}_0^H & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_0^T \end{pmatrix} \begin{pmatrix} \mathbf{R}_{zz} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{zz}^* \end{pmatrix} \begin{pmatrix} \mathbf{V}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_0^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_0 \end{pmatrix} \quad (5.50)$$

Since the eigenvalues of a matrix are unique, we have $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Thus, propriety implies balanced eigenvalues, which was also shown by [148]. Working from (5.38) we note that

$$\mathbf{V}^H \boldsymbol{\Gamma} \mathbf{V} = \boldsymbol{\Sigma} \quad (5.51)$$

when $\Sigma_1 = \Sigma_2$. It follows that $\mathbf{V}_0 = \mathbf{V}_1$ and $\mathbf{V}_2 = \mathbf{0}$, as in (5.50) and still satisfying the WL operator defined by (5.38). Therefore,

$$\mathbf{U} = \mathbf{V}\mathbf{T} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \\ \mathbf{U}_1^* & \mathbf{U}_2^* \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{V}_1 & j\mathbf{V}_1 \\ \mathbf{V}_1^* & -j\mathbf{V}_1^* \end{pmatrix} \quad (5.52)$$

which is what we set out to prove. Thus, the stated conditions of $\Sigma_1 = \Sigma_2$ and $\mathbf{U}_2 = \pm j\mathbf{U}_1$ are necessary for propriety. \square

The next corollary connects Theorem 5.3.1 to a real time-domain signal, by simple application of the DFT.

Corollary 5.3.2. *A real random vector \mathbf{y} is spectrally proper if and only if $\Sigma_1 = \Sigma_2$ and $\Phi_2 = \pm\mathcal{H}\{\Phi_1\}$, where \mathcal{H} is the Hilbert transform operating on columns, and \pm is column-wise.*

Proof. From (5.42), we have the DFT relation

$$\mathbf{F}^H \Phi_1 = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_1^* \end{pmatrix} \quad (5.53)$$

and by the definition of the Hilbert transform (5.2),

$$\mathbf{F}^H \mathcal{H}\Phi_1 = \begin{pmatrix} -j\mathbf{U}_1 \\ j\mathbf{U}_1^* \end{pmatrix}. \quad (5.54)$$

Therefore, $\mathbf{U}_2 = -j\mathbf{U}_1$ is equivalent to Φ_2 being the Hilbert transform of Φ_1 . If $\mathbf{U}_2 = +j\mathbf{U}_1$, then Φ_1 is the Hilbert transform of Φ_2 . Either case signifies a quadrature phase differential of $\pi/2$ radians.

Since the DFT coefficients are exactly determined by the system of equations (5.34), \mathbf{y} is spectrally proper if and only if \mathbf{z} is proper. Therefore, \mathbf{y} is proper if and only if its principal components are balanced and in Hilbert-transform pairs. \square

These derivations also lead to:

Corollary 5.3.3. *Only a quadrature process can be proper. That is, \mathbf{y} consisting of principal components in Hilbert-transform pairs. Propriety then results when $\Sigma_1 = \Sigma_2$.*

This corollary is important because it finally defines the entire set of proper signals, and hence, the complement set of improper signals. Figure 5.3 displays the resulting taxonomy of second-order random processes, while the algebra developed here makes our claims complete.

5.3.4 Synthetic Examples

What do the taxonomic categories “look” like? Is it possible to recognize the categories without performing PCA? We will briefly address this question by presenting several synthetic examples. Let \mathbf{y} be one N -length realization of a real random process. After observing L realizations, let \mathbf{Y} be the $N \times L$ matrix of side-by-side observations. We estimate the time-domain covariance as

$$\hat{\mathbf{R}}_{yy} = L^{-1} \mathbf{Y} \mathbf{Y}^T \quad (5.55)$$

which converges to the true \mathbf{R}_{yy} as L goes to infinity. The bifrequency covariance estimate is therefore

$$\hat{\mathbf{\Gamma}} = L^{-1} \underline{\mathbf{Z}} \underline{\mathbf{Z}}^H = L^{-1} \mathbf{F}^H \mathbf{Y} \mathbf{Y}^T \mathbf{F} \quad (5.56)$$

which approaches the true augmented covariance $\mathbf{\Gamma}$ as L goes to infinity. In the following, we will observe the appearance of both covariance functions for the taxonomic categories of sinusoidal, quadrature, and non-quadrature.

The first example appears in Figure 5.7, which displays $\hat{\mathbf{R}}_{yy}$ and $\hat{\mathbf{\Gamma}}$ for a balanced sinusoidal (WSS) process and an unbalanced (polarized) process. In the WSS case, the time-domain covariance is Toeplitz, with a diagonal frequency-domain equivalent. This reflects that fact that the frequency increments of a WSS process are proper and uncorrelated, where the diagonal values are the power spectrum of the process. The polarized process, however, has nonzero complementary variances in the upper-right and bottom-left quadrants of the augmented frequency covariance matrix. In the time-domain, the process exhibits an interesting forward-backward correlation pattern, the significance of which is as yet unknown.

The second example appears in Figure 5.8, where the same process from Figure 5.7 has been windowed in the time domain with a Hamming window. The process has quadrature principal components, like the Gabor function (5.14). When the eigenvalues are

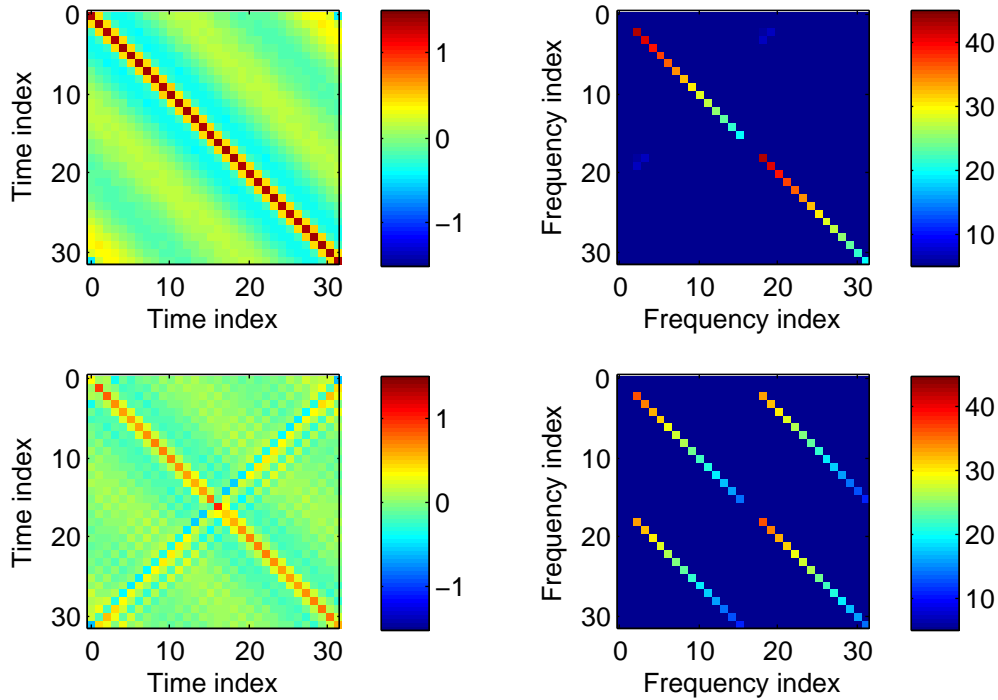


Figure 5.7: Time-domain (left) and dB-magnitude frequency-domain (right) covariance matrices for two sinusoidal random processes. Top: balanced, WSS. Bottom: unbalanced, improper polarized.

balanced, the spectral covariance is strictly proper yet non-diagonal, corresponding to a slowly-modulated time-domain covariance. Unbalanced eigenvalues, however, induces complementary correlation in the frequency domain as well as a highly disruptive modulation in the time-domain.

Our final example is in Figure 5.9, which features two non-quadrature processes. Without quadrature, there is no concept of balance or imbalance in the eigenvalues. Each process is improper in the frequency domain, which is consistent with the theoretical prediction that all non-quadrature processes are improper. The top example is composed by the Haar discrete-wavelet transform (DWT) basis, whereas the bottom example is the Daubechies-4

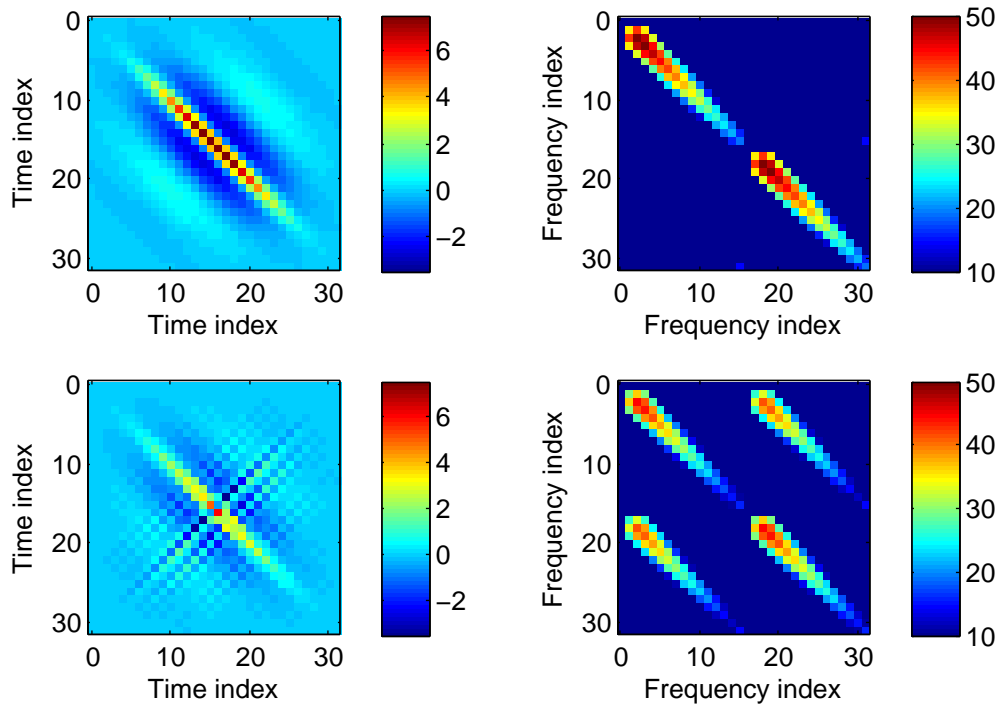


Figure 5.8: Time-domain (left) and dB-magnitude frequency-domain (right) covariance matrices for two quadrature random processes. Top: balanced, proper. Bottom: unbalanced, improper.

minimum-phase DWT basis.⁴

⁴Recently, others have reported on a Hilbert-transform pairing relationship within the Daubechies wavelet class [150] as well as non-minimum-phase wavelets [151]. The dual-tree complex wavelet transform (DT-CWT) is another example of constructing nearly orthogonal quadrature bases from classic DWT bases [152]. Implications of impropriety in these cases is a potential avenue for future work.

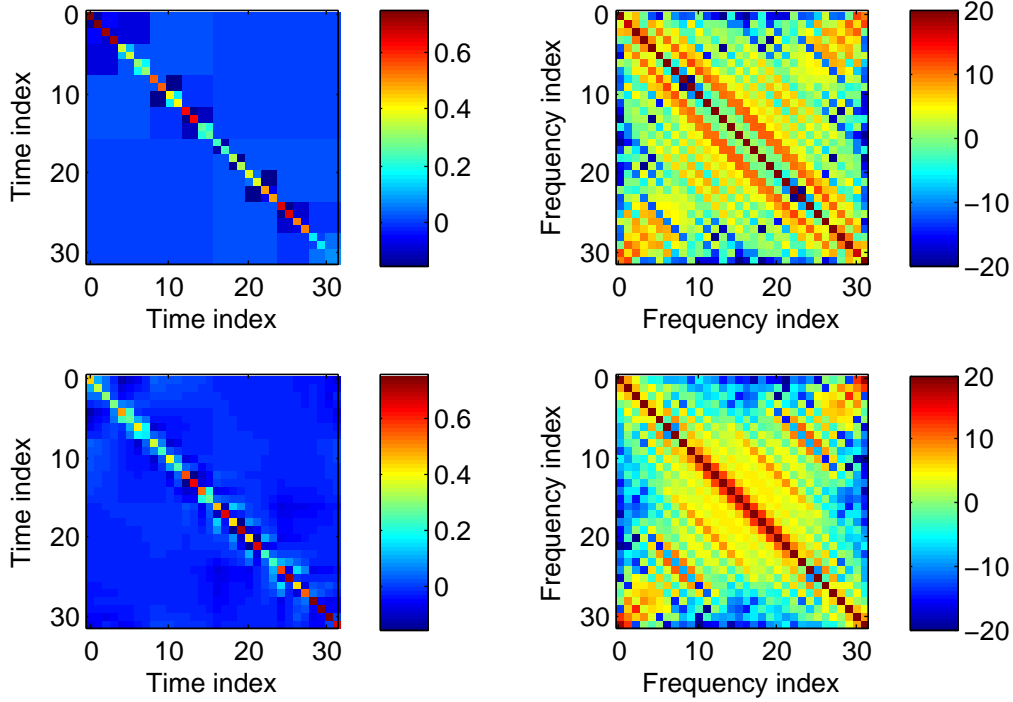


Figure 5.9: Time-domain (left) and dB-magnitude frequency-domain (right) covariance matrices for two non-quadrature random processes. Top: Haar basis, improper. Bottom: Daubechies-4 basis, also improper.

5.3.5 Eigenvalue Pairing for Separation of Signals

Our frequency-domain derivations rely heavily on the augmented covariance factorization originally from Schreier and Scharf [148]. Their paper showed that the factorization has the form

$$\mathbf{\Gamma} = \mathbf{V}\mathbf{T} \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{pmatrix} \mathbf{T}^H \mathbf{V}^H. \quad (5.57)$$

With respect to $\mathbf{\Gamma}$, the order of eigenvalues is completely arbitrary. However, Schreier and Scharf proposed arranging the eigenvalues such that $\sigma_{1,1}^2 \geq \sigma_{1,2}^2 \geq \sigma_{2,1}^2 \geq \sigma_{2,2}^2 \geq \dots \geq \sigma_{N,1}^2 \geq \sigma_{N,2}^2$. In other words, define $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ as the even- and odd-numbered entries in the list of descending-magnitude eigenvalues. This ordering ensures that if \mathbf{z} is proper, then $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$.

We instead recommend ordering eigenvalues based not on relative magnitude, but according to the characteristics of their respective eigenvectors. Specifically, quadrature pairing has priority. If the k th column of \mathbf{U}_1 has a quadrature counterpart, then that counterpart should be the k th column of \mathbf{U}_2 . Of course, quadrature pairing only makes sense if \mathbf{z} are the DFT coefficients of a real random process \mathbf{y} .

The importance of quadrature pairing is clear in the following example. A basic theoretical test for improper PCA is that it should be able to separate proper and improper signals in a mixture. Suppose $y[n]$ is a mixture given by

$$y[n] = y_{prop}[n] + y_{imp}[n] \quad (5.58)$$

where $y_{prop}[n]$ is spectrally proper $y_{imp}[n]$ is spectrally improper. Let us assume the signals are orthogonal, with respective KL decompositions

$$\begin{aligned} y_{prop}[n] &= s_1\phi_1[n] - s_2\widehat{\phi}_1[n] \\ y_{imp}[n] &= s_3\phi_3[n] - s_4\phi_4[n] \end{aligned} \quad (5.59)$$

where $\phi_3[n]$ and $\phi_4[n]$ may or may not be in quadrature. Using the ordering of [148], it is possible that σ_1^2 could be paired with the wrong eigenvalue. For example, s_3 may have the largest power followed by s_1 , which would pair together σ_1^2 and σ_3^2 . To separate proper from improper signals, however, we must instead pair $\phi_1[n]$ with its Hilbert transform, and thus σ_1^2 with σ_2^2 .

5.4 Conclusion

Proceeding by example and then algebraically, this chapter showed how both propriety and impropriety result in terms of paired principal components. The most fundamental result is that propriety is equivalent to quadrature component pairs with balanced power or variance. Thus, only a quadrature process can be proper. An example of a quadrature pair is the cosine and sine of the same frequency. Balanced sinusoidal components is equivalent to wide-sense stationarity, whereas unbalanced sinusoidal components lead to the new class of polarized random processes with phase alignment in the time domain. Also important is the fact that a complex modulator is meaningful only for an unbalanced quadrature

process. These results follow from a taxonomy of second-order processes, which is the main contribution made by this chapter. As a framework of eigenfunctions for temporal alignment, this taxonomic approach is clearly useful for synchronous analysis of rhythmic signals. It is possible to draw a connection to complex subband demodulation, as shown in Appendix E.

Chapter 6

REAL-WORLD EXAMPLES

6.1 Introduction

There is a compelling theoretical connection between spectral impropriety, a statistical property, and coherent phasing of carriers in real-valued signals. Modulation of this phasing characteristic is possibly important for representing “rhythmic” signals in which timing is structured yet partially random. There is no established definition for rhythm, but loosely, it is systematic temporal organization of signal-specific cues [153]. For example, speech is rhythmic at both glottal and syllabic time scales. In this Chapter, our first real-world example will be to detect impropriety in speech subbands. The implication is that, although a random process, speech harmonics have a statistically measurable phase reference. This approach could generalize to the syllabic time scale, but we will not discuss that here.

Another example of a rhythmic signal is cavitation noise produced by an underwater propeller. The signal is audibly rhythmic yet its spectrum is broadband. We will show that cavitation subbands are indeed improper, which suggests the existence of coherent phasing of high-frequency modulations, upwards to hundreds or thousands of Hertz. It is possible that this phase relation is synchronized to the much lower-frequency beating of the propeller rhythm. This is an interesting finding but the physical origin of impropriety is currently unknown. Our intent is simply to demonstrate its existence, and leave interpretation for future work.

In order to detect the presence of impropriety with confidence, we will assess statistical significance based on a hypothesis test. Under the null hypothesis, that the data is proper, an estimator has some distribution which is either known in closed-form or simulated by Monte Carlo analysis. This will allow us to establish a threshold for 2% probability of false-detection of impropriety.

We should emphasize that the following experiments detect impropriety in *real, scalar*

data. The internal timing of a signal manifests as improper statistics in the frequency domain or analytic signal. This is fundamentally different from the analysis of two-dimensional data such as planar wind velocity [124], and from nonlinear cross-correlation of distinct real signals [154]. For our purpose, impropriety arises from the internal coherent structure of a time series $y(t)$. Complex numbers are not merely a convenient 2D algebra, but rather an algebra of sinusoidal functions.

6.2 A Motivating Example

Before proceeding to data analysis, it is perhaps useful to summarize the main ideas of the previous chapters. Let us analyze one more synthetic example, which appears in the top panel of Figure 6.1. A power spectral estimate for $y[n]$, computed over 2 seconds of data, also appears in the bottom panel. The process is bandpass with center frequency around 1000 Hz. Although qualitatively sinusoidal, the amplitude and phase seem to vary randomly with about 200 Hz bandwidth (measured between -40 dB points). To what extent is this variation predictable? More fundamentally, what is the rank, or minimal representation for this signal? Is it possible to reduce $y[n]$ to a *single carrier frequency* with an elliptically distributed modulator?

The signal in Figure 6.1 is composed as follows:

$$y[n] = \sum_i f[n - iN/2]y_i[n] \quad (6.1)$$

where $f[n]$ is a N -length Hamming window. We call N the fundamental period of the process. The subcomponent $y_i[n]$ is defined by

$$y_i[n] = s_1[i] \cos(2\pi k_0 n/N + \theta_0) - s_2[i] \sin(2\pi k_0 n/N + \theta_0) \quad (6.2)$$

where $s_1[i]$ and $s_2[i]$ are mutually independent, real, white Gaussian processes. From Chapter 5, $y_i[n]$ is improper, for some i , if and only if $\sigma_1^2[i] \neq \sigma_2^2[i]$.

We argue that $y[n]$ is a “rank-1 process.” It consists of one carrier frequency, $\omega_0 = 2\pi k_0/N$, weighted by the elliptically distributed, complex modulator $s_1 + js_2$. “Rank” is a deliberate reference to the principal components analysis in Chapter 6. There are technically

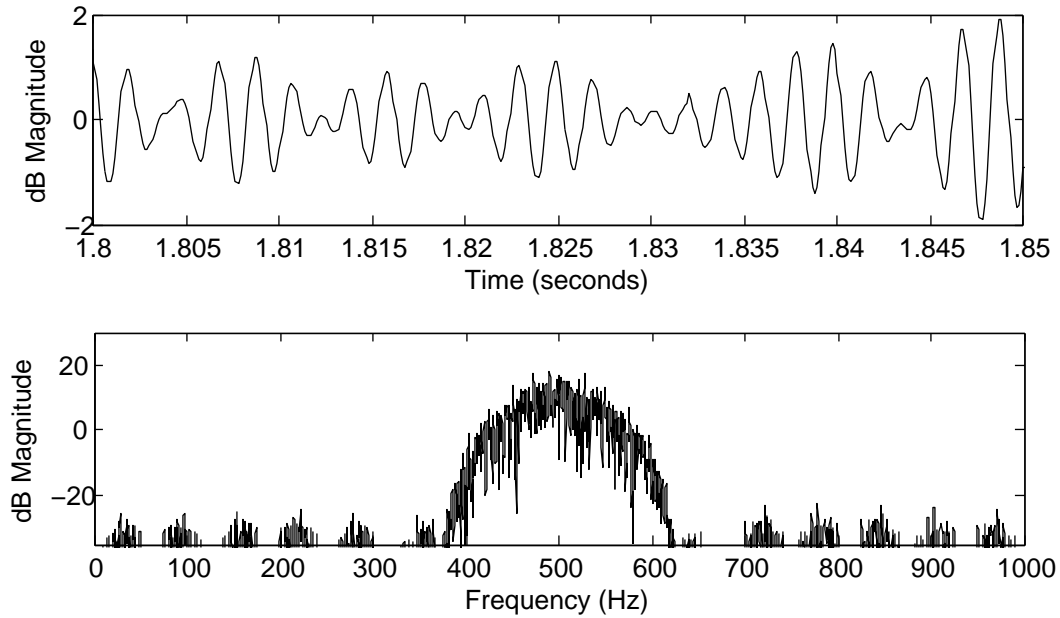


Figure 6.1: Time-domain view (top) of the random sinusoidal signal, which some would possibly call “quasi-periodic,” and average power spectral estimate (bottom) using a long-term periodogram.

other parameters on the side, namely the window $f[n]$ and the frame-rate $2/N$, but these can be attributed to the signal basis for $y[n]$.

Coherent demodulation reduces $y[n]$ to its essential components. First, square-law spectral analysis reveals a prominent spike at 2000 Hz in the Fourier transform of $y^2[n]$ (see the sideband plot of Figure 6.2). From Section 4.5 we ascertain that $y[n]$ is spectrally improper as a result of a 1000-Hz carrier. Furthermore, the 62.5-Hz spikes, seen at both baseband and sideband, correspond to the frame-rate. Let $Y[n, k]$ denote the discrete STFT of $y[n]$, using a N -length rectangular analysis window. Since $Y[n, k]$ is locked to the fundamental frequency, we call it a coherent STFT. Averaging $|Y[n, k]|^2$ and $Y^2[n, k]$ over n yields Hermitian and complementary spectral estimates. Both estimates appear in Figure 6.3, showing a distinct sharpness at 1000 Hz in the complementary spectrum. This is due to the elliptical distribution of $Y[n, k_0]$ for k_0 corresponding to 1000 Hz.

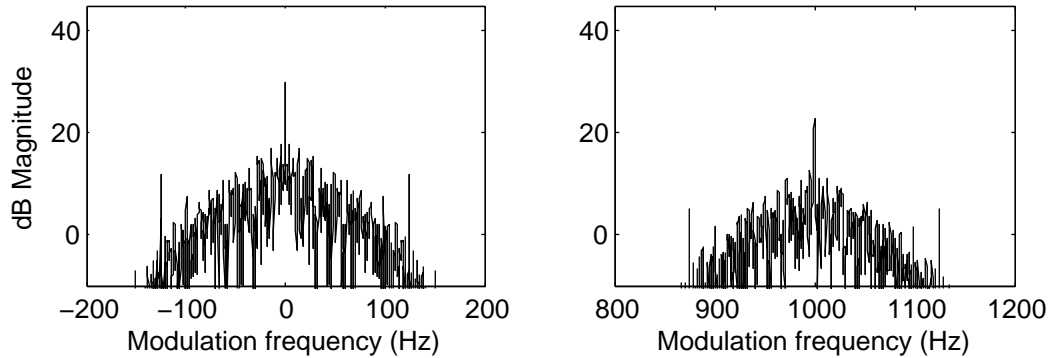


Figure 6.2: Square-law spectral analysis of the random sinusoidal signal, showing the base-band spectrum (left) and sideband spectrum (right). These are equivalent to the Hermitian and complementary envelope spectra for the analytic signal.

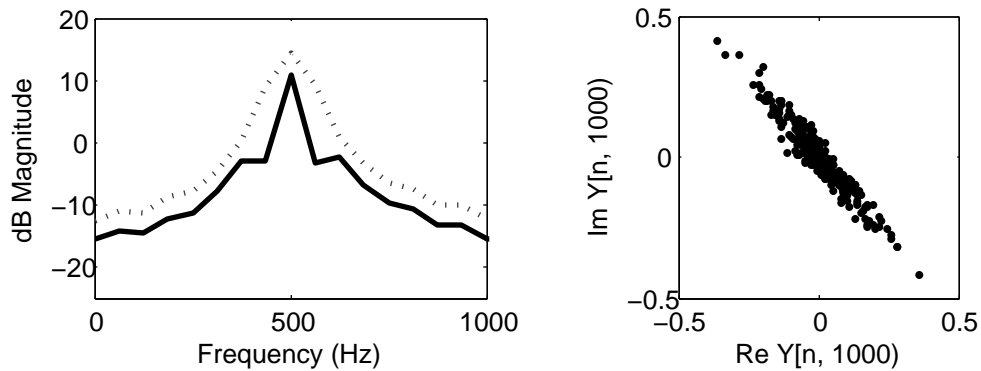


Figure 6.3: Left: synchronous Hermitian power spectrum (dotted) overlaid with the synchronous complementary spectrum (solid), both found through averaging across T -length STFT frames. Right: complex-plane scattergram for the STFT bin corresponding to 1000 Hz.

We have shown how a coherent STFT can reduce $y[n]$ to a nominal frequency and a complex modulator. This illustrates how the theory of spectral impropriety is a compact representation for demodulating random carriers. The reduction in bandwidth seen here – from 200 Hz to a single frequency – is much like dimensional reduction. Thus, coherent estimation minimizes the representation of an improper signal.

The rest of this chapter is devoted to detecting impropriety in two kinds of real-world signals: speech and underwater propeller noise. First, we require statistical tests to measure the significance of measured impropriety, as described next.

6.3 Hypothesis Testing for Subband Impropriety

We will use two estimators for detecting statistical impropriety. The first is the univariate noncircularity coefficient (NC), denoted $|\hat{\gamma}_z|$, which applies to a scalar random process $z[n]$. An example of such a process is a complex subband signal, with n as the discrete time variable. The second estimator is the multivariate generalized likelihood ratio (GLR), denoted \hat{L}_z , which applies to a multidimensional or vector random process, $\mathbf{z}[n]$. The n th vector can correspond to the n th cycle of a periodically-correlated subband signal, or to the n th frame of an STFT.

Details about these estimators can be found in Appendix C. The following discussion summarizes how these estimators perform in practice, based on the hypothesis-test framework developed in [155][156][157].

The most important thing to realize is that $|\hat{\gamma}_z|$ and \hat{L}_z are themselves random variables with variance depending on the sample size of estimation. We refer to the sample size as N , which is the length of $z[n]$ or $\mathbf{z}[n]$ such that $0 \leq n < N$. Nominally, the true NC $|\gamma_z|$ is identically zero if and only if $z[n]$ is proper. For finite N , however, the sample average $|\hat{\gamma}_z|$ can be nonzero even when $z[n]$ is truly proper.

Figure 6.4 illustrates the variability of $|\hat{\gamma}_z|$. One trial consists of generating two independent, N -length, white Gaussian sequences, one proper and the other improper with true NC equal to 0.6. Over 2000 trials, the normalized histograms estimate the probability distributions of $|\hat{\gamma}_z|$ for $N = 15$ and for $N = 45$. As seen, there is significant overlap between the proper and improper distributions for $N = 15$, indicating a low reliability for detecting impropriety when the true NC ≤ 0.6 . The situation improves substantially for $N = 45$, however. Asymptotically, the distributions become infinitely narrow and centered at the true NC values of zero and 0.6.

Walden and Rubin-Delanchy [158] were perhaps the first to use Monte Carlo tests (as in Figure 6.4) to estimate the distribution of $|\hat{\gamma}_z|$ under the null hypothesis H_0 . This refers

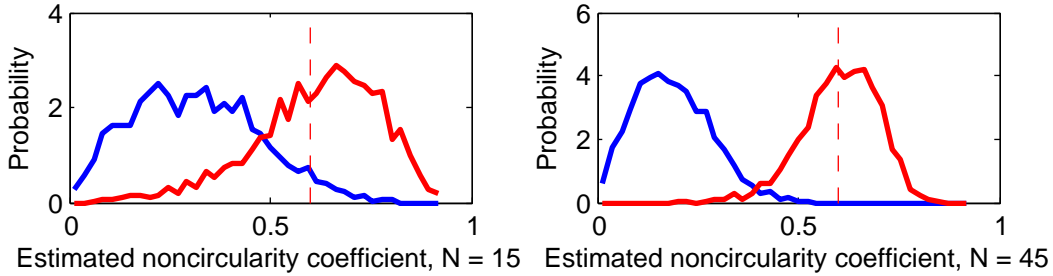


Figure 6.4: Normalized histograms for 2000 independent trials of computing $|\hat{\gamma}_z|$ for a proper (blue) and an improper (red) complex r.v. The dashed lines indicate the true NC for the improper sequence. Left: $N = 15$ samples, right: $N = 45$ samples.

to the hypothesis test for impropriety, defined as

$$\begin{aligned}
 H_0 & : z \text{ is proper, i.e., } \rho_z^2 = 0 \\
 H_1 & : z \text{ is improper, i.e., } |\rho_z^2| > 0
 \end{aligned} \tag{6.3}$$

where ρ_z^2 is the complementary variance of $z[n]$. Moving beyond Monte Carlo tests, Delmas *et al.* [157, 159] derived asymptotic distributions for the noncircularity coefficient under H_0 and H_1 , as functions of N . They found that the NC converges to Rayleigh and Gaussian variables for proper and improper cases, respectively (see Appendix C). It is therefore possible to view impropriety detection as a separation problem between two classes with N -dependent distributions.

The degree of separation is a function of the true impropriety of $z[n]$. We demonstrate in Figure 6.5, which displays estimator distributions for $N = 100$ and multiple degrees of impropriety. The text annotations indicate the true $|\gamma_z|$ used to synthesize the trials. The dashed vertical line shows the threshold for statistical significance. This means a measured impropriety of 0.29 or greater has a 2% of occurring under the null hypothesis. Therefore, for $N = 100$, an improper sequence with true $|\gamma_z| \leq 0.29$ is indistinguishable from a proper sequence. As suggested in Figure 6.4, larger N will cause increased class separability and therefore a lower minimum detectable impropriety.

The multivariate GLR behaves in the same basic fashion as the univariate NC, except

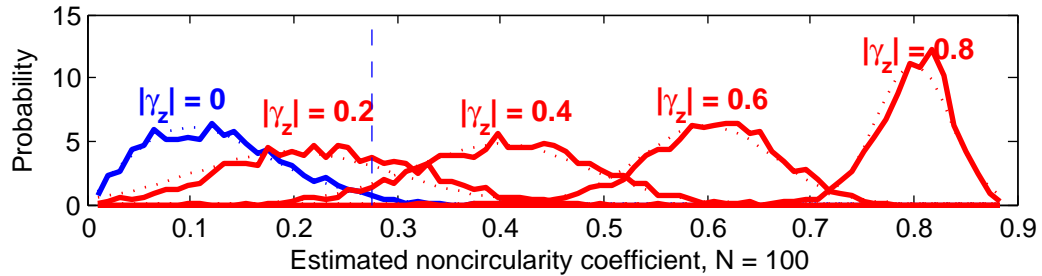


Figure 6.5: Asymptotic distributions (dotted) overlaid with Monte Carlo histograms (solid) for five values of $|\gamma_z|$. The H_0 , or proper, distribution is in blue. The 2% null-rejection threshold of 0.29 is indicated by the vertical dashed line.

asymptotic expressions are more difficult to come by. So for the GLR, we use Monte Carlo tests to determine the null distribution and null-rejection threshold.

6.4 Improperity Detection in Speech

In this section, we show that speech is highly improper and therefore coherently modulated. This is based on the fact that voiced speech can be represented as a sum of modulated harmonics (see Section 2.2). Harmonics are clearly visible in a narrowband spectrogram, as in Figure 6.6. Rather than use an STFT with fixed window length, it is more appropriate to use adaptive bandpass filters which track each harmonic over time. The resulting subband signals are nonstationary due to syllabic articulation, but we assume the modulators are slowly-varying. This allows calculation of the NC within windows 40-160 milliseconds long.

Given a continuous-time, real speech signal $y(t)$, let us define the complex subbands as

$$Y[n, k] = \int g(nT - t) y(t) e^{-j2\pi[kF_0(t)]t} dt \quad (6.4)$$

where $F(t)$ is the fundamental frequency in Hz, and $g(t)$ is the impulse response of a lowpass filter. As in Chapter 4.2, we assume the model

$$Y[n, k] = M[n, k] \cdot C[n, k] \quad (6.5)$$

where $M[n, k]$ are complex modulators and $C[n, k]$ are improper stochastic carrier processes. Unlike in previous chapters, we cannot assume $C[n, k]$ is stationary in n , because the car-

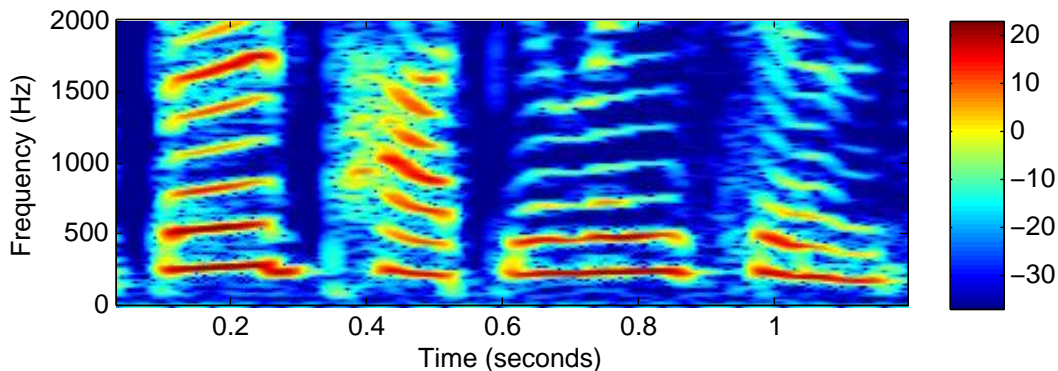


Figure 6.6: Narrowband spectrogram of female speech, “bird populations,” on a dB color scale.

rier phase is likely to have a random starting angle after each silent or unvoiced interval. However, it may be reasonable to assume $C[n, k]$ is stationary within continuously voiced segments.

Therefore, the proposed impropriety detector is the time-varying NC,

$$|\hat{\gamma}_k[n]| = \frac{|\sum_i Q[n-i] \cdot Y^2[i, k]|}{\sum_i Q[n-i] \cdot |Y[i, k]|^2} \quad (6.6)$$

where $Q[n]$ is a lowpass filter which computes a type of local average¹. Assuming $M[n, k]$ varies slowly in n , the bandwidth of $Q[n]$ should be less than or equal to the modulation bandwidth. If that is true, then

$$|\hat{\gamma}_k[n]| \approx \frac{|\sum_i Q[n-i] \cdot C^2[i, k]|}{\sum_i Q[n-i] \cdot |C[i, k]|^2} \quad (6.7)$$

which yields the time-varying noncircularity of $C[n, k]$, assuming it is constant within voiced segments.

To support the aforementioned assumptions, Figure 6.7 displays three basebanded harmonics². The ratio of real to imaginary amplitude is relatively stable within syllabic segments, compatible with locally stationary impropriety. The phase relation seems to change

¹the general form of this filter was introduced in Section 4.3.3 as a modulation-subspace projection.

²Carrier-frequency detection consisted of an initial harmonic pitch estimate followed by spectral center-of-gravity refinement of individual harmonics. See the Modulation Toolbox v2.1 [160] for code. “Pitch” is interpolated linearly through unvoiced parts.

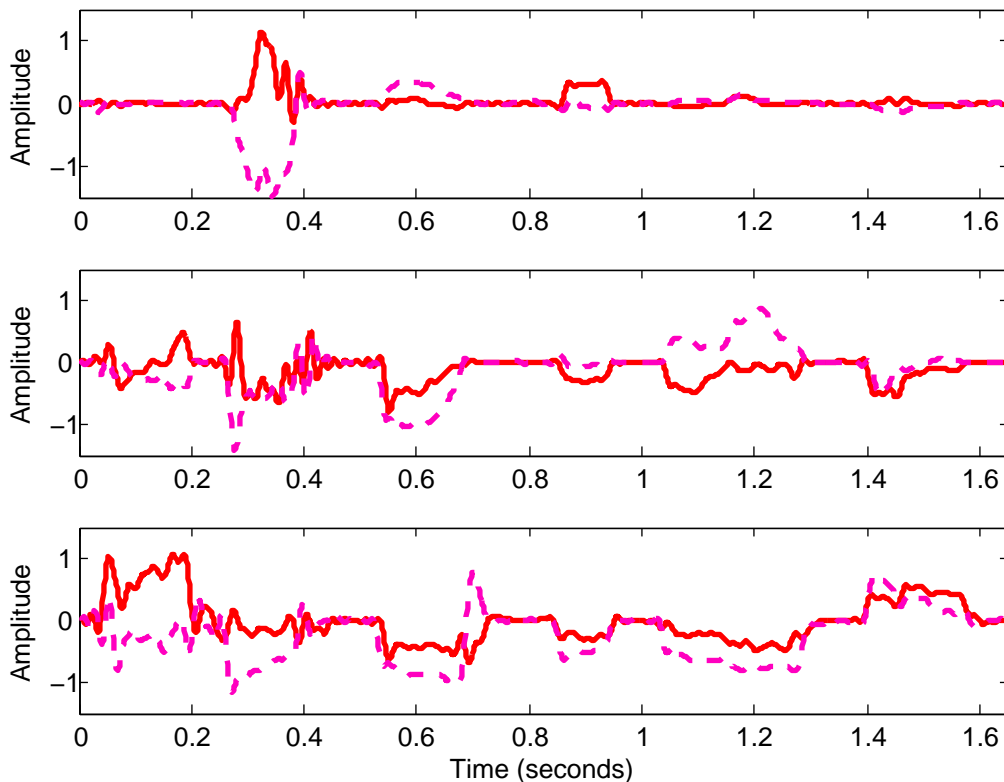


Figure 6.7: Real (red, solid) and imaginary (magenta, dashed) parts of three demodulated speech harmonics. The speech is the same as in Figure 6.6. Bottom: fundamental modulator. Middle: first harmonic. Top: second harmonic.

abruptly at syllabic onsets, however, which justifies the use of local averages. For instance, consider the middle plot. The real and imaginary parts are positively correlated between 0.6 seconds, yet negatively correlated around 1.2 seconds.

Despite the apparent nonstationarity of the basebanded harmonics, we can still detect impropriety with a relatively high degree of confidence. Figure 6.8 plots the time-varying NC for each harmonic, as defined in (6.6). The $Q[n]$ filter is a hamming window with one of three durations: 40, 80, and 160 milliseconds. At a sampling rate of 800 Hz, these correspond to averaging sample sizes of $N = 32$, $N = 64$, and $N = 128$. Using multiple estimators is a way

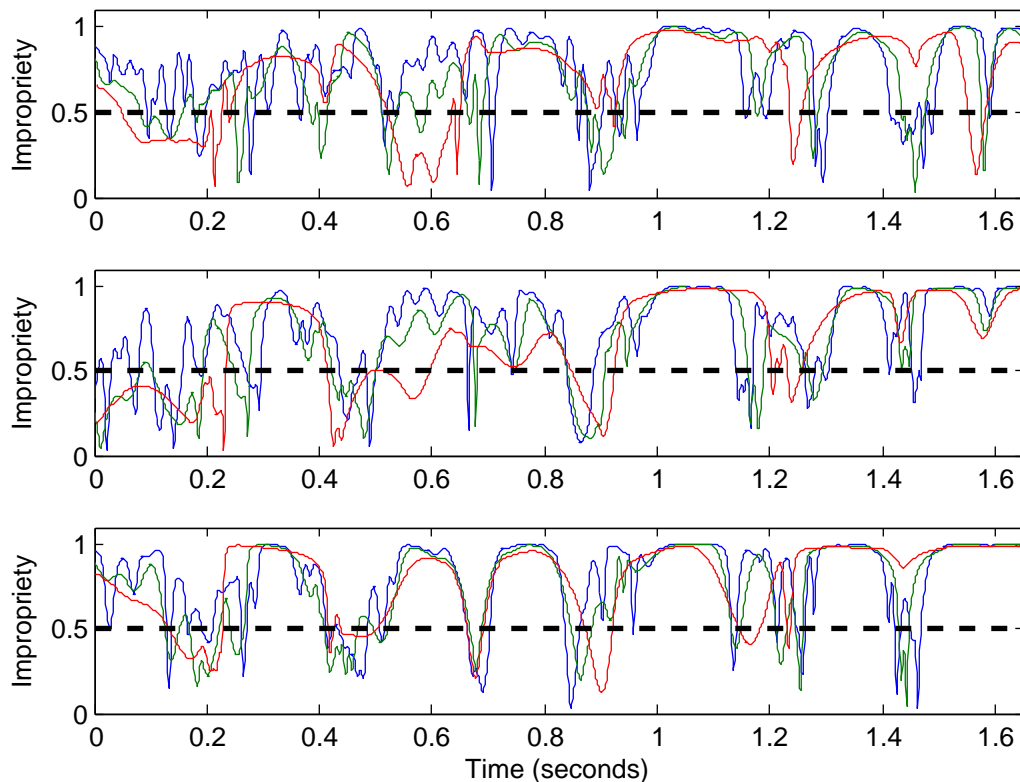


Figure 6.8: Local noncircularity coefficients for the three basebanded subbands in Figure 6.7. The three averaging lengths are 40 (blue), 80 (green), and 160 (red) milliseconds. The dashed lines indicate the null-rejection threshold, with a p-value of 2%, for the weakest estimator.

of dealing with the fact that the true modulation bandwidth or time-scale is unknown. The estimator with the weakest separability is $N = 32$, for which the 2% null-rejection threshold is 0.5 as indicated by the dashed lines. In each signal, all three detectors reveal high levels of local impropriety throughout the speech utterance. The cleanest detections occur in the bottom panel corresponding to the fundamental modulator. One can easily follow the syllabic rhythm of the female speaker’s phrase, “bird populations,” where unvoiced parts fall beneath the threshold and thus register as proper.

Finally, it is interesting to examine the square-law modulation spectra of the speech sig-

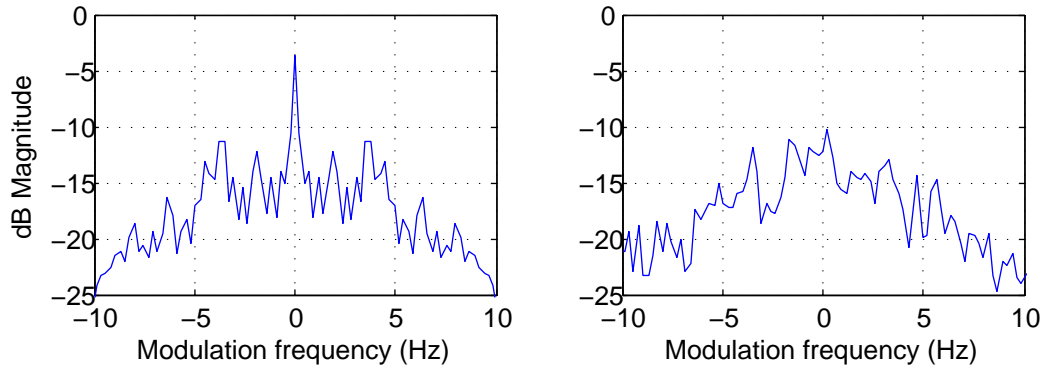


Figure 6.9: Square-law modulation spectra, averaged over all harmonics. Left: Hermitian, showing syllabic-like peaks at 1.9 and 3.6 Hz. Right: Complementary, showing possible traces of syllabic-like modulations.

nal. From Figure 6.8, the existence of impropriety implies that the complementary spectrum is meaningful. The *average* Hermitian and complementary square-law spectra are

$$\begin{aligned}
 D^{(H)}[\lambda] &= \sum_k \left| \sum_n |Y[n, k]|^2 e^{-j2\pi\lambda n/N} \right| \\
 D^{(C)}[\lambda] &= \sum_k \left| \sum_n Y^2[n, k] e^{-j2\pi\lambda n/N} \right|
 \end{aligned} \tag{6.8}$$

In Figure 6.9, the Hermitian modulation spectrum shows clear peaks at 3.6 Hz and possibly 1.9 Hz. This structure is commonly interpreted as syllabic frequencies [21, 25]. The new, complementary spectrum is less clear, but there are possible indications of syllabic-frequency content, particularly at -3.5 Hz. We hypothesize that the complementary modulation spectrum is still distorted by the abrupt phase transitions observed in Figure 6.7. In other words, the signal model put forward by Chapter 4, with a continuously periodic excitation, is not accurate for speech in general.

Speech is nevertheless often modeled as a sinusoidal signal, which makes it a natural candidate for spectral impropriety and coherent demodulation. In fact, the synthetic tonal example in Section 6.2 is an analogue for harmonic speech analysis. Next, we generalize the concept of impropriety beyond tonal signals, with the example of underwater propeller noise.

6.5 DEMONgram Analysis of Propeller Cavitation Noise

Propeller cavitation, introduced in 2.3.1, is modulated, broadband noise prominent in the acoustic signature for surface ships [64]. As shown in this section, however, cavitation noise can be improper even though it is non-tonal. We argue that this is an interesting phenomenon which merits future work in 1) determining the physical origins of impropriety, and 2) how to use complementary statistics for improved detection and classification in passive sonar applications.

Relating to point (1) above, the mathematical formalisms presented earlier in this thesis suggest possible interpretations of impropriety in cavitation noise. In an abstract sense, Chapter 5 shows how impropriety arises from non-sinusoidal yet quadrature-paired carriers. This, at least, is precedent of a non-tonal model for impropriety. Possibly more relevant is the signal model presented in Chapters 3 and 4, in which a periodically-correlated, random pulse train excites a linear system to produce an improper signal. Morozov [161] modeled cavitation as a random sequence of resonant pulses corresponding to individual bubble collapses. Also, Ida [162] recently studied the physics of pulses and inverted reflections between cavitating microbubbles. We merely hypothesize that, in cavitation noise, spectral impropriety is connected to possible coherent phasing of pulsatile activity.

For now, we are only concerned with the demonstration of statistically significant impropriety in cavitation signal subbands. In the next subsection, we describe the signal model by which we apply the multivariate GLR detector for impropriety.

6.5.1 Subband Product Model

Let $y[n']$ be real-valued hydrophone recording of underwater propeller noise, sampled at rate f_s Hz. We will focus on one complex subband given by

$$y_0[n] = \sum_{n'} g[nR - n'] y[n'] e^{-j2\pi n' f_0 / f_s} \quad (6.9)$$

where $g[n]$ is a lowpass filter and f_0 is the chosen center frequency of the subband in Hz. The downsampling interval R should be small enough compared to the bandwidth of $g[n]$ so as to prevent aliasing. As in the multiband DEMON method [82], we assume that the

bandwidth is narrow enough to ensure $y_0[n]$ is spectrally flat. In other words, it will appear as a white process for a large enough R . Therefore, the second-order characteristics are

$$E\{|y_0[n]|^2\} = \sigma_0^2[n] \quad \text{and} \quad E\{y_0^2[n]\} = \rho_0^2[n] \quad (6.10)$$

which are the Hermitian and complementary envelopes of the subband.

Due to the rotation of the propeller blades, cavitation noise is often assumed to be periodically correlated, or PC (see Section 2.3.1). We assume the product model

$$y_0[n] = m_0[n] \cdot c_0[n] \quad (6.11)$$

where $m_0[n]$ is a generally complex, periodic modulator, and $c_0[n]$ is a complex random process. Conventional DEMON deals with $\sigma_0^2[n]$ exclusively, which is real-valued and obliterates modulator phase. Here, we observe the complementary relation

$$\rho_0^2[n] = m^2[n] \cdot \rho_C^2[n] \quad (6.12)$$

where $\rho_C^2[n]$ is the instantaneous complementary variance of $c_0[n]$. If the carrier is improper, then the phase of the modulator is recoverable via $\angle m_0[n] = 1/2(\angle \rho_0^2[n] - \angle \rho_C^2[n])$. Our new contribution in subband DEMON analysis is the detection of impropriety, and hence the possibility of a complex modulator.

At this point we must be clear about the role played by the carrier in this model. To distinguish it from the modulator, we formally define

$$E\{|c_0[n]|^2\} = 1 \quad \text{and} \quad E\{c_0^2[n]\} = \rho_C^2[n] \quad (6.13)$$

so that the power fluctuation of the subband, $\sigma_0^2[n] = |m_0[n]|^2$ is attributed solely to the modulator. However, the detection of complex modulation depends on the existence of impropriety, embodied by the carrier phase reference $\rho_C^2[n]$.

The final important assumption is second-order periodicity. This applies to the modulator as well as the carrier. We assume there exists an integer N_m such that

$$\sigma_0^2[n + kN_m] = \sigma_0^2[n] \quad \text{and} \quad \rho_0^2[n + kN_m] = \rho_0^2[n] \quad (6.14)$$

for integer k . This implies that $m_0[n]$ is N_m -periodic and that $c_0[n]$ is periodically correlated with the same period.

6.5.2 Frequency-Sweep to Find Improper Subbands

It is of interest to quantify subband impropriety as a function of center frequency f_0 . Given a PC subband signal as defined in the previous subsection, we propose the vectorization

$$z[n, k] = y_0[kN_m + n], \quad 0 \leq n < N_m \quad (6.15)$$

so that n is the temporal position within a cycle, and k indicates which cycle. This array is equivalently represented as the vector process

$$\mathbf{z}[k] = z[n, k]. \quad (6.16)$$

The GLR can now be readily computed for $\mathbf{z}[k]$, with dimension N_m and sample size N , where $0 \leq k < N$. Refer to Appendix C for details. The GLR is a function of the center-frequency, $\widehat{L}_y(f_0)$, which allows a frequency-sweep for detecting improper subbands in $y(t)$.

Note that in the following examples, we will use the diagonal GLR, described in Appendix C, which assumes independence between the dimensions. This estimator is perhaps weaker than the full GLR because it neglects possible improper cross-correlations between dimensions of $\mathbf{z}[k]$. However, the diagonal estimator has far less variance for finite data, which is why we deploy it here.

Figure 6.10 presents frequency-sweep impropriety results for three data sources. In each case, we used conventional DEMON to estimate the propeller rates. which are 1.96, 6.96, and 1.53 Hz from top to bottom. Every subband is maximally downsampled, so the subband bandwidth is also the sampling rate of $y_0[n]$. Therefore, $N_m = B/F_m$, where B is the subband bandwidth and F_m is the propeller-rate, both in Hz. Also, $N' = T_{tot}F_m$, where T_{tot} is the total time duration of the subband, which is 15 seconds for every plot. Based on the dimension N_m and sample size N' , the dashed line represents the null-rejection threshold for a p-value of 2%. GLR values above this line are considered statistically significant indicators of subband impropriety. Information about the source data³:

- Top: passing merchant ship noise downloaded from [163], $F_m = 1.96$ Hz. In processing, $B = 200$ Hz, $N_m = 102$, $N' = 30$.

³We acknowledge the contribution of tanker and self-noise from Brad Hanson, at the Marine Mammal Program at the NOAA Northwest Fisheries Science Center, Seattle, WA. Used with permission.

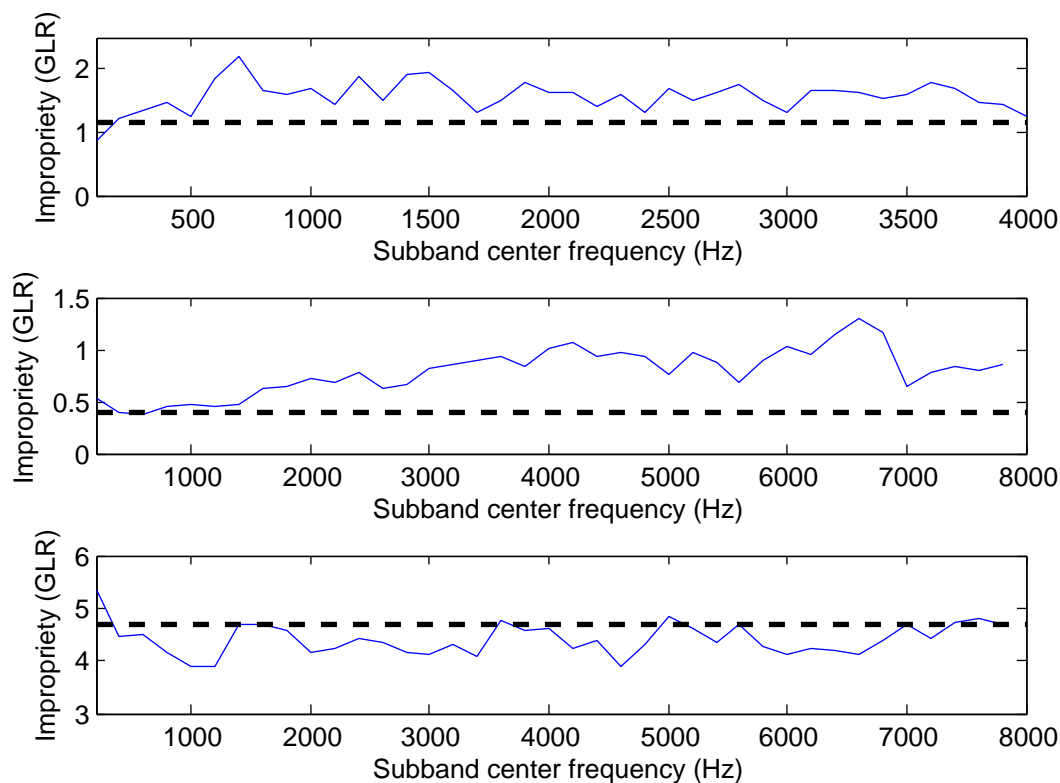


Figure 6.10: Impropriety measure as a function of subband center frequency, for three different ships and recording conditions. See text for details.

- Middle: passing tanker noise, $F_m = 6.96$ Hz. In processing, $B = 400$ Hz, $N_m = 115$, $N' = 104$.
- Bottom: self-noise of a research vessel, $F_m = 1.53$ Hz. In processing, $B = 400$ Hz, $N_m = 261$, $N' = 23$.

By self noise we mean the propeller noise of the vessel towing the hydrophone. The other two sources were recorded from an unknown distance.

As seen in the figure, impropriety is strongly apparent in the merchant and tanker examples. Furthermore, the degree of impropriety is dependent on the subband center

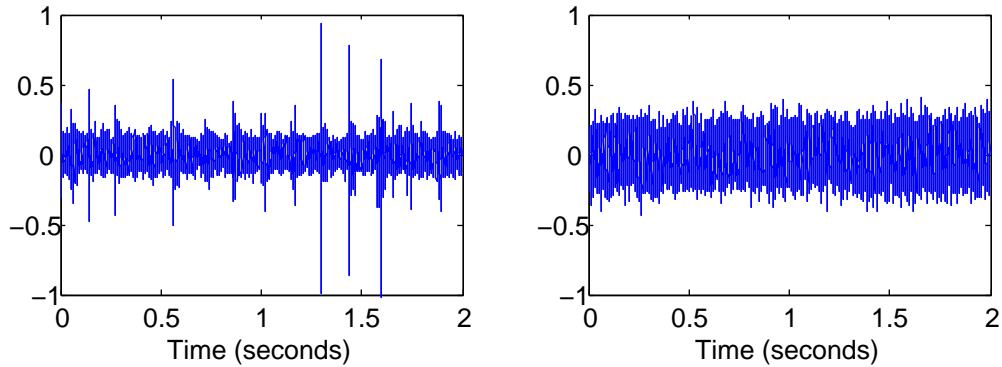


Figure 6.11: Sound pressure waveforms for the tanker (left) and self-noise (right) recordings used in Figure 6.10. Hard modulation bursts are prominent in the tanker data, suggesting a possible source of impropriety.

frequency. Measured impropriety also depends on subband bandwidth, partly because the GLR is a better detector with smaller N_m (see Appendix C). Despite trying several different bandwidths, the self-noise is consistently proper. A possible reason for this is the self-noise does not have obvious propeller slapping sounds, unlike the tanker noise recorded by the same hydrophone. It is possible that impropriety is related to consistent phasing of “hard” modulations which appear as localized bursts. Refer to Figure 6.11 for a visual comparison of the tanker and self-noise sound pressure waveforms. The exact relation of impropriety to the dimensions and speed of a propeller is one direction for future research.

6.5.3 The Complementary DEMONgram

Another form of analysis is a so-called “waterfall” DEMONgram display [66]. Familiar to naval sonar operators, the DEMONgram displays the DEMON spectrum as a function of time and modulation-frequency. The existence of impropriety leads to our new contribution, the complementary DEMONgram.

The computation of the DEMONgram proceeds as follows:

1. For some f_0 , filter the real-valued hydrophone data $y(t)$ into the complex subband $y_0[n]$, as in (6.9).

2. Segment $y_0[n]$ into overlapping frames, several seconds long.
3. For each frame, calculate the Hermitian and complementary square-law (DEMON) modulation spectrum.
4. Compute statistical significance of impropriety (and hence the complementary modulation spectrum) using the GLR and the periodic vectorization of (6.15).

DEMONgram displays appear in Figures 6.12, 6.14, and 6.16 for the merchant, tanker, and self-noise recordings, respectively. In all cases, the frame size is 15 seconds with a 5-second skip. The subband width is 200 Hz for merchant, and 400 Hz for tanker and self-noise. The GLR measurements, as function of frame offset, appear in Figures 6.13, 6.15, and 6.17, showing the null distribution in grayscale, GLR values in red, and the 2% p-value threshold as a dashed line. As seen, the merchant and tanker signals are thoroughly improper over the 90-second duration, whereas the self-noise data is consistently proper. Impropriety manifests as visible patterns in the complementary DEMONgrams, which are periodic at the propeller-rate in Figures 6.12 and 6.14 for the merchant and tanker.

The complementary modulation spectrum of cavitation noise has some interesting, and currently unexplained, properties. Whereas the Hermitian spectrum contains peaks at integer multiples of the shaft-rate, the complementary spectrum is less obvious. Figures 6.12 and 6.14 contain shallow ripples spaced at the respective rates of 1.96 and 6.96 Hz. The evidence of periodic structure suggests that the complementary modulation spectrum may contain useful information, such as features for vessel identification.

What is unexplained, however, is why the complementary spectrum is extremely diffuse. The DEMONgrams in the figures have limited scope, but the ripples in Figures 6.12 and 6.14 both spread uniformly throughout the bandwidth of the subband. The Hermitian modulation spectrum, on the other hand, is always concentrated in a lowpass fashion. Returning to our subband signal model,

$$y_0[n] = m_0[n] \cdot c_0[n] \tag{6.17}$$

the assumption that $m_0[n]$ is lowpass means $\rho_C^2[n]$ is a broadband signal. However, recall our definition where $\sigma_C^2[n] = 1$. What kind of random process has a flat Hermitian variance

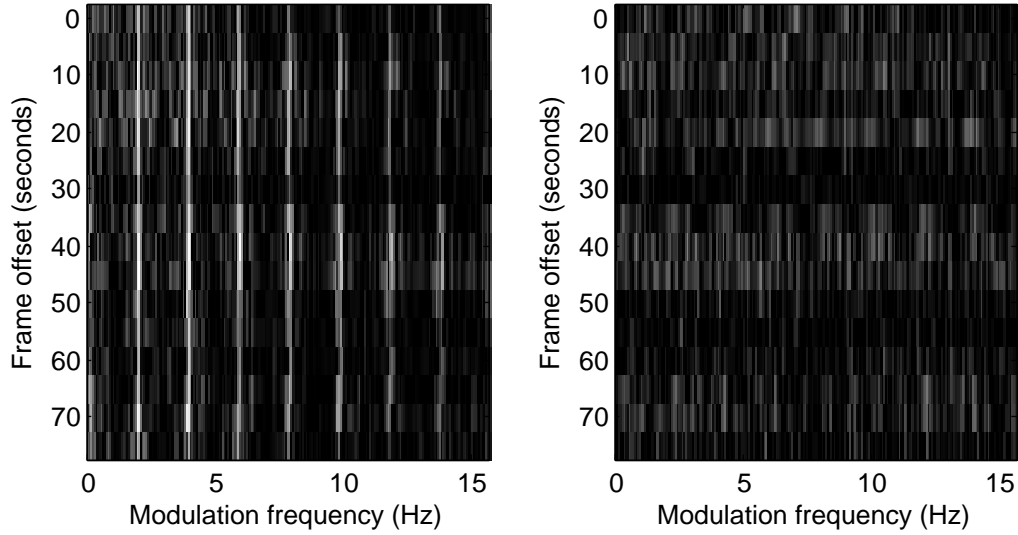


Figure 6.12: Hermitian (left) and complementary (right) DEMONgrams for merchant data, center-frequency = 450 Hz, bandwidth = 200 Hz, frame-length = 15 seconds. Impropriety is detected, as seen by complementary modulations.

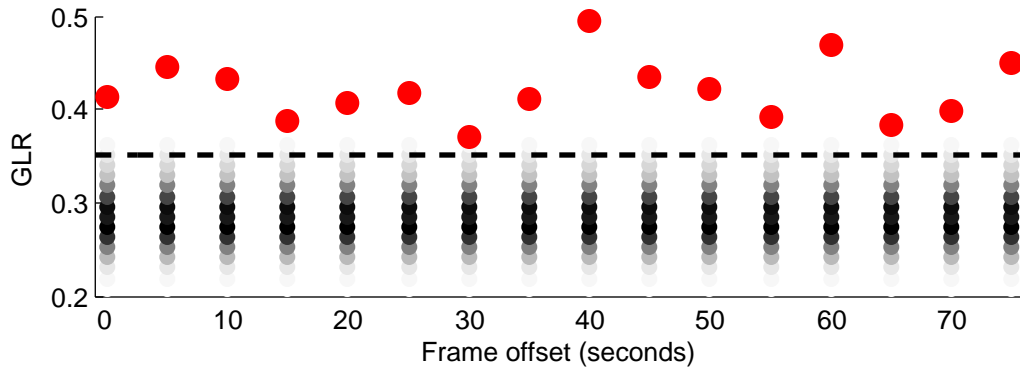


Figure 6.13: Frame-by-frame GLR (red dots) for the subband observed in Figure 6.12, with null distribution in grayscale and the 2% p-value threshold shown as a dashed line.

yet a broadband complementary variance? This is an interesting question, possibly related to the underlying physics of cavitation, but is beyond the scope of the present work.

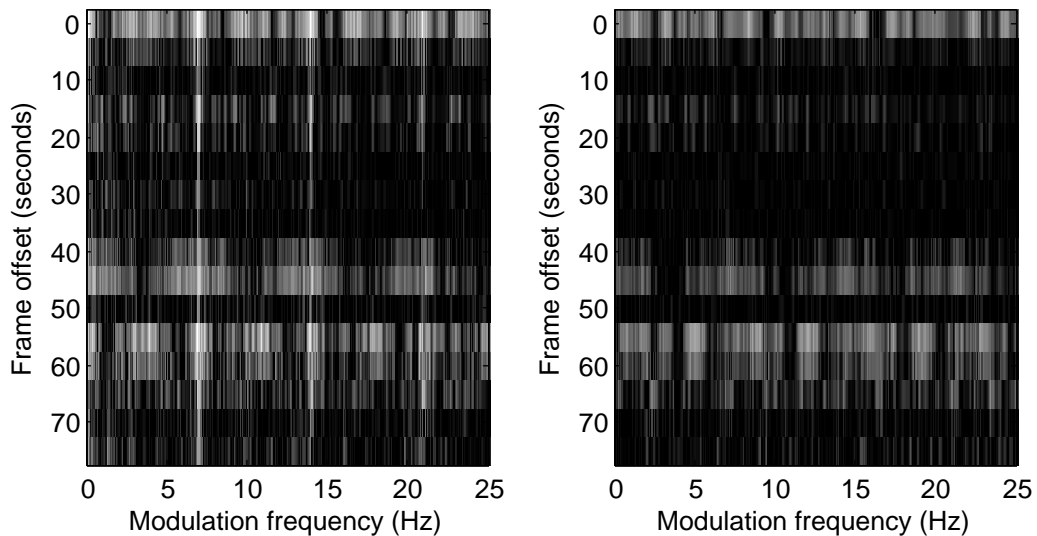


Figure 6.14: Hermitian (left) and complementary (right) DEMONgrams for tanker data, center-frequency = 2250 Hz, bandwidth = 400 Hz, frame-length = 15 seconds. Impropriety is detected, as seen by complementary modulations.

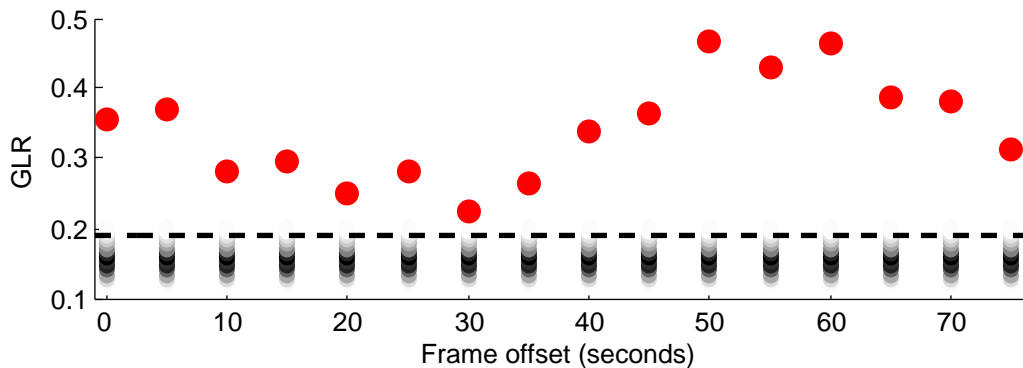


Figure 6.15: Frame-by-frame GLR (red dots) for the subband observed in Figure 6.14, with null distribution in grayscale and the 2% p-value threshold shown as a dashed line.

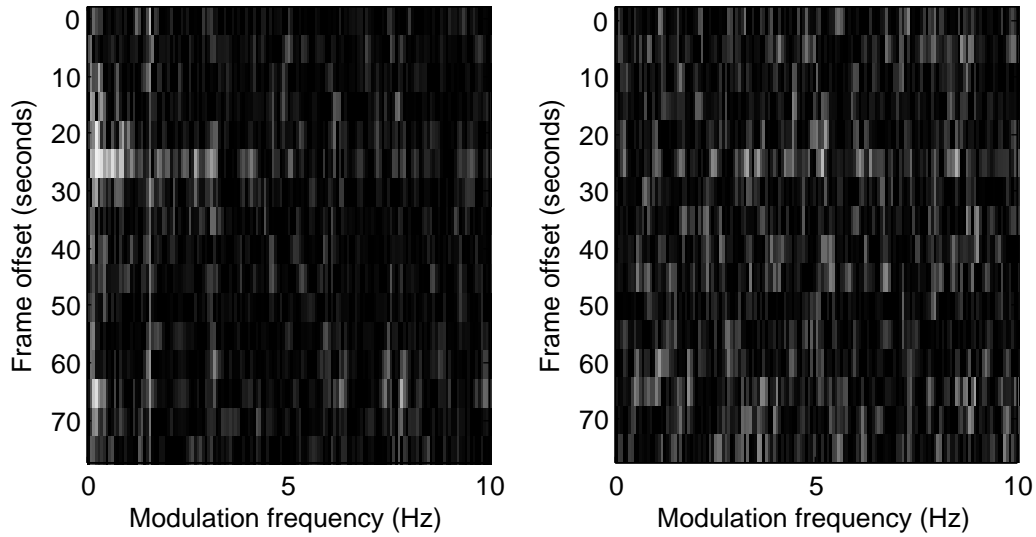


Figure 6.16: Hermitian (left) and complementary (right) DEMONgrams for self-noise data, center-frequency = 5000 Hz, bandwidth = 400 Hz, frame-length = 15 seconds. No inpropriety is detected.

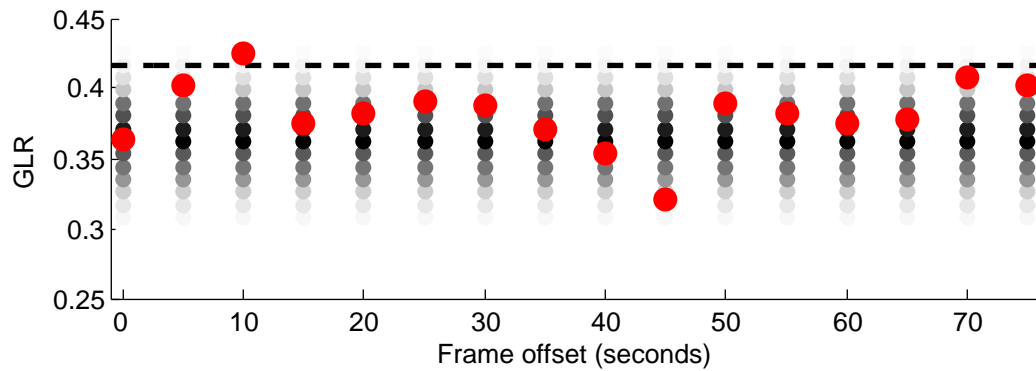


Figure 6.17: Frame-by-frame GLR (red dots) for the subband observed in Figure 6.16, with null distribution in grayscale and the 2% p-value threshold shown as a dashed line.

6.6 Coherent Enhancement of Propeller Cavitation Noise

To recap previous sections, we have demonstrated the statistical existence of impropriety in real-world signals. Impropriety is a necessary condition for coherent modulation and complex modulators. As seen in DEMONgram analysis, the complementary modulation spectrum for cavitation noise can be difficult to analyze or decode. In this section, we present an alternative method which makes fewer assumptions on the data. For instance, what if subbands are not the best way for extracting improper components? We propose augmented PCA as a means for learning the optimal signal space with respect to impropriety. Following the taxonomy of random processes in Chapter 5, we can partition the principal components into coherent and non-coherent subspaces. This allows a new type of filtering and possible signal enhancement. We will verify this approach with preliminary results on the merchant and tanker signals from the previous section.

6.6.1 STFT Covariance Factorization

Given a real discrete signal $y[n']$, let $Y[n, k]$ denote its short-time Fourier transform (STFT), defined as

$$Y[n, k] = \sum_{n'} g[nR - n'] y[n'] e^{-j2\pi kn'/K}, \quad 1 \leq k < K/2 - 1 \quad (6.18)$$

where $g[n]$ is a K -length, lowpass window, K is the (even) number of subbands spaced uniformly around the unit circle, and R is a downsampling factor. We will use $R = K/2$. Note that $Y[n, k]$ contains frequency samples only from the *upper* unit circle, excluding subbands at $k = 0$ and $k = K/2$. (One could instead use the first $K/2$ coefficients of the modified DFT defined in Section 5.3.)

As in Section 6.5, we treat $Y[n, k]$ as a complex random process with periodic correlation (PC). The difference here is that it is a vector process,

$$\mathbf{y}[n] = Y[n, k] \quad (6.19)$$

where the n th vector is the DFT at time n . Its augmented form is

$$\underline{\mathbf{y}}[n] = \begin{pmatrix} \mathbf{y}[n] \\ \mathbf{y}^*[n] \end{pmatrix}. \quad (6.20)$$

From this it follows that the augmented covariance matrix, combined with the assumption of periodic correlation, satisfies

$$E\{\underline{\mathbf{y}}[n]\underline{\mathbf{y}}^H[n]\} = \mathbf{\Gamma}[n] = \mathbf{\Gamma}[n + iN_m] \quad (6.21)$$

for integer i and integer period N_m . We should be clear that $\mathbf{\Gamma}[n]$ is the covariance between frequencies, or subbands, within a temporal neighborhood in the n frame. Cyclic estimation exploits periodicity by averaging over I cycles, as in

$$\widehat{\mathbf{\Gamma}}[n] = I^{-1} \sum_{i=0}^{I-1} \underline{\mathbf{y}}[n + iN_m]\underline{\mathbf{y}}^H[n + iN_m], \quad 0 \leq n < N_m. \quad (6.22)$$

Next, it is desirable to find the principal components, or basis functions. From Section 5.3, factorization of $\mathbf{\Gamma}[n]$ leads to the synthesis formula

$$Y[n, k] = \sum_{p=1}^P s_{p,1}[n] U_{p,1}[n, k] - s_{p,2}[n] U_{p,2}[n, k] \quad (6.23)$$

where $P = \lceil (K - 2)/2 \rceil$, both $s_{p,1}[n]$ and $s_{p,2}[n]$ are real, independent, Gaussian random variables for every n , and $U_{p,1}[n, k]$ and $U_{p,2}[n, k]$ are complex, deterministic, and orthogonal functions in k for every n . Equivalently, we have the entirely real, time-domain sum-of-products

$$y[n, n'] = \sum_{p=1}^P s_{p,1}[n] \Phi_{p,1}[n, n'] - s_{p,2}[n] \Phi_{p,2}[n, n'] \quad (6.24)$$

where n' is the local time variable. Due to periodic correlation, the principal components are also periodic in n . Thus, the number of unique basis functions is $2PN_m = (K - 2)N_m$. Finally, the eigenvalues of the decomposition are the variances

$$\sigma_{p,1}^2[n] = E\{s_{p,1}^2[n]\}, \quad \sigma_{p,2}^2[n] = E\{s_{p,2}^2[n]\}. \quad (6.25)$$

STFT analysis, or subband analysis, assumes the Fourier basis for expanding a signal into meaningful components. Augmented PCA instead learns the best basis in terms of decorrelating the STFT frames. As we learned in Chapter 5, coherent impropriety is determined by quadrature pairing of principal components. In the present framework, pairing occurs within the n th frame, and amounts to associating each function of k with a pair number p . Next, we discuss quadrature pairing as a means toward designing practical filters for extracting coherent signal components.

6.6.2 Quadrature Pairing and Coherence Measure

From the taxonomy of second-order processes (Section 5.2), there are different types of impropriety. For component pairs in quadrature, eigenvalue imbalance leads to impropriety. Such pairs are also coherent in the sense that their analytic signal is complex-modulated. To find evidence of complex modulation in propeller noise, we use the following.

For some pair p , $U_{p,1}[n, k]$ and $U_{p,2}[n, k]$ are in quadrature when

$$U_{p,2}[n, k] = \pm j U_{p,1}[n, k] \quad (6.26)$$

which comes from the definition of the Hilbert transform. Due to covariance estimation error, however, a component pair may not be in perfect quadrature. We define the Q-score as

$$Q_p[n] = 2 \left| \text{Imag} \left\{ \sum_k U_{p,1}^*[n, k] U_{p,2}[n, k] \right\} \right| \quad (6.27)$$

which ranges between zero and one. Only a true quadrature pair will have a Q-score of one, which follows from three facts. Let $\mathbf{u} = u[k]$ and $\mathbf{v} = v[k]$ be two distinct principal components for the same n .

- $\mathbf{u}^H \mathbf{v} + (\mathbf{u}^H \mathbf{v})^* = 0$ for $\mathbf{u} \neq \mathbf{v}$, guaranteed by the orthogonality of augmented components. Therefore, $\mathbf{u}^H \mathbf{v}$ is purely imaginary.
- $\mathbf{u}^H \mathbf{u} + (\mathbf{u}^H \mathbf{u})^* = 1$, and the same for \mathbf{v} , due to normality of augmented components. Therefore, $\mathbf{u}^H \mathbf{u} = \mathbf{v}^H \mathbf{v} = 1/2$.
- $\mathbf{u}^H \mathbf{v} = \pm j/2$ if and only if $\mathbf{v} = \pm j \mathbf{u}$.

A greedy algorithm for finding quadrature pairs is to arbitrarily pick one eigenvector, compute its Q-score with every other eigenvector, and form a pair with the largest Q-score. Then, repeat this process for all remaining eigenvectors. This is obviously non-optimal but will suffice for the time being. The coherence, or eigenvalue-imbalance, for the p th pair is

$$\gamma_p[n] = \frac{|\sigma_{p,1}^2[n] - \sigma_{p,2}^2[n]|}{\sigma_{p,1}^2[n] + \sigma_{p,2}^2[n]} \quad (6.28)$$

$Q \backslash \gamma$	<i>Low</i>	<i>High</i>
<i>Low</i>	Non-quadrature (improper)	Non-quadrature (improper)
<i>High</i>	Non-coherent quadrature (proper)	Coherent quadrature (improper)

Figure 6.18: Coherence chart for principal component pairs. A high Q-score corresponds to a quadrature pair, for which γ is meaningful as a measure of noncircularity or coherence.

which is the noncircularity coefficient for the complex random variable $s_{p,1}[n] + js_{p,2}[n]$. As such, it ranges from zero to one. If the Q-score is large for a given p , then $\gamma_p[n]$ is a measure for coherence. See also Figure 6.18. We are less interested in low Q-scores, because non-quadrature processes are always improper and beyond the proper/improper dichotomy.

6.6.3 Eigenfilters for Subspace Projection

With user-defined Q-score and coherence thresholds, it is possible to partition the principal component pairs into three categories: (1) non-coherent quadrature, (2) coherent quadrature, and (3) non-quadrature. Each category occupies an orthogonal signal subspace, with a corresponding projection operator or eigenfilter.

An eigenfilter is defined as

$$\mathbf{W}[n] = \mathbf{U}[n]\mathbf{\Lambda}[n]\mathbf{U}^H[n] \quad (6.29)$$

where \mathbf{U} is the concatenation of augmented, frequency-domain principal components satisfying

$$\mathbf{U}[n] = \begin{pmatrix} \mathbf{U}_1[n] & \mathbf{U}_2[n] \\ \mathbf{U}_1^*[n] & \mathbf{U}_2^*[n] \end{pmatrix} \quad (6.30)$$

and $\mathbf{\Lambda}[n]$ is a diagonal matrix of the form

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_1[n] & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_1[n] \end{pmatrix} \quad (6.31)$$

where $\mathbf{\Lambda}[n]$ consists of ones and zeros. The purpose of $\mathbf{\Lambda}[n]$ is to attenuate principal component pairs, without altering their internal balance. Setting diagonal elements of $\mathbf{\Lambda}_1[n]$ to zero results in a projection which is orthogonal to the attenuated component pair. Therefore, a quadrature eigenfilter will have $\mathbf{\Lambda}_1[n]$ with diagonal entries $\Lambda_p[n] = 1$ for p where $Q_p[n] > Q_{threshold}$, and zero otherwise.

The eigenfilter is a widely linear (WL) operator. Therefore, the filtered STFT results from the n -varying, matrix multiplication

$$\tilde{\mathbf{y}}[n] = \mathbf{W}[n \bmod N_m] \mathbf{y}[n]. \quad (6.32)$$

Finally, inverse STFT synthesis yields the filtered signal $\tilde{y}[n]$ from $\tilde{\mathbf{y}}[n] = \tilde{Y}[n, k]$. For this reason, we advise using a perfect-reconstruction STFT algorithm⁴.

6.6.4 Eigenfiltering Enhancement Results

The following is a presentation of preliminary results in applying eigenfilters to propeller cavitation data. We will study two signals, the merchant and tanker recordings which were found to be empirically improper in Section 6.5. In both cases, we find that the quadrature subspace also tends to be improper and hence coherent. By isolating the coherent subspace in these signals, eigenfiltering seems to enhance the resulting modulation spectrum. Since the algorithm is cyclic, we require initial knowledge of the propeller rate obtained by conventional DEMON analysis. However, the resulting modulation enhancement could possibly lead to improved classification based on a higher-quality ship signature.

⁴Even with perfect-reconstruction, it is still not guaranteed that $\tilde{Y}[n, k]$ is the STFT of $\tilde{y}[n]$. Our STFT-filtering approach could perhaps be improved by incorporating STFT consistency constraints on the inverse algorithm. See [164].

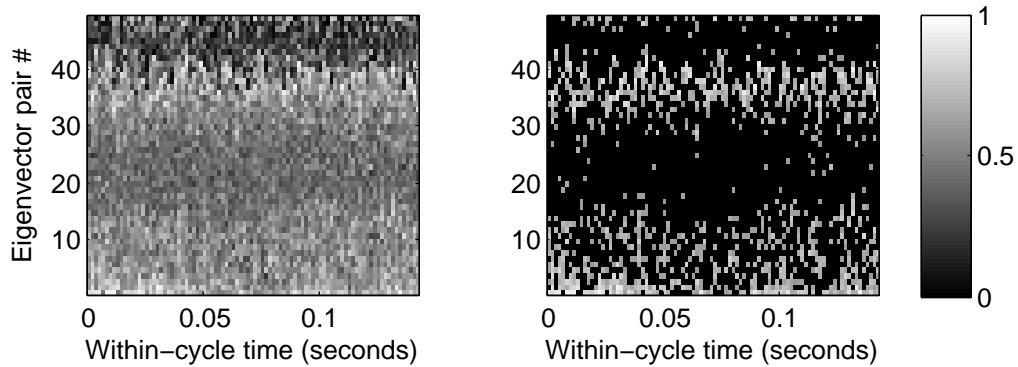


Figure 6.19: Q-scores for tanker data, before (left) and after (right) thresholding above 0.6.

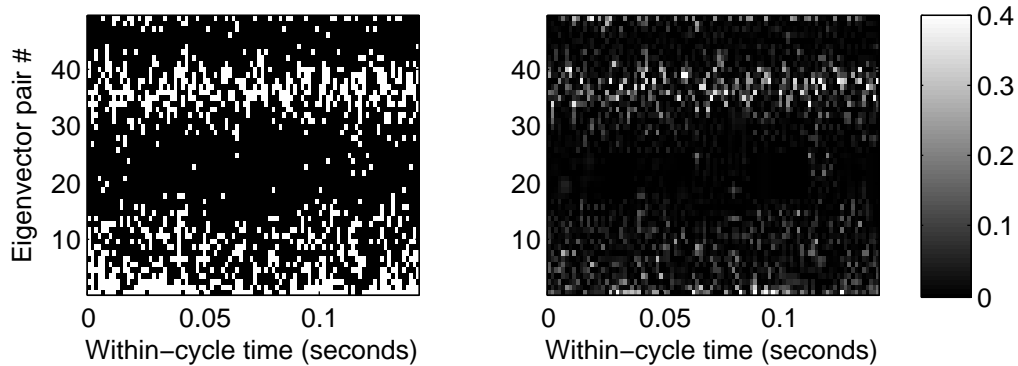


Figure 6.20: Thresholded Q-scores (left) compared to coherence scores (right) for tanker data.

Let us begin with the tanker data, with a propeller rate of 6.96 Hz and a sampling rate of 32 kHz. For this demonstration, the STFT $Y[n, k]$ consists of 100 subbands spaced uniformly around the unit circle, based on a 100-point Hamming window with bandwidth of 320 Hz. For sixty seconds of data, the cyclic covariance matrices $\hat{\Gamma}[n]$ result from (6.22) with $I = 417$ full revolutions of the propeller. Figure 6.19 displays the Q-scores of the resulting eigenvectors, before and after thresholding at $Q_{thresh} = 0.6$. From Figure 6.20, we can see that the quadrature eigenvectors also tend toward coherence.

Therefore, the subspace spanned by high-Q principal components is arguably the coher-

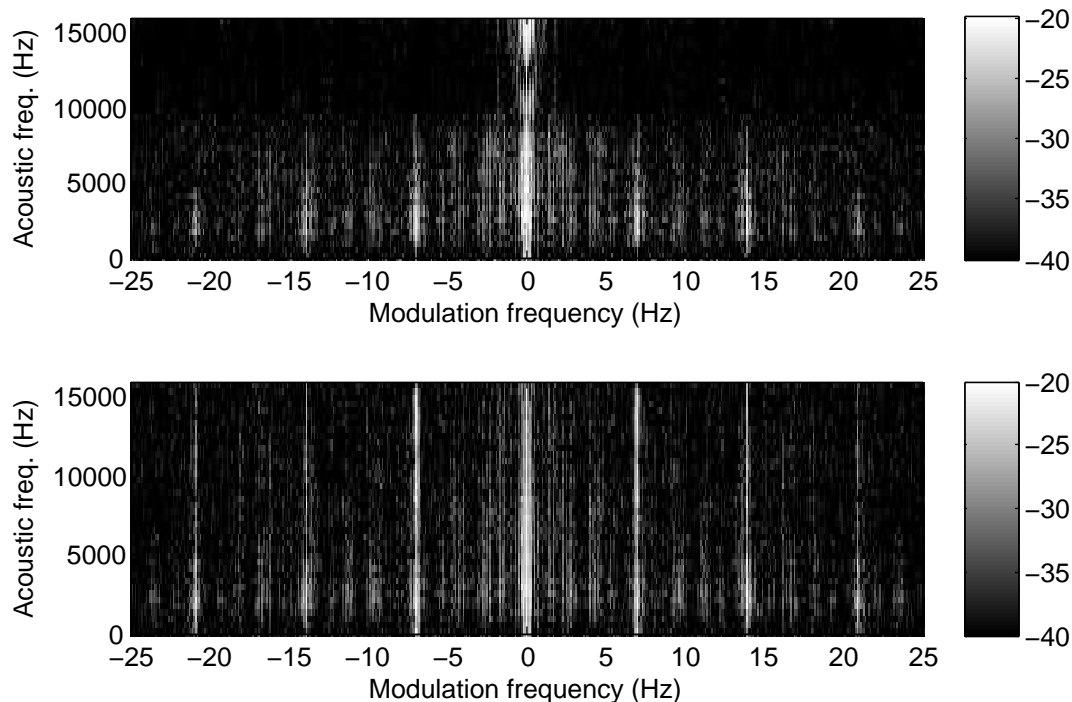


Figure 6.21: Hilbert-envelope multiband modulation spectra for tanker data, before (top) and after (bottom) quadrature-subspace projection.

ent subspace of the tanker data. Further filtering could also involve coherence thresholding, but at the risk of overstating the statistical significance of the measured coherence scores. We instead limit our analysis to the Q-score threshold of 0.6. The corresponding eigenfilter has an audible de-noising effect on the data. Figure 6.21 displays multiband modulation spectra [82] before and after eigenfiltering, using the Hilbert envelope⁵ for subband demodulation. Although the modulation spectra are non-coherent, the eigenfilter can be thought of as isolating the “coherently-modulated subspace” of the tanker data. This subspace is clearly important, as evidenced by the visible enhancement of the modulation-frequency

⁵In practice, the Hilbert envelope is often cleaner than the square-law envelope in the modulation-frequency domain. The nonlinear compression of the square-root might have a rudimentary noise-suppression property.

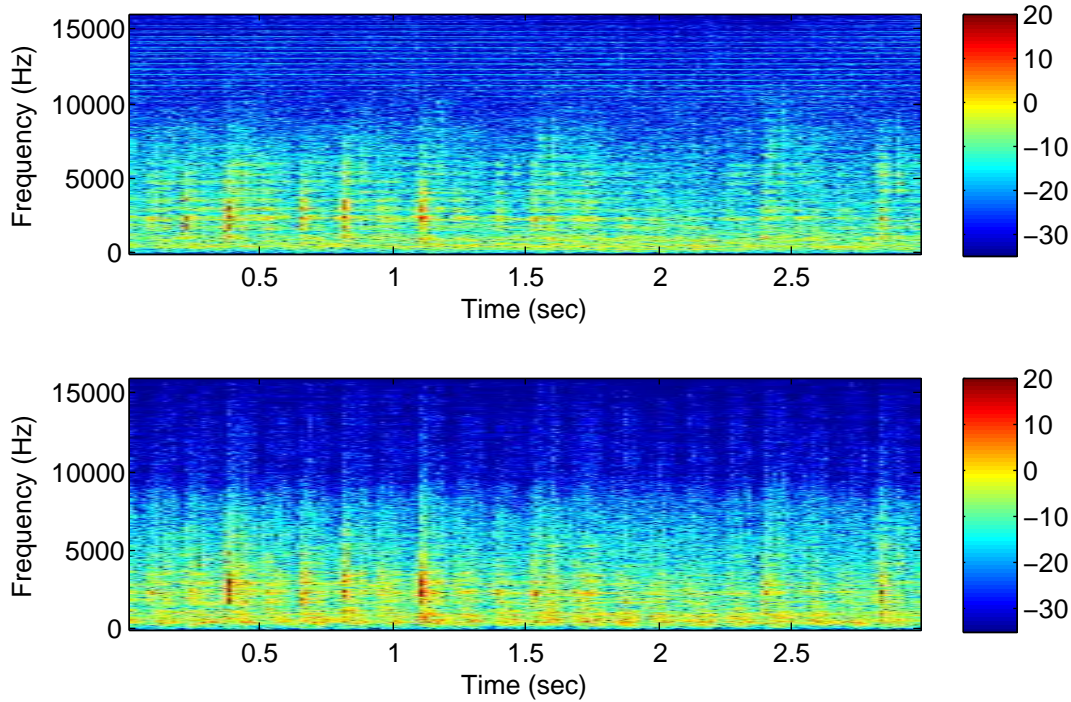


Figure 6.22: Narrowband spectrograms for tanker data, before (top) and after (bottom) quadrature-subspace projection.

domain.

Figure 6.22 also shows narrowband spectrograms for the original and filtered signal. We observe two effects. The first is that the eigenfilter has removed the spectral lines above 10 kHz, which are narrowband components varying slowly over time, evidently in a non-synchronous manner. Second, the propeller modulations appear extended across frequency and also more evident at some points in time, such as between 2 and 2.5 seconds.

We repeated the eigenfiltering procedure on merchant ship data, with a propeller rate of 1.96 Hz over 90 seconds. In Figure 6.23, it again appears that quadrature components also tend to be coherent. Using the same Q-score threshold of 0.6, the modulation spectrum changes somewhat after filtering, seen in Figure 6.24. As in the tanker data, modulations are more clear for high acoustic frequencies, especially above 6 kHz.

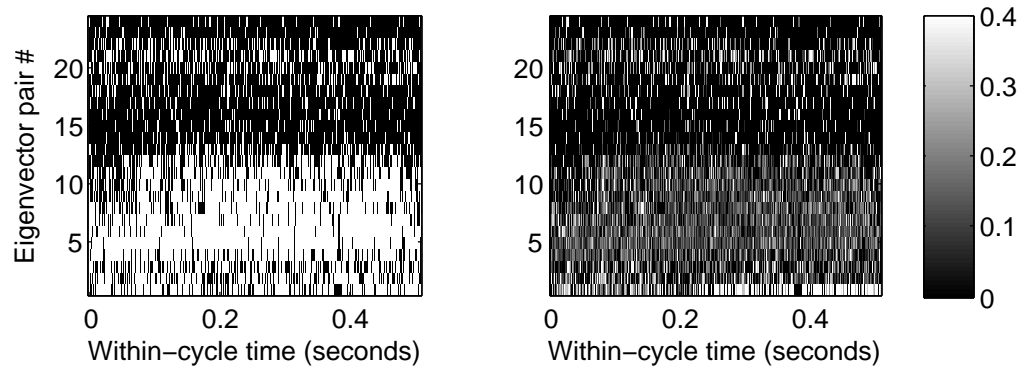


Figure 6.23: Thresholded Q-scores (left) compared to coherence scores (right) for merchant data.

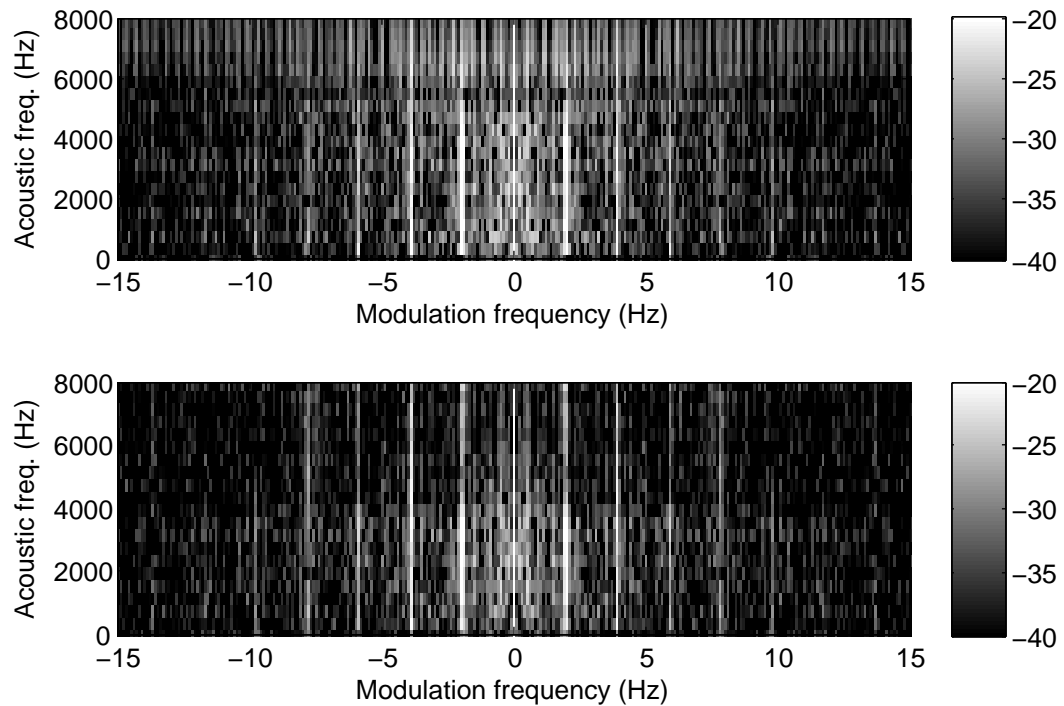


Figure 6.24: Hilbert-envelope multiband modulation spectra for merchant data, before (top) and after (bottom) quadrature-subspace projection.

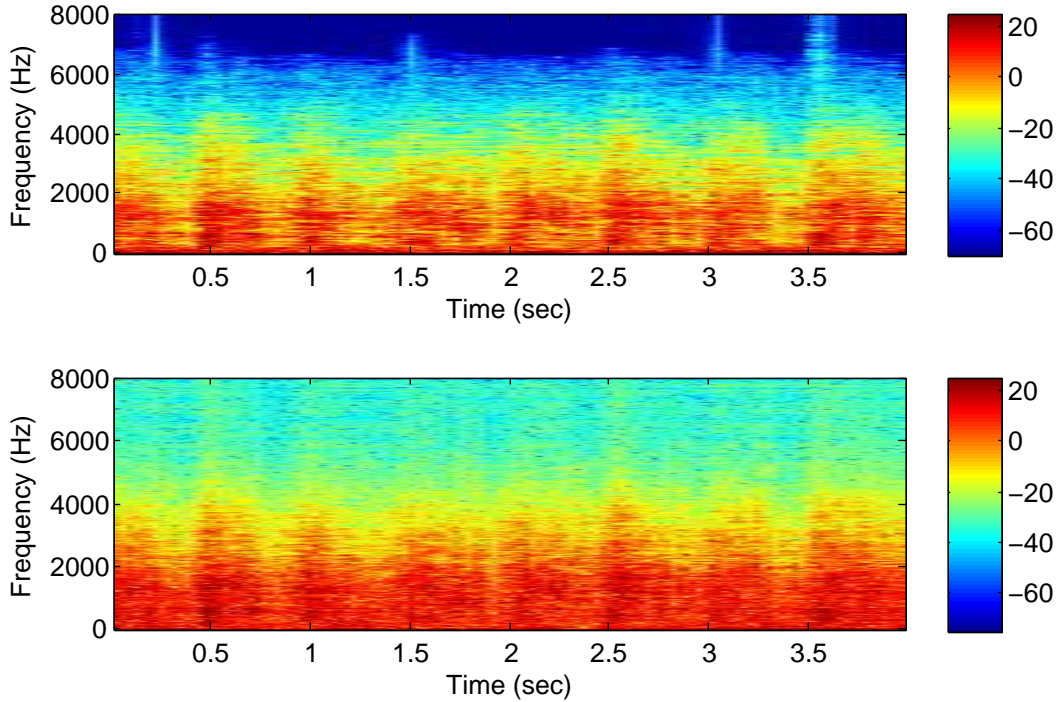


Figure 6.25: Narrowband spectrograms for merchant data, before (top) and after (bottom) quadrature-subspace projection.

In the spectrogram analysis in Figure 6.25, however, the region above 6 kHz is most devoid of energy pre-filtering. One reason for this is the effect of an anti-aliasing filter. Therefore, the bandwidth expansion seen in the post-filtered spectrogram is somewhat suspect. Perhaps there is a leakage effect in the estimation of eigenvectors, which could explain the similar bandwidth expansion in tanker Figures 6.21 and 6.22. It is currently unknown whether or not the “enhancement” of quadrature-subspace projection is a real effect, or a processing artifact. Possibly related is the effect of estimation noise in the covariances $\hat{\Gamma}[n]$, which will inevitably perturb the resultant eigenvectors and hence the quadrature subspace. Matrix perturbation is a well-studied phenomenon (e.g., see [165]). More investigation is required on this matter.

6.7 Conclusion

Based on previous chapters, we hypothesized the existence of spectral impropriety in rhythmic signals in which modulations are systematic yet randomly timed. Speech subbands are highly improper, as long as the subband filters track the harmonic frequencies. Cavitation noise is also measurably improper in its subbands. However, the resulting complementary modulation spectrum suggests that cavitation is perhaps not best represented by time-frequency subbands. To address this shortcoming, we proposed a generalized detection method based on short-time, augmented PCA. The basis functions of cavitation, although not necessarily sinusoidal, tended to be paired coherently in quadrature. From this observation, we designed eigenfilters for isolating the coherent subspace of a cavitation signal, synchronized to its propeller rate. The preliminary results suggest a possible de-noising effect, although it is unclear whether or not this is a genuine improvement or a processing artifact.

Chapter 7

CONCLUSION

Based on a theory of spectral impropriety, we have defined coherent demodulation for Gaussian random processes. In the following, we review major results and briefly discuss remaining questions and possibilities for future work.

7.1 Main Results

We first presented periodically-correlated noise as an elementary signal model. Impropriety in the frequency domain has a clear relationship to the Fourier coefficients of the periodic modulator. The sufficient statistics for this process are the Hermitian power spectrum and the complementary spectrum. We showed that the latter must be estimated coherently, using subband filters synchronized to the periodicity of the modulator. The resulting subbands are stationary and improper, which realizes our theme of estimation through removal of nonstationarity. Impropriety is therefore a potential reference for detecting the rhythm of a signal.

Having established these tools, we defined a signal model for representing both long-term modulations and short-term carrier rhythm. The signal is the output of a linear system driven by periodically correlated noise. Passing the signal through a coherent analysis filterbank yields subbands which are modulated and improper. This multiplicative subband model represents a time-varying transfer function relationship between the carrier spectrum and the signal's complementary spectrum. We derived conditions for the separability of carriers and modulators, and demonstrated coherent estimators for synthetic examples. Our intent in these derivations was to show that complex subband demodulation is equivalent with the assumption of a periodically-correlated carrier structure.

The above signal model is not a necessary condition for synchronous demodulation, however. Using periodic modulation as a prototype system, we showed that a linear operator

in the time domain is generally widely-linear in the frequency domain. Therefore, the subband formalism is merely one interpretation of coherence. Perhaps our most fundamental result is the taxonomic classification of paired principal components. From the algebraic structure of impropriety, we proved three important results:

- Only a quadrature process can be proper.
- A quadrature pair of principal components is improper if and only if their eigenvalues are imbalanced.
- An improper quadrature process is complex-modulated in a meaningful way, whereas proper quadrature is strictly non-coherent.

These observations reveal the fundamental nature of the Hilbert transform in impropriety and coherence.

Finally, we demonstrated how to detect and measure impropriety in real-world data using the generalized likelihood ratio. For speech, we found that pitch-tracking subbands are highly improper when averaged with a window between 50 and 200 milliseconds. We also tested propeller cavitation noise, assuming periodic correlation within subbands. Although cavitation is broadband and non-tonal, we detected impropriety in two out of three recordings.

7.2 Future Work and Open Questions

Overall, these results establish tools for the synchronous estimation, analysis, and modification of signals based on their internal rhythm. For the problems posed in this thesis, there is still work to be done. First, we have yet to discover the physical correspondent for impropriety observed in cavitation noise. The complementary modulation spectra tend to be broadband with spread peaks, which is not in accordance with the subband model postulated in Chapters 3 and 4. Loosely speaking, we may attribute the modulation-bandwidth spread to the phase modulation of the carrier, which may in some way relate to short-time instability of aggregate bubble formation on the propellers.

Second, there is a similar problem with speech subbands, which seem to abruptly switch phase at the onset of voiced segments. As a result, the complementary modulation spectrum appears spread. It is possible that modulation-frequency is not the best signal basis for analyzing speech. Rather, a temporal signal basis may be more appropriate, synchronized to an information-theoretic concept of syllabic rhythm. This is a potentially rich field with applications in speech and speaker recognition, as well as speech enhancement in noise and interference. Whereas the focus of this dissertation is impropriety of the carriers, an interesting extension would be to model impropriety of the long-term modulations. For syllabic frequencies around 2-8 Hz, random phase may have a large effect on the representation of information in speech.

Third, complementary processing has yet to be proven for effectively demodulating a speech signal. Ideally, such a demodulator would require the speech signal as the only input, and from there internally determine the pitch and complex modulators based on a theory of impropriety. The objective would be to produce a model residual which is comparatively small and devoid of speech intelligibility cues.

Finally, the principal components theory of impropriety suggests a possible new form of signal decomposition based on quadrature pairs of basis functions. It remains to be seen if quadrature constraints can be used for dictionary learning and sparse coding techniques of signals such as speech and audio.

BIBLIOGRAPHY

- [1] H. Dudley, “The automatic synthesis of speech,” *Proc. Natl. Acad. Sci. USA*, vol. 25, no. 7, pp. 377383, July 1939.
- [2] T. Kailath, “Measurements on time-variant communication channels,” *IRE Trans. Information Theory*, vol. 8, no. 5, pp. 229–236, 1962.
- [3] G. Matz, F. Hlawatsch, and W. Kozek, “Generalized evolutionary spectral analysis and the weyl spectrum of nonstationary random processes,” *IEEE Trans. Signal Processing*, vol. 45, no. 6, pp. 1520–1534, June 1997.
- [4] G. Matz and F. Hlawatsch, “Nonstationary spectral analysis based on time-frequency operator symbols and underspread approximations,” *IEEE Trans. Information Theory*, vol. 52, no. 3, pp. 1067–1086, March 2006.
- [5] P. Bello, “Measurement of random time-variant linear channels,” *IEEE Trans. Information Theory*, vol. 15, no. 4, pp. 469–475, July 1969.
- [6] Q. Li and L. Atlas, “Coherent modulation filtering for speech,” in *Proc. IEEE ICASSP*, April 2008, pp. 4481–4484.
- [7] P. Clark and L. Atlas, “Time-frequency coherent modulation filtering of nonstationary signals,” *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4323–4332, Nov. 2009.
- [8] B.J. King and L. Atlas, “Single-channel source separation using complex matrix factorization,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2591–2597, Nov. 2011.
- [9] H. Cramér, “On some classes of nonstationary stochastic processes,” in *4th Berkeley Symp. Math., Statist., Probability*, Los Angeles, CA, 1961, vol. 2, Univ. California Press.
- [10] Y. Grenier, “Time-dependent ARMA modeling of nonstationary signals,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 31, no. 4, pp. 899–911, Aug 1983.
- [11] H. Dudley, “Remaking speech,” *J. Acoust. Soc. Am.*, vol. 11, pp. 165, 1939.
- [12] R.K. Potter, “Visible patterns of sound,” *Science*, vol. 102, no. 2654, pp. 463–470, 1945.

- [13] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, Nov. 1946.
- [14] J. Dugundji, "Envelopes and pre-envelopes of real waveforms," *IEEE Trans. Inf. Theory*, vol. 4, no. 1, pp. 53–57, March 1958.
- [15] E. Bedrosian, "A product theorem for hilbert transforms," *Proceedings of the IEEE*, vol. 51, no. 5, pp. 868 – 869, May 1963.
- [16] J.R. Carson, "Notes on the theory of modulation," *Proc. IRE*, vol. 10, pp. 57–64, 1922.
- [17] A.W. Rihaczek and E. Bedrosian, "Hilbert transforms and the complex representation of real signals," *Proceedings of the IEEE*, vol. 54, no. 3, pp. 434 – 435, March 1966.
- [18] A. R. Møller, "Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat," *Acta Physiologica Scandinavica*, vol. 86, no. 2, pp. 223–238, 1972.
- [19] B. Delgutte, B.M. Hammond, and P.A. Cariani, "Neural coding of the temporal envelope of speech: Relation to modulation transfer functions," in *Psychophysical and Physiological Advances in Hearing: Proceedings of the 11th International Symposium on Hearing*, A.R. Palmer, A. Rees, A.Q. Summerfield, and R. Meddis, Eds., Wurr, London, 1997, pp. 595–603.
- [20] T. Houtgast and H. J. M. Steeneken, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," *J. Acoust. Soc. Am.*, vol. 54, no. 2, pp. 557–557, 1973.
- [21] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.*, vol. 77, no. 3, pp. 1069–1077, 1985.
- [22] R. R. Riesz, "Differential intensity sensitivity of the ear for pure tones," *Phys. Rev.*, vol. 31, pp. 867–875, May 1928.
- [23] N.F. Viemeister, "Temporal factors in audition: a systems analysis approach," in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds., 1977.
- [24] N.F. Viemeister, "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. of Am.*, vol. 66, no. 5, pp. 1364–1380, 1979.
- [25] R. Drullman, J.M. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1053–1064, 1994.

- [26] R.V. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, no. 5234, pp. 303–304, 1995.
- [27] M.S. Vinton and L.E. Atlas, "Scalable and progressive audio codec," *Proc. IEEE ICASSP*, vol. 5, pp. 3277–3280, 2001.
- [28] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Trans. Speech, Audio Process.*, vol. 2, no. 4, pp. 578–589, Oct 1994.
- [29] T.G. Stockham, T.M. Cannon, and R.B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proc. of the IEEE*, vol. 63, no. 4, pp. 678 – 692, April 1975.
- [30] R. Schwartz, T. Anastasakos, F. Kubala, J. Makhoul, L. Nguyen, and G. Zavaliagkos, "Comparative experiments on large vocabulary speech recognition," in *Proc. Workshop on Human Language Technology*, Stroudsburg, PA, USA, 1993, HLT '93, pp. 75–80, Association for Computational Linguistics.
- [31] S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 34, no. 1, pp. 52 – 59, Feb 1986.
- [32] S. Greenberg and B.E.D. Kingsbury, "The modulation spectrogram: In pursuit of an invariant representation of speech," *Proc. IEEE ICASSP*, vol. 3, pp. 1647–1650, 1997.
- [33] B. E. D. Kingsbury, N. Morgan, and S. Greenberg, "Robust speech recognition using the modulation spectrogram," *Speech Communication*, vol. 25, no. 1-3, pp. 117 – 132, 1998.
- [34] Y.-H.B. Chiu and R.M. Stern, "Minimum variance modulation filter for robust speech recognition," in *Proc. IEEE ICASSP, Taipei*, April 2009, pp. 3917 –3920.
- [35] O. Ghitza, "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," *J. Acoust. Soc. Am.*, vol. 110, no. 3, pp. 1628–1640, 2001.
- [36] D. Vakman, "On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 791–797, April 1996.
- [37] L. Cohen, P. Loughlin, and D. Vakman, "On an ambiguity in the definition of the amplitude and phase of a signal," *Elsevier Signal Process.*, vol. 79, pp. 301–307, June 1999.

- [38] C.P. Clark, “Effective coherent modulation filtering and interpolation of long gaps in acoustic signals,” Master’s thesis, University of Washington, 2008.
- [39] P.J. Loughlin and B. Tacer, “On the amplitude- and frequency-modulation decomposition of signals,” *J. Acoust. Soc. Am.*, vol. 100, no. 3, pp. 1594–1601, 1996.
- [40] A. Papoulis, “Random modulation: A review,” *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 31, no. 1, pp. 96–105, 1983.
- [41] S.O. Rice, “Mathematical analysis of random noise,” *The Bell System Technical Journal*, vol. 23-24, 1944.
- [42] L. Atlas and C. Janssen, “Coherent modulation spectral filtering for single-channel music source separation,” *Proc. IEEE ICASSP*, vol. IV, pp. 461–464, 2005.
- [43] S. Schimmel and L. Atlas, “Coherent envelope detection for modulation filtering of speech,” *Proc. IEEE ICASSP*, vol. 1, pp. 221–224, March 18-23 2005.
- [44] P. Clark and L. Atlas, “A sum-of-products model for effective coherent modulation filtering,” *Proc. IEEE ICASSP, Taipei*, pp. 4485–4488, 2009.
- [45] J. L. Flanagan, D. I. S. Meinhart, R. M. Golden, and M. M. Sondhi, “Phase vocoder,” *J. Acoust. Soc. of Am.*, vol. 38, no. 5, pp. 939–940, 1965.
- [46] L. Mandel, “Complex representation of optical fields in coherence theory,” *J. Opt. Soc. Am.*, vol. 57, no. 5, pp. 613–617, 1967.
- [47] S. Disch and B. Edler, “Multiband perceptual modulation analysis, processing and synthesis of audio signals,” in *IEEE ICASSP, Taipei*, April 2009, pp. 2305–2308.
- [48] S.M. Schimmel and L.E. Atlas, “Target talker enhancement in hearing devices,” in *Proc. IEEE ICASSP*, April 2008, pp. 4201–4204.
- [49] S. M. Schimmel, *Theory of Modulation Frequency Analysis and Modulation Filtering, with Applications to Hearing Devices*, Ph.d. dissertation, University of Washington, 2007.
- [50] B. King and L. Atlas, “Coherent modulation comb filtering for enhancing speech in wind noise,” in *Proc. Int. Workshop on Acoustic Echo and Noise Control*, 2008.
- [51] X. Li, K. Nie, L. Atlas, and J. Rubinstein, “Harmonic coherent demodulation for improving sound coding in cochlear implants,” *Proc. IEEE ICASSP*, pp. 5462–5465, mar. 2010.

- [52] C.S. Ramalingam and R. Kumaresan, “Voiced-speech analysis based on the residual interfering signal canceler (risc) algorithm,” in *Proc. IEEE ICASSP*, April 1994, vol. 1, pp. I/473 –I/476.
- [53] R. Kumaresan, C.S. Ramalingam, and A. Rao, “Risc: An improved costas estimator-predictor filter bank for decomposing multicomponent signals,” in *IEEE Seventh SP Workshop on Statistical Signal and Array Processing*, June 1994, pp. 207 –210.
- [54] M. A. Brennan, P. E. Souza, L. E. Atlas, and A. R. Greenhall, “Coherent versus incoherent modulation filtering,” *J. Acoust. Soc. Am.*, vol. 127, no. 3, pp. 1902–1902, 2010.
- [55] R. McAulay and T. Quatieri, “Speech analysis/synthesis based on a sinusoidal representation,” *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 34, no. 4, pp. 744 – 754, aug 1986.
- [56] L. Girin, M. Firouzmand, and S. Marchand, “Perceptual long-term variable-rate sinusoidal modeling of speech,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 3, pp. 851 –861, March 2007.
- [57] G. Sell and M. Slaney, “Solving demodulation as an optimization problem,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 2051 –2066, Nov. 2010.
- [58] G. Sell and M. Slaney, “The information content of demodulated speech,” in *Proc. IEEE ICASSP*, Dallas, TX, March 2010, pp. 5470–5473.
- [59] R. Turner and M. Sahani, “Probabilistic amplitude demodulation,” in *Independent Component Analysis and Signal Separation*, Mike Davies, Christopher James, Samer Abdallah, and Mark Plumbley, Eds., vol. 4666 of *Lecture Notes in Computer Science*, pp. 544–551. Springer Berlin / Heidelberg, 2007.
- [60] R.E. Turner and M. Sahani, “Demodulation as probabilistic inference,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2398 –2411, November 2011.
- [61] P. Clark, G. Sell, and L. Atlas, “A novel approach using modulation features for multiphone-based speech recognition,” in *Submitted to IEEE ICASSP, Prague*, 2011.
- [62] J. Ville, “Theory and application of the notion of the complex signal,” *Cables et Transmission*, vol. 2A, pp. 61–74, 1948, Translated in 1958 from the original French by I. Belin.

- [63] J. Le Roux, H. Kameoka, N. Ono, S. Sagayama, and A. de Cheveigné, “Modulation analysis of speech through orthogonal FIR filterbank optimization,” in *Proc. IEEE ICASSP, Las Vegas*, 2008, pp. 4189–4192.
- [64] L. M. Gray and D. S. Greeley, “Source level model for propeller blade rate radiation for the world’s merchant fleet,” *J. Acoust. Soc. Am.*, vol. 67, no. 2, pp. 516–522, 1980.
- [65] J.G. Lourens and J.A. du Preez, “Passive sonar ML estimator for ship propeller speed,” *IEEE Journal of Oceanic Engineering*, vol. 23, no. 4, pp. 448–453, Oct 1998.
- [66] A. Kummert, “Fuzzy technology implemented in sonar systems,” *IEEE J. Oceanic Engineering*, vol. 18, no. 4, pp. 483–490, Oct 1993.
- [67] R.O. Nielsen, “Cramer-Rao lower bounds for sonar broad-band modulation parameters,” *IEEE J. Oceanic Engineering*, vol. 24, no. 3, pp. 285–290, July 1999.
- [68] E. Parzen and M.S. Shiren, “Analysis of a general system for the detection of amplitude modulated noise,” Technical report 24, Columbia University, August 1954.
- [69] E. G. Gladyshev, “Periodically and almost-periodically correlated random processes with a continuous time parameter,” *Theory of Probability and its Applications*, vol. 8, no. 2, pp. 173–177, 1963.
- [70] Y.P. Dragan, “Spectral properties of periodically correlated random processes,” *Otbor i Peredacha Informatsii*, vol. 30, pp. 16–24, 1971.
- [71] Y.P. Dragan, “Properties of readings of periodically correlated random processes,” *Otbor i Peredacha Informatsii*, vol. 33, pp. 9–12, 1972.
- [72] W.A. Gardner, “Exploitation of spectral redundancy in cyclostationary signals,” *IEEE Signal Processing Magazine*, vol. 8, no. 2, pp. 14–36, April 1991.
- [73] R. F. Dwyer, “Fourth-order spectra of gaussian amplitude-modulated sinusoids,” *J. Acoust. Soc. Am.*, vol. 90, no. 2, pp. 918–926, 1991.
- [74] G. B. Giannakis and G. Zhou, “Harmonics in multiplicative and additive noise: parameter estimation using cyclic statistics,” *IEEE Trans. Signal Processing*, vol. 43, no. 9, pp. 2217–2221, Sept 1995.
- [75] G. Zhou and G. B. Giannakis, “Harmonics in multiplicative and additive noise: performance analysis of cyclic estimators,” *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1445–1460, June 1995.

- [76] G. Zhou and G. B. Giannakis, "Harmonics in Gaussian multiplicative and additive noise: Cramer-Rao bounds," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1217–1231, May 1995.
- [77] K.S. Voychishin and Y.P. Dragan, "Elimination of rhythm for periodically correlated random processes," *Otbor i Peredacha Informatsii*, vol. 33, pp. 12–16, 1972.
- [78] L. A. Zadeh, "Frequency analysis of variable networks," *Proc. IRE*, vol. 38, no. 3, pp. 291 – 299, mar. 1950.
- [79] H.L. Hurd, "Spectral coherence of nonstationary and transient stochastic processes," in *Spectrum Estimation and Modeling, 1988., Fourth Annual ASSP Workshop on*, Aug 1988, pp. 387–390.
- [80] A. A. Kudryavtsev, K. P. Luginets, and A. I. Mashoshin, "Amplitude modulation of underwater noise produced by seagoing vessels," *Acoustical Physics*, vol. 49, no. 2, pp. 184–188, March 2003, Translated from *Akusticheskii Zhurnal*, Vol. 49, No. 2, 2003, pp. 224228.
- [81] G.-Z. Shi and J.-C. Hu, "Ship noise demodulation line spectrum fusion feature extraction based on the wavelet packet," in *Proc. 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2007*.
- [82] P. Clark, I. Kirsteins, and L. Atlas, "Multiband analysis for colored amplitude-modulated ship noise," in *Proc. 2010 IEEE ICASSP, Dallas, TX, March 2010*.
- [83] S.M. Schimmel, L.E. Atlas, and K. Nie, "Feasibility of single channel speaker separation based on modulation frequency analysis," *Proc. IEEE ICASSP*, vol. 4, pp. 605–608, April 2007.
- [84] M.B. Priestley, "Evolutionary spectra and non-stationary processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 27, no. 2, pp. pp. 204–237, 1965.
- [85] W.D. Mark, "Spectral analysis of the convolution and filtering of non-stationary stochastic processes," *Journal of Sound and Vibration*, vol. 11, no. 1, pp. 19 – 63, 1970.
- [86] R. Silverman, "Locally stationary random processes," *Information Theory, IRE Transactions on*, vol. 3, no. 3, pp. 182 –187, September 1957.
- [87] T. Kailath, *Lectures on Communication System Theory*, chapter Ch. 6: Channel Characterization: Time-Variant Dispersive Channels, pp. 95–123, McGraw-Hill, 1961.

- [88] W. H. Huggins, "System-function analysis of speech sounds," *J. Acoust. Soc. Am.*, vol. 22, no. 6, pp. 765–767, 1950.
- [89] B. S. Atal and J. R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bitrates," in *Proc. IEEE ICASSP*, 1982.
- [90] M. Schroeder and B. Atal, "Code-excited linear prediction(celp): High-quality speech at very low bit rates," in *IEEE ICASSP*, apr 1985, vol. 10, pp. 937 – 940.
- [91] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveign, "Restructuring speech representations using a pitch-adaptive timefrequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," *Speech Communication*, vol. 27, no. 3-4, pp. 187 – 207, 1999.
- [92] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno, "Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation," in *IEEE Proc. ICASSP*, 312008-april4 2008, pp. 3933 –3936.
- [93] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Communications Systems*, vol. 11, no. 4, pp. 360 –393, December 1963.
- [94] G.E. Pfander and D.F. Walnut, "Measurement of time-variant linear channels," *IEEE Trans. Information Theory*, vol. 52, no. 11, pp. 4808 –4820, Nov 2006.
- [95] W. Kozek, F. Hlawatsch, H. Kirchauer, and U. Trautwein, "Correlative time-frequency analysis and classification of nonstationary random processes," in *Proc. IEEE-SP Int. Symp. Time-Frequency and Time-Scale Analysis*, Oct. 1994, pp. 417 –420.
- [96] A. W. Rihaczek, "Signal energy distribution in time and frequency," *IEEE Trans. Information Theory*, vol. 14, no. 3, pp. 369 – 374, may 1968.
- [97] P.J. Schreier and L.L. Scharf, "Stochastic time-frequency analysis using the analytic signal: Why the complementary distribution matters," *IEEE Trans. Signal Processing*, vol. 51, no. 12, pp. 3071–3079, Dec. 2003.
- [98] M. Loève, *Probability Theory*, Princeton, N.J., Van Nostrand, 3rd edition, 1963.
- [99] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941 –981, jul 1989.
- [100] Y. Zhao, L.E. Atlas, and II Marks, R.J., "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 38, no. 7, pp. 1084 –1091, July 1990.

- [101] W. Martin and P. Flandrin, “Wigner-ville spectral analysis of nonstationary processes,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 33, no. 6, pp. 1461 – 1470, Dec. 1985.
- [102] I. Kirsteins, P. Clark, and L. Atlas, “Maximum-likelihood estimation of propeller noise modulation characteristics,” in *Proc. Underwater Acoustic Measurements*, June 2011.
- [103] S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 28, no. 4, pp. 357–366, Aug 1980.
- [104] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, “Complex nmf: A new sparse representation for acoustic signals,” in *IEEE ICASSP, Taipei*, April 2009, pp. 3437 –3440.
- [105] A. Ozerov, P. Philippe, F. Bimbot, and R. Gribonval, “Adaptation of bayesian models for single-channel source separation and its application to voice/music separation in popular songs,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1564 –1578, July 2007.
- [106] E.M. Grais and H. Erdogan, “Single channel speech music separation using nonnegative matrix factorization and spectral masks,” in *17th International Conf. Digital Signal Processing*, July 2011, pp. 1 –6.
- [107] M. Vetterli and D. Le Gall, “Perfect reconstruction fir filter banks: some properties and factorizations,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 1057 –1071, July 1989.
- [108] M. Portnoff, “Time-frequency representation of digital signals and systems based on short-time fourier analysis,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 28, no. 1, pp. 55 – 69, Feb. 1980.
- [109] R. Crochiere, “A weighted overlap-add method of short-time fourier analysis/synthesis,” *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 28, no. 1, pp. 99 – 102, feb 1980.
- [110] B. Picinbono, “On circularity,” *IEEE Trans. Signal Processing*, vol. 42, no. 12, pp. 3473 –3482, dec 1994.
- [111] B. Picinbono and P. Bondon, “Second-order statistics of complex signals,” *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 411–420, Feb 1997.

- [112] F.D. Neeser and J.L. Massey, “Proper complex random processes with applications to information theory,” *IEEE Trans. Information Theory*, vol. 39, no. 4, pp. 1293–1302, Jul 1993.
- [113] A. van den Bos, “The multivariate complex normal distribution—a generalization,” *IEEE Trans. Information Theory*, vol. 41, no. 2, pp. 537–539, March 1995.
- [114] E. Ollila, “On the circularity of a complex random variable,” *IEEE Signal Processing Letters*, vol. 15, pp. 841–844, 2008.
- [115] W. M. Brown and R. B. Crane, “Conjugate linear filtering,” *Proc. IEEE Trans. Information Theory*, vol. 15, no. 4, pp. 462–465, jul 1969.
- [116] B. Picinbono and P. Chevalier, “Widely linear estimation with complex data,” *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 2030–2033, aug 1995.
- [117] B. Picinbono, “Wide-sense linear mean square estimation and prediction,” in *Proc. IEEE ICASSP*, May 1995, vol. 3, pp. 2032–2035 vol.3.
- [118] P. J. Nahin, *An Imaginary Tale: The Story of $\sqrt{-1}$* , Princeton University Press, 1998.
- [119] B. Rivet, L. Girin, and C. Jutten, “Log-rayleigh distribution: A simple and efficient statistical representation of log-spectral coefficients,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 796–802, March 2007.
- [120] P.J. Schreier, L.L. Scharf, and C.T. Mullis, “Detection and estimation of improper complex random signals,” *IEEE Trans. Information Theory*, vol. 51, no. 1, pp. 306–312, Jan. 2005.
- [121] J. Navarro-Moreno, M.D. Estudillo-Martinez, R.M. Fernandez-Alcala, and J.C. Ruiz-Molina, “Estimation of improper complex-valued random signals in colored noise by using the hilbert space theory,” *IEEE Trans. Information Theory*, vol. 55, no. 6, pp. 2859–2867, June 2009.
- [122] G. Gelli, L. Paura, and A.R.P. Ragozini, “Blind widely linear multiuser detection,” *IEEE Communications Letters*, vol. 4, no. 6, pp. 187–189, jun 2000.
- [123] H. Gerstacker, R. Schober, and A. Lampe, “Receivers with widely linear processing for frequency-selective channels,” *IEEE Trans. Communications*, vol. 51, no. 9, pp. 1512–1523, Sept. 2003.
- [124] A. Kuh and D. Mandic, “Applications of complex augmented kernels to wind prediction,” in *Proc. IEEE ICASSP*, 2009.

- [125] S. Javidi, D. P. Mandic, and A. Cichocki, "Complex blind source extraction from noisy mixtures using second-order statistics," *IEEE Trans. Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1404–1416, July 2010.
- [126] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Processing*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.
- [127] W. A. Gardner, "Cyclic wiener filtering: theory and method," *IEEE Trans. Communications*, vol. 41, no. 1, pp. 151–163, Jan. 1993.
- [128] D. P. Percival and A. T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press, 1993.
- [129] L. Cerrato and B. Eisenstein, "Deconvolution of cyclostationary signals," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 25, no. 6, pp. 466–476, Dec. 1977.
- [130] J. Kormylo and J. Mendel, "Identifiability of nonminimum phase linear stochastic systems," *IEEE Trans. Automatic Control*, vol. 28, no. 12, pp. 1081–1090, Dec. 1983.
- [131] W.A. Gardner, "A new method of channel identification," *IEEE Trans. Communications*, vol. 39, no. 6, pp. 813–817, June 1991.
- [132] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Trans. Information Theory*, vol. 40, no. 2, pp. 340–349, March 1994.
- [133] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: a frequency-domain approach," *IEEE Trans. Information Theory*, vol. 41, no. 1, pp. 329–334, Jan. 1995.
- [134] M. G. Hall, A. V. Oppenheim, and A. S. Willsky, "Time-varying parametric modeling of speech," *Signal Processing*, vol. 5, no. 3, pp. 267–285, 1983.
- [135] M. V. Mathews, Joan E. Miller, and Jr. E. E. David, "Pitch synchronous analysis of voiced sounds," *J. Acoust. Soc. Am.*, vol. 33, no. 2, pp. 179–186, 1961.
- [136] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, vol. 9, no. 5-6, pp. 453–467, 1990, Neuropeech '89.
- [137] H. Yang, W.B. Kleijn, E.F. Deprettere, and H. Chen, "Pitch synchronous modulated lapped transform of the linear prediction residual of speech," in *Signal Processing Proceedings, 1998. ICSP '98*, 1998, pp. 591–594 vol.1.

- [138] S. Kim, T. Eriksson, H.-G. Kang, and D. H. Youn, "A pitch synchronous feature extraction method for speaker recognition," in *Proc. IEEE ICASSP*, May 2004, vol. 1, pp. I – 405–8 vol.1.
- [139] M. Portnoff, "Short-time fourier analysis of sampled speech," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 3, pp. 364 – 373, June 1981.
- [140] M. Portnoff, "Time-scale modification of speech based on short-time fourier analysis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 3, pp. 374 – 390, June 1981.
- [141] R. M. Fano, "Short-time autocorrelation functions and their power spectra," *J. Acoust. Soc. Am.*, vol. 22, no. 5, pp. 546–550, Sept 1950.
- [142] B. Picinbono, "Quadratic filters," in *Proc. IEEE ICASSP*, May 1982, vol. 7, pp. 298 – 301.
- [143] B. Picinbono and P. Devaut, "Optimal linear-quadratic systems for detection and estimation," *IEEE Trans. Information Theory*, vol. 34, no. 2, pp. 304 –311, March 1988.
- [144] R.J. McAulay and T.F. Quatieri, "Pitch estimation and voicing detection based on a sinusoidal speech model," vol. 1, pp. 249–252, Apr 1990.
- [145] N. Abu-Shikhah and M. Deriche, "A robust technique for harmonic analysis of speech," vol. 2, pp. 877–880, 2001.
- [146] D. Tjøstheim, "Spectral generating operators for non-stationary processes," *Advances in Applied Probability*, vol. 8, no. 4, pp. pp. 831–846, 1976.
- [147] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2001.
- [148] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 714 – 725, 2003.
- [149] M.D. Yacoub, G. Fraidenraich, and J.C.S. Santos Filho, "Nakagami-m phase-envelope joint distribution," *Electronics Letters*, vol. 41, no. 5, pp. 259 – 261, March 2005.
- [150] D.B.H. Tay, "Daubechies wavelets as approximate Hilbert-pairs?," *IEEE Signal Processing Letters*, vol. 15, pp. 57 –60, 2008.

- [151] D.B.H. Tay and J. Zhang, “On Hilbert-pairs from non-minimum phase Daubechies filters,” in *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS)*, June 2010, pp. 1619 –1622.
- [152] N. Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234 – 253, 2001.
- [153] A. Patel, *Music, Language, and the Brain*, chapter 3, Oxford University Press, Inc., 2008.
- [154] D. P. Mandic, S. Javidi, G. Souretis, and V. S.L. Goh, “Why a complex valued solution for a real domain problem,” in *2007 IEEE Workshop Machine Learning for Signal Processing*, aug. 2007, pp. 384 –389.
- [155] E. Ollila and V. Koivunen, “Generalized complex elliptical distributions,” in *Proc. 3rd Sensor Array Multichannel Signal Processing Workshop*, Sitges, Spain, July 2004.
- [156] P. J. Schreier, L. L. Scharf, and A. Hanssen, “A generalized likelihood ratio test for impropriety of complex signals,” *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 433–436, July 2006.
- [157] J.-P. Delmas, A. Oukaci, and P. Chevalier, “Asymptotic distribution of glr for impropriety of complex signals,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 3594 –3597.
- [158] A.T. Walden and P. Rubin-Delanchy, “On testing for impropriety of complex-valued gaussian vectors,” *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 825 –834, March 2009.
- [159] J.-P. Delmas, A. Oukaci, and P. Chevalier, “On the asymptotic distribution of glr for impropriety of complex signals,” *Signal Processing*, vol. 91, no. 10, pp. 2259 – 2267, 2011.
- [160] L. Atlas, P. Clark, and S. Schimmel, “Modulation Toolbox Version 2.1 for MATLAB,” <http://isdl.ee.washington.edu/projects/modulationtoolbox/>, Sept 2010.
- [161] V. P. Morozov, “Cavitation noise as a train of sound pulses generated at random times,” *Soviet Physics - Acoustics*, vol. 14, no. 3, pp. 361–365, Jan.-March 1969.
- [162] M. Ida, “Bubble-bubble interaction: A potential source of cavitation noise,” *Phys. Rev. E*, vol. 79, no. 1, pp. 016307, Jan 2009.
- [163] University of Connecticut, “Bioacoustics Laboratory Research Page,” <http://www.bioacoustics.uconn.edu/research.html>, September 2009.

- [164] J. Le Roux, N. Ono, and S. Sagayama, “Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction,” in *Proc. SAPA 2008 ISCA Workshop on Statistical and Perceptual Audition*, Sept. 2008, pp. 23–28.
- [165] D.W. Tufts, “The effects of perturbations on matrix-based signal processing,” in *Fifth ASSP Worksho on Spectrum Estimation and Modeling*, Oct. 1990, pp. 159–162.
- [166] D. Slepian, “On bandwidth,” *Proc. of the IEEE*, vol. 64, no. 3, pp. 292–300, March 1976.
- [167] D. Slepian, “Some comments on fourier analysis, uncertainty and modeling,” *SIAM Review*, vol. 25, no. 3, pp. pp. 379–393, 1983.
- [168] B. Picinbono, “Second-order complex random vectors and normal distributions,” *IEEE Trans. Signal Processing*, vol. 44, no. 10, pp. 2637–2640, Oct 1996.
- [169] D.H. Brandwood, “A complex gradient operator and its application in adaptive array theory,” *IEEE Proc. Communications, Radar and Signal Processing, part F*, vol. 130, no. 1, pp. 11–16, feb. 1983.
- [170] H. Kramer and M. Mathews, “A linear coding for transmitting a set of correlated signals,” *IRE Trans. Information Theory*, vol. 2, no. 3, pp. 41–46, september 1956.
- [171] R. Plomp, L. C. W. Pols, and J. P. van de Geer, “Dimensional analysis of vowel spectra,” *J. Acoust. Soc. Am.*, vol. 41, no. 3, pp. 707–712, 1967.

Appendix A

SPECTRAL REPRESENTATION OF HARMONIZABLE RANDOM PROCESSES

A.1 Introduction

Given a random process $x(t)$, it is often desirable to study it in the frequency domain. In principle, this requires that the Fourier transform $X(\omega)$ exist in some sense. In practice, this requires that we sample $x(t)$ over a finite period of time and compute the discrete Fourier transform (DFT). The resulting DFT should relate logically to the underlying Fourier transform $X(\omega)$. For completeness, the requisite Fourier theory should apply to stationary as well as nonstationary processes. In this appendix we review Loève's theory of "harmonizable processes" as the underpinning for the notation and Fourier relations used throughout this dissertation.

First, we must understand the problem. If the realizations of a random process $x(t)$ have finite energy, then the Fourier transform is

$$X(\omega) = \int x(t)e^{-j\omega t} dt \quad (\text{A.1})$$

where $X(\omega)$ is a complex random process. The inverse, or synthesis, formulation¹ is

$$x(t) = \int X(\omega)e^{j\omega t} d\omega \quad (\text{A.2})$$

which means $X(\omega)$ preserves all of the information in $x(t)$. The assumption of finite energy is limiting, however. Stationary processes, for example, contain infinite energy. In that case, (A.1) will not converge and $X(\omega)$ does not exist. The power spectrum gives a frequency decomposition for the process, but only in an energetic sense. What we require is a linear Fourier representation for signal analysis in the frequency domain.

¹For this appendix, we dispense with the $1/2\pi$ scale factor in order to smoothly bridge the concepts of Riemann and Riemann-Stieltjes integrals. In continuous time and frequency, the $1/2\pi$ is a holdover from the variable change $\omega = 2\pi f$.

As a work-around, the rest of this dissertation uses the following convention. Let $x'(t)$ be a random, possibly stationary, Gaussian process. We define $x(t) = w_T(t)x'(t)$, where $w_T(t)$ is a rectangular window of length T . We assume that T is “long” in the sense that it encompasses our interest in the signal. In practice T is perhaps several seconds. Conceptually, it could be a human lifetime or the lifespan of our sun (which ever makes you more mathematically comfortable). When we say $x(t)$ is “stationary” or “nonstationary,” we mean it is effectively stationary or nonstationary within the window. Most importantly, we assume that $x(t)$ has finite energy within the interval. This means the Fourier relations (A.1) and (A.2) hold, where the time integral has implicit bounds defined by $w_T(t)$, i.e., $X(\omega) = \int_T x(t) \exp(-j\omega t) dt$.

The rest of this appendix justifies why (A.1) and (A.2) accurately represent $x(t) = w_T(t)x'(t)$. This begins with the frequency transform for $x'(t)$, which is generally nonstationary but may be stationary in the exact sense of its autocovariance $R_{xx}(t, \tau) = R_{xx}(\tau)$ for all $-\infty < t < \infty$. For this we require the harmonizable representation.

A.2 Harmonizable Processes

To address the shortcoming of the finite-energy Fourier integral, Loève introduced a new spectral integral for “harmonizable” random processes. We will explain by example, but the reader is encouraged to consult Loève [98], and Percival and Walden [128] for more detailed treatments.

A.2.1 Example 1

Let $x'(t) = \sum_k c_k e^{j\omega_k t}$, defined for all t , where ω_k are discrete frequencies and c_k are independent, complex random variables. (We owe this example to Chapter 4 of [128].) Even though (A.1) does not converge for this process, an obvious spectral representation is simply the set of ordered pairs (ω_k, c_k) . Let $F'(\omega)$ be the cumulative spectrum

$$F'(\omega) = \begin{cases} 0, & \omega = -\infty \\ F'(\omega_k - d\omega) + c_k, & \omega = \omega_k \end{cases} \quad (\text{A.3})$$

which we visualize as a discontinuous complex process with steps of size c_k at frequencies $\omega = \omega_k$. We define the “increment process” as

$$dX(\omega) = F'(\omega) - F'(\omega - d\omega) \quad (\text{A.4})$$

so that $dX(\omega_k) = c_k$ and 0 everywhere else. Therefore we may write the Fourier synthesis as

$$x(t) = \sum_k dX(\omega_k) e^{j\omega_k t}. \quad (\text{A.5})$$

A.2.2 Example 2

We note that $x'(t)$ in the previous example is a wide-sense stationary process for any number of c_k 's as long as they are independent. Now suppose we let the spacing between frequencies shrink to zero. This is desirable for defining a process which is full-band in some sense, such as a “white” process which nominally contains equal amounts of every frequency. The resulting synthesis integral is

$$x'(t) = \int dX(\omega) e^{j\omega t} \quad (\text{A.6})$$

where we understand (A.6) to be a Riemann-Stieltjes integral with increment process $dX(\omega)$. It behaves like the limit of (A.5), except $dX(\omega)$ is infinitesimal in order to distribute the power² of $x'(t)$ over all frequencies from $-\infty < \omega < \infty$. A process satisfying (A.6) is called harmonizable by Loève's definition [98]. The cumulative spectrum is given by

$$F'(\omega) = \int_{-\infty}^{\omega} dX(u). \quad (\text{A.7})$$

When $x(t)$ is stationary and Gaussian, as in both examples above, $dX(\omega)$ is an orthogonal process (akin to independent c_k 's). Therefore, the autocovariance of $x'(t)$ satisfies

$$R_{xx}(\tau) = \int E\{dX(\omega)dX^*(\omega)\} e^{j\omega\tau}. \quad (\text{A.8})$$

This is the famous Wiener-Khinchine theorem for wide-sense stationary processes, where the power spectrum $S_{xx}(\omega)$ is related to the variance at each frequency via

$$S_{xx}(\omega)d\omega = E\{dX(\omega)dX^*(\omega)\} \quad (\text{A.9})$$

²Although $x'(t)$ is possibly infinite in energy, we assume that it has finite power, or variance, for every value of t . That is, $E\{x^2(t)\} < \infty$.

whenever $R_{xx}(\omega)$ has finite energy such that $S_{xx}(\omega)$ is bounded.

In general, the bivariate autocovariance is

$$R_{xx}(t, \tau) = \int \int E\{dX(\omega - \nu)dX^*(\omega)\}e^{j\omega\tau - j\nu t} \quad (\text{A.10})$$

where the bifrequency covariance $E\{dX(\omega_1)dX^*(\omega_2)\}$ is the Loève spectrum $d\Gamma_{xx}(\omega_1, \omega_2)$. It follows that nonstationarity in the time domain is equivalent to cross-frequency correlation in the linear frequency domain. When $x'(t)$ is stationary, $d\Gamma_{xx}(\omega_1, \omega_2)$ is zero everywhere except for the line $\omega_1 = \omega_2$, which corresponds to the power spectrum.

The exact nature of $dX(\omega)$ is sometimes difficult to visualize. When $x'(t)$ has finite energy, we have $dX(\omega) = X(\omega)d\omega$, where the infinitesimal $d\omega$ scales $X(\omega)$ to match the infinitesimal $dX(\omega)$. To illustrate, we consider again the cumulative spectrum $F'(\omega)$. When $dX(\omega) = X(\omega)d\omega$, the fundamental theorem of calculus gives $X(\omega)$ as the derivative of $F'(\omega)$, or

$$X(\omega) = \lim_{d\omega \rightarrow 0} \frac{F'(\omega + d\omega) - F'(\omega)}{d\omega} = \lim_{d\omega \rightarrow 0} \frac{dX(\omega)}{d\omega}. \quad (\text{A.11})$$

However, $X(\omega)$ exists only when the limit exists. If $F'(\omega)$ is discontinuous, then (A.11) does not converge. This explains Example 1, where the spectrum could only be described in terms of $dX(\omega)$ and a Riemann-Stieltjes integral.

It is common in engineering to use generalized functions instead of Riemann-Stieltjes integrals. The Dirac delta function $\delta(\omega)$ is accordingly defined as the derivative of a step function. This allows the discrete spectrum of Example 1 to be expressed as an impulse train, or $X(\omega) = \sum_k c_k \delta(\omega - \omega_k)$. Then, we may express $x'(t)$ in terms of (A.2) provided that we define

$$\int \delta(\omega - u)A(u)du = A(\omega). \quad (\text{A.12})$$

The delta function therefore expresses the stepwise Riemann-Stieltjes integration in the language of more familiar Riemannian integration. This may suffice for Example 1, which consists of a countable number of frequencies, but the picture becomes more complicated in Example 2.

For a fullband, stationary random process, we imagine $dX(\omega)$ as an orthogonal process everywhere in ω . This means $dX(\omega)$ and $dX(\omega + d\omega)$ are orthogonal, or statistically unrelated, regardless of how small $d\omega$ becomes without being zero. What does such a process

look like? It can not have smoothness at any scale, since that would imply correlation across frequencies. Instead, one might think of $dX(\omega)$ as being discontinuous everywhere in ω . The generalized function approach would, in this case, require an infinitely dense train of Delta functions. The Riemann-Stieltjes integral is a formal resolution of the seeming contradiction of orthogonality at every point in a continuum of frequencies ³.

In the next section, we show that windowing in time converts the harmonizable spectrum $dX(\omega)$ to a finite-energy spectrum $X(\omega)d\omega$, in a way that justifies our conceptualization of random processes existing within T -length windows.

A.3 Time- and Frequency-Concentrated Processes

Let $x'(t)$ be a harmonizable process. Let $w_T(t)$ be a finite-energy window with Fourier transform $W(\omega)$. In general, the window need not be rectangular. The windowed process is

$$x(t) = w_T(t)x'(t) = \int \int W(\omega)e^{j\omega t}d\omega dX(u)e^{jut}. \quad (\text{A.13})$$

Substitution of variables gives the frequency-domain convolution

$$x(t) = \int \int W(\omega - u)dX(u)e^{j\omega t}d\omega. \quad (\text{A.14})$$

We claim that the convolution $\int W(\omega - u)dX(u) = X(\omega)$ such that

$$x(t) = \int X(\omega)e^{j\omega t}d\omega. \quad (\text{A.15})$$

If $F(\omega)$ is the cumulative spectrum of $x(t)$, then (A.15) requires $F(\omega)$ to be absolutely continuous with derivative $X(\omega)$. That is,

$$\begin{aligned} X(\omega) &= \lim_{d\omega \rightarrow 0} \frac{F(\omega + d\omega) - F(\omega)}{d\omega} \\ &= \lim_{d\omega \rightarrow 0} \frac{\int [F_W(\omega + d\omega - u) - F_W(\omega - u)] dX(u)}{d\omega} \end{aligned} \quad (\text{A.16})$$

where $F_W(\omega)$ is the cumulative spectrum of $w_T(t)$, and

$$F(\omega) = \int_{-\infty}^{\omega} \int W(\omega - u)dX(u) = \int F_W(\omega - u)dX(u). \quad (\text{A.17})$$

³We have this problem in the time domain as well, with the equally troublesome concept of continuous-time white noise.

Expression (A.16) needs to converge for $X(\omega)$ to exist. As $d\omega$ approaches zero, the difference $F_W(\omega + d\omega - u) - F_W(\omega - u)$ will also approach zero if $F_W(\omega)$ is a continuous function (e.g., the cumulative integral of a sinc function when $w_T(t)$ is rectangular). This does not constitute a rigorous proof, so we will leave the convergence of (A.16) as a conjecture.

It is desirable to retain orthogonality in $X(\omega)$ when $x(t)$ is effectively stationary within the window of length T . That is to say, when $x'(t)$ is stationary. Assuming conjecture (A.16) holds, the frequency-domain correlation is

$$E\{X(\omega_1)X^*(\omega_2)\} = \int \int W(\omega_1 - u)W^*(\omega_2 - v)E\{dX(u)dX^*(v)\}. \quad (\text{A.18})$$

As discussed in the previous section, stationarity of $x'(t)$ implies that the Loève spectrum $d\Gamma_{xx}(\omega_1, \omega_2)$ is zero everywhere except for the line $\omega_1 = \omega_2$. We may instead write

$$E\{X(\omega_1)X^*(\omega_2)\} = \int W(\omega_1 - u)W^*(\omega_2 - u)S_{xx}(u)du \quad (\text{A.19})$$

where $S_{xx}(\omega)$ is the power spectrum of $x'(t)$. Note that, assuming $W(\omega)$ is lowpass, the kernel $W(\omega_1 - u)W^*(\omega_2 - u)$ is approximately zero when the relative delay $\omega_1 - \omega_2$ is greater than $2\pi/T$. Therefore, a very large T ensures that $E\{X(\omega_1)X^*(\omega_2)\} \approx 0$ except when ω_1 and ω_2 are very close. It can also be shown that $E\{X(\omega_1)X^*(\omega_2)\} = 0$ for $\omega_1 - \omega_2 = 2\pi k/T$ when $W(\omega)$ is a sinc function corresponding to rectangular $w_T(t)$. In this case, Fourier-series coefficients for $x(t)$, with period T , are completely uncorrelated.

Finally, we note that a similar argument can be made with respect to windowing in the frequency domain, or bandwidth restriction. The resulting $x(t)$ becomes continuous in the time domain. Dual use of (A.1) and (A.2) requires continuity in both domains. Of course, a signal can not be both time- and band-limited, but the preceding arguments also apply to non-rectangular windows with infinitely long, yet decaying tapers. The resulting process is time- and frequency-concentrated rather than strictly limited. We treat bandwidth and duration as “principal quantities” [166] invariant to times and frequencies at plus or minus infinity (see also Slepian [167] for the maximization of time and frequency concentration). This is how we justify (A.1) and (A.2) for Fourier decomposition and synthesis of random processes.

Appendix B

CORRELATIVE PROPERTIES OF CYCLIC DEMODULATION

In the following, we discuss the correlative structure of $X[n, \omega]$ and the conditions under which it is stationary and white in n . The difference between this subsection and the previous one is that we allow the stationary excitation noise $w(t)$ to be colored. In general,

$$E\{w(t)w(t + \tau)\} = R_{ww}(\tau) \quad (\text{B.1})$$

which is proportional to $\delta(\tau)$ when $w(t)$ is white.

First, we prove that $X[n, \omega]$ is stationary in n . From (3.15), the Hermitian autocovariance of the subband centered on ω is

$$\begin{aligned} R_{x|\omega}[n, p] &= E\{X[n, \omega]X^*[n + p, \omega]\} \\ &= \iint m(\tau)g(nT - \tau)m(\tau')g(nT + pT - \tau')\dots \\ &\quad R_{ww}(\tau - \tau')e^{-j\omega(\tau - \tau')}d\tau d\tau' \end{aligned} \quad (\text{B.2})$$

for integer p . The above expression is constant¹ in n because $m(\tau + nT) = m(\tau)$. By the same argument, the complementary autocovariance is also constant in n because the expression is much the same:

$$\begin{aligned} C_{x|\omega}[n, p] &= E\{X[n, \omega]X[n + p, \omega]\} \\ &= \iint m(\tau)g(nT - \tau)m(\tau')g(nT + pT - \tau')\dots \\ &\quad R_{ww}(\tau - \tau')e^{-j\omega(\tau + \tau')}d\tau d\tau' \end{aligned} \quad (\text{B.3})$$

where $e^{-j\omega(\tau + \tau')}$ replaces $e^{-j\omega(\tau - \tau')}$. By substituting $R_{ww}(\tau - \tau')$ with $\delta(\tau - \tau')$, integrating out τ' , and setting $p = 0$, we arrive at (3.15).

¹The time shift of the window relative to the modulator induces a linear phase term in the Fourier transform, but a synchronous shift of nT evaluates to zero phase on the harmonics of $m(t)$.

For colored $w(t)$, the integrals are more difficult. Assuming that $m(t)$ varies slowly over the correlation length of $w(t)$, we approximate (B.3) and (B.2) as

$$C_{x|\omega}[p] \approx S_{ww}^{(H)}(\omega) \int m^2(\tau)g(-\tau)g(pT - \tau)e^{-j2\omega\tau}d\tau \quad (\text{B.4})$$

$$R_{x|\omega}[p] \approx S_{ww}^{(H)}(\omega) \int m^2(\tau)g(-\tau)g(pT - \tau)d\tau. \quad (\text{B.5})$$

The assumption of a short correlation length was essential for the multiband DEMON estimator [82], and we tied this to Silverman's locally stationary relation [86] and Kailath's rectangular underspread criterion [2] in Section 2.3.2. Setting $p = 0$ gives $C_{x|\omega}[0] \approx S_{ww}^{(H)}(\omega)\rho_X^2(\omega)$ and $R_{x|\omega}[0] \approx S_{ww}^{(H)}(\omega)\tilde{\sigma}_m^2$, where $\tilde{\sigma}_m^2$ is the modulator energy within the window $g(-\tau)$. From this it is clear that the complementary spectral estimator is strongly related to the modulation spectrum of $m(t)$, whereas the Hermitian estimator primarily reflects the power spectral structure of $w(t)$.

It is also worth noting that the true complementary spectrum of $x(t)$ changes form. Instead of (3.12), we have

$$\begin{aligned} S_{xx}^{(C)}(\omega) &= \frac{1}{2\pi} \int M(\omega - u)M(\omega + u)S_{ww}^{(H)}(u) du \\ &= \frac{1}{2\pi} \int M(2\omega - u)M(u)S_{ww}^{(H)}(u - \omega) du. \end{aligned} \quad (\text{B.6})$$

The time-frequency uncertainty relation dictates that $S_{ww}^{(H)}(\omega)$ is a broad function if $R_{ww}(\tau)$ is narrow. Using the same assumption as in the previous paragraph, the theoretical complementary spectrum is a distorted version of (3.12), or

$$S_{xx}^{(C)}(\omega) \approx S_{ww}^{(H)}(\omega) \int m^2(t)e^{-j2\omega t}dt. \quad (\text{B.7})$$

It is hard to establish the exact relationship between the asymptotic, estimated spectrum in (B.4) and the theoretical approximation given above. As seen in the synthetic examples in Figure B.1, the correspondence can be rather close.

Referring again to (B.2) and (B.3), it is possible to ensure that the STFT subbands are individually white as well as stationary. This is equivalent to $R_{x|\omega}[p]$ and $C_{x|\omega}[p]$ both vanishing for $p \neq 0$. By inspection, this is achieved under the following condition:

$$g(t)g(t + \tau + pT) = 0 \quad \text{for } \tau \text{ such that } R_{ww}(\tau) \neq 0. \quad (\text{B.8})$$

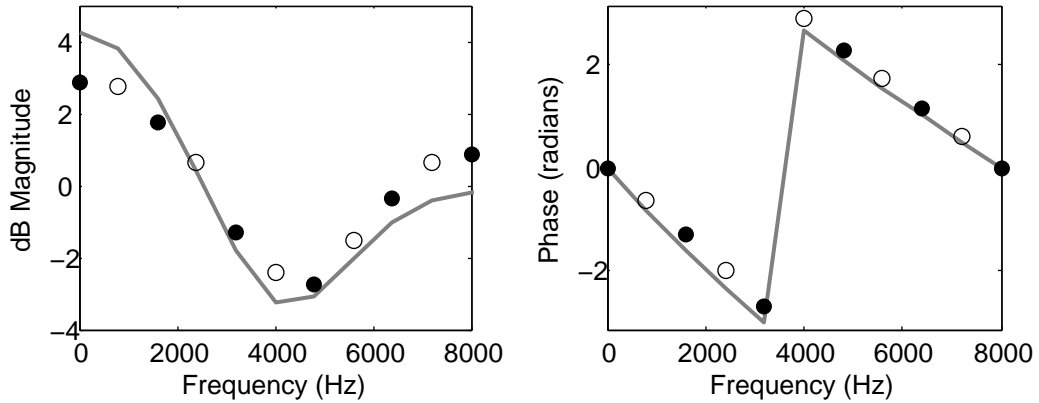


Figure B.1: Magnitude (left) and phase (right) of the complementary spectrum for the same signal as in Figure 3.3, except with a colored noise source. The gray curves, derived analytically via (B.7), predict somewhat the measured subband complementary variances (circles). The correlation length of the signal is very short, around 0.5 millisecond, in order to make the predictions close.

Suppose the correlation length L is finite, so that $R_{ww}(|\tau|) = 0$ for $|\tau| > L$. Then the above condition holds for all $p \neq 0$ as long as the window duration is less than or equal to $T - L$. This is clearly problematic since the window will not be long enough to capture the full period of $m(t)$. Furthermore, the previous section showed that a window length of $2T$ is actually desirable to estimate $S_{xx}^{(C)}(\omega)$ at intervals of $\omega_0/2$. A remedy is to downsample even further in time, as in

$$X[n, \omega] = \int x(\tau)g(nDT - \tau)e^{-j\omega\tau} d\tau \quad (\text{B.9})$$

where D is the smallest integer such that $(D - 2)T \geq L$. In practice, this may not be worthwhile due to the reduction of cycles available in empirical averages. See Figure B.2 for a visual reference.

If $w(t)$ is white, however, then $L = 0$ and the window skip rate can be $D = 1$ for a T -length widow or $D = 2$ for a $2T$ -length window, and still yield white subbands. The primary benefit of assuming a white noise excitation, however, is the expressions in the previous subsection become exact rather than the approximations derived here. Of course, truly white processes are rare in nature, so we must find some way to handle non-white

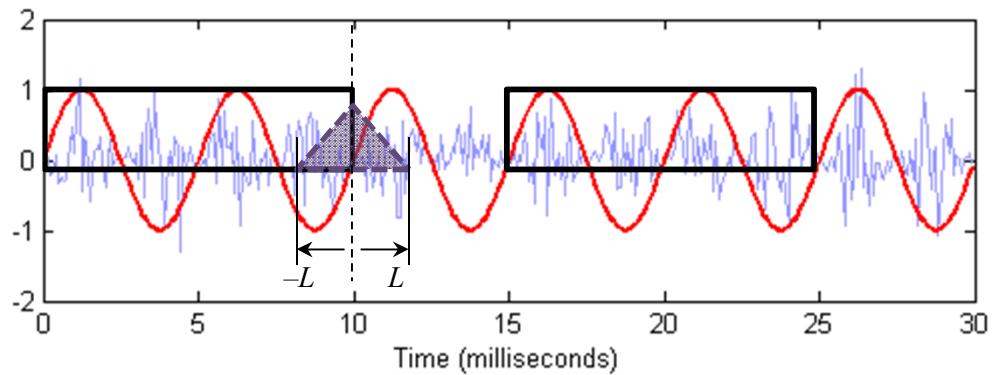


Figure B.2: A demonstration of synchronized filtering of a periodically correlated process, with $D = 3$. The modulator is shown in red, in this case a sinusoid. The process $x(t)$ is shown in light blue. Successive window placement is chosen such that samples in one frame are uncorrelated with samples in an adjacent frame, as dictated by the correlation length (illustrated by the triangle at the edge of the first frame).

signals. In Section 3.4 we propose a new signal model which represents an observed signal as the output of a linear system driven by white, modulated noise.

Appendix C

HYPOTHESIS TESTING FOR IMPROPRIETY**C.1 Scalar Noncircularity Coefficient**

A univariate, Gaussian random variable (r.v.), denoted z , has only three nontrivial statistics. We assume the mean to be zero, which leaves the Hermitian and complementary variances σ_z^2 and ρ_z^2 . To measure the degree of impropriety of z , the noncircularity coefficient (NC) is

$$|\gamma_z| = \frac{|\rho_z^2|}{\sigma_z^2} \quad (\text{C.1})$$

which is the magnitude of the noncircularity quotient [114]. As a number from zero to one, the NC naturally represents the degree of impropriety in z . It is zero for the proper case of a circular distribution in the complex plane, greater than zero for an improper ellipse, and unity for the maximally improper case of a line.

In practice we wish to determine whether or not an N -length sequence $z[n]$ is improper. If $z[n]$ is white and identically distributed over n , then we treat $z[n]$ as N realizations of the random variable z . The estimate for the NC is therefore

$$|\hat{\gamma}_z| = \frac{|\sum_{n=0}^{N-1} z^2[n]|}{\sum_{n=0}^{N-1} |z[n]|^2}. \quad (\text{C.2})$$

As pointed out in Section 6.3, the estimator $|\hat{\gamma}_z|$ is itself a random variable with nonzero variance when N is finite. Any $|\gamma_z| > 0$ nominally indicates impropriety, but $|\hat{\gamma}_z|$ can be nonzero even if $z[n]$ is truly proper.

Under the hypothesis-test framework, there are two hypotheses:

$$\begin{aligned} H_0 & : z \text{ is proper, i.e., } \rho_z^2 = 0 \\ H_1 & : z \text{ is improper, i.e., } |\rho_z^2| > 0. \end{aligned} \quad (\text{C.3})$$

As demonstrated empirically in Figure 6.5, the distribution of $|\hat{\gamma}_z|$ varies as a function of the true impropriety. Asymptotically, H_0 and H_1 converge to Rayleigh and Gaussian variables

[157, 159], with distributions specified by

$$\begin{aligned} H_0 : \quad |\hat{\gamma}_z| &\longrightarrow \mathcal{R} \left(N^{-1/2} \right) \\ H_1 : \quad |\hat{\gamma}_z| &\longrightarrow \mathcal{N} \left(|\gamma_z|, \sigma_\gamma^2 \right) \end{aligned} \quad (\text{C.4})$$

where $\sigma_\gamma^2 = |N|^{-1}(1 - 2|\gamma_z|^2 + |\gamma_z|^4)$, and \longrightarrow indicates convergence in distribution. These expressions will be inaccurate for very small N , but we have found that they are close approximations for N as little as 10.

Refer to [157] for exact expressions for probability of detection and false-alarm, and to [159] for receiver-operating characteristic (ROC) curves. A simpler approach, adopted in Chapter 6, is to find a detection threshold with a specified p-value with respect to the null distribution. That is, find $|\gamma_{thresh}|$ such that the probability of $|\hat{\gamma}_z| > |\gamma_{thresh}|$ is 2% under the null-hypothesis. If the data is truly proper, then the NC will falsely detect impropriety 1 out of 50 times.

C.2 Multivariate Generalized Likelihood Ratio

For a K -dimensional multivariate r.v., \mathbf{z} , [155] and [156] independently defined the generalized likelihood ratio (GLR) as

$$\lambda_z = \frac{\max p(\mathbf{z} | \mathbf{C}_{zz})}{\max p(\mathbf{z} | \mathbf{C}_{zz} = \mathbf{0})} \quad (\text{C.5})$$

where $\mathbf{C}_{zz} = E\{\mathbf{z}\mathbf{z}^T\}$ is the complementary covariance. This expression is the ratio of the maximum likelihood of \mathbf{z} with unconstrained \mathbf{C}_{zz} , to the maximum likelihood of \mathbf{z} with \mathbf{C}_{zz} constrained to be zero. It is unity when \mathbf{z} is proper, and approaches infinity when \mathbf{z} is maximally improper (*i.e.*, $|\mathbf{C}_{zz}| = \mathbf{R}_{zz}$, where \mathbf{R}_{zz} is the Hermitian covariance).

Let us again assume we have a sequence of N independent realizations, denoted $\mathbf{z}[n]$. Each index n is another K -dimensional vector drawn randomly from a distribution. We use the modified GLR [157]

$$L_z = 2 \log \lambda_z = -\frac{4}{N} \log \det \left(\mathbf{I} - \hat{\mathbf{R}}_{zz}^{-1} \hat{\mathbf{C}}_{zz} \hat{\mathbf{R}}_{zz}^{-*} \hat{\mathbf{C}}_{zz}^* \right) \quad (\text{C.6})$$

where $\hat{\mathbf{R}}_{zz}$ and $\hat{\mathbf{C}}_{zz}$ are the estimated Hermitian and complementary covariance matrices of $\mathbf{z}[n]$. The logarithm has the appealing property of returning zero when \mathbf{z} is proper. Numbers

larger than zero indicate impropriety. In practice, we have found the following formulation, due to [156], is more numerically stable:

$$\lambda_z = \left(\prod_{k=1}^K (1 - \hat{r}_k^2) \right)^{-2/N} \quad (\text{C.7})$$

where r_k are real-valued canonical correlations, or singular-values of the matrix

$$\hat{\mathbf{R}}_{zz}^{-1/2} \hat{\mathbf{C}}_{zz} \hat{\mathbf{R}}_{zz}^{-T/2}. \quad (\text{C.8})$$

The resulting GLR is

$$L_z = 2 \log \lambda_z = -\frac{4}{N} \sum_{k=1}^K \log (1 - \hat{r}_k^2) \quad (\text{C.9})$$

The asymptotic distributions of L_z is apparently difficult to derive. Delmas *et al.* [157] claim the following:

$$H_0 : L_z \longrightarrow \chi_{K(K+1)}^2 \quad (\text{C.10})$$

which is a chi-squared distribution with $K(K+1)$ degrees of freedom. The p-value for rejecting the null hypothesis can then be computed for a given dimensionality K and sample size N . We have been unable to verify this equation in practice, however. It is possible that the convergence is very slow and requires a large N before the chi-square approximation becomes accurate. Regardless, we use Monte Carlo simulations to estimate the null distribution and p-value threshold.

C.3 A Special Case: Diagonal Covariance

It is sometimes advantageous to assume that the dimensions of \mathbf{z} are independent. This is the same as \mathbf{R}_{zz} and \mathbf{C}_{zz} both being diagonal matrices. Note that this has nothing to do with samples in n , since we always assume $\mathbf{z}[n]$ is independent and identically distributed.

Diagonal covariance allows some simplifications of the GLR. Specifically,

$$\lambda_z = \left(\prod_{k=1}^K (1 - |\hat{\gamma}_k|^2) \right)^{-2/N} \quad (\text{C.11})$$

where $|\hat{\gamma}_k|$ is the noncircularity coefficient for the k th dimension of \mathbf{z} . That is,

$$|\hat{\gamma}_k| = \frac{|\sum_{n=0}^{N-1} z^2[n, k]|}{\sum_{n=0}^{N-1} |z[n, k]|^2}. \quad (\text{C.12})$$

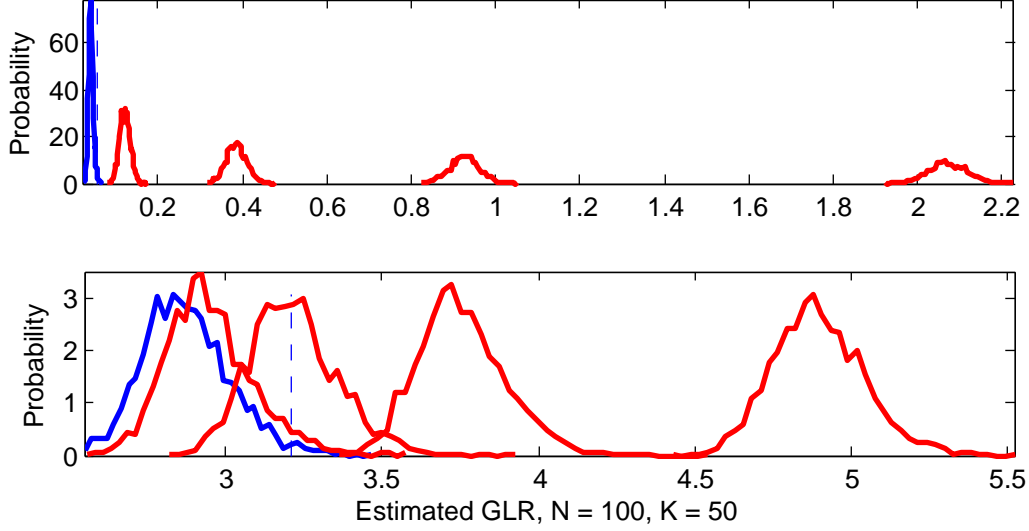


Figure C.1: Monte Carlo histograms for five values of L_z measured with additive proper noise. The H_0 , or proper, distribution is in blue. The 2% p-value threshold of 0.29 is indicated by the vertical dashed line. Top: Diagonal covariance GLR. Bottom: Full-covariance GLR.

Expression (C.11) follows from the fact that

$$\widehat{\mathbf{R}}_{zz}^{-1/2} \widehat{\mathbf{C}}_{zz} \widehat{\mathbf{R}}_{zz}^{-T/2} = \text{diag} \left\{ \frac{\widehat{C}_{zz}[k]}{\widehat{R}_{zz}[k]} \right\} \quad (\text{C.13})$$

when the covariances are diagonal matrices with diagonal entries given by $\widehat{C}_{zz}[k]$ and $\widehat{R}_{zz}[k]$. The singular-value decomposition then returns canonical correlations $\widehat{r}_k = |\widehat{C}_{zz}[k]| / \widehat{R}_{zz}[k]$. The resulting modified GLR is

$$L_z = 2 \log \lambda_z = -\frac{4}{N} \sum_{k=1}^K \log (1 - |\widehat{\gamma}_k|^2). \quad (\text{C.14})$$

In practice, the diagonal covariance GLR has only $2K$ parameters and therefore has better asymptotics. This means the H_0 and H_1 distributions are better separated for smaller N . Refer to Figure C.1 for a comparison between diagonal and full GLRs.

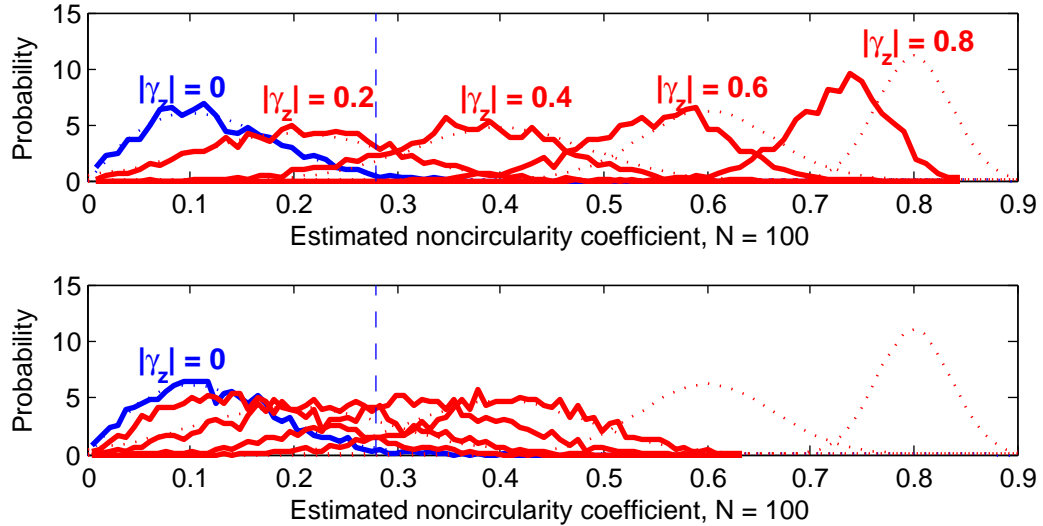


Figure C.2: Monte Carlo histograms (solid) for five values of $|\gamma_z|$ measured with additive proper noise. The H_0 , or proper, distribution is in blue. The 2% p-value threshold of 0.29 is indicated by the vertical dashed line. Top: SNR = 10 dB. Bottom: SNR = 0 dB.

C.4 Effect of SNR

An unavoidable aspect of detecting impropriety is the detrimental effect of additive, proper noise. For instance, the theoretical univariate NC is defined as

$$\gamma_z \triangleq \frac{|\rho_z^2|}{\sigma_z^2} \quad (\text{C.15})$$

where ρ_z^2 and σ_z^2 are the complementary and Hermitian variance. If $z[n]$ is in fact a sum of an improper noise, $z_{imp}[n]$, and a proper noise, $z_{prop}[n]$, then the latter will dilute the influence of the former. This is because

$$\gamma_z \triangleq \frac{|\rho_z^2|}{\sigma_z^2} = \frac{|\rho_{imp}^2|}{\sigma_{prop}^2 + \sigma_{imp}^2} = \gamma_{imp} \left(\frac{\sigma_{prop}^2}{\sigma_{prop}^2 + \sigma_{imp}^2} \right) \leq \gamma_{imp} \quad (\text{C.16})$$

assuming $z_{imp}[n]$ and $z_{prop}[n]$ are independent. Thus, additive proper noise makes an improper signal appear less improper.

In practice, this effectively raises the threshold of minimum detectable impropriety via hypothesis test. Figure C.2 repeats the NC distribution analysis of Figure 6.5, except

this time with two levels of additive noise. The dotted distributions are the asymptotic distributions for $z_{imp}[n]$, whereas the solid lines correspond to the measurement for $z[n] = z_{imp}[n] + z_{prop}[n]$. For 10 dB SNR, the effective NC distributions shift partially leftward. For 0 dB SNR, the shift is more severe, to the point where only the estimator distributions for $|\gamma_{imp}| \geq 0.6$ are at least half beyond the null-distribution cutoff. In other words, only $|\gamma_{imp}| \geq 0.6$ can be reliably detected, which is a 50% reduction from the infinite-SNR case in 6.5.

Appendix D

DERIVATION OF THE IMPROPER LIKELIHOOD FUNCTION

Let $z[n]$ be a complex, Gaussian time series with assumed signal model

$$z[n] = m[n] \cdot c[n] + \varepsilon[n] \quad (\text{D.1})$$

where $m[n]$ is complex and N -periodic, $c[n]$ is improper and stationary, and $\varepsilon[n]$ is proper and white. Without loss of generality, let $c[n]$ be normalized such that its Hermitian variance is unity, and complementary variance is ρ_C^2 . The Hermitian variance of $\varepsilon[n]$ is σ^2 . Our goal in this section is to derive the log-likelihood function for this process, in order to define the maximum-likelihood estimator for the complex modulator $m[n]$.

D.1 Derivation of the Complex Gaussian Log-Likelihood

Since the modulator is periodic, we need only define the N samples of one period of the modulator. Let \mathbf{z} denote the N -dimensional vector process consisting of N -length cycles from $z[n]$. Its $2N$ -dimensional augmented form is

$$\underline{\mathbf{z}} = \begin{pmatrix} \mathbf{z} \\ \mathbf{z}^* \end{pmatrix} \quad (\text{D.2})$$

with associated augmented covariance

$$\mathbf{\Gamma}_{zz} = \begin{pmatrix} \mathbf{R}_{zz} & \mathbf{C}_{zz} \\ \mathbf{C}_{zz}^* & \mathbf{R}_{zz}^* \end{pmatrix}. \quad (\text{D.3})$$

From [168] and [113], the Gaussian likelihood of \mathbf{z} is

$$p(\underline{\mathbf{z}}) = \pi^{-N} [\det \mathbf{R}_{zz} \det \mathbf{P}]^{-1/2} \exp \left(-\frac{1}{2} \underline{\mathbf{z}}^H \mathbf{\Gamma}_{zz}^{-1} \underline{\mathbf{z}} \right) \quad (\text{D.4})$$

where \mathbf{P} is the Schur complement of $\mathbf{\Gamma}_{zz}^*$, or the conditional covariance of \mathbf{z} conditioned on \mathbf{z}^* , is defined as

$$\mathbf{R}_{zz}^* - \mathbf{C}_{zz}^H \mathbf{R}_{zz}^{-1} \mathbf{C}_{zz}. \quad (\text{D.5})$$

The log-likelihood is therefore

$$-\log p(\mathbf{z}) = N \log \pi + \frac{1}{2} \log \det \mathbf{R}_{zz} + \frac{1}{2} \log \det \mathbf{P} + \mathbf{z}^H \mathbf{\Gamma}_{zz}^{-1} \mathbf{z}. \quad (\text{D.6})$$

To simplify the log-likelihood, we assume that $z[n]$ is white with correspondingly diagonal covariance matrices \mathbf{R}_{zz} and \mathbf{C}_{zz} . Therefore, the augmented covariance $\mathbf{\Gamma}_{zz}$ consists of four diagonal quadrants. We may now proceed to simplify (D.6) on a term-by-term basis.

Since \mathbf{R}_{zz} is diagonal, its determinant is the product of its diagonal entries. Therefore,

$$\log \det \mathbf{R}_{zz} = \sum_{n=0}^{N-1} \log (|m[n]|^2 + \sigma^2). \quad (\text{D.7})$$

The Schur complement \mathbf{P} is also diagonal. Therefore,

$$\log \det \mathbf{R} = \sum_{n=0}^{N-1} \log [(1 - |\rho_C^2|^2) |m[n]|^4 + 2\sigma^2 |m[n]|^2 + \sigma^4] - \log (|m[n]|^2 + \sigma^2) \quad (\text{D.8})$$

The third term requires some more effort. Invoking a common quadratic identity,

$$\mathbf{z}^H \mathbf{\Gamma}_{zz}^{-1} \mathbf{z} = \text{trace} \{ \mathbf{\Gamma}_{zz}^{-1} \mathbf{z} \mathbf{z}^H \}. \quad (\text{D.9})$$

Since we are assuming the $z[n]$ is white, we can replace the outer-product matrix with its maximum-likelihood estimate

$$E\{\mathbf{z} \mathbf{z}^H\} = \hat{\mathbf{\Gamma}}_{zz} = \begin{pmatrix} \hat{\mathbf{R}}_{zz} & \hat{\mathbf{C}}_{zz} \\ \hat{\mathbf{C}}_{zz}^* & \hat{\mathbf{R}}_{zz}^* \end{pmatrix}. \quad (\text{D.10})$$

where each quadrant is diagonal, defined as the cyclic estimators

$$\hat{R}_{zz}[n] = I^{-1} \sum_{i=0}^{I-1} |z[n + iN]|^2, \quad \hat{C}_{zz}[n] = I^{-1} \sum_{i=0}^{I-1} z^2[n + iN] \quad (\text{D.11})$$

for I full cycles of data. Next, the closed-form expression for $\mathbf{\Gamma}_{zz}^{-1}$ follows from block-matrix inversion, which gives

$$\mathbf{\Gamma}_{zz}^{-1} = \begin{pmatrix} \mathbf{P}^{-1} & -\mathbf{R}_{zz}^{-1} \mathbf{C}_{zz} \mathbf{P}^{-1} \\ -\mathbf{P}^{-1} \mathbf{C}_{zz}^* \mathbf{\Gamma}_{zz}^{-1} & \mathbf{P}^{-1} \end{pmatrix}. \quad (\text{D.12})$$

It is helpful to use the identity

$$\mathbf{P} = (|\mathbf{R}_{zz}|^2 - \mathbf{C}_{zz}^H \mathbf{C}_{zz}) \mathbf{R}_{zz}^{-1}. \quad (\text{D.13})$$

Substituting terms and carrying out some tedious algebra eventually yields

$$\text{trace} \{ \mathbf{\Gamma}_{zz}^{-1} \mathbf{z} \mathbf{z}^H \} = \sum_{n=0}^{N-1} \frac{2 (|m[n]|^2 + \sigma^2) \widehat{R}_{zz}[n] - 2 \text{Re} \{ (\rho_C^2 m^2[n])^* \widehat{C}_{zz}[n] \}}{(1 - |\rho_C^2|^2) |m[n]|^4 + 2\sigma^2 |m[n]|^2 + \sigma^4}. \quad (\text{D.14})$$

Recognizing that $R_{zz}[n] = |m[n]|^2 + \sigma^2$ and $C_{zz}[n] = \rho_C^2 m^2[n]$, we simplify (D.7) as

$$\log \det \mathbf{R}_{zz} = \sum_{n=0}^{N-1} \log R_{zz}[n] \quad (\text{D.15})$$

and simplify (D.8) as

$$\log \det \mathbf{R} = \sum_{n=0}^{N-1} \log \left(R_{zz}^2[n] - |C_{zz}[n]|^2 \right) - \log R_{zz}[n] \quad (\text{D.16})$$

and simplify (D.14) as

$$\text{trace} \{ \mathbf{\Gamma}_{zz}^{-1} \mathbf{z} \mathbf{z}^H \} = \sum_{n=1}^{N-1} \frac{2R_{zz}[n] \widehat{R}_{zz}[n] - 2 \text{Re} \{ C_{zz}^*[n] \widehat{C}_{zz}[n] \}}{R_{zz}^2[n] - |C_{zz}[n]|^2}. \quad (\text{D.17})$$

Substituting these three terms back into the log-likelihood gives the full log-likelihood expression

$$\begin{aligned} -2 \log p(\mathbf{z}) = 2N \log \pi &+ \sum_{n=0}^{N-1} \log \left(R_{zz}^2[n] - |C_{zz}[n]|^2 \right) \dots \\ &+ \sum_{n=0}^{N-1} \frac{2R_{zz}[n] \widehat{R}_{zz}[n] - 2 \text{Re} \{ C_{zz}^*[n] \widehat{C}_{zz}[n] \}}{R_{zz}^2[n] - |C_{zz}[n]|^2} \end{aligned} \quad (\text{D.18})$$

where $R_{zz}[n]$ and $C_{zz}[n]$ are functions of $m[n]$, ρ_C^2 , and σ^2 , and $\widehat{R}_{zz}[n]$ and $\widehat{C}_{zz}[n]$ are cyclic estimates of the envelopes of $z[n]$.

D.2 Gradient of the Log-Likelihood

To solve for the optimal estimate of $m[n]$, we must differentiate the log-likelihood with respect to the complex modulator at each point within the cycle. We define the objective function $J(\mathbf{z}|\mathbf{m})$ as

$$J(\mathbf{z}|\mathbf{m}) = \sum_{n=0}^{N-1} J_n(m[n]) \quad (\text{D.19})$$

where

$$J_n(m[n]) = \log \left(R_{zz}^2[n] - |C_{zz}[n]|^2 \right) + \frac{2R_{zz}[n]\widehat{R}_{zz}[n] - 2\text{Re} \left\{ C_{zz}^*[n]\widehat{C}_{zz}[n] \right\}}{R_{zz}^2[n] - |C_{zz}[n]|^2}. \quad (\text{D.20})$$

This is essentially the log-likelihood, without the constant $2N \log \pi$ term which will not affect the derivative.

Now, $J_n(m)$ is not differentiable with respect to the complex variable m in the Cauchy-Riemann sense. In fact, no real-valued function of m can be, because any mapping from the complex plane to the real line is non-holomorphic. We can instead treat $J_n(m)$ as a function, for n held constant, plotted over the Cartesian plane of the real and imaginary parts of m . According to Brandwood [169], the zeros of $\nabla J_n(m)$ are also the zeros of $\partial J_n(m)/\partial m^*$. The partial derivative with respect to m^* is

$$\frac{\partial J_n(m[n])}{\partial m^*[n]} = m[n]\alpha_n(|m[n]|) + m[n]^*\beta_n(|m[n]|) + m^3[n]\gamma_n(|m[n]|) \quad (\text{D.21})$$

where, for $r > 0$,

$$\alpha_n(r) = \left[(8 + j4\rho_C^2)r^6 + (4 - j2\rho_C^2)(3\sigma^2 - \widehat{R}_{zz}[n])r^4 + \dots \right. \\ \left. \left(\sigma^4(6 + j\rho_C^2) - 2\sigma^2(2 + j\rho_C^2)\widehat{R}_{zz}[n] \right) r^2 + \sigma^4(\sigma^2 - \widehat{R}_{zz}[n]) \right] \quad (\text{D.22})$$

and

$$\beta_n(r) = j\rho_C^2\sigma^2 \left(r^2 + \frac{\sigma^2}{2} \right) \quad (\text{D.23})$$

and

$$\gamma_n(r) = r^3\rho_C^2 \left[(2 + j\rho_C^2)r^2 + \sigma^2 \right]. \quad (\text{D.24})$$

The solution for $m[n]$ at each point in time is the complex value which sets the partial derivative to zero. We have argued elsewhere that the zeros of $\partial J_n(m)/\partial m^*$ are at the origin as well as at $\pm m[n]$. This explains the 180-degree ambiguity seen in Figure 4.8.

Appendix E

AUGMENTED PCA DEMODULATION

Chapter 5 derived the taxonomy of principal components as related to impropriety and coherence. The formulas apply to finite-length signals, which are compatible with the frames of a short-time Fourier transform (STFT). In the following, we will show that principal component analysis (PCA) defines modulators and carriers by solving an eigenvalue problem within each STFT frame. By orthogonalizing the carriers, the resulting decomposition is locally optimal in the sense of decomposing the signal variance. From this we will see that, under certain conditions, augmented PCA demodulation amounts to complex subband demodulation.

This should not be confused with dimensional analysis of a spectrogram, which has several precedents in the literature. Kramer and Mathews [170] used PCA to compress vocoder channels, and Plomp *et al.* [171] analyzed the dimensionality of critical-band spectrograms of vowel sounds. Davis and Mermelstein [103] later used the discrete cosine transform as a rough approximation to the principal components observed empirically by Plomp *et al.* Each case used PCA to decompose short-time power spectra. In the following, we are strictly interested in the principal components of linear, short-time Fourier transforms.

E.1 Short-Time Augmented PCA

Let $y(t)$ be a continuous-time random process bandlimited to the interval $|\omega| < \pi N/T$ for some integer N . The discrete STFT, using a modified Fourier transform (similar to (5.34)) is

$$Y[n, k] = \begin{cases} \int y(\tau)g(nT - \tau)e^{-j\pi\tau/T}e^{-j2\pi k\tau/T}d\tau, & 0 \leq k < N/2 \\ Y^*[n, -k], & N/2 \leq -k < 0 \end{cases} \quad (\text{E.1})$$

The STFT is complex over time index n and frequency index k . We can also treat $Y[n, k]$ as a random vector time-series

$$\underline{\mathbf{z}}[n] = Y[n, k] \quad (\text{E.2})$$

where the n th vector is the Fourier transform of one time frame. Each vector is in conjugate-augmented form because $y(t)$.

Straightforward application of the formulations from Section 5.2 gives

$$Y[n, k] = \sum_{p=1}^N s_p[n] U_p[n, k], \quad 0 \leq k < N/2 \quad (\text{E.3})$$

where $s_p[n]$ is a real random process uncorrelated across p and generally nonstationary in n . The array $U_p[n, k]$ contains the p th complex principal component for the n th frame. The decomposition only guarantees orthogonality in the p -dimension:

$$\begin{aligned} \delta[p - q] R_{ss}[n_1, n_2] &= E\{s_p[n_1] s_q[n_2]\}, \\ \delta[p - q] R_{uu}[n_1, n_2] &= \sum_{k=0}^{N/2-1} U_p^*[n_1, k] U_q[n_2, k] + U_p[n_1, k] U_q^*[n_2, k]. \end{aligned} \quad (\text{E.4})$$

We define a second vector time-series which undoes the Fourier transform for each frame,

$$\mathbf{y}[n] = \mathbf{F} \mathbf{z}[n] \quad (\text{E.5})$$

which is random and real. Each vector corresponds to the discrete samples of $y(t)$ within one analysis frame. It has the PCA decomposition

$$y[n, \tau] = \sum_{p=1}^N s_p[n] \Phi_p[n, \tau] \quad (\text{E.6})$$

where $\Phi_p[n, \tau]$ is the time-domain correspondent to $U_p[n, k]$, and $y[n, \tau] = \mathbf{y}[n]$ denotes the n th frame with local time τ .

E.2 Time-Varying Covariance Estimation

So far we have assumed knowledge of the carrier components. In reality, they come from factorizing an estimate of a covariance matrix. Covariance estimation poses a difficult problem because the covariance is itself time-varying. In Section 4.3, we showed how assumptions of periodic or smooth modulators can allow time-varying estimation of second-order statistics. (A generalization also appears in Section 4.3.3.) The following applies these same principles to the more estimation of a more general, local covariance.

First, we find the autocovariance of $y(t)$ from (E.13) as

$$E\{y(t_1)y(t_2)\} = \sum_n \sum_{p=1}^N \sigma_p^2[n] \Phi_p[n, t_1] \Phi_p[n, t_2] f(t_1 - nT) f(t_2 - nT) \quad (\text{E.7})$$

which follows from orthogonality in n and p . By the definition of principal components,

$$E\{y(t_1)y(t_2)\} = \sum_n R_n(t_1, t_2) f(t_1 - nT) f(t_2 - nT) \quad (\text{E.8})$$

where $R_n(t_1, t_2)$ is the local covariance function in the vicinity of global time $t = nT$. The eigenvalues and eigenfunctions of $R_n(t_1, t_2)$ are, respectively, $\sigma_p^2[n]$ and $\Phi_p[n, t]$. As before, we assume that only one realization of $y(t)$ is available, which means the covariances must be estimable by time averaging.

E.2.1 Cyclic Averaging

Cyclic averaging is possible in the periodic case. Assume $\sigma_p^2[n]$ and $\Phi_p[n, t]$ are each periodic in n , with integer period $N_m \geq 1$. The covariance estimate is therefore

$$\widehat{R}_n(t_1, t_2) = I^{-1} \sum_{i=0}^{I-1} y(t_1 + T[n + iN_m]) y(t_2 + T[n + iN_m]), \quad 0 \leq n < N_m \quad (\text{E.9})$$

where I is the total number of periods. Cyclic averaging thus requires knowledge of the modulation period TN_m , which can be estimated via DEMON spectral analysis. As a result of periodic eigenvalues, so too are the corresponding coherent envelopes defined in (E.24).

Periodicity assumes the modulators and carriers are synchronized to the same fundamental period TN_m . Note also that stationarity, or constancy in n , satisfies the periodicity relation for all $N_m > 0$. Thus, cyclic averaging also applies to the case where $\sigma_p^2[n]$ is periodic and $\Phi_p[n, t] = \Phi_p(t)$. This is the situation in Section 4.3, wherein the signal model predicted stationary carriers and periodic modulators within subbands.

E.2.2 Lowpass Averaging

Another estimation method lowpass averaging, which assumes that $\sigma_p^2[n]$ and $\Phi_p[n, t]$ are smooth in n so as to appear nearly constant within a window of integer length N_m . This is essentially a “quasi-stationary” assumption on the system components of $y(t)$, but we must

remember that $y(t)$ is still highly nonstationary due to the local phasing of components $\Phi_p[n, t]$ in t .

A lowpass average is a quadratic filter that projects the bilinear product $y(t_1)y(t_2)$ onto a low-frequency subspace. The covariance estimate is of the form

$$\widehat{R}_n(t_1, t_2) = \sum_i Q[i] y(t_1 + T[n + i]) y(t_2 + T[n + i]) \quad (\text{E.10})$$

where $Q[i]$ is a discrete lowpass filter with unit sum. As in the periodic case, the modulators and carriers may vary at different rates. The lowpass assumption is only that the cutoff frequency of $Q[i]$ is greater than or equal to the maximum rate of change, in n , for $\sigma_p^2[n]$ and $\Phi_p[n, t]$.

The point we wish to make is that demodulation is a joint estimation problem for the long-term evolution of both carriers and modulators. As in Chapter 4, the process $y(t)$ is nonstationary at two time scales. The advantage of PCA demodulation is that the two scales are separable as time variables n and t . Covariance estimation first occurs in n , which allows local basis estimation in t . The resulting principal components $\Phi_p[n, t]$ then determine exact modulator amplitudes via (E.18) and (E.24). Consistent with Section 4.3, we have given two examples of time-varying estimation. However, the problem is generally one of fitting a basis to the second-order statistics of $y(t)$.

E.3 Sum-of-Products Form

For our purposes, the above formulations are important because they define how to “demodulate” the signal $y(t)$ in a very general sense. First, we have to derive a sum-of-products synthesis equation for $y(t)$ based on its principal components. For convenience, let us assume the sampling interval T is equal to the length of the analysis window $g(t)$. In this case, there is no overlap between windows. Thus we can synthesize¹ $y(t)$ by inverting the modified DFT and overlap-adding successive frames,

$$y(t) = \sum_n f(t - nT) e^{j\pi t/T} \sum_{k=0}^{N-1} Y[n, k] e^{j2\pi kt/T} \quad (\text{E.11})$$

¹This framework can obviously extend to overlapping windows, in which case $f(t)$ is the synthesis window counterpart to $g(t)$. See [108, 109, 107] and Section 2.3.3.

where $f(t) = g^{-1}(t)$ for $-T/2 < t < T/2$ and zero everywhere else. This is a Fourier-series synthesis, defining

$$e^{j\pi t/T} \sum_{k=0}^{N-1} Y[n, k] e^{j2\pi kt/T} = y[n, t] \quad (\text{E.12})$$

which is T -periodic in t . Similarly periodizing the short-time principal components yields

$$y(t) = \sum_n \sum_{p=1}^N s_p[n] \Phi_p[n, t] f(t - nT) \quad (\text{E.13})$$

where $\Phi_p[n, t]$ is also T -periodic in t due to Fourier-series extrapolation. The window $f(t - nT)$, however, truncates the series to one period for each n .

At this point, we are close to representing $y(t)$ as a sum of products conducive to coherent demodulation. We make two assumptions. First, $s_p[n]$ is uncorrelated in n , which is true when T is greater than the correlation duration of $y(t)$. Therefore,

$$s_p[n] = \lambda_p(nT) \cdot w_p[n] \quad (\text{E.14})$$

where $w_p[n]$ is white and stationary Gaussian noise, and $\lambda_p(nT)$ is a real modulating function satisfying $\lambda_p^2(nT) = \sigma_p^2[n]$. The second assumption is that $\lambda_p(t)$ is smooth enough in t such that it appears almost constant over an interval of length T . This allows the sum-of-products formulation

$$y(t) \approx \sum_p \lambda_p(t) \cdot x_p(t) \quad (\text{E.15})$$

where $x_p(t)$ is a random process with local principal components $\Phi_p[n, t]f(t - nT)$ at time $t = nT$, or

$$x_p(t) = \sum_n w_p[n] \Phi_p[n, t] f(t - nT). \quad (\text{E.16})$$

We refer to $\lambda_p(t)$ as the p th real modulator and $x_p(t)$ as the p th real carrier stream consisting of local components $\Phi_p[n, t]f(t - nT)$. Each stream is independent from every other stream, so (E.15) represents an orthogonal modulation decomposition. When $\Phi_p[k, t]$ are bandpass with different center frequencies depending on k , then (E.15) corresponds to modulated subbands akin to those in Chapters 3 and 4.

E.4 Generalized Coherent Demodulation

As in previous chapters, coherent demodulation requires knowledge of the carrier processes. Based on the synthesis equation (E.13), and using the fact that $\Phi_p[n, t]$ is orthonormal at each time instant n , we obtain

$$s_p[n] = \int y(t)\Phi_p[n, t]g(t - nT)dt. \quad (\text{E.17})$$

The result is an AM-WSS random process, which we assumed earlier is uncorrelated in n . Therefore, the non-negative modulator can be found via DEMON square-law as

$$\lambda_p(nT) = E \left\{ \left(\int y(t)\Phi_p[n, t]g(t - nT)dt \right)^2 \right\}^{1/2}. \quad (\text{E.18})$$

This is the real modulator for the p th independent carrier stream. It is real and non-negative by definition, and is essentially the square-root of the p th, time-varying, short-time PCA eigenvalue.

It is possible to define a complex modulator within the PCA framework, as long as the principal components form quadrature pairs. As we found in Section 5.2, quadrature pairs are still orthogonal even though one is the Hilbert transform of the other. For $N/2$ pairs,

$$\Phi_{p,2}[n, t] = \widehat{\Phi}_{p,1}[n, t], \quad 1 \leq p \leq N/2 \quad (\text{E.19})$$

or equivalently,

$$U_{p,2}[n, k] = -jU_{p,1}[n, k], \quad 1 \leq p \leq N/2, \quad 0 \leq k < N/2. \quad (\text{E.20})$$

In this case, the synthesis equation becomes

$$y(t) = \sum_n \sum_{p=1}^{N/2} \left(s_{p,1}[n]\Phi_p[n, t] - s_{p,2}[n]\widehat{\Phi}_p[n, t] \right) f(t - nT). \quad (\text{E.21})$$

The analytic signal² is therefore

$$y_a(t) = \sum_n \sum_{p=1}^{N/2} (s_{p,1}[n] + js_{p,2}[n]) \Phi_{p,a}[n, t] f(t - nT) \quad (\text{E.22})$$

²This assumes that $f(t)$ and the principal components are disjoint in the frequency domain, as required by Bedrosian's product relation. Since $f(t)$ is lowpass, we assume the principal components are primarily high-frequency signals.

which is a complex sum-of-products with complex random modulators $Y_p[n] = s_{p,1}[n] + js_{p,2}[n]$. These are detected via

$$Y_p[n] = \int y_a(t)\Phi_{p,a}^*[n,t]g(t-nT)dt, \quad 1 \leq p \leq N/2 \quad (\text{E.23})$$

with corresponding Hermitian and complementary square-law envelopes

$$m_p^{(H)}(nT) = E\{|Y_p[n]|^2\}, \quad m_p^{(C)}(nT) = E\{Y_p^2[n]\} \quad (\text{E.24})$$

reminiscent of Section 4.5. The above envelopes are the sum and difference, respectively, of the PCA eigenvalues $\sigma_{p,1}^2[n]$ and $\sigma_{p,2}^2[n]$.

We have shown that, with quadrature pairing of principal components, PCA demodulation is equivalent to complex demodulation. If $\Phi_p[n,t]$ corresponds to bandpass functions, with a different center frequency for each p , then $m_p^{(H)}(t)$ and $m_p^{(C)}(t)$ are coherent subband envelopes. This justifies the claim that PCA demodulation generalizes the subband approaches taken in Chapters 3 and 4. It is important to note, however, that the definition of coherent envelopes requires the association of $\Phi_p[n_1,t]$ with $\Phi_p[n_2,t]$ and so on. PCA gives eigenvectors sorted arbitrarily in p , so in practice one will have to associate components over time n by sorting in p accordingly. If the components are indeed bandpass, then it may be desirable to associate by nearest center frequency.

E.5 Summary

It is possible to apply widely linear PCA to STFT frames, as a local estimate of principal components. The full scope of this approach is to generalize demodulation via abstracted versions of modulators and carriers which are possibly not associated with subbands. Carriers are instead sequences of principal component functions synchronized with local features in $y(t)$, where the “feature size” is the time interval T . The corresponding modulators evolve on a time scale considerably greater than T . Estimation generally requires joint estimation of modulators and carriers, since modulators are defined with respect to carriers yet also affect the time-variation of the local covariance function of $y(t)$. Although a hard problem, covariance estimation is possible under assumptions of periodicity or smooth variation over

time. Ultimately, a taxonomic pairing of STFT principal components is generally useful for discovering carriers in a signal.

VITA

Pascal Clark was born in Wyoming, where he prospected for dinosaur fossils in his youth. After moving to Texas and then Washington state, he entertained notions of becoming a writer. This changed after a programming project involving fractals and animated plant growth turned him to science and engineering. He completed his BSEE, MSEE, and PhD at the University of Washington, Seattle, obtaining the latter in early 2012. As of this writing, he is starting a post-doctoral research position at Johns Hopkins University.