

**EMPIRICAL ANALYSIS FOR NON-STATIONARY SIGNAL  
DE-NOISING, DE-TRENDING AND DISCRIMINATION  
APPLICATIONS**

by

Muhammad Farhat Kaleem, B.Sc., M.Sc.,  
University of Engineering & Technology, Lahore, Pakistan, 1996,  
Hamburg University of Technology, Hamburg, Germany, 2000.

A dissertation  
presented to Ryerson University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
in the Program of  
Electrical and Computer Engineering.

Toronto, Ontario, Canada, 2014

© Muhammad Farhat Kaleem, 2014.

## **Author's Declaration**

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

# **Abstract**

## **EMPIRICAL ANALYSIS FOR NON-STATIONARY SIGNAL DE-NOISING, DE-TRENDING AND DISCRIMINATION APPLICATIONS**

© Muhammad Farhat Kaleem, 2014.

Doctor of Philosophy in the Program of Electrical and Computer Engineering,  
Ryerson University.

This dissertation focuses on the study and development of methods for empirical analysis of non-stationary signals in the context of de-noising, de-trending and discrimination applications. For this purpose, Empirical Mode Decomposition (EMD), which is a relatively new signal decomposition technique, is chosen as the starting point. EMD does not rely on any fixed basis, but instead defines a signal adaptive decomposition methodology. The use of EMD for signal de-noising and de-trending is demonstrated through formulation of a methodology for mental task classification using EEG signals. Furthermore, a methodology for analysis and classification of pathological speech signals is developed, whereby a high classification accuracy through use of meaningful instantaneous features is demonstrated. Following this, a novel modification of EMD, named Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS), is proposed. EMD-MPS allows a time-scale based decomposition of signals, which is not possible using the original EMD algorithm. The EMD-MPS algorithm is defined, and its properties empirically established, thereby validating the expected behaviour of EMD-MPS. Importantly, EMD-MPS is shown to provide new insight into the decomposition behaviour of the original EMD algorithm. Also, a novel hierarchical decomposition methodology, which uses the time-scale based decomposition of EMD-MPS to divide a signal into selected frequency bands, is developed and illustrated using synthetic and real world signals. EMD-MPS is also used for time-scale based de-noising and de-trending of signals, first demonstrated using synthetic and real signals, and then validated by practical applications such as mental task classification and seizure detection. An empirical sparse dictionary learning framework based on EMD with application to signal classification is then proposed and developed in the dissertation. As part of this framework, a discriminative dictionary learning algorithm is developed, and characteristics of the empirical dictionary established. The utility of the proposed framework for signal classification is demonstrated using EEG signals. The proposed framework is then applied for automated seizure detection using long-term EEG recordings, and the results are used to discuss the potential and implications

for patient-specific dictionaries, as well as the associated advantages of the framework when using long-term data.



# Acknowledgments

First of all, I would like to sincerely thank both of my supervisors and mentors, Dr. Sridhar Krishnan and Dr. Aziz Guergachi, for their support, and most importantly, patience, with all aspects of my PhD studies. For this, I would always be indebted.

I would also like to thank Dr. Karthi Umapathy, who answered many of my questions over the years, and Dr. Behnaz Ghoraani, collaborating with whom was a pleasure and a beneficial learning experience. Thanks are also due to Dr. Shengkun Xie, for both the technical and general discussions. I also want to thank all the friends and colleagues at Signal Analysis Research (SAR) group.

Finally, I would like to express my gratitude for the love and support extended to me by my family. I hope my having achieved this milestone will make them proud and happy, and feel relieved as well.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Signal Categories . . . . .	3
1.2.1	Deterministic and Non-Deterministic . . . . .	3
1.2.2	Stationary and Non-Stationary . . . . .	4
1.2.3	Linear and Non-Linear Signals . . . . .	6
1.2.4	Real-World Signals . . . . .	7
1.3	Analysis of Non-Stationary Signals . . . . .	7
1.3.1	Signal Representation Tools . . . . .	7
1.3.2	Signal Analysis Methods . . . . .	11
1.3.3	Feature Analysis . . . . .	17
1.4	Contributions . . . . .	18
1.5	Structure of Dissertation . . . . .	20
<b>2</b>	<b>Empirical Mode Decomposition: An Empirical Approach for Signal Analysis</b>	<b>23</b>
2.1	Introduction . . . . .	23
2.2	Empirical Mode Decomposition . . . . .	26
2.3	Empirical Mode Decomposition Algorithm . . . . .	27
2.3.1	Instantaneous Measurements . . . . .	29
2.3.2	Marginal Spectrum . . . . .	32
2.4	De-noising and De-trending based on Empirical Mode Decomposition . . . . .	32
2.4.1	Experiment: Mental Task Classification using EEG Signals . . . . .	38
2.5	Feature Analysis using EMD . . . . .	46

2.5.1	Experiment 1: Pathological Speech Signal Analysis and Classification	47
2.5.2	Experiment 2: Telephone-Quality Pathological Speech Classification	65
2.6	Chapter Summary	69
<b>3</b>	<b>Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS)</b>	<b>71</b>
3.1	Introduction	71
3.2	Variations of Empirical Mode Decomposition	73
3.3	Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS)	78
3.3.1	EMD-MPS Method	79
3.3.2	Important Properties of EMD-MPS	82
3.4	Hierarchical Decomposition using Empirical Mode Decomposition-Modified Peak Selection	99
3.4.1	Illustration of Hierarchical Decomposition using Synthetic and Real Signals	101
3.5	Chapter Summary	107
<b>4</b>	<b>De-Noising and De-Trending with Empirical Mode Decomposition-Modified Peak Selection</b>	<b>108</b>
4.1	Background	108
4.2	Time-scale based De-noising and De-trending	112
4.2.1	De-noising and De-trending of Synthetic Signals	112
4.2.2	De-noising and De-trending of Real-life Signals	115
4.2.3	Discussion	119
4.2.4	Experiment: Mental Task Classification using EEG Signals	120
4.2.5	Experiment: EEG Seizure Detection and Epilepsy Diagnosis	128
4.3	Chapter Summary	135
<b>5</b>	<b>Empirical Sparse Dictionary Learning</b>	<b>136</b>
5.1	Background	136
5.2	Discriminative Dictionary Learning Algorithm	140
5.2.1	Learning the Trained Dictionary	140
5.2.2	Some Characteristics of the Learned Dictionary	142
5.2.3	Signal Classification Using the Trained Dictionary	150
5.3	Experiments	151

5.3.1	Scenario I . . . . .	152
5.3.2	Scenario II . . . . .	152
5.3.3	Scenario III . . . . .	153
5.3.4	Discussion . . . . .	154
5.4	Conclusions . . . . .	155
5.5	Automatic Seizure Detection Based on Empirical Dictionary Learning Using Long-term EEG Data . . . . .	155
5.5.1	Data . . . . .	156
5.5.2	Data Processing For Dictionary Creation and Learning . . . . .	159
5.5.3	Dictionary Formation and Training . . . . .	162
5.5.4	Results . . . . .	163
5.5.5	Discussion: From Global to Patient-Specific Dictionaries . . . . .	164
5.6	Chapter Summary . . . . .	168
<b>6</b>	<b>Conclusion</b>	<b>169</b>
<b>A</b>	<b>EMD: Algorithm Implementation Issues and Characteristics</b>	<b>175</b>
A.1	Algorithm Implementation Issues . . . . .	175
A.2	Characteristics of Empirical Mode Decomposition . . . . .	178
<b>B</b>	<b>EMD-MPS: De-noising and De-trending</b>	<b>183</b>
B.1	De-noising and De-trending of Synthetic Signals . . . . .	183
B.2	De-noising and De-trending of Real-life Signals . . . . .	186
	<b>References</b>	<b>190</b>

# List of Figures

1.1	Illustration of different signal categories: (a) Deterministic: 5 cycles of sinusoidal signal with frequency 5 Hz and an amplitude of 1. (b) Deterministic and non-stationary: a chirp signal with an amplitude of 1 and frequency linearly increasing from 0 to 50 Hz. (c) Non-deterministic and stationary: 3 signals representing 3 different trials of rolling a dice each second, with the value observed taken as the signal amplitude. (d) Non-deterministic and non-stationary: Mean temperature recorded in Toronto from January 1, 2008 till December 31, 2008, with the data obtained from the National Climate Data and Information Archive provided by Environment Canada [1]. . . . .	5
1.2	(a) Signal containing two sinusoids corrupted with zero-mean random noise. The two sinusoids, having frequencies of 50 and 120 Hz, occur between 0.2 and 0.4 seconds, and between 0.6 and 0.8 seconds, respectively; (b) Representation of the signal in the frequency domain, using the Fourier Transform. The frequency domain representation indicates the presence of the 50 and 120 Hz components; (c) Representation of the signal in the time-frequency domain, using the Short-Time Fourier Transform (with Kaiser window having length of 256 samples and overlap of 220 samples). The time-frequency representation not only indicates the presence of the 50 and 120 Hz components, but also their temporal locations in the signal, between 0.2 and 0.4 seconds, and 0.6 and 0.8 seconds respectively; (d) Representation of the linear chirp signal in Figure 1.1 in the time-frequency domain, showing the linear increase in frequency with time. . . . .	10
1.3	Block diagram of the dissertation. The connections between the Chapters are shown in solid lines. The main practical contributions of each Chapter are shown with dashed lines. . . . .	21

2.1	Chapter 2 contains details about the Empirical Mode Decomposition (EMD) method, as well as practical applications based on de-noising/de-trending and feature analysis using EMD. . . . .	24
2.2	Demonstration of EMD: (a) Signal containing two quadratic chirps. This signal is decomposed using EMD. (b) The first IMF, containing the chirp with the higher instantaneous frequencies. (c) The second IMF, containing the chirp with the lower instantaneous frequencies. (d) The time-frequency plot represented by the Hilbert spectrum, $H[\omega, n]$ , obtained using the steps defined in Section 2.3.1. . . . .	30
2.3	An original EEG signal used in the experiment. . . . .	40
2.4	Mean of IMFs during partial reconstruction; mean changes significantly from 0 after IMF number $l = 5$ . . . . .	41
2.5	(A De-trended EEG signal (top), and Trend of the EEG signal (bottom). . .	42
2.6	Average correct classification percentage for all subjects using high and low frequency energies, listed as HFE (left bars) and LFE (right bars), respectively. 46	
2.7	Block diagram for pathological speech classification described in Experiment 1. The dashed lines represent feedback links used to determine speech segment length and number of IMFs (Section 2.5.1.7). . . . .	50
2.8	800 ms portions of randomly chosen normal (upper figure) and pathological (lower figure) speech signals and their corresponding first 10 IMFs. . . . .	51
2.9	Time-Frequency plot of IMF 7 of both normal (upper figure) and pathological (lower figure) speech signals, showing a more even instantaneous frequency structure for normal speech signals. . . . .	55
2.10	Figures showing marginal spectrum $h(\omega)$ of IMFs 4, 5, 6 and 7 of both normal (solid lines) and pathological (dashed lines) speech signals. These figures illustrate that the marginal spectrum for pathological speech has higher amplitude beyond the frequency thresholds for the respective IMFs as compared to marginal spectrum for normal speech. . . . .	57
3.1	Chapter 3 describes Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS), which is a novel modification of EMD, and discusses the method's decomposition properties. . . . .	72

3.2	Mean spectra (power spectral density in $dB$ on y-axis, and $\log_2(frequency)$ on x-axis) of $\tau$ -functions $T_1$ to $T_5$ , obtained using three different values of $\tau$ for 2500 fGn sequences with value of $H = 0.5$ . Unlabeled numbers in the figures represent the indices of $\tau$ -functions. . . . .	84
3.3	Center frequency of $\tau$ -functions $T_2, T_3, T_4$ obtained using different values of $\tau$ plotted against the $\tau$ -function index $i$ , using different values of $H$ (top: $H = 0.2$ , center: $H = 0.5$ , bottom: $H = 0.8$ ). The slopes of the least-square fits and the values of $\tau$ are indicated on the plots. . . . .	85
3.4	Normalized power spectra (with power spectral density in $dB$ on y-axis, and $\log_2(frequency)$ on x-axis) of 3 $\tau$ -functions according to Equation 3.5. The $\tau$ -functions were obtained using a value of $\tau = 5$ . The spectra for $H = 0.2$ were obtained using $C = \beta^{\alpha/2}$ in Equation 3.5. . . . .	87
3.5	Center frequency of $\tau$ -function $T_2$ ( $\log_2$ values) obtained for different values of $\tau$ , plotted against the index $j$ . . . . .	89
3.6	Probability density function (PDF) estimates of 2500 realizations of three $\tau$ -functions of fractional Gaussian noise with $H = 0.5$ obtained using a Gaussian smoothing function estimate. The dashed lines represent the average PDF of all the PDFs of individual $\tau$ -functions. . . . .	91
3.7	Probability density function (PDF) $\tau$ -function $T_1^5$ for different lengths of the fractional Gaussian noise signal with $H = 0.5$ . As the length of the noise sequence increases, the PDF curve starts changing to multi-modal. . . . .	92
3.8	Relationship between decomposed frequencies and $\tau$ for different values of $H$ . . . . .	95
3.9	Value of the performance measure in Equation 3.11 plotted against $f_\tau/f_2$ . . . . .	97
3.10	Hierarchical decomposition of a signal into a hierarchy of three levels a, b and c. The values of $\tau$ given by $\tau_a, \tau_b$ and $\tau_c$ determine the hierarchy of decomposition. . . . .	100
3.11	Synthetic signal (plot at the bottom) containing 3 sinusoids at frequency values 16, 64 and 256 cycles/second (top three plots). . . . .	101
3.12	Power spectral density plot of the synthetic signal shown in Figure 3.11 which illustrates the three frequency components in the signal. . . . .	102
3.13	Power spectral density plots of the $\tau$ -functions $T_1^{\tau_a}, T_2^{\tau_a}, T_1^{\tau_b}$ and $T_2^{\tau_b}$ obtained by hierarchical decomposition of the synthetic signal shown in Figure 3.11. . . . .	103

3.14	EEG signal segment of length four seconds (1024 samples) used to demonstrate hierarchical decomposition to separate frequency components in the frequency range $3 \leq f \leq 29$ . . . . .	104
3.15	Power spectral density plots of the $\tau$ -functions $T_1^{\tau_a}$ (a), $T_1^{\tau_b}$ (b) and $T_2^{\tau_b}$ (c) obtained by hierarchical decomposition of the EEG signal shown in Figure 3.14. . . . .	106
4.1	Chapter 4 focuses on de-noising and de-trending using the time-scale based decomposition made possible by the EMD-MPS method. Practical applications of time-scale base de-trending in terms of mental task classification and seizure detection are also presented. . . . .	109
4.2	De-noising of a sinusoidal signal contaminated with additive white Gaussian noise (AWGN) using EMD-MPS. (a) Signal contaminated with AWGN: SNR = 20dB. (b) Extracted noise in $\tau$ -function $T_1$ . (c) Recovered signal in $\tau$ -function $T_2$ superimposed on the original signal. (d) Value of correlation coefficient ( $\tau$ -function $T_2$ and original sinusoidal signal) plotted against SNR of test signal to quantify de-noising performance of EMD-MPS. . . . .	113
4.3	De-noising of a sinusoidal signal contaminated with AWGN using EMD. Severe mode-mixing results even at a relatively high SNR of 16 dB. . . . .	114
4.4	Demonstration of EMD mode-mixing, making removal of intermittently occurring low amplitude sinusoidal signals impossible. (a) Sinusoidal signal with intermittently occurring low amplitude components of higher frequency. (b) Mode-mixed IMFs resulting from EMD decomposition of signal in (a). . . . .	115
4.5	Removal of intermittently occurring low amplitude sinusoidal signals from the larger amplitude sinusoid using EMD-MPS. (a) Sinusoidal signal with intermittently occurring low amplitude components of higher frequency. (b) De-noised signal (dashed line) using EMD-MPS compared with the original signal (solid line). . . . .	115
4.6	Time-scale based de-trending of S&P 500 index data from November 6, 2001 to October 11, 2011, showing the monthly trend in (a) and the 3-monthly trend in (b). . . . .	116



4.7	De-trending EEG signals using EMD-MPS. (a) De-trended EEG signals from 4 selected channels. (b) Marginal spectrum of the de-trended signal and the trend, showing clear separation between the spectra. . . . .	117
4.8	De-trending for baseline wander removal of a 15000 sample portion of Holter ECG signal. The upper graph shows the original signal along with the extracted trend (baseline wander), whereas the lower graph shows the de-trended ECG signal. . . . .	118
4.9	Schematic representation of the methodology described in this experiment. .	121
4.10	EEG signal from Subject 1 performing Task 2 (top). $\tau$ -function $T_1$ , representing de-trended signal (middle). $\tau$ -function $T_2$ , representing the trend (bottom). These $\tau$ -functions have been obtained using a value of $\tau = 31.25$ ( $\hat{\tau} = 14$ ), corresponding to a frequency separation value of $F = 8$ Hz. . . . .	122
4.11	Classification accuracy for all subjects corresponding to $\hat{\tau} = 14$ (left bars) and $\hat{\tau} = 28$ (right bars). . . . .	127
4.12	Diagram representing the methodology described in this experiment. . . . .	129
4.13	EEG signal from an epilepsy patient from set C (top). $\tau$ -function $T_1$ , representing de-trended signal (middle). $\tau$ -function $T_2$ , representing the trend (bottom). These $\tau$ -functions have been obtained using a value of $\tau = 21.6$ ( $\hat{\tau} = 9.5$ ) corresponding to a frequency separation value $F = 8$ Hz. . . . .	131
5.1	Chapter 5 presents details of the novel empirical sparse dictionary learning framework, which is based on signal decomposition using EMD. The use of the framework for signal classification is also described, including seizure detection using long-term data. . . . .	137
5.2	Increase in size of learned dictionary plotted against the number of iterations in Algorithm 3, showing expected and actual increase in dictionary size using examples described in Section 5.3. . . . .	143
5.3	Increase in computational time of Algorithm 3 with increase in number of iterations. The solid line represents a least-squares fit. The values have been obtained using data described in Section 5.3.1. . . . .	144

5.4	Increase in computational time for projecting a fixed number of signals against the trained dictionary $\mathbf{D}_{Train}^C$ , with an increase in dictionary size, represented here by the number of iterations for Algorithm 3. The solid line represents a least-squares fit, and the values have been obtained using data described in Section 5.3.1. . . . .	145
5.5	Original signal in dashed line superimposed with the reconstructed signal obtained using the relation in Equation 5.5. For the plot in (a), the trained dictionary was obtained after one iteration of Algorithm 3, and after 10 iterations for the plot in (b). . . . .	147
5.6	Decrease in reconstruction error with an increase in size of the trained dictionary, as indicated by the iteration number. . . . .	147
5.7	Original signal in dashed line superimposed with the reconstructed signal obtained using the relation in Equation 5.5. For the plot in (a), the trained dictionary was obtained after one iteration of Algorithm 3, whereas for the plot in (b) the trained dictionary was obtained after 10 iterations. . . . .	148
5.8	Decrease in reconstruction error with an increase in size of the trained dictionary, as indicated by the iteration number. . . . .	148
5.9	Examples of dictionary atoms from a trained dictionary $\mathbf{D}_{Train}^C$ learned after 10 iterations of Algorithm 3. . . . .	149
5.10	Scatter plot of projection coefficients of test signals from two classes, D (shown as crosses) and E (shown as circles). Clear separation of the signals in the feature space is visible. . . . .	151
5.11	Example of dictionary atoms consisting of IMFs of decomposed signals taken from the trained dictionary described in Section 5.3.2 . . . . .	153
5.12	EEG signals from a 27 seconds seizure segment from a recording of patient number 1. The channels, from top to bottom, are, respectively: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ. . . . .	158
5.13	EEG signals from a 27 seconds non-seizure segment from a recording of patient number 1. The channels, from top to bottom, are, respectively: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ. . . . .	158

5.14	Block diagram representing the data processing for seizure signals from all 15 patients. . . . .	161
5.15	Block diagram representing the data processing for non-seizure signals from all 15 patients. . . . .	162
5.16	Examples of dictionary atoms in the trained dictionary $\mathbf{D}_{Train}^C$ . . . . .	163
A.1	Illustration of mode-mixing: (a) A 10 Hz sinusoidal signal containing intermittently occurring 50 and 100 Hz components; (b) Envelope formation using cubic spline interpolation. The dashed black lines represent the upper and lower envelopes, and the red line represents the average of the envelopes; (c) The first IMF obtained after the sifting process, showing mode-mixing, as it contains multiple oscillatory modes; (d) The residue obtained after the first IMF has been extracted, from which further IMFs will be extracted by the sifting process. The residue illustrates that mode-mixing is going to propagate to other IMFs as well. . . . .	179
B.1	Non-seasonal and seasonal time-series with exponential trends, represented by the expression in Eq. B.1. (a) $y^1(t)$ . (b) Extracted exponential trend from $y^1(t)$ compared with the original trend $T_1(t)$ . (c) $y^2(t)$ . (d) Extracted exponential trend from $y^2(t)$ compared with the original trend $T_1(t)$ . . . . .	184
B.2	Time-series consisting of an AR(2) process superimposed on a piecewise linear trend, represented by the expression in Eq. B.2. (a) $y^3(t)$ . (b) Extracted piecewise linear trend from $y^3(t)$ compared with the original trend $T_2(t)$ . . . . .	185
B.3	Time-scale based de-trending of S&P 500 index data from August 1, 2001 to May 3, 2010, showing the 3-monthly trend in (a), 6-monthly trend in (b) and 9-monthly trend in (c). . . . .	187
B.4	Pathological speech signal de-noising using EMD-MPS. (a) A 0.5 second segment of a pathological speech signal. (b) $\tau$ -function $T_1$ containing oscillatory components above a frequency threshold of 2 kHz. (c) $\tau$ -function $T_2$ containing the de-noised signal. (d) Marginal spectrum of $T_1$ and $T_2$ showing separation of the spectra of the noise and de-noised signal. . . . .	188

B.5	Baseline wander removal from 25000 sample long segments of Holter ECG signals using EMD-MPS de-trending. The upper graphs in each figure represent the original ECG segments, whereas the lower graphs represent the de-trended ECG segments. . . . .	189
-----	---	-----

# List of Tables

2.1	Classification Accuracy For Classification Of Mental Tasks For All Subjects .	44
2.2	Thresholds for calculating the sum of marginal spectrum as a discriminative feature. . . . .	58
2.3	Classification accuracy (rounded to nearest whole number) obtained using different continuous speech portion lengths. . . . .	60
2.4	Classification accuracy (rounded to nearest whole number) obtained using different number of IMFs. . . . .	61
2.5	$p$ -values obtained by performing unpaired $t$ -tests of the null hypothesis that the feature values obtained for each of 10 IMFs of normal and pathological signals used in the study have the same mean. . . . .	62
2.6	Classification Results using Linear Discriminant Analysis with 10-fold cross-validation. TP:True Positive, TN:True Negative, FP:False Positive, FN:False Negative . . . . .	63
2.7	Average classification accuracy with different levels of noise added . . . . .	68
2.8	Classification results for one run of experiment in terms of measures defined in [2]. SNR: signal-to-noise ratio. Acc: Accuracy. Sens: Sensitivity. Spec: Specificity. Pos Pred: Positive Predictivity. Neg Pred: Negative Predictivity. Acc(%):(TP+TN)/TP+TN+FP+FN. Sens(%):TP/(TP+FN). Spec(%):TN/(TN+FP). Pos Pred(%):TP/(TP+FP). Neg Pred(%):TN/(TN+FN). . . . .	68
3.1	Illustration of the convergence of EMD algorithm due to doubling of $\tau$ after each IMF extraction for a signal with length $N = 1024$ samples . . . . .	94
3.2	Separation of the frequency components of the synthetic signal shown in Figure 3.11 using hierarchical decomposition . . . . .	102

3.3	Hierarchical decomposition of the EEG signal segment shown in Figure 3.14 using two-level hierarchical decomposition . . . . .	105
4.1	Classification Accuracy for Classification of Mental Tasks For All Subjects using 1-NN Classifier and 10-fold Cross-Validation. . . . .	125
4.2	Classification accuracy for the 3 scenarios (Section 4.2.5.2) using 1-NN clas- sifier and 10-fold cross-validation . . . . .	133
5.1	Gender and Ages of Patients. . . . .	156
5.2	Number of Seizures Used Per Patient. . . . .	159
5.3	Seizure detection performance using data presented in Table 5.1. . . . .	165
B.1	Objective measures of the quality of trends extracted using EMD-MPS from synthetic signals $y^1(t)$ , $y^2(t)$ and $y^3(t)$ . . . . .	186

# Chapter 1

## Introduction

### 1.1 Background

The field of signal processing is replete with different techniques for automated signal analysis. Automated signal analysis allows researchers better understanding of the underlying mechanisms and processes of the systems from which signals have been generated. For example, automated analysis of financial time-series using signal processing algorithms can be used to predict price evolution and stock classification for designing diversified investment portfolios [3]. Signal processing techniques also allow utilization of signals to ascertain the state of the underlying system generating the signals. As an example in this regard, we may refer to use of signal processing techniques for pathological speech classification [4], whereby speech produced by damaged or malfunctioning human speech organs can be automatically differentiated from healthy speech with high accuracy.

While the two examples in the previous paragraph were presented in the context of automated signal analysis using signal processing techniques, the same tasks are also performed manually. For example, a financial analyst might visually examine graphs of the daily prices of stocks of interest to predict future movement of a stock, or a medical practitioner might infer the level of pathology present in a patient's voice by listening to her speech. Yet another example could be that of a medical expert examining hours long Electroencephalogram (EEG) signals to identify the onset and duration of seizure activity. As may be ascertained

from these few examples, there are a number of issues associated with manual methods of signal analysis. These may be categorized as follows [5]:

- **Subjective:** Manual methods of signal analysis are subjective to the expertise of the person performing the analysis, whereas automated methods have no such limitation.
- **Error prone:** There are more chances of error in manual signal analysis, especially if a task is repetitive, or involves a large amount of data. On the other hand, automated signal analysis does not have such constraints.
- **Laborious:** Automated methods take the labour out of signal analysis, while at the same time minimizing the chances of error. A good example here is analysis of data in the form of several hours of EEG signals, with just one or two epileptic seizures of a few seconds. The brief duration of the portion of interest in the signal makes it difficult to find manually, leading to delayed or erroneous diagnosis.
- **Impractical:** In many cases, the signal analysis task can only be performed using automated means. For example, information about the frequency content of a human speech signal can only be obtained using transformation of the signal into the frequency domain, and not by visual analysis of a time representation of the signal.

In this context, the development of advanced signal processing techniques for automated signal analysis represents an important research task. As already mentioned, these techniques provide objectivity, reliability, ease of use, and practicality in different signal analysis situations. Another important advantage of automated signal processing lies in the reusability of the same techniques in different situations across domains. For example, automated signal processing techniques developed for analyzing geophysical data may be carried over to the domain of biomedical engineering. However, successful application of such techniques also requires a clear understanding of the different types of signals which will be analyzed by the techniques. This is important since the automated signal analysis method must adequately



account for the peculiarities of the signal being analyzed. Therefore knowledge of the different signal categories is important for successful application of an adequate automated signal analysis technique.

## 1.2 Signal Categories

The signals encountered by a signal processing researcher range from simple, predictable structures having well-defined mathematical properties, to complex variations of either time, frequency or space which are not possible to confine within a mathematical formulation. However, this wide variety of signals can be grouped into well-defined and meaningful categories. These categories, and the classification of signals within these categories, may be listed as follows:

### 1.2.1 Deterministic and Non-Deterministic

1. **Deterministic:** Signals which are deterministic can generally be described by a mathematical formulation, and the signal structure is known at all times, including the past, present and future. A basic example of a deterministic signal is a sinusoidal signal of the form  $y(t) = \sin(2\pi ft)$ , where  $y$  represents the value of the signal at time  $t$ , and  $f$  represents the frequency of the signal, or the number of signal cycles in one second.
2. **Non-deterministic:** Signals whose time and frequency values cannot be predicted can be categorized as non-deterministic signals. As an example of a non-deterministic signal, we may consider the value recorded after rolling a dice every second, with this value representing the signal amplitude. In this case, the values of the signal amplitude at different instants of time are unpredictable, and it is not possible to formulate a mathematical expression for this type of signal.

### 1.2.2 Stationary and Non-Stationary

1. **Stationary:** Signals may be classified as stationary if the signal statistics, such as mean and variance, do not change with time. We may here mention the dice rolling example presented in the previous subsection, where the probability of the signal having an amplitude value between 1 and 6 at a particular time instant is known to be  $\frac{1}{6}$ . This type of signal can be considered a stationary signal. According to the traditional definition [6], a signal  $X(t)$  can be considered stationary in the wide-sense if the following hold, for all values of time  $t$ :

$$\begin{aligned} E(|X(t)|^2) &< \infty, \\ E(X(t)) &= m, \\ C(X(t_1), X(t_2)) &= C(X(t_1 + \tau), X(t_2 + \tau)) = C(\tau_1 - \tau_2) \end{aligned} \tag{1.1}$$

where  $E(\cdot)$  represents the expected value, and  $C(\cdot)$  is the covariance function.

2. **Non-stationary:** If signal statistics are time-variant, the signals can be classified as non-stationary signals. An example here is a signal which represents the daily mean temperature of a city, since the temperature value at a particular time instant cannot be assigned a fixed probability value.

Given the categorization presented, signals may then be classified into the following four types, with an example signal from each type shown in Figure 1.1.

1. **Deterministic and stationary:** A sinusoidal signal of the type  $y(t) = \sin(2\pi ft)$  belongs to this type, and an example is shown in Figure 1.1 (a).
2. **Deterministic and non-stationary:** An example is a linear chirp signal, which is a sinusoidal signal whose instantaneous frequency varies linearly with time. Such an example signal is shown in Figure 1.1 (b).
3. **Non-deterministic and stationary:** A signal formed by considering the number obtained by rolling a dice at specific instants of time is an example, which is illustrated

in Figure 1.1 (c).

4. **Non-deterministic and non-stationary:** A good example is the daily mean temperature of a city, and such a signal is shown in Figure 1.1 (d).

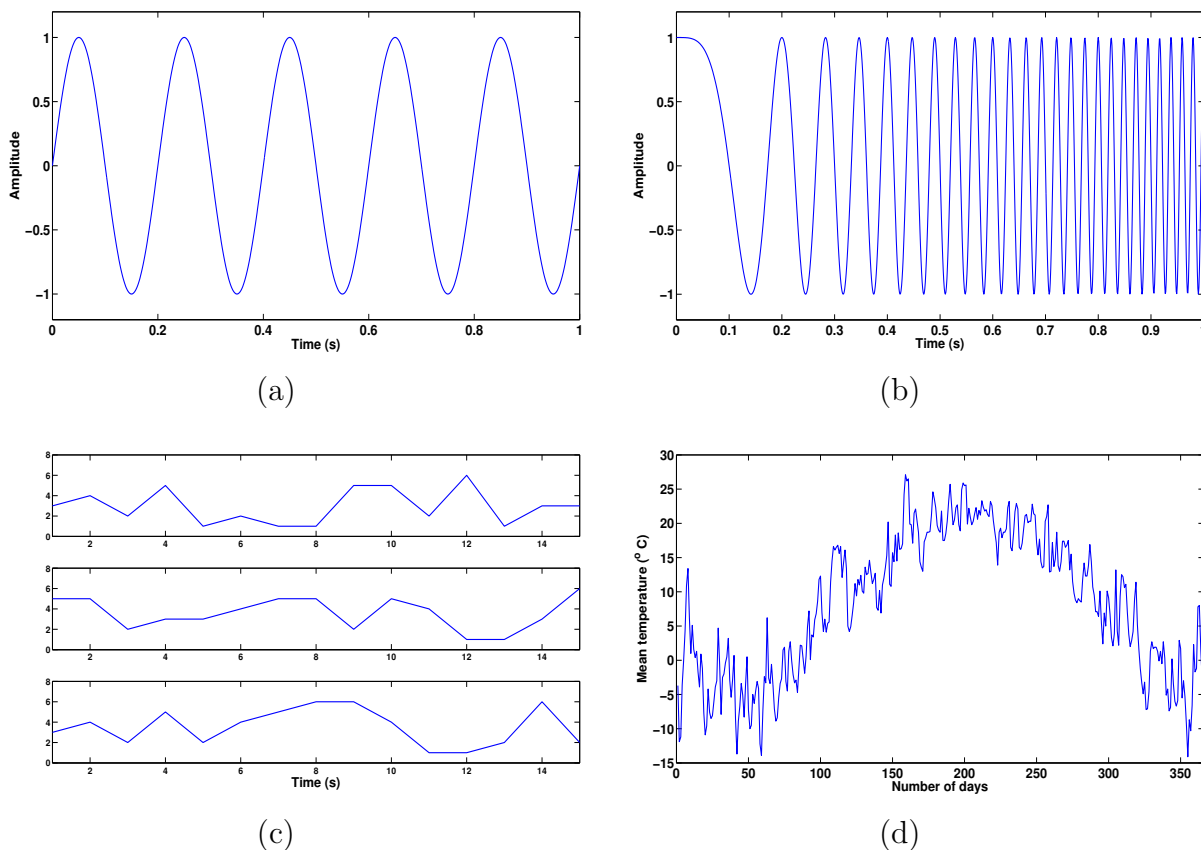


Figure 1.1: Illustration of different signal categories: (a) Deterministic: 5 cycles of sinusoidal signal with frequency 5 Hz and an amplitude of 1. (b) Deterministic and non-stationary: a chirp signal with an amplitude of 1 and frequency linearly increasing from 0 to 50 Hz. (c) Non-deterministic and stationary: 3 signals representing 3 different trials of rolling a dice each second, with the value observed taken as the signal amplitude. (d) Non-deterministic and non-stationary: Mean temperature recorded in Toronto from January 1, 2008 till December 31, 2008, with the data obtained from the National Climate Data and Information Archive provided by Environment Canada [1].

### 1.2.3 Linear and Non-Linear Signals

In addition to the signal categories and the types of signals characterized according to the categorization mentioned in the previous sections, another important aspect of signal characterization relates to whether the signals have been generated from linear or non-linear systems and processes. In general, the terms linear and non-linear signals are used extensively in engineering literature, specially in context of signal analysis techniques that deal with such signals [7]. In this regard, we may refer to a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  being linear iff  $\forall x, y, n \in \mathbb{R}, f(x + y) = f(x) + f(y)$ , and  $f(nx) = nf(x)$ . As pointed out in [7], any signal which does not conform to this form might be thought of as a non-linear signal, for which, however, there is no standard definition. Therefore it is prudent to consider linear and non-linear signals as those which arise from linear and non-linear systems, respectively. For example, non-linear signals may be considered as those generated from underlying dynamical systems which obey non-linear equations. As an example of a non-linear system, we may consider the classic Duffing equation, which has been considered previously in the context of signal processing methods for non-linear signals [8]. The Duffing equation can be written as:

$$\frac{d^2x}{dt^2} + x(1 - \varepsilon x^2) = b \cos vt \quad (1.2)$$

This equation can be seen as representing a non-linear spring having a variable spring constant given by  $1 - \varepsilon x^2$ , where  $\varepsilon$  is a small parameter. When  $\varepsilon$  has a value equal to zero, the equation represents a simple oscillator with a constant period. However, for the case when  $\varepsilon$  is non-zero, the spring constant becomes a function of position, and the equation represents a non-linear oscillator having variable frequency within one oscillation cycle, thereby representing the case of intra-wave frequency modulation. This then represents a typical non-linear system exhibiting non-linear effects represented by harmonics and inter-modulations [8][9]. An example of a signal  $x(t)$  representing intra-wave frequency modulation is presented in [10], and can be written as follows:

$$x(t) = \cos(2\pi t) + 0.5 \cos(2\pi 0.7t) \quad (1.3)$$

The signal in Equation 1.3 can be considered as a single component arising through heterodyning of two closely spaced tones. As pointed out in [10], heterodyning is a routine occurrence in non-linear and physiological systems (occurring, for example, at the contact between the skin and metal electrodes).

It should also be pointed out that the term non-linear signal used in this dissertation has been adopted mainly for the ease of use, and actually refers to a signal generated from an underlying non-linear process or system.

### 1.2.4 Real-World Signals

Most of the signals acquired in the real world, whether through physical measurements, or numerical modeling, represent data from non-stationary and non-linear processes. This holds across diverse domains, e.g. geophysical data analysis [11], structural engineering [12], machine health monitoring [13], biomedical engineering [9], financial data analysis [14], to name just a few. Therefore, it is essential that automated signal analysis methods used for real-world signals are suitable for analysis of non-stationary and non-linear signals.

## 1.3 Analysis of Non-Stationary Signals

In this Section, important requirements for automated signal analysis methods for non-stationary and non-linear signals will be mentioned, followed by a brief review of some representative signal analysis methods. However, before doing that, it is important to characterize the different methods for signal representation, as signal analysis methods also utilize these different forms of signal representation.

### 1.3.1 Signal Representation Tools

There are three main ways to represent a given signal, as described below:

1. **Time:** Signal representation in terms of its amplitude values at each instant of time is the time domain representation of the signal. All the example signals shown in Figure

1.1 are signal representations in the time domain.

2. **Frequency:** The time domain representation of the signals shows the variation of the signal amplitude with time, and it also gives visual clues as to the rate of change of the signal amplitude with time, which is an indication of the signal frequency. For example, in case of the chirp signal shown in Figure 1.1 (b), it can be observed that the signal amplitude varies slowly at the start of the signal, indicating a low frequency, and then the rate of signal amplitude variation increases with time, representing an increase in frequency. Most signals, however, will have not just one frequency component, but many different frequency components, and it is not possible to ascertain the frequency content of a signal just by visual inspection. Therefore the signal needs to be represented in the frequency domain, which is the second way to represent a signal.

The frequency domain representation of a signal shows the relative energy of the frequency components present in the signal, and is also commonly referred to as the spectrum of the signal. The most common way of obtaining the signal spectrum is by using the Fourier spectral analysis, whereby a signal is expanded into a family of an infinite number of sinusoidal functions, and the energy at a particular sinusoidal function represents the signal energy at that particular frequency, resulting in a global energy-frequency representation. The sinusoidal functions are not localized in time, and the Fourier spectrum defines a signal's spectral components along with their corresponding time-invariant amplitudes and phases. In this regard, Figure 1.2 (a) shows the time domain representation of a signal containing two sinusoids corrupted with zero-mean random noise. The length of the signal is 1 second, and the sampling frequency is 1000 Hz. The two sinusoids occur at different temporal locations. The frequency domain representation of this signal, obtained by taking the Fourier transform of the signal, is shown in Figure 1.2 (b). The frequency domain representation clearly shows the presence of the two frequency components, at 50 Hz and 120 Hz. Also, the higher frequency component has an amplitude of unity, whereas the lower frequency component has a lower amplitude with a value of 0.7, which is also reflected in the energy-frequency

distribution represented by the signal spectrum.

3. **Time-Frequency:** Although the signal spectrum obtained using the Fourier transform helps find the frequency components in the signal, the temporal location of the frequency components cannot be ascertained using the Fourier transform. The one second long example signal shown in Figure 1.2 (a) has two sinusoids occurring at different temporal locations, the 50 Hz sinusoid between 0.2 and 0.4 seconds, and the 120 Hz sinusoid between 0.6 and 0.8 seconds, but the frequency domain representation of this signal, shown in Figure 1.2 (b), does not have any information about the temporal location of the sinusoids. In order to represent frequency components of signals in terms of their temporal distribution in a signal, the third way of signal representation is used, which is the time-frequency representation of a signal. A time-frequency representation of a signal is a simultaneous representation of the signal's instantaneous energy distribution over time and frequency. This representation of the example signal of Figure 1.2 (a) is shown in Figure 1.2 (c). In this figure, the  $x$  and  $y$  axes represent the signal's time and frequency values respectively, and the third dimension, which is the intensity of the plot, represents the energy of the signal at each frequency and time instant. A darker point on the graph represents a higher energy content at that particular frequency and time instant. The time-frequency representation of Figure 1.2 (c) clearly shows the temporal location and frequency characteristics of the two sinusoidal components in the signal. Additionally, the time-frequency plot in Figure 1.2 (d) shows the time and frequency characteristics of the chirp signal of Figure 1.1 (c), and the linear increase in signal frequency from a value of 0 Hz to 50 Hz is clearly indicated.

As mentioned in this Section, meaningful analysis of signals requires signal representation in both time and frequency domains. The frequency domain representation is obtained using Fourier spectral analysis, which provides a global energy-frequency distribution, an example of which was provided in Figure 1.2 (b). However, the Fourier method requires the signals to be linear and stationary. This is because the Fourier spectrum represents uniform

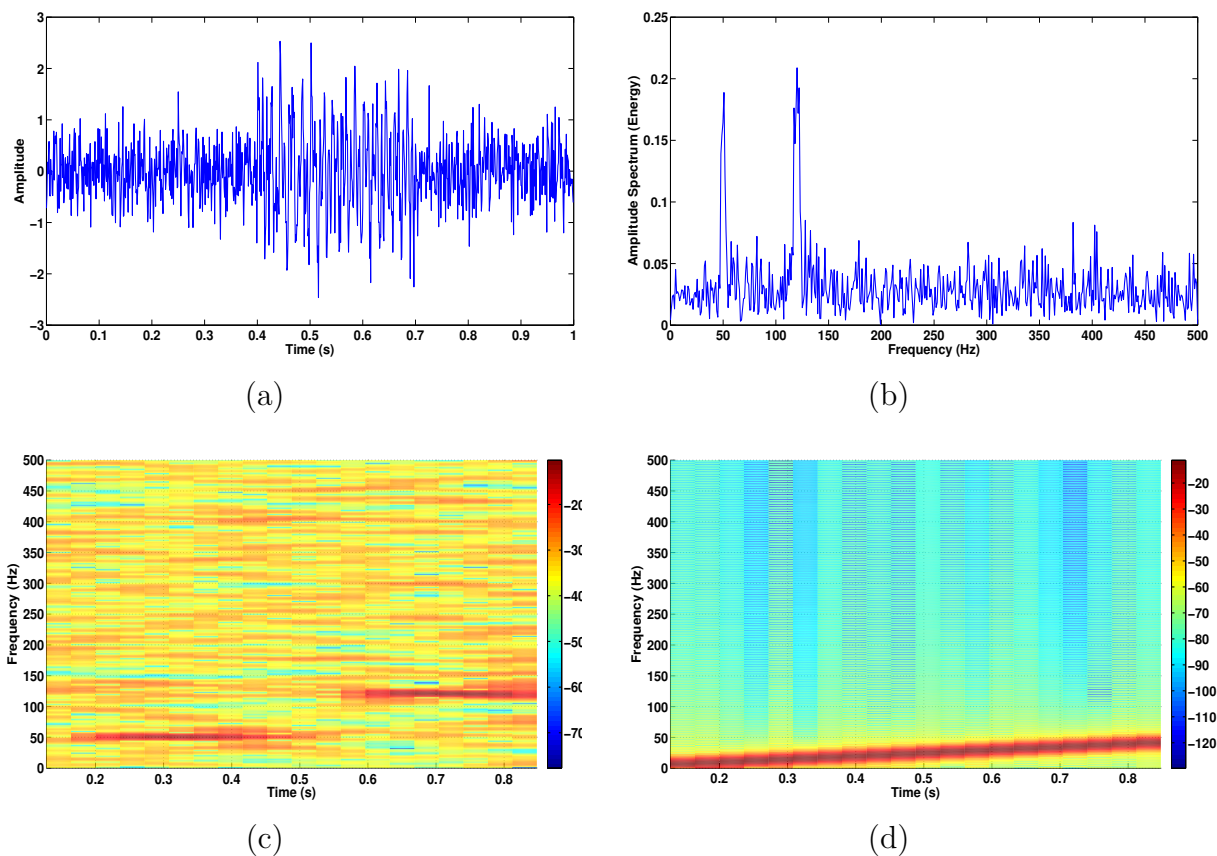


Figure 1.2: (a) Signal containing two sinusoids corrupted with zero-mean random noise. The two sinusoids, having frequencies of 50 and 120 Hz, occur between 0.2 and 0.4 seconds, and between 0.6 and 0.8 seconds, respectively; (b) Representation of the signal in the frequency domain, using the Fourier Transform. The frequency domain representation indicates the presence of the 50 and 120 Hz components; (c) Representation of the signal in the time-frequency domain, using the Short-Time Fourier Transform (with Kaiser window having length of 256 samples and overlap of 220 samples). The time-frequency representation not only indicates the presence of the 50 and 120 Hz components, but also their temporal locations in the signal, between 0.2 and 0.4 seconds, and 0.6 and 0.8 seconds respectively; (d) Representation of the linear chirp signal in Figure 1.1 in the time-frequency domain, showing the linear increase in frequency with time.

harmonic components globally, so many additional harmonic components are needed for signals which are globally time-varying, or non-stationary in nature, leading to energy being spuriously spread over a wide frequency range. Also, since the Fourier transform is based on a linear superposition of trigonometric functions, modeling non-linear effects in signals,



such as deformed wave profiles, requires additional harmonic components. Thus the Fourier spectrum of non-stationary and non-linear signals will have spurious harmonic components that cause spreading of energy over the spectrum, resulting in a misleading, and physically non-meaningful energy-frequency distribution. Therefore the used signal analysis methods should be able to take the signal non-stationarity and non-linearity into account for accurate and meaningful signal analysis.

### 1.3.2 Signal Analysis Methods

A number of methods have been devised for signal analysis, in both time and frequency domains. Many of these methods have been designed to incorporate signal non-stationarity aspects as well. Most of these methods are applicable to signals from linear systems only, whereas only a few methods are suitable for analysis of signals from non-linear systems. At the same time, it is important to realize that analysis of non-linear signals is more successful when performed with methods suitable for non-linear signal analysis [15].

There are linear and non-linear signal processing techniques reported in literature which extract parameters from signals being analyzed in order to characterize the signals. Some examples of these techniques, as applied to Electromyogram (EMG) signals, which are generated by non-linear physiological processes, are provided in [16]. The linear techniques mentioned in [16] include the root-mean square value of the signals, the peak and median frequencies of the power spectrum of the signals obtained using the Fourier transform, and the autocorrelation zero-crossings. The non-linear techniques discussed include estimates of the signals' maximal Lyapunov exponent, correlation dimension and the sample entropy. Not surprisingly, it was reported in [16] that the non-linear techniques performed better than the linear techniques in characterizing the non-linear signals.

Another type of signal analysis methods consists of parametric techniques which involve fitting the signal under analysis to a particular model. These techniques are available in the time domain, and the model, once found, can be mapped to the frequency domain. Some linear methods for non-stationary signals include time-varying auto-regressive (AR),

auto-regressive moving average (ARMA) and auto-regressive with exogenous input (ARX). However, these methods cannot reproduce non-linear effect such as harmonics and inter-modulations in the frequency domain [9]. For this purpose, the work in [9] models non-stationary and non-linear dynamic systems using polynomial time-varying non-linear auto-regressive with exogenous input (TV-NARX) models, and the identified models are then transformed to the frequency domain using the concept of time-varying generalized frequency response functions (TV-GFRFs). Although methods such as these provide good results, these require fitting the signals being analyzed to a model, which may not be always possible, or may require tuning the parameters of the parametric models or addition of parameters to the models. Additionally, as mentioned in [9], obtaining the time-frequency representation can require assuming the input signal as linear, so as to allow the use of linear parametric approaches.

### 1.3.2.1 Signal decomposition methods

The approaches presented in the previous two paragraphs either extract parameters from signals being analyzed, or require estimation, and possible adjustment, of models in order to match the signals to the analysis methodology. However, real world non-stationary and non-linear signals have important information hidden at intrinsic time and frequency scales, and the signal analysis methods should provide a meaningful way of reproducing this information in the time and frequency domains. For this purpose, methods which involve signal decomposition are particularly relevant. A number of methods which allow analysis through signal decomposition, in either the time or frequency domains, or the time-frequency domain, have been proposed. Here we review some of the most widely used methods, and then mention conditions which should be fulfilled by signal decomposition methods for non-stationary and non-linear analysis.

**Short-Time Fourier Transform:** A very basic, but widely used method, based on the Fourier transform, is called the Short-Time Fourier Transform (STFT). The STFT basically introduces time dependency in the Fourier transform, by means of calculating the Fourier

transform of short-time windows over the signal, and obtaining a time-dependent spectrum, or time-frequency distribution, by sliding the window along the time axis. However, the STFT assumes signal stationarity over the window length, which is often difficult to justify in real signals, as alignment of window length with stationary time-scales cannot be guaranteed. Also, for proper localization of an event in time, a narrow window is needed, whereas proper frequency resolution requires a longer length window, leading to conflicting requirements which are not easy to resolve.

**Wavelet Transform:** A more flexible technique is available in the form of the Wavelet transform, which can also be seen as Fourier spectral analysis but with an adjustable window. Due to its precise mathematical formulation, the wavelet technique, defined in continuous and discrete representations, is widely used for signal analysis in a broad range of domains. For wavelets, the basis functions are called mother wavelets, which can be stretched and compressed (dilated and scaled), and the dilated and scaled versions are slid in time across the signals to be analyzed. This has the effect of correlating the signal under analysis with the dilated and scaled versions of the mother wavelet, and results in the amplitude spectrum of the signal in time and frequency domains, representing the energy of the signal at a particular frequency and time instant. A drawback of the wavelet analysis relates to the same uncertainty in simultaneous time and frequency resolution as the Fourier analysis. For example, the effect of an event local in time will be reflected in the high frequency range. However, this will lead to issues if the local effect occurs in the low frequency range. Another important drawback of the wavelet approach is its non-adaptive nature. This is because the mother wavelet, once chosen, is used for all subsequent analysis, and therefore the choice of the mother wavelet has significant effect on the overall signal analysis. The fixed nature of the mother wavelet may not be able to track the time-varying nature of the signal at every instant of time, or the intra-wave frequency modulations occurring intermittently. Since the wavelet analysis is a linear technique, the results of the analysis for non-linear signals might not be always meaningful.

**Principal Component Analysis:** Another useful signal decomposition technique is Principal Component Analysis (PCA), also known as Karhunen-Loeve Transform or Empirical Orthogonal Expansion Function. In PCA, no assumptions about the shape of the basis function are made, nor is the expansion basis fixed, but is instead derived from the signal under analysis. Therefore PCA is adaptive, as it depends on the data to derive its modes, or decomposed components. However, the decomposition of the signal into modes, or principal components, is based on statistical properties of the signal, as the method extracts the modes that account for most of the variance in the signal being analyzed, and there is no suggestion of the frequency content of the signal in the modes. Further, as pointed out in [6], the PCA decomposition is not unique, and a single component from a non-unique decomposition, even for an orthogonal basis, is not physically meaningful.

**Empirical Mode Decomposition:** A relatively recent technique for signal analysis through decomposition of the signal is called Empirical Mode Decomposition (EMD) [6]. The EMD technique has gained much popularity, and has been widely used for signal analysis in a broad range of domains, such as biological and medical sciences, engineering, geology, astronomy, finance, as well in the visual arts (e.g. [17][18][19][20][21] [22][23][24][25] [26][4][27]). EMD is a fully data adaptive decomposition technique, and is therefore suitable for non-stationary and non-linear signal analysis. EMD considers a signal at the level of the most local oscillation, that between two consecutive extrema, and defines an algorithm to adaptively extract the different oscillatory modes present in the signal. The algorithm defines steps for successively separating the signal variations from the signal mean, though the mean is considered as an instantaneous mean, and not as a global mean. The extracted modes are called Intrinsic Mode Functions (IMFs), and satisfy certain properties. Since the extraction of the modes is determined by the signal characteristics, the decomposed modes are also generally meaningful. In this regard, decomposition of EEG signals into IMFs representing the alpha rhythm (8-13 Hz) and eye movement artifacts ( $\leq 2$  Hz) was demonstrated in [10] as an example of the physical meaningfulness of the decomposed modes. Importantly, EMD

requires no assumptions and requirements (e.g. harmonics) on the nature of oscillations as well. Decomposition of a signal using EMD can be seen as in Equation 1.4, where a signal  $x(t)$  is decomposed into  $k$  IMFs  $c_k(t)$ , and a residue  $r_K(t)$ , which is a monotonic function from which no further modes can be extracted.

$$x(t) = \sum_k c_k(t) + r_K(t) \quad (1.4)$$

Once a signal has been decomposed using EMD into IMFs, Hilbert transform can be applied to the IMFs to obtain the instantaneous amplitude (energy) and frequency values, which allow the construction of an energy-time-frequency distribution. This energy-time-frequency distribution is now generally termed as the Hilbert-Huang spectrum [10]. Given the adaptive nature of the decomposition, and the ability to obtain instantaneous energy and frequency values, this time-frequency representation preserves the time localities of events. Also, since EMD decomposes a signal into components which are in the time domain and have the same length as the original signal, the varying frequency in time is preserved [28], thereby allowing capture of non-linear effects in signals. In certain cases, the non-stationarity and non-linearity might require processing the signals in the time domain [21], thereby making EMD a natural choice for a signal analysis method. On the other hand, the EMD technique is defined by an algorithm, and does not have a precise mathematical formulation, although its characteristics have been extensively studied and validated through numerical means. Even though the technique is empirical, the knowledge of its properties and its extensive use in a wide range of domains underscores the propriety of using empirical approaches for signal analysis.

Here it is also pertinent to consider the conditions a basis must fulfil to represent non-stationary and non-linear signals, as defined in [6]. These may be identified as:

1. **Completeness:** This condition relates to the precision of the decomposition, as to whether the original signal can be reconstructed back from the decomposed components.

2. **Orthogonality:** This condition is relevant with respect to the positivity of energy, and to avoid energy leakage. However, the orthogonality condition for non-linear signals generally holds only locally. Depending on the non-linear time and frequency variations, it may be that neighboring components have same spectral structures at different temporal locations, but no components will have the same spectral structures at the same temporal instant.
3. **Locality:** The locality requirement is important for non-stationary signals, as for such signals there is no fixed time scale, and the events of interest have to be identified at the temporal locations of their occurrence.
4. **Adaptivity:** The adaptivity condition is relevant for both non-stationary and non-linear signals, and especially non-linear signals, given that adapting to the local variations of the signals being analyzed is important for a meaningful decomposition, and also for a spectral representation that is reflective of the signals' oscillatory structures. A fixed basis cannot be expected to fit all temporal and spectral variations in non-stationary and non-linear signals.

From the methods discussed here, the Wavelet and EMD methods are good candidates for fulfilling these conditions. In fact, there are many similarities between the wavelet and EMD techniques as well [29][30]. For example, the EMD and wavelet (Discrete Wavelet Transform (DWT)) decompositions may be seen as follows:

$$\begin{aligned}
 x(t) &= \sum_k c_k(t) + r_K(t) \text{ (EMD)} \\
 x(t) &= \sum_k d_k(t) + a_K(t) \text{ (DWT)}
 \end{aligned} \tag{1.5}$$

where  $d_k(t)$  and  $a_K(t)$  are the details and approximation components obtained after wavelet decomposition of signal  $x(t)$ . As pointed out in [30], both EMD and wavelets lead to decomposition into components which may be seen as a fluctuation and a trend. The major difference is that the scales are pre-determined for DWT and adaptive or data-driven for

EMD. Thus the main compromise involved when using wavelets is due to the lack of adaptivity, resulting from the fixed and *a priori* nature of the mother wavelet. In addition, as explained in [10], integral transforms (e.g. Fourier, wavelet) involve a trade-off between temporal and spectral resolution due to the uncertainty principle, and the resolution of the Fourier and wavelet methods also depends on the data length. EMD, on the other hand, is not compromised by these issues [10].

### 1.3.3 Feature Analysis

An important part of automated signal analysis is feature extraction and classification. The process of feature extraction refers to obtaining representative parameters from signals being analyzed using a signal analysis method. The number of features is generally much less than the dimensions of the signal being analyzed, and the dimension reduction results in a significant reduction in the computational requirements of the signal analysis task. If the features extracted from a signal are meaningful and representative of the signal, they can be used to effectively describe a signal, but with much reduced dimensions. This is also relevant in terms of characterizing a signal in terms of its membership of a particular class, which can then be used for automatic classification of signals, which is an important aspect of automated decision making algorithms which depend on accurate and effective signal analysis. It is therefore possible to identify the following properties for the extracted features:

- **Representative and Meaningful:** The extracted features should be representative of the signals from which these have been extracted. This also allows the features to reflect the common characteristics existing in signals belonging to the same class. Also, the features should be meaningful, as opposed to abstract, which allows a physical meaning to be assigned to the features based on the signal characteristics. For example, in case of biomedical signals, meaningful features are relevant for understanding the behaviour of the underlying physiological processes that generate the signals.

- **Discriminative:** In addition to being representative of the peculiar patterns occurring in individual signals, features should also be able to characterize signal properties common to a class of signals.
- **Localized:** The locality of features refers to their ability to model representative and discriminatory patterns in signals in both time and frequency domains, and is particularly relevant for non-stationary and non-linear signals. Indeed, extraction of such features also requires a signal analysis technique which is local and adaptive, as previously mentioned.

Also connected to feature analysis is the process of classification, which refers to a scheme that assigns signals to different classes. Classification schemes have two stages, the training stage and the testing stage. In the training stage, a set of training signals with labeled class information (in case of supervised learning) are used to train the classification algorithm. In the testing stage, the trained classification algorithm is then used to classify any new signal with unknown class information (a test signal) as belonging to a particular class of signals. It is also important to realize that the quality of the extracted features will have a significant effect on classification accuracy, just as the choice of a proper signal processing technique will positively influence the extracted features according to the desirable properties presented previously. Also important in this respect is the fact that if a signal processing technique allows the extraction of features that are representative, meaningful, discriminative and localized, the classification of signals may be possible in the feature space, without having to use a classification scheme. This scenario would thus represent a highly desirable outcome of signal analysis and the consequent feature extraction.

## 1.4 Contributions

In the previous Sections it has been shown that the analysis of non-stationary and non-linear signals requires signal processing techniques which can deal with signal non-stationarity and non-linearity. From among the presented techniques, Empirical Mode Decomposition (EMD)



stands out due to the locality and adaptivity of the decomposition. At the same time, EMD is defined by an algorithm, not by a mathematical formulation, though its decomposition properties are well-understood through numerical, but empirical methods. Also, while the EMD technique decomposes signals according to the local time scales at each step of the algorithm, the decomposition itself cannot be controlled, which is to say that it is not possible to selectively decompose certain time scales. Addressing this drawback can lead to a time-scale based decomposition which retains all the advantages of EMD, but adds much more flexibility to the decomposition process.

The goal of this dissertation is to develop techniques and methodologies aimed at empirical analysis for non-stationary signal de-noising, de-trending and discrimination applications using biomedical signals such as pathological speech and EEG signals as real world examples.

1. A novel modification of the EMD algorithm was proposed, which is named Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS). Through this modification, a time-scale based decomposition is possible. The properties of EMD-MPS were also established empirically, thereby validating the expected behaviour of the decomposition method. Importantly, it was shown how decomposition behaviour of EMD could be formulated in terms of EMD-MPS, thereby providing novel insight into EMD.
2. Using EMD-MPS, a novel hierarchical decomposition methodology was proposed, and illustrated with synthetic and real world signals. The hierarchical decomposition uses the time-scale based decomposition of EMD-MPS, and can decompose a signal into specified frequency bands.
3. De-noising and de-trending of signals is an important signal processing task. EMD-MPS was used for a novel time-scale based de-noising and de-trending of signals, demonstrated using synthetic and real world biomedical signals, and also for practical applications such as mental task classification using Electroencephalogram (EEG) signals and seizure detection and epilepsy diagnosis.
4. A novel empirical sparse dictionary learning framework was proposed. As part of the

framework, a discriminative dictionary learning algorithm was developed, and characteristics of the empirical dictionary established. The use of the empirical sparse dictionary learning framework for signal classification was demonstrated using EEG signals. Furthermore, the framework was applied for automated seizure detection using long-term EEG recordings, and the implications of this application for creating patient-specific dictionaries were described.

5. EMD was used for analysis and classification of pathological speech signals, and meaningful instantaneous features in the time and frequency domains were extracted from a large subset of intrinsic mode functions obtained after decomposition of the speech signals. The use of true instantaneous features obtained from the intrinsic mode functions distinguishes this work from previous similar studies.

## 1.5 Structure of Dissertation

This dissertation is organized in six chapters. The overall structure of the dissertation, and the flow between the chapters, is shown diagrammatically in Figure 1.3. The brief description of each Chapter is as follows:

**Chapter 2: Signal analysis using Empirical Mode Decomposition (EMD).** Chapter 2 describes the EMD method, as well as its important properties. De-noising and de-trending in the context of EMD is also discussed in this Chapter. The use of EMD for signal and feature analysis is described in terms of real world examples using pathological speech classification and mental task classification using EEG signals.

**Chapter 3: Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS).** In Chapter 3, the novel modification of the EMD algorithm, named EMD-MPS, is described, along with discussion of the method's decomposition properties. Also included in this Chapter is the description of a hierarchical decomposition method based on EMD-MPS.

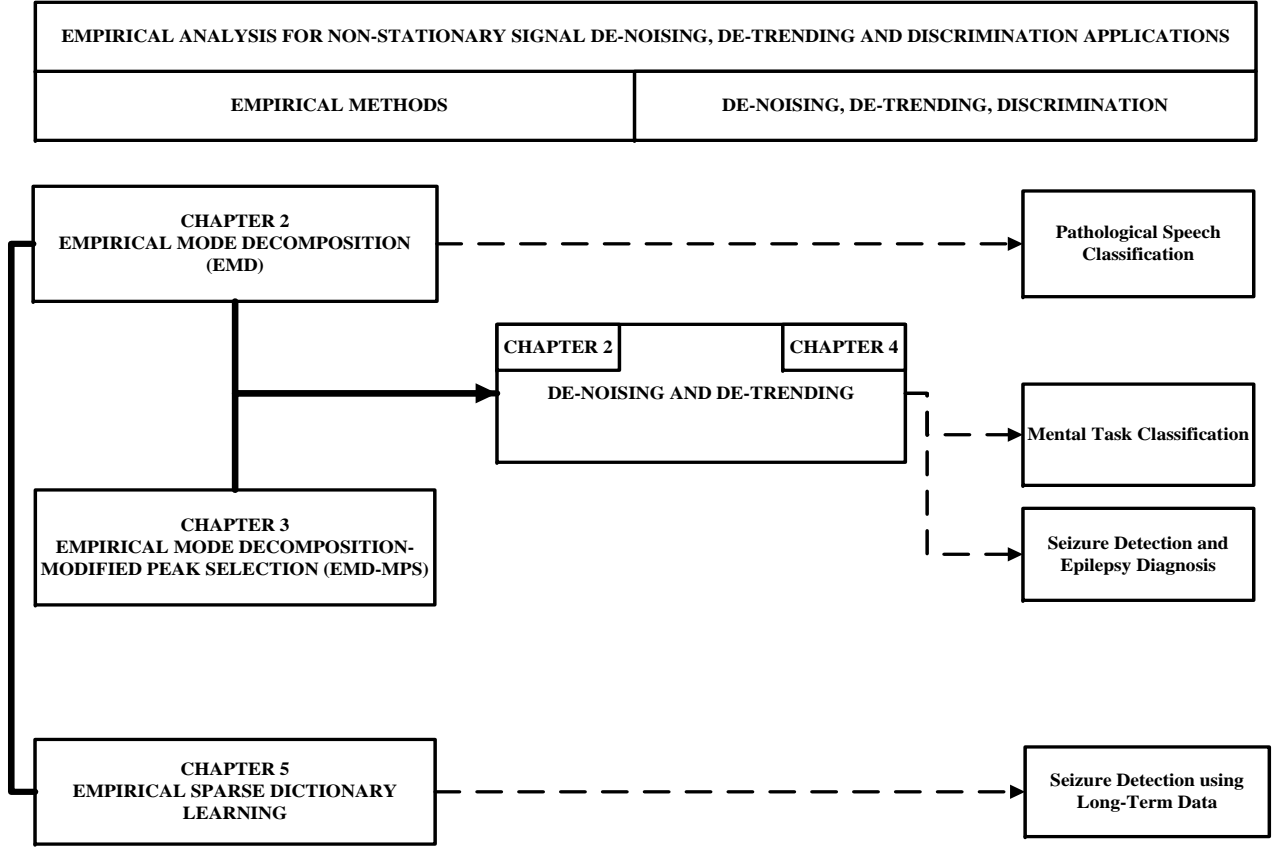


Figure 1.3: Block diagram of the dissertation. The connections between the Chapters are shown in solid lines. The main practical contributions of each Chapter are shown with dashed lines.

**Chapter 4: De-noising and de-trending using EMD-MPS.** The focus of Chapter 4 is de-noising and de-trending using the novel time-scale based decomposition using EMD-MPS, which is demonstrated using synthetic and real world signals. In addition, the Chapter includes discussion of the practical application of time-scale based de-trending in terms of feature extraction for mental task classification and seizure detection and diagnosis.

**Chapter 5: Empirical sparse dictionary learning.** Chapter 5 presents the novel empirical sparse dictionary learning framework, which is based on signal decomposition using EMD. The discriminative dictionary learning algorithm developed as part of the framework

is described, and the properties of the learned sparse dictionary are discussed in the Chapter as well. The use of the framework for signal classification is described, and the Chapter also includes a complete section on using the framework for automatic seizure detection with long-term EEG records, and how the application of the framework may evolve towards patient-specific dictionaries.

**Chapter 6: Conclusions and future work.** The dissertation is concluded in Chapter 6, which also includes a discussion of the directions for further research resulting from this dissertation.

# Chapter 2

## Empirical Mode Decomposition: An Empirical Approach for Signal Analysis

### 2.1 Introduction

Most signals encountered in real life are non-stationary and non-linear in nature, and the analysis of such signals requires signal processing techniques that can take the signal non-stationarity and non-linearity into account. Depending on the type of signals being analyzed, it is important to use appropriate tools for analysis. Indeed, it has been shown in [15] that for linear data, linear signal processing techniques outperform methods meant for non-linear signal analysis. Therefore appropriate choice of the signal processing technique has important implications for the robustness of the signal analysis methodology. A number of widely-used signal processing techniques were discussed in Chapter 1, and the efficacy of the Empirical Mode Decomposition (EMD) technique for analysis of non-stationary and non-linear signals was pointed out. Empirical approaches, however, lack the mathematical formulation which exists in signal analysis methods having a pre-defined basis. In general, data analysis techniques have a mathematical formulation for the pre-defined basis, as well as for the associated properties of convergence, completeness, uniqueness and orthogonality

[31]. On the other hand, a fixed basis might fit the data at one temporal location, but not at another in a non-stationary signal. Similar problems will result in a signal with non-linear waveform characteristics, where using a fixed signal basis can lead to spurious harmonics.

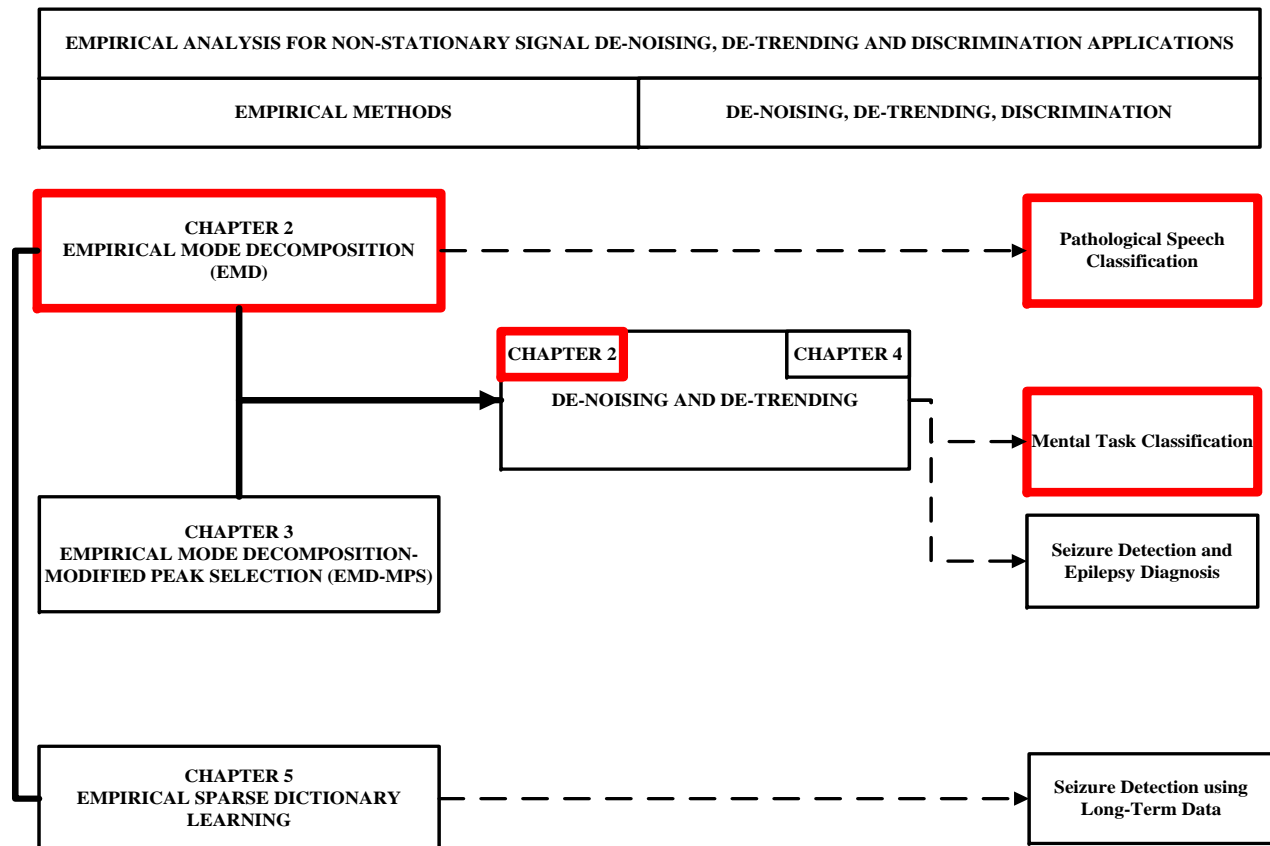


Figure 2.1: Chapter 2 contains details about the Empirical Mode Decomposition (EMD) method, as well as practical applications based on de-noising/de-trending and feature analysis using EMD.

In contrast, the EMD technique uses an adaptive basis which is derived from the data, thereby allowing a robust analysis of non-stationary and non-linear signals. However, there is no analytic definition of the adaptive basis and the associated properties. At the same time, it is important to keep in view the fact that the characteristics and properties of EMD are well-understood through empirical and numerical means. The shortcomings of the algorithm, and their causes, are also known, thereby allowing researchers to make an informed decision

about application of EMD to a particular signal processing scenario. Furthermore, there have been a number of attempts to formalize the EMD algorithm (e.g. [32][33], along with many others), but in general these require certain constraints to be met, or parameters to be set, which again would depend on the signals to be analyzed, and thus still require an empirical formulation for an optimal decomposition. None of these analytical formulations of EMD have found much acceptance, and the original EMD algorithm is still widely used in a broad range of applications.

As was previously mentioned in Chapter 1, EMD is particularly suited for the analysis of non-stationary and non-linear signals due to the data-adaptive nature of the method. This dissertation takes the view that the benefits of using an empirical method which uses an adaptive basis derived from the data and hence can fit the non-stationary and non-linear manifestations that may occur in the signals, while also providing a high time and frequency resolution as well as means for de-noising and de-trending, which are important signal processing operations, far outweigh any implications of a lack of mathematical formulation of the method. Also, since the properties of EMD are well-understood, as are the issues arising under certain conditions, it is possible to confidently apply the method for robust analysis of non-stationary and non-linear signals.

In this Chapter, the EMD technique will first be described, followed by a discussion of some important aspects of EMD. Discussion of de-noising and de-trending in the context of EMD will form an important part of this Chapter, which includes an experiment describing a framework for mental task classification based on the de-noising and de-trending capabilities of EMD. This will be followed by discussion of feature analysis in the context of EMD, which also includes an experiment for classification of pathological speech signals using novel temporal and spectral features obtained using an EMD-based methodology. The robustness of the EMD-based methodology is then further demonstrated by classification of telephone-quality pathological speech signals with much higher accuracy than previous results mentioned in literature.

## 2.2 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) considers signals at the level of their local oscillations, and hence allows signals to be seen as faster oscillations superimposed on slower oscillations; the EMD algorithm allows iterative separation of the faster oscillations, and the algorithm terminates when no more oscillations remain to be separated. Looking at a signal at the local level means that we may consider a signal  $x(t)$  between two local extrema, for example two minima at times  $t_-$  and  $t_+$ , as discussed in [34]. Within this interval, we can define a high frequency part  $d(t)$  which represents the oscillation terminating at the two minima and passing through the maxima that necessarily exist between them [34]. Similarly, we can also identify a corresponding low frequency part  $m(t)$ , such that the signal at the local level may be defined as:

$$x(t) = d(t) + m(t) \text{ for } t_- \leq t \leq t_+ \quad (2.1)$$

The EMD algorithm defines the steps to separate  $d(t)$ , and then iterate on the the residual  $m(t)$ . This may be explained step-wise as follows [34]. Consider the original signal  $x(t)$  as:

$$x(t) = d_1(t) + m_1(t) \quad (2.2)$$

After extraction of  $d_1(t)$ , the first residual is decomposed as:

$$m_1(t) = d_2(t) + m_2(t) \quad (2.3)$$

and the complete steps of the algorithm can be written as:

$$\begin{aligned} x(t) &= d_1(t) + m_1(t) \\ &= d_1(t) + d_2(t) + m_2(t) \\ &\dots \\ &= \sum_{k=1}^K d_k(t) + m_K(t) \end{aligned} \quad (2.4)$$



where  $m_K(t)$  represents the final residue containing no oscillations. It can be seen from Equation 2.4 that the algorithm decomposes the signal  $x(t)$  into  $K$  oscillatory modes and a residue. The number  $K$  is not known a priori, but is determined by the algorithm [7], and depends on the characteristics of the signal.

It is also important to point out that the high and low frequency qualification given to the oscillations applies at a local level only, hence is different from pre-determined sub-band filtering. In fact, the selection of the oscillatory modes corresponds to an automatic and adaptive time-variant filtering [6][34]. The extracted oscillatory modes are called *intrinsic mode functions*, which satisfy two simple conditions described in the next Section.

## 2.3 Empirical Mode Decomposition Algorithm

As described in the previous Section, EMD is an adaptive technique that allows decomposition of signals into intrinsic mode functions. An intrinsic mode function (IMF) satisfies the following two conditions [6]:

1. The number of maxima, which are strictly positive, and the number of minima, which are strictly negative, for each IMF, are either equal, or differ at most by one.
2. The mean value of the envelope, as defined by the maxima and the minima, for each IMF, is zero, or very close to zero.

In practice, the second condition above is usually considered in an approximate sense, such that the mean value of the envelope  $|(E_U + E_L)/2| < \varepsilon$ , where  $\varepsilon > 0$ , and  $\|\cdot\|$  denotes some norm, such as the absolute maximum  $\|\cdot\|_\infty$  or the mean square value  $\|\cdot\|_2$  [7]. Also, as pointed out in [35], the first condition is redundant unless the function is discontinuous, and if a function satisfies the second condition, it also satisfies the first, since there must be a zero between a peak and a valley, given that the envelopes are symmetric.

The technique for decomposition of the signals into IMFs is known as sifting, a brief description of which follows. It needs to be mentioned here that in the subsequent sections,

the notation used conforms to discrete-time signal processing. Furthermore, although most of the time the notation used for EMD is for the continuous-time domain, there is precedent of using discrete-time notation in the context of EMD [36], which we have adopted for the subsequent sections in this Chapter, and also for the EMD algorithm shown in Algorithm 1.

1. For a given discrete time signal  $x[n]$ , all the local minima and maxima of  $x[n]$  are identified.
2. The upper envelope  $E_U$  is calculated by using a cubic spline to connect all the local maxima. Similarly, the lower envelope  $E_L$  is calculated from the local minima. The upper and lower envelopes should cover all the data in  $x[n]$  between them.
3. The mean  $E_{mean} = (E_U + E_L)/2$  of the upper and lower envelopes is calculated, and  $x[n]$  is updated to  $c'[n]$  by subtracting the mean from it  $x[n] \leftarrow x[n] - E_{mean}$ .
4. The previous three steps are executed till  $c'[n]$  is reduced to an IMF  $c_1[n]$ , which conforms to the properties of IMFs described previously. The first IMF contains the highest oscillation frequencies found in the original signal  $x[n]$ .
5. The first IMF  $c_1[n]$  is subtracted from  $x[n]$  to get the residue  $r_1[n]$ .
6. The residue  $r_1[n]$  is now taken as the starting point instead of  $x[n]$ , and the previously mentioned steps are repeated to find all the IMFs  $c_i[n]$  so that the final residue  $r_K$  either becomes a constant, a monotonic function, or a function with a single maximum and minimum from which no further IMF can be extracted.

Therefore, at the end of the decomposition, we can represent  $x[n]$  as the sum of  $K$  IMFs and a residue  $r_K$ :

$$x[n] = \sum_{i=1}^K c_i[n] + r_K[n] \quad (2.5)$$

The EMD algorithm, as explained in this Section, is also shown in Algorithm 1. Also, a demonstration of EMD to decompose a signal consisting of two quadratic chirps is shown in Figure 2.2.

---

**Algorithm 1** Empirical Mode Decomposition Algorithm

---

- 1: Find the location of all extrema (minima and maxima) of  $x[n]$
  - 2: Interpolate between all the maxima to obtain the upper signal envelope  $E_U$ , and the minima to obtain the lower signal envelope  $E_L$
  - 3: Compute the local mean  $m[n] = (E_U + E_L)/2$
  - 4: Obtain the oscillatory mode  $c'[n]$  by subtracting the local mean from the signal,  $c'[n] = x[n] - m[n]$
  - 5: If  $c'[n]$  fulfils the conditions of an intrinsic mode function (IMF), designate it as an IMF  $c[n]$ ; otherwise let  $x[n] = c'[n]$  and iterate the algorithm from Step 1.
- 

An interpretation of the sifting process as an amplitude modulation/frequency modulation (AM/FM) decomposition is given in [35]. Seen this way, the sifting process is a means of removing the non-symmetry between the upper and lower envelopes, such that the original signal is transformed into an AM signal. This results in considering the obtained IMF as an AM signal. Further, since the instantaneous frequency of the IMF may change from instant to instant, the IMF is an AM/FM signal, with application of EMD to a signal resulting in a set of AM/FM signals. Still keeping in view the interpretation of EMD as an AM/FM decomposition, the envelope can be seen as slowly varying as compared to the signal  $x[n]$ . This means that the bandwidth of the envelope is a fraction of the central frequency of  $x[n]$ . Therefore the sifting process can be seen as the removal of the lower frequency components, while leaving the higher frequency components for the next iterations of the algorithm ultimately resulting in an IMF. This can also explain the appearance of IMFs in a higher to lower frequency order. In the same context, EMD may also be seen as a time-frequency decomposition.

Further discussion of some important EMD algorithm implementation issues and characteristics of EMD are provided in Appendix A.

### 2.3.1 Instantaneous Measurements

The IMFs obtained through decomposition of a signal using EMD lend themselves well to calculation of the instantaneous amplitude and instantaneous frequency through application of the Hilbert transform on each IMF [6]. First, the analytic signal  $z_i$  corresponding to each

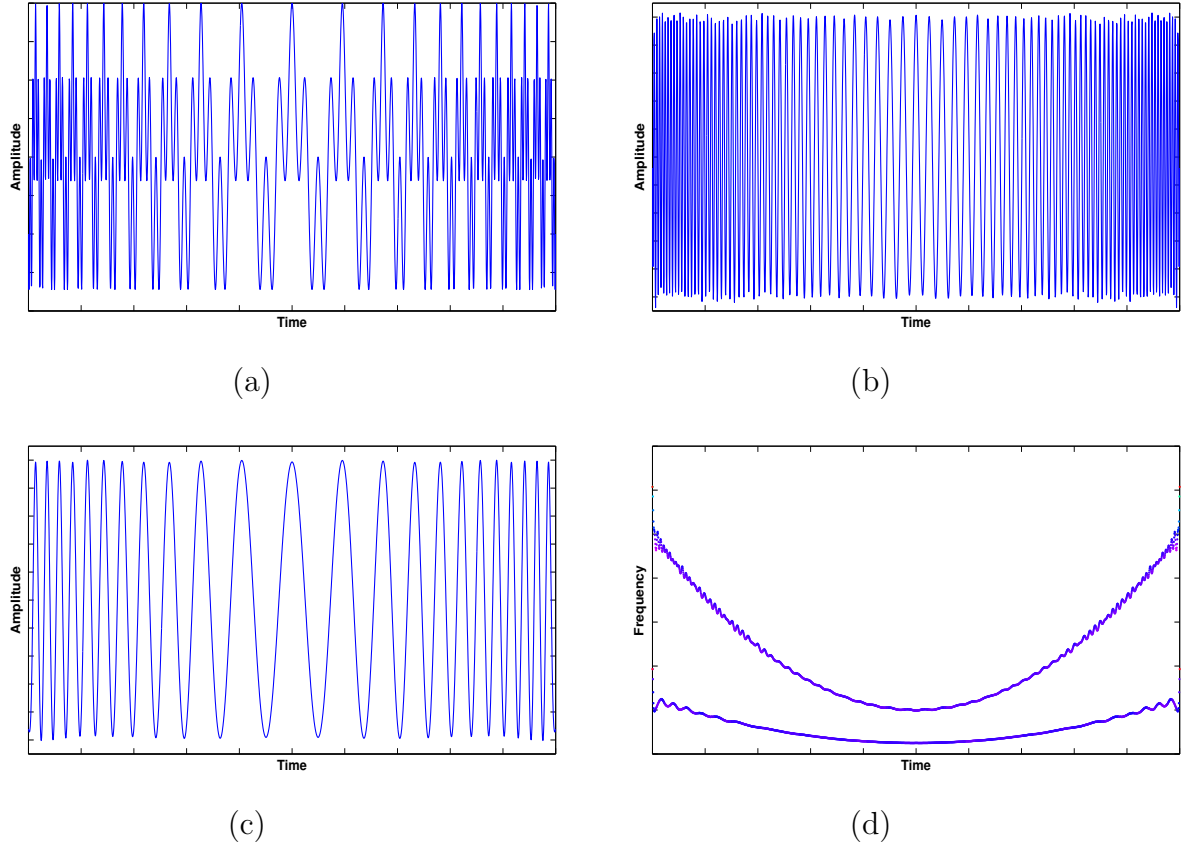


Figure 2.2: Demonstration of EMD: (a) Signal containing two quadratic chirps. This signal is decomposed using EMD. (b) The first IMF, containing the chirp with the higher instantaneous frequencies. (c) The second IMF, containing the chirp with the lower instantaneous frequencies. (d) The time-frequency plot represented by the Hilbert spectrum,  $H[\omega, n]$ , obtained using the steps defined in Section 2.3.1.

IMF is obtained, as shown below:

$$z_i[n] = c_i[n] + jH(c_i[n]), \quad (2.6)$$

where  $H(c_i[n])$  is the Hilbert transform of an IMF  $c_i[n]$ . Writing Equation 2.6 in the polar form we have:

$$z_i[n] = a_i[n]e^{j\theta_i[n]}, \quad (2.7)$$

where  $a_i$  represents the instantaneous amplitude, and  $\theta_i$  the instantaneous phase, corresponding to IMF  $c_i$ . The instantaneous amplitude and instantaneous phase are given by:

$$a_i[n] = \sqrt{c_i^2[n] + H^2(c_i[n])} \quad (2.8)$$

$$\theta_i[n] = \tan^{-1} \frac{H(c_i[n])}{c_i[n]} \quad (2.9)$$

The instantaneous frequency,  $\omega_i$ , can then be obtained as :

$$\omega_i[n] = \frac{d\theta_i[n]}{dn} \quad (2.10)$$

It is important to realize here that these instantaneous values represent the amplitude and frequency varying trigonometric function, as given in Equation 2.7, that best fits the original signal at the local level, with each IMF representing the signal at the local level corresponding to a particular time-scale. The instantaneous frequency represents the time-varying value of a frequency component occurring at a time instant corresponding to a local time-scale extracted in an IMF, whereas the instantaneous amplitude represents the energy value of the frequency component at that time instant. In this way, the instantaneous amplitude acts as a weighting coefficient for the instantaneous frequency values.

Once the Hilbert transform has been applied to all IMFs, and the instantaneous amplitude and frequency values calculated, the original signal  $x[n]$  can be expressed as follows:

$$x[n] = \text{Re} \sum_{i=1}^K a_i[n] \exp(j \sum \omega_i[n] dn) \quad (2.11)$$

Equation 2.11 is a representation of the instantaneous amplitude (hence instantaneous energy) and frequency contribution by the locally occurring frequency components in each IMF. The values of the amplitude  $a_i[n]$  and frequency  $\omega_i[n]$  plotted against time result in a time-frequency representation  $H[\omega, n]$  of the signal  $x[n]$ , which is now called the Hilbert-Huang spectrum [10]. The use of these steps to obtain a time-frequency representation of a

signal after decomposition using EMD is demonstrated in Figure 2.2.

### 2.3.2 Marginal Spectrum

From the Hilbert-Huang spectrum  $H[\omega, n]$ , it is also possible to calculate the marginal spectrum,  $h(\omega)$ , which measures the total amplitude (and hence energy) contribution from each frequency value over the whole signal length. The marginal spectrum is defined as:

$$h(\omega) = \sum_{n=1}^N H[\omega, n] \quad (2.12)$$

## 2.4 De-noising and De-trending based on Empirical Mode Decomposition

Real world time-series data is often non-stationary and noisy, and effective extraction of the desired information requires pre-processing of data to reduce noise. De-noising refers to removing the noise from the signal of interest, and effective de-noising requires noise removal without affecting the desired signal. Therefore development of effective methods of signal de-noising form an important aspect of signal processing research. Since signals of interest may have a broad-band spectrum that overlaps with the spectrum of noise, linear filtering methods are not considered effective for non-stationary and non-linear time-series, as filtering methods only work for the cases where noise has time or frequency scales different from those of the signal of interest [37]. Application of linear filtering methods to non-stationary and non-linear processes will effectively distort the signal of interest [38]. Therefore, for real-world signals, data-adaptive de-noising methods have proven more suitable. From such de-noising methods, wavelet-based de-noising schemes, also termed wavelet shrinkage or wavelet thresholding [39], are more popular.

Wavelet shrinkage de-noising methods work on the following principle : first, the wavelet transform of the data is taken; second, thresholding of wavelet coefficients is performed; third, inverse wavelet transform is taken to obtain an estimate of the signal of interest.

Wavelet shrinkage de-noising methods do not make any assumptions about the nature of the signals, and permit discontinuities and spatial variation as well [38]. However, these methods require a choice of the wavelet and an appropriate number of resolution levels, as well as choice of an appropriate thresholding algorithm. Adaptive filtering methods such as the one proposed in [38] have shown to perform better at de-noising non-linear time-series than wavelet shrinkage de-noising. However, the adaptive method in [38] requires a time-scale that is  $n + 1$  sample points, as it uses time-series segmentation, and there are also two free parameters in the method, which have to be determined. Therefore the de-noising does not work at the very local time-scale, that of two consecutive extrema, and requirement for the selection of free parameters compromises full adaptivity to the time-series being de-noised.

In the same context, de-trending refers to extracting the trend of a time-series, where trend may be seen as a function with monotone variations [40][41]. Trend extraction is also an important aspect of signal processing, and has applications in diverse domains, such as finance [42] and climate data studies [40], among others. Trend itself can have important meaning, e.g. in finance and climate data studies, but it might also be required to remove the trend in order for meaningful analysis of the de-trended data, for example in case of spectral analysis, where the results of the analysis might become skewed by non-zero mean and effect of other trend terms if the data being analyzed is not de-trended [40]. Simple methods of trend extraction, e.g. de-trending data by subtraction of linear fits to the data, are only suitable for linear and stationary data. Trend extraction methods based on moving averages require a pre-determined time-scale for the averaging operation, making them unsuitable for non-stationary data. More involved trend extraction methods, such as regression analysis or Fourier-based filtering are often based on stationarity and linearity assumptions [40].

Similarly, for non-linear regression analysis, it is not always possible to justify application of non-linear regression formula globally to a non-stationary time-series. Also, trend estimation via regression works if a model can be applied to the data under analysis, whereas a model might not be available for real-world data, or might not be appropriate for such data. For non-stationary and non-linear time-series, the trend should be seen as intrinsic

to the data, and not an external imposition on the data which can be removed by the previously mentioned mechanisms. Extraction of an intrinsic trend, which itself is adaptive to the data, will require a data adaptive method for trend extraction. Also, unlike noise, which represents the unwanted part of the data, trend is composed of useful information that conveys a meaning. In a time-series, however, meaning is imparted to the data by its time-scale. For non-stationary and non-linear signals, the time-scale of importance is the most local time-scale. Therefore any method of trend extraction should be able to consider data at its local time-scales. Another importance of associating a time-scale with the trend relates to giving a meaning to the trend in terms of the associated time-scale, for example a six-monthly trend, or an yearly trend in terms of financial or climate data. This requires the trend extraction method to be able to extract the trend adaptively at different time-scales.

De-noising and de-trending can also be seen as closely related together. Trend, as defined in the last paragraph, represents the slower oscillations in data, therefore de-trending can be seen as extraction of the slower oscillations in the data, whereas de-noising can be seen as removal of the faster oscillations. Another way of looking at this is using an example as presented in [43], whereby a white noise process superimposed on an oscillatory component with frequency  $f_0$  is considered. Depending on the observation scale  $T$ , an oscillatory component can be seen as an oscillation if  $T \gg \frac{1}{f_0}$ , or as a trend if  $T \ll \frac{1}{f_0}$ . This is especially relevant for decomposition-based approaches, such as EMD, which decompose a signal into faster and slower oscillatory components.

As was previously described in Section 2.1, EMD considers signals at the levels of their local oscillations, and decomposes signals into IMFs by iteratively extracting the faster oscillations from the local lower frequency trend. Given the iterative signal decomposition, the lower index IMFs contain the faster oscillatory components, whereas the higher index IMFs contain the slower oscillatory components. Also, a perfect reconstruction of the decomposed signal  $x(t)$  is possible, as shown in Equation 2.4, by adding the detail components  $d_k(t)$ ,  $k = 1, \dots, K$  and the residue  $m_K(t)$  [34]. This property of perfect reconstruction also allows partial construction of the signal, whereby IMFs of different, but adjacent, indices may



be combined. This can be relevant, for example, in the case where the higher index IMFs representing slower oscillations are combined to extract the trend, or the lower index IMFs representing faster oscillations are combined to remove noise [34]. The idea of partial reconstruction is the fundamental concept behind de-noising and de-trending approaches based on EMD, and many variations of this idea have been proposed. Some of these approaches are discussed next.

The approaches for de-noising and de-trending based on partial reconstruction of IMFs can be categorized based on classification of IMFs as consisting of noise or trend. This classification is done as follows:

1. Signal-specific methods
2. Statistical approaches
3. Energy and Energy-Ratio approaches

**Signal-specific methods:** An example of a problem-specific method for de-noising is described in [21], where a method for removing wow noise from ground penetrating radar data is described. Wow noise is an electromagnetic interference which is non-linear in nature, hence EMD is a good method for removal of this type of noise. The study in [21] finds that the wow noise is decomposed into the last two IMFs, and a method based on amplitude modulation of these IMFs is described, whereby the time-varying components in these IMFs are scaled to near zero amplitude. The reconstructed signal is then not degraded. The study in [21] also compares the EMD-based method with a method based on residual median filter, and concludes that the EMD method outperforms the method based on the median filter, which is also not data adaptive as it requires setting the size of the window for the median operation. A method for speech signal enhancement is described in [44], where speech signals are first decomposed into IMFs and each IMF is then compared against a filtered reference noise. Signals are reconstructed based on this comparison, with the reconstructed signal containing reduced noise. Another method for musical noise removal from speech is presented

in [45], which considers most of the musical noise as being contained in the first IMF, allowing signal de-noising after partial reconstruction. While partial signal reconstruction in general is done based on analysis of IMFs, the study in [46] determines the instantaneous frequency of IMFs, and then removes IMFs with unwanted frequencies before signal reconstruction. The utility of this method is demonstrated in case of ocular artifacts removal from EEG signals.

**Statistical approaches:** A method for assigning statistical significance to IMFs from noisy data with respect to the information content contained in the IMFs is described in [37]. The study in [37] concludes that IMFs are normally distributed, and the product of energy density of IMFs and their corresponding averaged period is a constant, with the energy density function being chi-squared distributed. Based on this statistical analysis, a method for assigning statistical significance to IMFs with respect to their information content is described. Once the IMFs containing the meaningful and desired part of the signal are identified by this method, partial reconstruction can be performed to de-noise the data or extract the trend. A similar method based on IMF statistics is described in [34], which also allows identification of IMFs as being composed of noise or the desired signal. In the study presented in [47], statistical characteristics of zero-crossing intervals and zero-crossing amplitudes of IMFs are studied, which allow inference of statistical significance of IMFs, so as to ascertain which IMFs contain the signal and which the noise, so as to de-noise the signal using partial reconstruction. A de-noising scheme for speech signal noise reduction, which is similar to wavelet shrinkage de-noising, is presented in [48]. In this approach, IMFs are thresholded using a shrinkage function, and the signal is then reconstructed. The method is compared against different wavelet-shrinkage based methods, and found to perform better. A method for de-noising and characterization of heart sounds is presented in [49], whereby noise is assumed to be concentrated in the first IMF, and the Euclidean distance of the other IMFs with the first IMF is used as a distance metric to select IMFs containing the relevant signal, and to exclude IMFs containing noise.

**Energy and energy-ratio approaches:** The energy approach for partial signal reconstruction was initially proposed in [34], and the main idea of the approach is as follows. The IMFs are considered zero-mean by definition (Section 2.3), and the empirical mean of the partial reconstruction,  $\hat{y}_D(t)$ , is calculated, where  $\hat{y}_D(t)$  is given by the following equation for IMFs  $d_k(t)$ :

$$\hat{y}_D(t) = \sum_{k=1}^D d_k(t) \quad (2.13)$$

The IMF index  $D$  at which the mean of  $\hat{y}_D(t)$  departs significantly from zero is used to identify the IMFs which contain the trend. Partial reconstruction of IMFs with indices  $D$  and above leads to extraction of the trend of the signal. This approach is further refined using fractional Gaussian noise analysis in [43], and is supplemented by another trend estimation approach called the ratio approach. In the ratio approach, the zero-crossing number of  $i$ -th IMF is denoted by  $Z_i$ , and  $R_i = \frac{Z_{i-1}}{Z_i}$  for  $i \geq 2$ . According to the filter-bank property of EMD (Section A.2),  $R_i \approx 2$  holds for generic broadband processes. Therefore, any deviation from  $R_i \approx 2$  helps identify the IMF with index  $i$  containing a trend component. The study in [43] also presents a combination of the energy and ratio approaches as the energy-ratio approach, which allows trend estimation with more confidence. These EMD-based approaches are also compared in [43] to existing trend estimation approaches, namely the  $l_1$ -trend filtering and the Hodrick-Prescott filtering, and found to have similar performance without requiring any assumptions about the trend, or parameter estimation. An improvement to the energy-ratio approach by using noise-assisted EMD, (called Ensemble Empirical Mode Decomposition, or EEMD [50]), instead of EMD is proposed in [51] to reduce the effects of mode-mixing. The study in [51] further extends the EEMD-based energy-ratio approach by incorporating a seasonality checking criterion to cater for seasonal time-series de-trending, to overcome the limitations of the energy-ratio approach when de-trending seasonal time-series.

### 2.4.1 Experiment: Mental Task Classification using EEG Signals

In this Section a practical application taking advantage of the de-trending capability of EMD will be described. A novelty of this application is the use of a non-linear energy operator, namely the Teager-Kaiser energy operator, for feature extraction, as will be described in the next sections.

#### 2.4.1.1 Background

Identification of patterns in electroencephalography (EEG) signals corresponding to various types of mental tasks performed by human subjects can have important applications in the context of a brain-computer interface to take various actions, such as computer cursor movement or selection of a letter for a typing task [52]. EEG signals are known to be non-stationary and non-linear in nature [53][15], hence EEG signal analysis can benefit from analysis using EMD.

Another useful tool for non-stationary and non-linear signal analysis comes in the form of the Teager-Kaiser energy operator (TEO) [54]. The TEO is a non-linear operator that can be used for energy estimation of amplitude-frequency (AM-FM) modulated representations of non-stationary and non-linear signals. There is considerable research using EMD in conjunction with the TEO for non-stationary signal analysis, whereby a non-stationary signal is decomposed into AM-FM modulated representations, and the TEO is applied to these representations in various contexts, for example in creating energy-time-frequency representations [55] or feature extraction for classification of normal and abnormal biomedical signals [56].

The TEO was developed from the point of view of the energy required to generate a signal  $x[n]$ . This non-linear energy-tracking operator  $\Psi$  is given in its discrete form as [54]:

$$\Psi(x[n]) = x^2[n] - x[n+1] \cdot x[n-1] \quad (2.14)$$

The TEO is nearly instantaneous, given that only three samples are required for computing

the energy at a given time instant, as can be seen from Equation 2.14. This also makes the operator easy to implement efficiently.

In the next Sections, a novel method for mental task classification using EMD and the Teager energy operator for parameter extraction from EEG signals will be presented. This method achieves an average correct classification rate of more than 85% for classification between different tasks for each test subject. Furthermore, these results were obtained using a limited number of parameters extracted from the EEG signals.

#### 2.4.1.2 Data

The data used for this experiment is in the form of EEG signals recorded on 7 subjects, who performed five tasks each. These tasks were 1) *baseline task*, which required subjects to relax and think of nothing in particular; 2) *multiplication task*, involving mental multiplication of 2 multi-digit numbers; 3) *letter task*, in which subjects mentally composed a letter without vocalizing; 4) *rotation task*, requiring subjects to visualize a three dimensional object being rotated; finally, 5) *counting task*, in which subjects imagined a blackboard with numbers being written on it. The subjects performed the tasks in ten trials each over two days, and for each trial, EEG was recorded from six electrodes at positions ( $C_3, C_4, P_3, P_4, O_1, O_2$ ) for ten seconds at a frequency of 250 samples per second. Further details about the data are available in [52], and the data is also available for public download<sup>1</sup>.

#### 2.4.1.3 EMD and de-trending

EEG data of each subject, per trial, consists of 5 tasks. For each of these tasks, there are 6 EEG signals corresponding to the 6 electrodes, as mentioned in the previous subsection. Each EEG signal is of 10 seconds duration, and contains 2500 samples. For this experiment, the data of the first and last seconds was discarded, and hence 2000 samples from the second till the ninth seconds were used. EMD was applied to each of 6 EEG signals, for each task. In general, each EEG signal was decomposed into 9 IMFs. The IMFs with the lower indices

---

<sup>1</sup><http://www.cs.colostate.edu/eeg/>

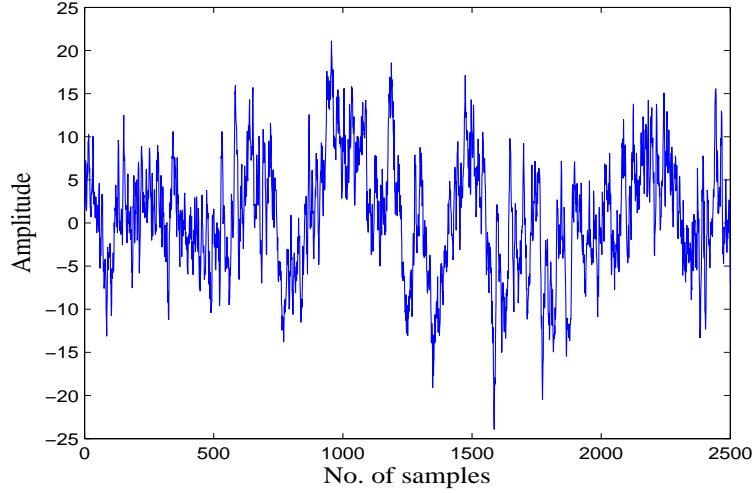


Figure 2.3: An original EEG signal used in the experiment.

correspond to high frequency oscillations, whereas those with higher indices correspond to the trend in the signal. Therefore it is possible to separate the trend from the original signal, and obtain a de-trended signal, using a simple method as described in [34]. This de-trending method falls under the energy approach described previously. Each IMF, by definition,  $c_i(t); i = 1$  to  $l$ , is a zero-mean function. Therefore, given a signal  $x(t)$ , a method for finding the de-trended signal  $y_d(t)$  is through partial reconstruction of the signal as

$$y_d(t) = \sum_{i=1}^l c_i(t) \quad (2.15)$$

and observing for which  $i = l$  the mean of  $y_d(t)$  varies significantly from zero. Figure 2.3 shows an original EEG signal used in the experiment. EMD is applied to this original signal, which is then partially reconstructed by adding up the IMFs, starting from lower indices. Figure 2.4 shows that the mean of  $y_d(t)$  changes significantly from zero after  $l = 5$ . It should be mentioned here that  $l = 5$  was obtained very consistently in tests using a large number of EEG signals from different subjects and different tasks.

Therefore the de-trended signal can be separated by partial reconstruction of the original signal through summing up IMFs 1 to 5, as shown in the upper plot of Fig 2.5. The trend

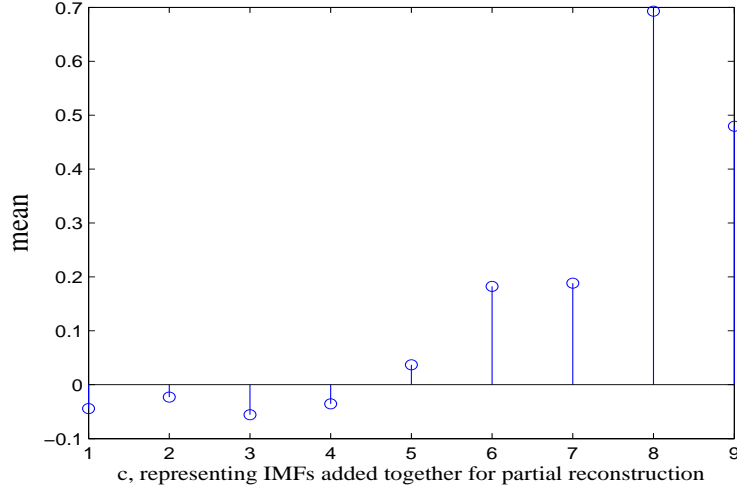


Figure 2.4: Mean of IMFs during partial reconstruction; mean changes significantly from 0 after IMF number  $l = 5$

can be separated from the original signal by summation of IMFs 6 to 9, illustrated in the lower plot of Figure 2.5.

#### 2.4.1.4 Calculation of Teager energy

For each subject, and for all 5 tasks performed by a subject, EMD was applied to each of the 6 EEG signals per task. The de-trended signal (sum of IMFs 1-5) and the trend (sum of IMFs 6-8, ignoring the last IMF, which is the residue) is obtained for each EEG signal using the method described in the Section 2.4.1.3. The next step in the process involves calculating the average Teager energy for each IMF separately, for both sets of IMFs, corresponding to the higher indices 1-5, and the lower indices 6-8. The average Teager energy,  $e_c$ , is calculated as [57]:

$$e_c = \frac{1}{N} \sum_{n=1}^N | \Psi[c_i(n)] | \quad (2.16)$$

where  $N$  is equal to the number of samples in the IMFs,  $\Psi[\cdot]$  is the discrete-time TEO and  $i = 1 - 5$  or  $i = 6 - 8$ .

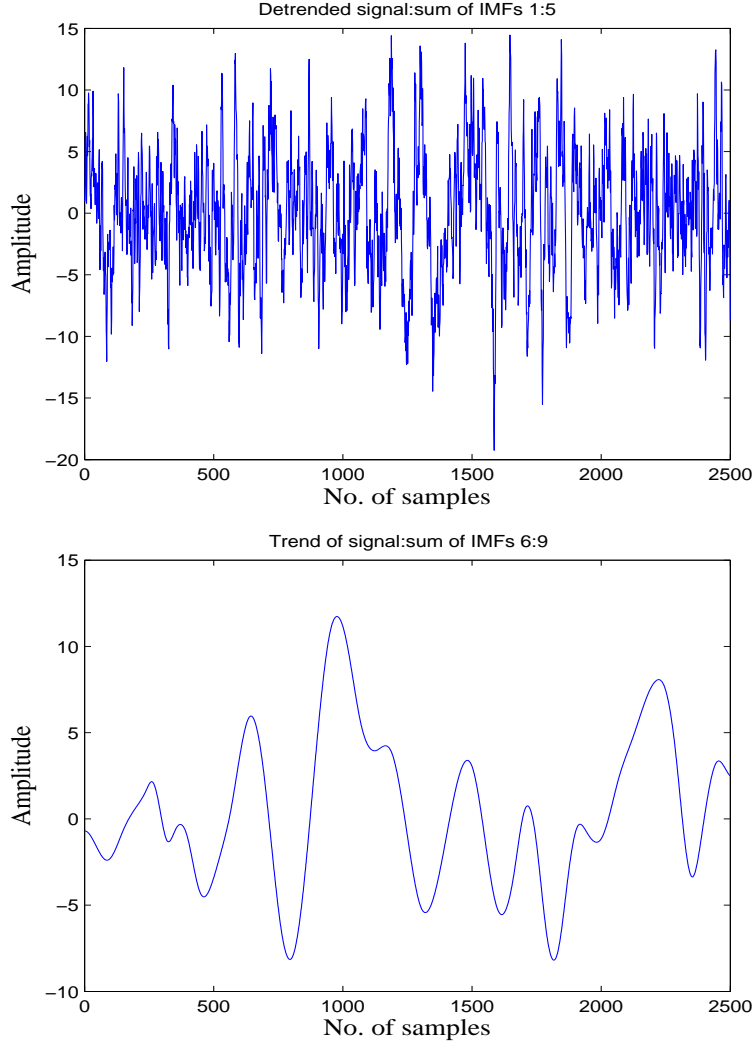


Figure 2.5: (A De-trended EEG signal (top), and Trend of the EEG signal (bottom).

#### 2.4.1.5 Parametrization and classification

The Teager energy is used for the parametrization of the EEG signals in the following way. For each task performed by a subject, we have a feature vector with 30 parameters (1 task  $\rightarrow$  6 EEG signals  $\rightarrow$  first 5 IMFs) consisting of the average Teager energy of the higher index IMFs, and another feature vector of dimension 18 (1 task  $\rightarrow$  6 EEG signals  $\rightarrow$  last 3 IMFs excluding the residue) corresponding to the lower index IMFs. These feature vectors are called high frequency energies (HFE) and low frequency energies (LFE) respectively, given



that the higher index IMFs correspond to higher oscillation rates in the original signal, and the lower index IMFs correspond to lower frequencies.

The two feature vectors obtained for each task were used to discriminate between the tasks for each subject using Linear Discriminant Analysis (LDA). For this purpose, both the high frequency energies (HFE) and the low frequency energies (LFE) were used, and the results obtained are discussed in Section 2.4.1.6.

Table 2.1: Classification Accuracy For Classification Of Mental Tasks For All Subjects

Task Combination	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7	
	H.F.E	L.F.E	H.F.E	L.F.E	H.F.E	L.F.E	H.F.E	L.F.E	H.F.E	L.F.E	H.F.E	L.F.E	H.F.E	L.F.E
base-count	80%	83.3%	70%	100%	90%	83.3 %	80%	83.3%	90%	83.3%	100%	83.3%	80%	100%
base-letter	70%	100%	80%	83.3%	90%	83.3%	90%	66.7%	90%	100%	90%	83.3%	70%	100%
base-math	80%	100%	70%	83.3%	80%	83.3%	80%	100%	70%	100%	90%	100%	100%	83.3%
base-rot	90%	66.7%	80%	100%	90%	100%	90%	100%	80%	66.7%	100%	100%	80%	100%
letter-count	80%	66.7%	80%	83.3%	90%	83.3%	80%	100%	90%	83.3%	100%	83.3%	80%	83.3%
letter-rot	100%	100%	80%	100%	80%	83.3%	80%	100%	90%	100%	100%	100%	100%	100%
math-count	70%	83.3%	70%	83.3%	70%	100%	80%	83.3%	80%	100%	90%	100%	100%	66.7%
math-letter	100%	100%	90%	100%	100%	83.3%	80%	100%	80%	66.7%	100%	66.7%	90%	100%
math-rot	80%	100%	60%	100%	90%	100%	80%	66.7%	90%	100%	90%	100%	100%	100%
rot-count	80%	100%	80%	100%	80%	100%	90%	83.3%	90%	66.7%	90%	100%	100%	100%
Average	83%	90%	76%	93.3%	86%	89.9%	83%	88.3%	85%	86.7%	95%	91.6%	90%	93.3%

H.F.E = High Frequency Energy; L.F.E. = Low Frequency Energy.

Tasks: base=baseline; math=multiplication; rot=rotation; count=counting;

#### 2.4.1.6 Results and discussion

The results of the analysis are presented in Table 2.1. The table lists the correct classification percentages for discrimination between tasks for each subject as obtained using both feature vectors, the high frequency energies and low frequency energies. Both feature vectors have an average correct classification rate above 80%, with the exception of subject 2 when using the high frequency energies. For many task combinations across all subjects, 100 % correct classification is obtained. At the same time, it is clear that the average correct classification rate using the low frequency energy feature vector is higher than that of the high frequency energy, except for subject 6, as shown in Figure 2.6. Also, the difference in the average correct classification rate using high and low frequency energies is significant in the case of subjects 1 and 2. Based on these results, the following conclusions can be drawn:

1. The correct classification rate obtained using the approach presented here is higher than or comparable to other approaches using the same data [58].
2. The dimensions of the feature vector used in this approach are low, especially using the low frequency energy feature vector.
3. The parameters of the feature vectors can be calculated easily and efficiently. Importantly, the whole EEG signal is used for extracting the parameters. This is in contrast to the approach presented in [58], where empirical mode decomposition was applied to 1 second segments of the EEG signal. This segmentation, or windowing, which can artificially remove the non-stationarity inherent in the data, does not benefit from the application of EMD.
4. The approach presented here also uses all the IMFs, by dividing them into two groups. The results show that the lower index IMFs, which capture the trend of the EEG signal, possess important discriminatory information. This is also in contrast to the works reported in [58] [59] where only the first 3 IMFs are used.

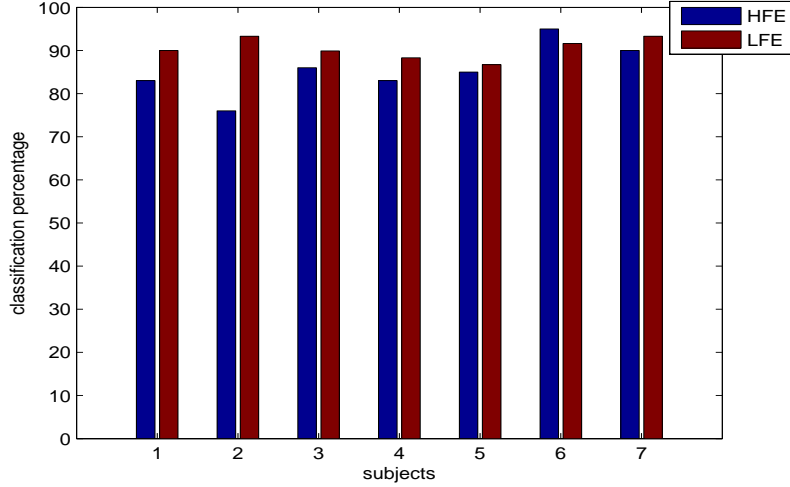


Figure 2.6: Average correct classification percentage for all subjects using high and low frequency energies, listed as HFE (left bars) and LFE (right bars), respectively.

In summary, this experiment presented a novel methodology for classification of mental tasks using EEG signals by application of EMD and the TEO. This methodology can be easily and efficiently implemented, and results in low-dimension feature vectors. It was also shown that both the de-trended EEG signal, as well as its trend, contain discriminatory information, which can be used for mental task classification.

## 2.5 Feature Analysis using EMD

The importance of feature analysis was discussed in Chapter 1, and it was mentioned that feature analysis is used to obtain representative parameters from signals being analyzed with the help of a signal analysis method. Feature analysis is an important aspect of automated decision making systems, and the quality of such systems depends on the features being representative and meaningful, as well as discriminative and localized, as discussed in Chapter 1.

EMD has been used extensively for feature analysis for automated decision making systems in a wide-range of domains, using various types of signals (e.g. [17][18][60][61][22][23][25][26][4], among many others). EMD decomposes signals based on their local time-

scales, and the characteristics of the signals hidden at these local time-scales are spread out into different IMFs. An important advantage of using EMD is that signal analysis can be performed in the time domain [18], or in the frequency domain using the methodology presented in Section 2.3.1, or in both domains [4], thereby allowing flexibility in feature analysis. In addition, the ability to calculate values of instantaneous energy and frequency from IMFs allows formulation of novel instantaneous features.

In the next Section, an experiment for pathological speech classification using EMD will be described. The use of EMD for decomposition of the speech signals and extraction of novel features from the obtained IMFs will be discussed. Since speech signals are non-stationary in nature [62], EMD is well-suited for analysis of speech signals.

## **2.5.1 Experiment 1: Pathological Speech Signal Analysis and Classification**

### **2.5.1.1 Background**

Pathological speech refers to speech problems that result from damage or malfunction of the human speech organs. These problems are exacerbated in people who use their voice professionally, for example professors, lawyers, actors and singers. The traditional way of diagnosing pathological speech is based on listening to a patient’s voice. However, such approaches are subjective to the training and expertise of the specialist performing the diagnosis [63]. Significant attention has thus been paid to objective assessment of speech pathology, making automated speech pathology detection an active field of research [64][65][66][67].

The main goal of automated pathological speech detection systems is to enable characterization of any input voice as either normal or pathological. These systems use signal processing tools, such as temporal, spectral and cepstral methods (e.g. [68][69][70]), or the more recently introduced joint time-frequency methods (e.g. [71][72]), as a means for accurate and discriminatory feature extraction. A classifier is then applied to the features thus calculated in order to discriminate between normal and pathological speech.

An important aspect to consider in pathological speech classification is whether the fea-

tures have been obtained from sustained vowels or continuous speech. Although sustained vowels offer a more controlled way of quantifying voice characteristics, and in general produce good classification results, they are not representative of speaking as they are of singing [62]. On the other hand, continuous speech signals capture important attributes of speech, such as rapid voice onset and termination, changes in voice frequency and amplitude, as well as sudden voice breaks. These attributes are also important in perception of voice quality of a speaker in everyday life [62], as has been assessed in recent works [71][72].

While extraction of features from sustained vowels and their effectiveness in classification of speech signals has been studied extensively in literature, similar studies using continuous speech signals are fewer in number. Moreover, the use of sustained vowels has been motivated by the resulting simple acoustic structure, hence resulting in simpler and consistent voice quality assessment [62]. Studies analyzing continuous speech signals rely in general on segmentation of speech to identify the voiced, unvoiced and silent periods in speech. This is required for features used to quantify periodicity and regularity in speech, such as the harmonic-to-noise ratio, cepstral peak prominence and pitch amplitude, which hold only for voiced regions of speech. Speech signals, however, have non-stationary dynamics, and stationarity of signals cannot be assumed over short time intervals. It is possible for important signal characteristics to be lost if segmentation occurs at intervals containing non-stationary dynamics [72]. One of the reasons for the scarcity of works on feature extraction from continuous speech is the challenge of segmenting the speech signal into voiced, unvoiced/silent periods prior to feature extraction [71]. The study in [62] has pointed out the challenge involved in the analysis of continuous speech due to the inherent non-stationarity of the signal.

This makes non-stationary signal analysis techniques, which do not require segmentation of the signals, and which can better reveal non-stationary behaviour of signals such as trends, discontinuities and repeated patterns, more attractive for the analysis and automatic classification of speech signals. Therefore, EMD was used for this experiment. Furthermore, EMD lends itself well to extraction of temporal and spectral descriptors from the original

signals, as it decomposes a signal into IMFs, which are functions of time from which spectral information may also be obtained. However, application of EMD to analysis of speech signals is scarce. A noise-enhanced algorithm of EMD has been used to extract the fundamental frequency from sustained vowels as described in [73], whereas EMD-based classification of normal and pathological voices is presented in [74]. However, the work in [74] is also based on sustained vowels, and not continuous speech, thereby not realizing the full potential of a non-stationary signal analysis technique like EMD.

With this background, the goal of this experiment is to apply EMD in its simplest form to randomly selected continuous speech portions of both normal and pathological signals so as to extract meaningful and well-defined temporal and spectral features. With an emphasis on simplicity and effectiveness of the proposed methodology, the efficacy of the features presented in this experiment are demonstrated by applying a simple classification scheme to classify normal and pathological signals with high accuracy. It should also be pointed out that there is no evidence of pathological voice classification using randomly selected continuous speech portions having been done before. In this context, this experiment can also be seen as a preliminary step towards text-independent pathological voice classification. The overall methodology of the experiment is shown in Figure 2.7. The flow of the methodology is indicated by the solid directional lines, whereas the dashed lines represent the feedback that was incorporated in selection of speech segment length and the number of IMFs to be included for further analysis (as described later in Section 2.5.1.7).

### 2.5.1.2 Speech data

The data used in this experiment has been obtained from the Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database from 1994 [75]. Comprehensive details of the database and a list of studies using this database have been presented in [65]. Many studies, however, have only used a subset of the database, whether sustained vowels (e.g. 53 normal and 173 pathological [63], 53 normal 53 pathological [74]) or continuous speech (e.g. 51 normal and 161 pathological [71][72][76]). This experiment also uses a subset of the database,

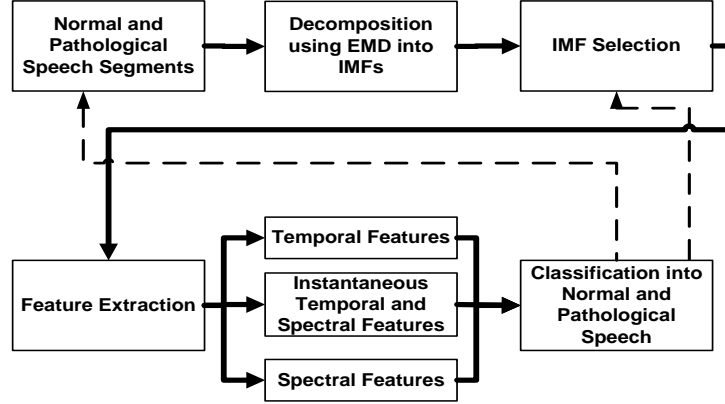


Figure 2.7: Block diagram for pathological speech classification described in Experiment 1. The dashed lines represent feedback links used to determine speech segment length and number of IMFs (Section 2.5.1.7).

consisting of continuous speech signals from 51 normal and 161 pathological speakers. This allows a good reference to compare the results of this experiment with those obtained using other approaches on the same database. All of the speakers spoke the same sentence, which is: "when the sunlight strikes rain drops in the air, they act like a prism and form a rainbow". The speech signals are sampled at 25 kHz and quantized at 16 bits/sample. Depending on the speaker, and also on the extent of voice pathology, the amount of time to speak the same sentence is different for different speakers. Therefore all speech signals, normal and pathological, are of different lengths.

### 2.5.1.3 Selection of speech portion length

Using the whole signal length for decomposition and feature extraction would have been computationally prohibitive, hence it was decided to use shorter portions of continuous speech selected randomly from different parts of the speech signals. The length of the speech portions was chosen based on the following design criteria:

1. Using a portion length long enough to incorporate actual speech, and not consisting entirely of unvoiced/silence period.



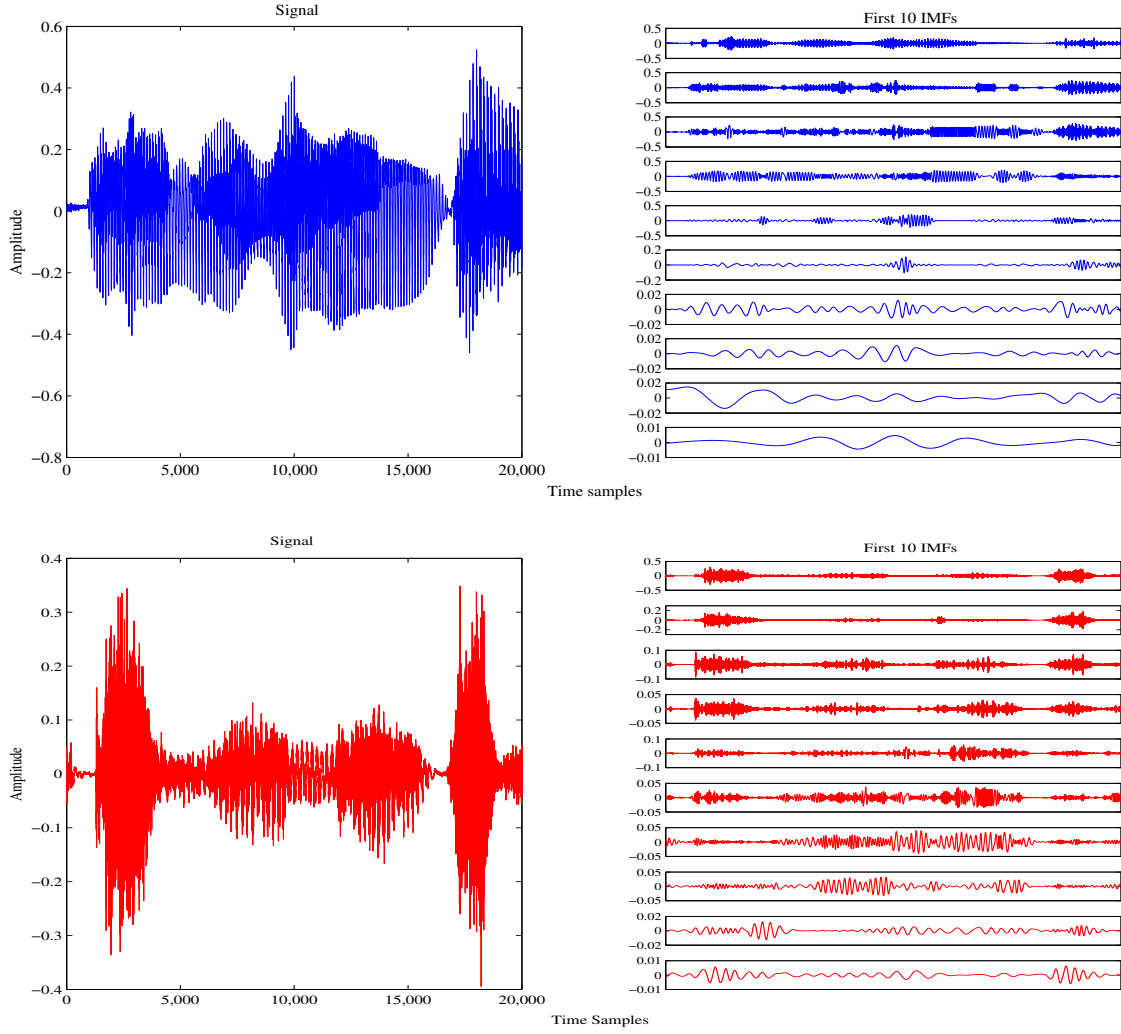


Figure 2.8: 800 ms portions of randomly chosen normal (upper figure) and pathological (lower figure) speech signals and their corresponding first 10 IMFs.

2. Having a portion length which preserves the non-stationarity present in the speech signals.
3. Achieving a fair compromise between classification accuracy and computational time required by the methodology.

Regarding Point 1, it has been suggested in [77], which uses the same speech database as in this work, that a good classification score between normal and pathological speech may be achieved by using the silence parts of the database, which demonstrates the differences

between the recordings. Therefore, the length of the speech portion should be a fair compromise that mitigates differences between normal and pathological speech that may arise due to reasons other than speech pathology, while at the same time reducing the overall computation time.

Point 2 relates to the non-stationarity aspects of portions of speech signals. Stationarity of speech signals is accepted over 10-30 ms segments [78], and any portion of continuous speech having a much longer length is able to capture the non-stationarity present in the signal. Though the 10-30 ms limit has been studied for normal speech, the presence of more transients, and larger spread of formants in pathological speech, as mentioned later in this Chapter (Section 2.5.1.5) and elsewhere (e.g. [72]), gives reasons for confident assumption of non-stationarity for pathological speech segments of length much greater than 10-30 ms, specially if the speech portions have been obtained from different parts of the signal. In the same context, confidence in the use of EMD as a means of temporal and spectral feature extraction can be had due to the demonstrated superiority (e.g. [36]) of time-frequency resolution of the EMD-based method, which is independent of the length of signal used.

To objectively select a speech portion length representing a compromise between classification accuracy and computational burden, as listed in Point 3, while also taking Points 1 and 2 into consideration, different lengths of the speech signals were chosen, and the classification accuracy checked for each length using the features described in Section 2.5.1.5. Based on this analysis, the results of which are presented in Section 2.5.1.7, a speech portion length of 800 ms was selected, which corresponds to 20,000 samples, for further analysis and feature extraction. These 800 ms portions were extracted randomly from different parts of the normal and pathological speech signals.

#### **2.5.1.4 Selection of IMFs**

For this experiment, the publicly available Matlab implementation [79] of EMD was used. Figure 2.8 shows 20,000 samples of randomly chosen continuous normal and pathological speech signals obtained from the database used in this experiment (Section 2.5.1.2) and first

10 of their corresponding IMFs obtained after application of EMD to the signals. The IMFs bring out the differing characteristics of the normal and pathological signals at different time-scales, which can be quantified in features useful for classification of normal and pathological speech.

Application of EMD to 20,000 samples of each speech signal results in the signals being decomposed into between 11 and 15 IMFs. The minimum number of IMFs was 11 for normal speech signals, and 13 for pathological speech signals. Since normal speech is in general more coherent than pathological speech, normal speech signals are expected to be decomposed faster than pathological speech.

In order to have uniformity in comparison between all speech signals, it was decided to use the same number of IMFs from all signals, making 11 the highest number of IMFs to be used for further analysis. Similar to the approach taken for selecting the length of the speech portion described in Section 2.5.1.3, the number of IMFs to be used for analysis and feature extraction was selected objectively. For this purpose, the accuracy of classification between normal and pathological speech signals was measured using features extracted (Section 2.5.1.5) from 11 IMFs obtained from both normal and pathological speech signals. Thereafter, the IMFs were excluded from feature extraction one-by-one, starting from IMF 11, and the classification accuracy measured. In the same way, the classification accuracy was measured with IMFs removed one-by-one, but starting from IMF 1.

Based on the results of this approach, more details of which are provided in Section 2.5.1.7, it was decided to use IMFs 1-10 for feature extraction, as the highest classification accuracy was obtained using this combination of IMFs.

#### **2.5.1.5 Features**

The first 10 IMFs obtained from each of the normal and pathological speech signals were carefully analyzed, and discriminative temporal and spectral features extracted from the IMFs. These features are described in this Section.

**Energy of intrinsic mode functions:** The energy of the analyzed IMFs  $c_i$  of the normal and pathological signals is selected as temporal feature. The energy,  $E_i$ , is calculated as:

$$E_i = \sum_{n=1}^N c_i^2[n], i = 1 \text{ to } 10 \quad (2.17)$$

where  $N$  is the length of the time samples.

Due to the normal speech being more coherent than pathological speech, the amplitude of normal speech signals in general has lesser variations than pathological signals. This difference is manifested in the decomposition process, whereby the amplitude variations of the pathological signals are spread over a larger number of IMFs as compared to normal speech signals, as can also be seen in Figure 2.8. Therefore the energy values  $E_i$  of IMFs  $i$  are expected to capture the spreading of the pathological signals' energy over a larger number of IMFs. It is also for this reason that using the energy values of the IMFs was found to be a better choice than using the energy of the signals themselves, as the distribution pattern of the signal energy into different IMFs is generally different for normal and pathological signals, which, together with the other features described next, helps increase the classification accuracy.

**Instantaneous joint time-frequency features:** Pathological speech is characterized by the more likely presence of transients and discontinuities, and more noisiness at higher frequencies. Normal speech generally has stronger and more distinguishable formant frequencies, whereas the formants in pathological speech are more spread and less structured. These differences are expected to be reflected in the instantaneous frequency and amplitude attributes, which can be extracted from IMFs as described previously in Section 2.3.1. It may also be re-iterated that the instantaneous amplitude represents the energy value of a particular frequency component at a given time instant. A time-frequency plot of IMF 7 from both normal and pathological speech signals is shown in Figure 2.9.

A set of two temporal features, and a set of two spectral features is calculated from the instantaneous amplitude and instantaneous frequency values obtained from each IMF. The

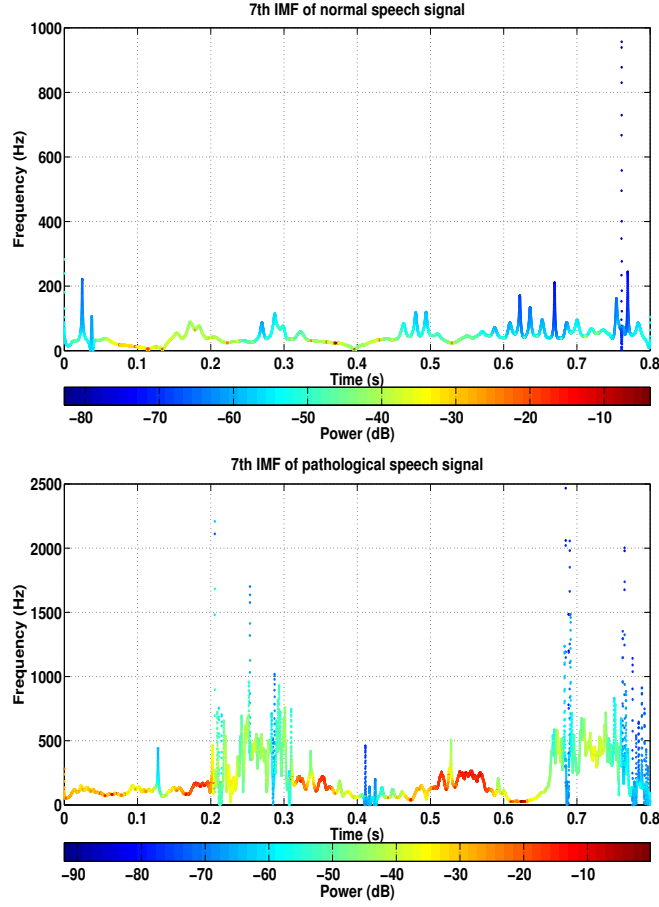


Figure 2.9: Time-Frequency plot of IMF 7 of both normal (upper figure) and pathological (lower figure) speech signals, showing a more even instantaneous frequency structure for normal speech signals.

set of instantaneous temporal features measures the following two values for each IMF  $i$ :

1. The spread of the instantaneous temporal energy,  $SP_i(ITE)$ , measured in terms of the instantaneous amplitude  $a_i$ . This is given by:

$$SP_i(ITE) = \frac{1}{N} \sum_{n=1}^N a_i^2[n] - \left( \frac{1}{N} \sum_{n=1}^N a_i[n] \right)^2 \quad (2.18)$$

2. The deviation of the instantaneous temporal energy,  $D_i(ITE)$ . This is given by:

$$D_i(ITE) = \frac{1}{N} \sum_{n=1}^N n \cdot a_i[n] \quad (2.19)$$

In both equations above,  $a_i$  represents the instantaneous amplitude of IMF  $i$ , and  $N$  is the length of the time samples.

Similarly, the set of instantaneous spectral features contains the following two feature values for each IMF  $i$ :

1. The spread of the values of the instantaneous frequency,  $SP(\omega_i)$ . This is given by:

$$SP(\omega_i) = \frac{1}{N} \sum_{n=1}^N \omega_i^2[n] - \left( \frac{1}{N} \sum_{n=1}^N \omega_i[n] \right)^2 \quad (2.20)$$

2. The deviation of the instantaneous spectral energy,  $D_i(ISE)$ .

$$D_i(ISE) = \frac{1}{N} \sum_{n=1}^N a_i[n] \omega_i[n] \quad (2.21)$$

In the two equations above,  $a_i$  and  $\omega_i$  represent the instantaneous amplitude and frequency, respectively, of IMF  $i$ , whereas  $N$  is the length of the time samples.

**Sum of marginal spectrum:** Due to the noisy and irregular nature of pathological speech, there are more components with higher energy at higher frequencies for pathological speech signals. This effect also depends on the level of speech pathology present in the speech, and can be exploited by extracting a feature from the marginal spectrum of the speech signals. As described in Section 2.3.2, the marginal spectrum represents the total amplitude (and hence energy) contribution from each frequency value over the whole signal length. It can be observed from Figure 2.10 that beyond a frequency threshold, values of which for each IMF are explained later in this Section, difference between the marginal spectrum amplitude for normal and pathological speech is discernable. Therefore the sum of the marginal spectrum,  $h(\omega)$ , above a frequency threshold, for IMFs 1 till 10, is used as a spectral feature, and is given by:

$$\sum_{f=f_{th_i}}^{f_{max_i}} h_i(\omega), i = 1 \text{ to } 10 \quad (2.22)$$

where  $f_{th_i}$  represents the frequency threshold for each IMF used in the feature,  $f_{max_i}$  is the maximum value of the frequency in each IMF, and  $\omega = 2\pi f$ . The selection of the frequency threshold for each IMF is described next.

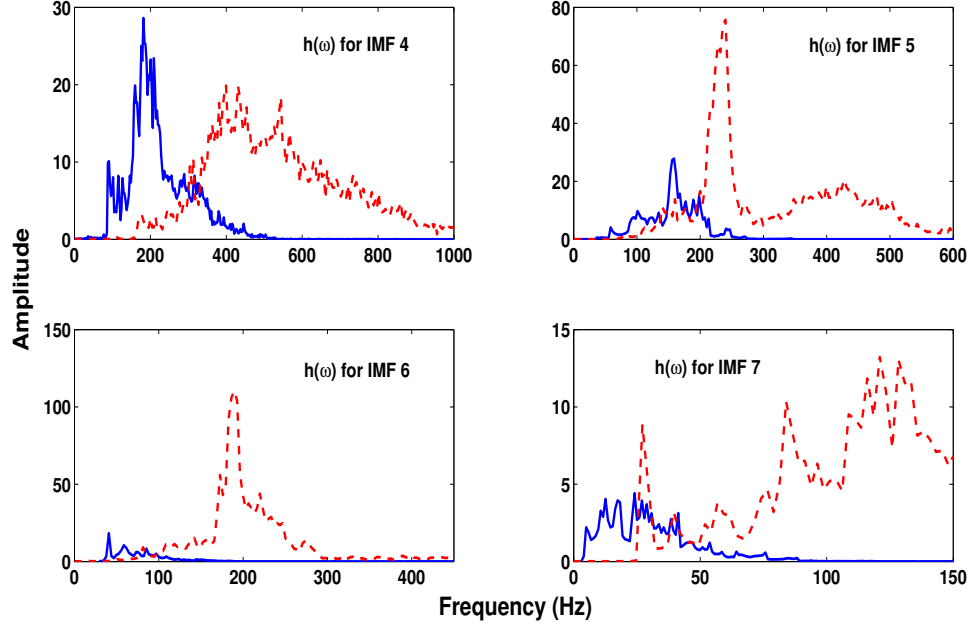


Figure 2.10: Figures showing marginal spectrum  $h(\omega)$  of IMFs 4, 5, 6 and 7 of both normal (solid lines) and pathological (dashed lines) speech signals. These figures illustrate that the marginal spectrum for pathological speech has higher amplitude beyond the frequency thresholds for the respective IMFs as compared to marginal spectrum for normal speech.

An experiment using a non-parametric statistical method, namely the Wilcoxon rank sum test [80], was designed to determine the frequency threshold beyond which there is discernable difference in the marginal spectrum of both normal and pathological speech signal. Firstly, frequency threshold values for each IMF based on visual analysis of the marginal spectrum plots of normal and pathological signals were selected. Then, four values, at fixed intervals, greater than each of these frequency thresholds, and four value less than the thresholds at the same intervals were used in the test to obtain that value of threshold which provides maximum separation between the marginal spectrum of normal and pathological speech signals. To carry out the test, first the sum of marginal spectrum was calculated

Table 2.2: Thresholds for calculating the sum of marginal spectrum as a discriminative feature.

IMF No.	Frequency threshold by visual inspection	Frequency threshold by statistical test
1	5000 Hz	5100 Hz
2	2100 Hz	2100 Hz
3	1550 Hz	1575 Hz
4	900 Hz	925 Hz
5	500 Hz	500 Hz
6	300 Hz	400 Hz
7	150 Hz	125 Hz
8	100 Hz	20 Hz
9	50 Hz	40 Hz
10	25 Hz	15 Hz

for each IMF for each of the 9 frequency threshold values for 25 randomly selected normal and pathological signals. Then the Wilcoxon rank sum test was performed using the sum of marginal spectrum for all the 25 normal and pathological speech signals for each frequency threshold value to test the separation between the sum of marginal spectrum values for normal and pathological speech. The frequency threshold value with the lowest  $p$ -value was then chosen as the threshold for that particular IMF. Table 2.2 lists the frequency thresholds decided by visual analysis of marginal spectrum plots, and the thresholds selected by the experiment described in this Section. These thresholds were used to obtain the feature values represented by Equation 2.22.

#### 2.5.1.6 Classification

In order to employ a simple classification scheme for classification of normal and pathological speech signals, a linear discriminant analysis (LDA) based classifier [81] was employed. A LDA-based classifier works in the following way: the feature vector is transformed into



canonical discriminant function values given by

$$f = a + b_1v_1 + b_2v_2 + b_3v_3 + b_4v_4 + b_5v_5 + b_6v_6 \quad (2.23)$$

where  $\{v\}$  represents the set of temporal and spectral features described in Section 2.5.1.5,  $\{b\}$  represents the coefficients, and  $a$  is a constant. The posterior probability of each sample that occurs in each of the two groups is calculated using the discriminant scores and the prior probability values of each group. Based on the posterior probability, the sample is assigned to the group with the highest posterior probability.

The classification accuracy for pathological speech classification was calculated using stratified 10-fold cross-validation on the whole speech data set. This technique randomly divides the training set into 10 disjoint subsets, where each subset has roughly equal size and same class proportions as in the training set. One subset is removed, and the classifier is trained using the other 9 subsets. The trained classifier is then used to classify the removed subset. This is repeated by removing each of the 10 subsets one at a time.

#### 2.5.1.7 Results

**Speech portion length and IMF number selection:** The analysis and feature extraction of normal and pathological signals previously described was based on 10 IMFs extracted from continuous speech portions of 800 ms length chosen randomly from different parts of the signals. This length for the speech portions was selected as a result of the objective methodology described in Section 2.5.1.3, which had the aim of finding a portion length offering a fair balance between classification accuracy and computational time. In this context, the classification accuracy obtained for different continuous portion lengths is shown in Table 2.3.

It can be seen from Table 2.3 that a portion length of 800 ms results in the highest classification accuracy compared to any smaller portion length. The decrease in accuracy for a shorter length of speech portion can be explained by the greater likelihood of silent periods, which have noise-like characteristics, affecting the correct classification of normal

Table 2.3: Classification accuracy (rounded to nearest whole number) obtained using different continuous speech portion lengths.

Portion length	Classification Accuracy
80 ms	83%
160 ms	82%
240 ms	83%
320 ms	87%
400 ms	87%
480 ms	85%
560 ms	87%
640 ms	84%
720 ms	89%
800 ms	96%
1200 ms	98%

signals. At the same time, a portion length of 1200 ms led to greater classification accuracy, due to the fact that more discriminatory characteristics of normal and pathological signals can be included in longer portion lengths, but at the cost of a higher computational time for signal decomposition and feature extraction.

The objective methodology for selection of the number of IMFs used for feature extraction was also described in Section 2.5.1.4. The focus of this methodology was to find the number of IMFs which lead to the highest classification accuracy. Table 2.4 lists the classification accuracy obtained for different number of IMFs used, with features extracted from IMFs 1-10 having the highest accuracy. The classification accuracy for less than 6 IMFs was lower than 90%. Also, removing the IMFs one-by-one, but starting from IMF 1, resulted in a classification accuracy of less than 90%. This demonstrates that the lower index IMFs contribute more to the discrimination than the higher index IMFs.

The empirical methods for selecting the speech portion length and the number of IMFs

Table 2.4: Classification accuracy (rounded to nearest whole number) obtained using different number of IMFs.

Number of IMFs	Classification Accuracy
1 - 11	93%
1 - 10	96%
1 - 9	94%
1 - 8	95%
1 - 7	92%
1 - 6	92%

works well for different signal lengths, as well as for different number of signals, thereby forming objective methodologies than can be used in different experimental setups.

**Effectiveness of extracted features:** To check the effectiveness of the extracted features in discriminating between normal and pathological speech, un-paired  $t$ -tests were used to test the null hypothesis that the feature values obtained from each of the 10 IMFs for both normal and pathological speech have the same mean. The  $p$ -values thus obtained are shown in Table 2.5 for each of the six features, and IMFs 1 to 10. It can be observed from Table 2.5 that the null hypothesis is rejected for all cases except for IMFs 9 and 10 for the feature  $E_i$ , and IMFs 4 and 10 for the feature  $SP_i(ITE)$ . The combination of all the extracted features determines the effectiveness in discrimination between normal and pathological speech.

**Classification results:** The accuracy of classification between normal and pathological speech signals is calculated with a LDA based classifier on the whole speech data set. The confusion matrix obtained using stratified 10-fold cross-validation on the speech data set is shown in Table 2.6. From this table it can be seen that out of 51 normal speech signals, 49 were correctly classified, whereas 2 were misclassified as pathological. Out of 161 pathological speech signals, 154 were correctly classified, and 7 were misclassified as normal. The overall total classification accuracy represented by these results was calculated using the definition of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) as

Table 2.5:  $p$ -values obtained by performing unpaired  $t$ -tests of the null hypothesis that the feature values obtained for each of 10 IMFs of normal and pathological signals used in the study have the same mean.

Feature	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5
$E_i$	0.00365	$3.85 \times 10^{-9}$	$1.38 \times 10^{-7}$	0.03045	0.01508
$D_i(ITE)$	$5.72 \times 10^{-5}$	0.01088	0.00575	0.00651	0.00156
$D_i(ISE)$	$4.22 \times 10^{-42}$	$2.95 \times 10^{-27}$	$1.78 \times 10^{-22}$	$4.94 \times 10^{-19}$	$3.60 \times 10^{-16}$
$h(\omega)$	$9.47 \times 10^{-22}$	$9.60 \times 10^{-10}$	$1.50 \times 10^{-17}$	$1.48 \times 10^{-24}$	$1.27 \times 10^{-22}$
$SP_i(ITE)$	0.00428	$9.57 \times 10^{-9}$	$1.64 \times 10^{-6}$	0.10234	0.01351
$SP(\omega_i)$	$1.44 \times 10^{-11}$	$9.12 \times 10^{-12}$	$2.93 \times 10^{-12}$	$2.60 \times 10^{-14}$	$1.75 \times 10^{-13}$

Feature	IMF 6	IMF 7	IMF 8	IMF 9	IMF 10
$E_i$	$4.29 \times 10^{-10}$	$3.43 \times 10^{-9}$	$4.84 \times 10^{-8}$	0.0526	0.75809
$D_i(ITE)$	0.00172	0.00589	0.00075	0.00788	0.00365
$D_i(ISE)$	$9.46 \times 10^{-16}$	$1.67 \times 10^{-17}$	$5.17 \times 10^{-19}$	$1.78 \times 10^{-20}$	$5.34 \times 10^{-17}$
$h(\omega)$	$2.23 \times 10^{-20}$	$4.58 \times 10^{-23}$	$6.34 \times 10^{-15}$	$1.41 \times 10^{-19}$	$1.61 \times 10^{-16}$
$SP_i(ITE)$	$1.03 \times 10^{-9}$	$1.03 \times 10^{-12}$	$4.97 \times 10^{-6}$	0.01415	0.08220
$SP(\omega_i)$	$8.49 \times 10^{-14}$	$1.04 \times 10^{-12}$	$9.82 \times 10^{-12}$	$1.43 \times 10^{-9}$	$1.72 \times 10^{-16}$

$(TP+TN)/(TP+TN+FP+FN)$  [2]. The classification accuracy thus obtained is 95.7%. The misclassified pathological signals were analyzed in detail, and it was found that these speech portions, extracted randomly from different parts of pathological signals, were more similar to normal speech in their characteristics, as demonstrated by a lack of noisiness, as well as of transients and discontinuities. Although the pathological signals from which these portions were extracted were perceptually quite similar to normal, we may speculate the resulting false negatives as a consequence of random selection of speech portions and a limited speech portion length. This, however, is corroborated by the finding discussed earlier in Section 2.5.1.7 that an increase in speech portion length leads to better classification accuracy.

To further ensure no bias in the classification accuracy due to the unequal class sizes, the experiment was repeated with 50 normal signals, as well as 50 pathological signals randomly selected from the 161 available. In this case, the classification accuracy using 10-fold cross-validation was 95.8%, thereby demonstrating the robustness of the methodology.

Table 2.6: Classification Results using Linear Discriminant Analysis with 10-fold cross-validation. TP:True Positive, TN:True Negative, FP:False Positive, FN:False Negative

		True Classification	
		Pathological	Normal
Predicted Classification	Pathological	154 (TP)	2 (FP)
	Normal	7 (FN)	49 (TN)

### 2.5.1.8 Discussion

This experiment presented a methodology based on EMD to automatically discriminate between normal and pathological speech. A main strength of this approach is the simplicity and effectiveness of the methodology with respect to decomposition, feature extraction, and classification. Also important in the methodology is the ability to extract meaningful features from both the time and frequency domains. In this regard, extraction of true instantaneous features at different time-scales distinguishes this methodology from other approaches. The efficacy of instantaneous features in capturing the discriminatory information hidden in normal and pathological speech signals is demonstrated by the high classification accuracy obtained in this experiment.

Another strength of the proposed approach lies in the use of continuous speech samples instead of sustained vowels. Compared to approaches that rely on sustained vowels or segmented portions of speech signals, this approach offers a more realistic automated assessment of voice quality, and demonstrates the advantage of using a non-stationary signal analysis technique to classify signals with non-stationary dynamics. An important point in this regard is that the use of continuous speech does not require an extra processing step to separate the sustained vowels from the speech. This is an advantage, as any error in the sustained vowel separation will propagate to the pathological speech classification stage. Due to the nature of pathological speech, which has a less periodic structure compared to normal speech, errors in sustained vowel detection and separation could be very common.

For this experiment, 800 ms portions of continuous speech were selected randomly from

different parts of the speech signals. This length of speech portion, which was selected objectively, corresponds to 20,000 samples and can effectively capture the non-stationarity present in speech. At the same time, extracting different portions from different parts of the signals contributes to the robustness of the methodology. A large subset of the IMFs was also selected for feature analysis using an objective methodology. This is important, since the useful temporal and spectral properties of signals are spread into a large subset of all IMFs.

The methodology of this experiment may also be compared favourably with another EMD-based approach for pathological speech classification [74] which uses speech signals from the same database as used in this experiment. The approach in [74], based on sustained vowels instead of continuous speech, uses only 3 IMFs, numbers 2-4, without a rationale for excluding the rest of the IMFs. Also, the approach in [74] uses the maximum power spectral density and the corresponding frequencies of these 3 IMFs as the feature vector. Since the use of EMD can result in mode-mixing, which is a known issue with EMD [50], features dependent on frequency values alone might not be robust to a wide range of speech signals. In case the noise-assisted EMD [50] is used to reduce the effect of mode-mixing, the methodology becomes computationally prohibitive. Furthermore, as already discussed previously, EMD is a non-stationary signal analysis technique, and should be taken advantage of for analyzing non-stationary signals, such as continuous portions of speech signals.

The results of this experiment can also be compared with some recent approaches using a similar methodology and the same speech database. The time-frequency-matrix-decomposition approach [72] using the same voice database and an LDA classifier achieves a correct classification result of 98.6% using speech segments of 80 ms. In comparison, the proposed methodology achieves comparable classification accuracy using text-independent portions of speech while using a less complex method.

Another approach based on adaptive time-frequency decomposition based on matching pursuit [71] using the same voice database achieves a correct classification result of 93.4% using an LDA classifier. Features extracted from the ambiguity domain [76], but using

segmented speech signals, reach a classification accuracy of 97.5%, again using the same database and classifier as used for this experiment. The work in [76] does not provide details about the segmentation.

An aspect of EMD decomposition that needs to be kept in view is the processing time required to extract all the IMFs. Since EMD is an iterative algorithm, the length of the signal to be decomposed adds to the processing time. In this regard, using shorter portions of speech signals is also an engineering compromise to keep a check on the processing time required by the decomposition process. In the methodology presented here, by far the largest chunk of the overall processing time is represented by the decomposition process, which consumes on average 10 seconds for each speech signal consisting of 20,000 samples. Extraction of the 6 features takes on average 2 seconds per speech signal. It should be mentioned here that these numbers have been calculated on a laptop computer (AMD processor, with 4 GB of RAM) running MATLAB software version 7.10.0, and are provided as a representative measure only. Other works to which this experiment has been compared to in this Section ([74][72][71][76]) do not provide estimates of processing times.

In the next Section, application of EMD for classification of telephone-quality pathological speech using a similar set of features will be described. This also demonstrates the robustness and efficacy of the extracted features by achieving high classification accuracy for pathological speech signals under different conditions.

## **2.5.2 Experiment 2: Telephone-Quality Pathological Speech Classification**

### **2.5.2.1 Background**

The methodology used for this experiment is similar to the methodology for the previous experiment described in Section 2.5.1. However, in this case, the speech signals represent telephone quality speech. Automation of speech pathology detection improves the accuracy of assessment, and allows long distance identification and monitoring of pathological speech for patients in distant geographical areas. In this regard, tele-medicine has an important

role to play, since transmitting the voice over telephone channels allows a low-cost alternative for automated voice pathology assessment and vocal quality monitoring [2]. However, the literature on telephone-quality pathological speech classification is sparse. The work described in [2] uses descriptors such as pitch and amplitude measures and harmonic-to-noise ratio extracted from sustained vowels to classify telephone-quality pathological speech with an accuracy of 74.2 %. In this Section, a methodology based on application of EMD on continuous speech samples to calculate features for classification between telephone quality normal and pathological speech is presented. Using four unique features, among them two based on instantaneous temporal and spectral measures not readily available in other popular approaches for speech analysis, telephone quality speech under different noise conditions will be classified with classification accuracy higher than that reported in [2]. In this context, the methodology presented in this Section elucidates the possibility of remote detection and assessment of voice pathology.

### 2.5.2.2 Data

For this experiment, the same database as described for the previous experiment in Section 2.5.1.2 is used. However, as described in the next Section, the data is manipulated to simulate telephone quality speech.

**Telephone-quality speech simulation:** In order to simulate telephone quality speech from the original data, distortions introduced by telephone transmission were incorporated into the original speech data, for both normal and pathological speech [2]. First, all the speech signals were down-sampled to 8 kHz, with the effective bandwidth being 4 kHz. Care was taken to low-pass filter the signals in order to prevent aliasing. The down-sampled data was then band-pass filtered with a linear filter having the frequency range 300 - 3400 Hz, which is the bandwidth of telephone transmission. The final manipulation of data included addition of additive white Gaussian noise (AWGN). The AWGN was added with 3 different signal-to-noise (SNR) ratios, which are 50 dB, 40 dB and 30 dB. This allowed testing of the proposed methodology with different levels of noise present in the telephone quality speech.



It should be mentioned that the 30 dB SNR was chosen to compare the obtained results with those presented in [2].

### 2.5.2.3 Decomposition

EMD was used to decompose the telephone-quality normal and pathological speech signals into IMFs. For this, the first 800 ms of each signal, which correspond to 6400 time samples, were used. The signals of length 800 ms are decomposed into between 12 and 14 IMFs. The IMFs resulting from the application of EMD to all speech signals were rigorously examined. It was found that IMFs 11 and higher exhibited insignificant temporal or spectral structure, with extremely low signal amplitudes and number of signal oscillations. Therefore, IMFs 1 till 10 for all normal and pathological signals were selected for further analysis and feature extraction.

### 2.5.2.4 Extraction and evaluation of features

For this experiment, the features extracted are similar to the features described previously in Section 2.5.1.5. These are described next.

**Energy of IMFs:** The energy,  $E_i$ , of the 10 IMFs (1 till 10) of the normal and pathological signals is selected as a feature. The energy of the IMFs represents a temporal feature, and is calculated as:

$$E_i = \sum_{n=1}^N c_i^2[n], i = 1 \text{ to } 10. \quad (2.24)$$

**Variance of instantaneous measurements:** As features, the variance,  $Var(a_i)$  (temporal feature), of the instantaneous amplitude  $a_i$ , and the variance,  $Var(\omega_i)$  (spectral feature), of the instantaneous frequency  $\omega_i$ , are obtained from each of the 10 IMFs. The instantaneous amplitude and instantaneous frequency are calculated by means of the analytic signal  $z_i$  as explained in Section 2.3.1.

Table 2.7: Average classification accuracy with different levels of noise added

SNR (dB)	30	40	50
Avg. classification accuracy	89.7 %	91.3 %	92.6 %

Table 2.8: Classification results for one run of experiment in terms of measures defined in [2]. SNR: signal-to-noise ratio. Acc: Accuracy. Sens: Sensitivity. Spec: Specificity. Pos Pred: Positive Predictivity. Neg Pred: Negative Predictivity. Acc(%):(TP+TN)/TP+TN+FP+FN. Sens(%):TP/(TP+FN). Spec(%):TN/(TN+FP). Pos Pred(%):TP/(TP+FP). Neg Pred(%):TN/(TN+FN).

SNR (dB)	Acc(%)	Sens(%)	Spec(%)	Pos Pred(%)	Neg Pred(%)
50	93.8	96.9	84.3	95.1	89.6
40	92.5	98.8	72.5	91.9	94.9
30	91.5	98.1	70.6	91.3	92.3

**Sum of marginal spectrum:** The sum of the marginal spectrum,  $h(\omega)$ , above a frequency threshold, for IMFs 1 till 6, is used as a spectral feature, and is given by:

$$\sum_{f=f_{th_i}}^{f_{max_i}} h_i(\omega), i = 1 \text{ to } 10 \quad (2.25)$$

where  $f_{th_i}$  represents the frequency threshold for each IMF used in the feature,  $f_{max_i}$  is the maximum value of the frequency in each IMF, and  $\omega = 2\pi f$ . The frequency thresholds in Hz, for IMFs 1 to 6, are, respectively, 2000, 1500, 1100, 900, 500, 300. The frequency thresholds were selected using the Wilcoxon rank sum statistical test as described in Section 2.5.1.5.

### 2.5.2.5 Classification results and discussion

For this experiment, 10 IMFs (1-10) of all the speech signals were used, and the feature vector was constructed from the features extracted from these IMFs. This way the feature vector consisted of 36 elements, given that 3 features were extracted from each of the 10 IMFs, and 1 feature from the first 6 IMFs. For classification, a linear discriminant analysis

(LDA) based classifier was employed. The average classification results for different levels of noise added to the telephone quality signals are presented in Table 2.7. Given the addition of AWGN to the speech signals after down-sampling and filtering, the experiment consisting of application of EMD to continuous speech samples (downsampled, filtered, noise-added), extraction of feature vector, and classification, was performed 5 times for each value of the SNR. The classification accuracy obtained for each value of SNR was then averaged over the 5 runs of the experiment. From Table 2.7, it can be observed that as the SNR is decreased from 50 dB to 30 dB, the average classification accuracy decreases by less than 3% only. The classification accuracy of 89.7% at SNR of 30 dB compares very favorably with the classification accuracy of 74.2 % presented in [2].

Furthermore, these results are illustrated using the measures defined in [2]. These measures are defined in terms of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). The classification results in terms of these measures for one run of the experiment for each value of SNR used are shown in Table 2.8. In terms of these measures as well, results using the proposed methodology show significant improvement. The decrease in classification accuracy with an increase in the noise level in the speech signals can be explained by the fact that addition of noise to the normal speech signals makes their characteristics similar to pathological speech, as also illustrated by the increase in false positives. In general, these results also demonstrate the efficacy of instantaneous descriptors extracted from signals, such as instantaneous amplitude and frequency, and in general, of EMD, which allows such descriptors to be extracted with ease. Also, since continuous speech segments were used, the methodology presented in this experiment can be carried over to an automated telephone-based pathological speech classification system.

## 2.6 Chapter Summary

This Chapter presented Empirical Mode Decomposition (EMD), which is an empirical approach for non-stationary and non-linear signal analysis. This was followed by discussion of using EMD for the important signal processing task of de-noising and de-trending, supported

by an experiment for mental task classification to demonstrate the practical application of EMD-based de-noising and de-trending. Feature analysis using EMD was also presented in this Chapter in the context of experiments using pathological speech that highlighted the efficacy of EMD in terms of novel feature calculation from both the time and frequency domains.

# Chapter 3

## Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS)

### 3.1 Introduction

As discussed previously, Empirical Mode Decomposition (EMD) is a technique for decomposition of non-stationary and non-linear signals into amplitude and frequency modulated waveforms called intrinsic mode functions (IMFs), which are obtained by adaptive extraction of all the oscillatory modes present in a signal. EMD is defined by an algorithm, and the IMFs are extracted through a process called sifting. The details of the EMD technique were presented in Chapter 2, along with a discussion of some of its important properties.

Although the original EMD technique as presented in [6] is widely used in a broad range of domains as discussed in the previous Chapter, a number of variations of the EMD technique have been proposed in literature. In general, the purpose of these variations is to overcome one or more of the issues identified in the original technique, or to enhance the algorithm in a particular way. One aspect of the decomposition, however, that is missing from the original EMD technique and its variations, relates to control of the decomposition process through a time-scale based decomposition. For instance, the EMD algorithm separates the

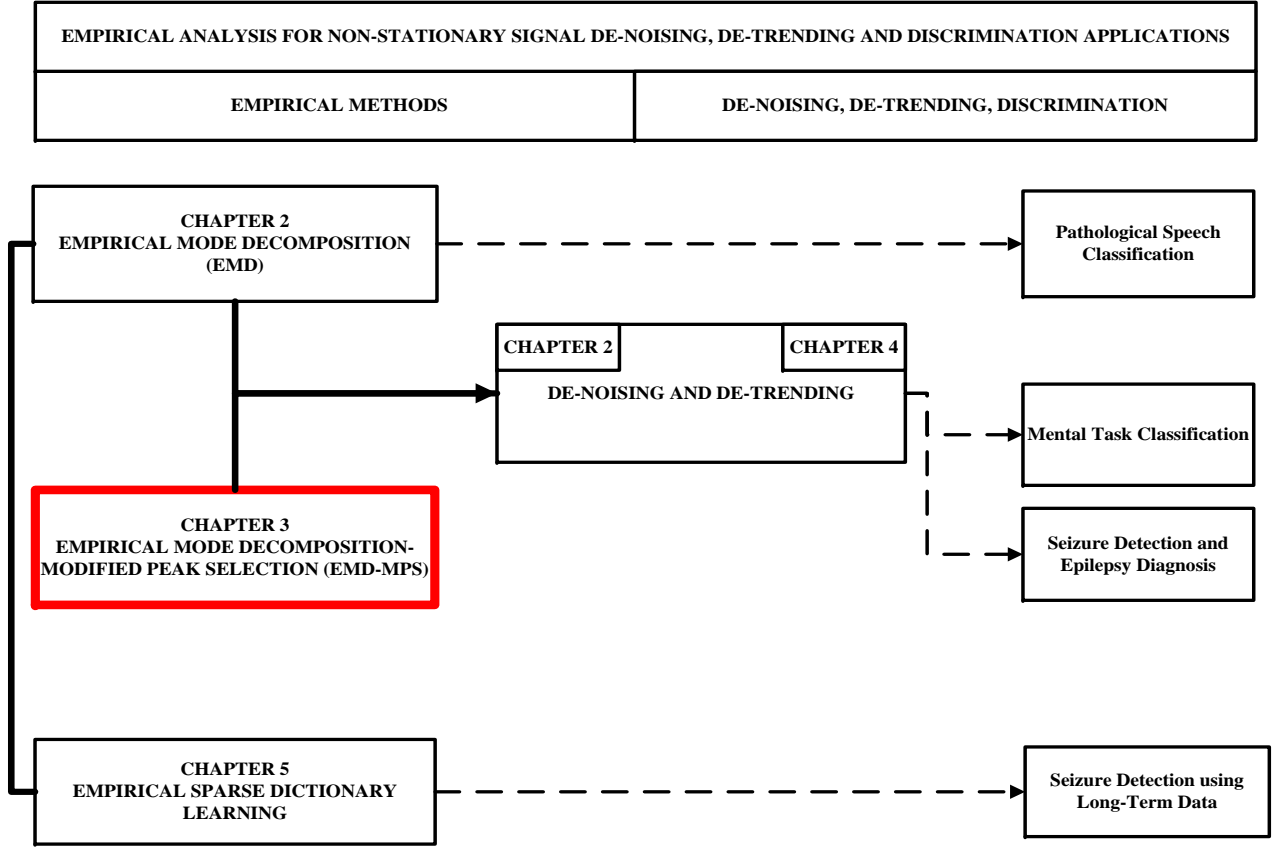


Figure 3.1: Chapter 3 describes Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS), which is a novel modification of EMD, and discusses the method’s decomposition properties.

different oscillatory components into IMFs based on their local time-scales, but the algorithm does not provide a way such that, based on a particular time-scale, some of the oscillatory components are decomposed, while the others are not.

In this Chapter, the proposed variations of the original EMD technique will be categorized first, with reference made to some of the important works in this regard. Then a variation of the EMD technique, developed in the context of this dissertation, will be presented. This technique, named Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS), allows a novel time-scale based decomposition. Some important properties of EMD-MPS will be established, along with highlighting how EMD-MPS provides a new insight into the

original EMD technique. The concept of hierarchical decomposition will also be presented, which is a novel methodology made possible through the use of EMD-MPS.

## 3.2 Variations of Empirical Mode Decomposition

As already mentioned, a number of variations of the original EMD algorithm have been proposed over the years in order to overcome the issues with the original algorithm. These variations may be categorized in terms of the issues with EMD the variations are designed to address. The categories may broadly be defined as follows:

- Mode-mixing
- Envelope formation
- Replacement of sifting process

These categories are not rigid in scope, and have been broadly defined, as a variation of EMD that replaces the traditional sifting process with another framework may provide an alternative formulation of the envelope and might prevent mode-mixing in certain scenarios as well.

**Mode-mixing:** The phenomenon of mode-mixing refers to an IMF containing signals of widely disparate time-scales, or a signal of the same time-scale appearing in different IMFs [50]. This may compromise the physical meaning of the IMFs and hence the robustness of the applications using EMD. The main reason for mode-mixing is intermittency in the signal being decomposed, which refers to component(s) of widely disparate time-scales occurring intermittently within the signal. An initial attempt to overcome mode-mixing due to intermittency was described in [8], whereby mode-mixing could be prevented using an intermittence test. However, the intermittence test required a subjectively selected time-scale, which infringed upon the adaptivity of the EMD technique. The intermittence test

was also not found to work well for the cases where components of different time-scale were not clearly separable, but mixed over a range continuously [50].

An important and widely used method to alleviate mode-mixing, called Ensemble Empirical Mode Decomposition (EEMD), was first proposed in [50]. The EEMD belongs to the category of noise-assisted data analysis methods, and takes advantage of the dyadic filter-bank property of EMD when applied to white noise [37][82]. In the EEMD method, multiple realizations of white noise with a small amplitude are added to the signal before decomposition using EMD. This way, the whole time-frequency space is uniformly populated by the added white noise, such that the resulting IMFs exhibit a dyadic filter bank structure, with the result that parts of the signal with different time-scales are automatically separated into the IMFs corresponding to the correct time-scale. Importantly, an ensemble mean of the IMFs resulting from different trials is calculated to cancel the effect of the added noise within each IMF, hence giving the method its name. There are, however, a number of issues associated with EEMD. For instance, as pointed out in [83], if a desired signal resides in multiple dyadic sub-bands, then a noise-assisted decomposition method such as EEMD will by design spread the signal into different IMFs, leading to unwanted mode-mixing. In addition, the optimal size of the ensemble and the amplitude of the added noise sequences are still open questions [84]. Also, for added noise to cancel out in the ensemble averaging process, the size of the ensemble has to be sufficiently large, which means that the EEMD technique is computationally prohibitive, as the decomposition has to be performed for every added noise sequence.

A variation of the EEMD technique, named Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is described in [85]. In contrast to the EEMD technique, where each signal with the added noise sequence is independently decomposed and then the resulting IMFs ensemble averaged, in CEEMDAN the first IMF is obtained in the same way as EEMD, but thereafter the noise addition and ensemble averaging is done over the residue obtained after extraction of each mode. This process is continued until all modes have been extracted and the algorithm terminates according to the stopping criterion.



Compared to EEMD, the CEEMDAN technique requires less than half the number of sifting operations and a smaller ensemble size, resulting in a reduction of the algorithm computation time. However, selection of the amplitude of the added noise is still a subjective choice.

Another interesting method to overcome mode-mixing is described in [86], whereby a masking signal is added to the signal under analysis to aid in the decomposition. This masking signal, which is a sinusoid, prevents the lower frequency components to be included in the IMF as mode-mixed components. Since the masking signal is known, it can be removed from the obtained IMF using the methodology defined in [86]. However, the method described in [86] can resolve mode-mixing for two sinusoids only, and the success of the method also depends on how accurately the higher frequency component in the signal is estimated.

A generalization of the masking signal approach is described in [87], whereby a high frequency frequency-modulated (FM) signal is used as a masking signal to prevent mode-mixing in amplitude-modulated/frequency-modulated (AM-FM) signals. This method, however, also requires estimation of the highest frequency component in the signal to be decomposed, and is only geared to signals that can be modeled as multi-component AM-FM signals.

A relatively different technique to weaken mode-mixing based on examining the bandwidth of the obtained IMFs is presented in [88]. In this technique, after an IMF has been obtained using the normal EMD sifting process, the sifting process is continued until a defined amplitude bandwidth of the IMF is smaller than a defined threshold, after which the sifting continues till the minimum of a defined frequency bandwidth is found. This approach is shown to weaken mode-mixing, but not resolve it comprehensively.

**Envelope formation:** The upper and lower envelopes, which are formed as part of the sifting process, are an indispensable part of the EMD algorithm used for extraction of the local mean. However, there is no formal definition of the envelope, with the cubic spline interpolation considered a simple, effective and popular choice for envelope formation [7]. However, the cubic spline is not necessarily an acceptable envelope, for reasons discussed

in [7]. A number of different attempts have been made at providing alternative envelope formation mechanisms, or a formal definition of the envelope.

The use of B-splines for envelope formulation is described in [31], due to the B-spline formulation being more amenable to mathematical study, with the work limited to showing decomposition results comparable to cubic spline envelope formation. A class of envelopes which minimize a quadratic penalty cost function involving cost parameters is proposed in [89]. In this formulation, the adjustment of the cost parameters allows adjustment of the frequency resolution of the IMFs. However, the described approach has a significant computational cost, and efficient and useful methods for changing the cost parameters have yet to be developed.

An alternative envelope formation approach, which uses four neighboring extrema in a particular time interval for envelope formation, is discussed in [90]. Although this approach is mathematically tractable, it is defined only for a two tone signal model. The work in [91] presents a methodology whereby instead of forming upper and lower envelopes using interpolation and calculating their average, a local mean curve of the signal is formed by interpolating data points that are local integral averages over segments between successive extrema of the signal. This sifting method, defined on the concept of the local integral mean, is shown to separate frequency components with a frequency ratio up to 0.8, which is an improvement over the frequency resolution power of EMD. However, the local integral mean based sifting is still dependent on the extrema of the signal, and is therefore susceptible to errors that may be induced due to noise present in signals.

A generalization of the local integral mean method is presented in [92], which uses a windowed average method for forming the local mean using a rectangular moving window. This method does not depend on discrete point detection or interpolation, and is more amenable to signals with noise, however the advantages of the method come at the cost of adaptability, as the method requires three parameters that have to be estimated, or supplied empirically.

Another windowed approach that replaces the cubic spline interpolation with raised cosine

interpolation is demonstrated in [93]. This method requires specifying the roll-off factor of the raise cosine pulse, as well as the window shape and size before the start of the decomposition process. Except for the interpolation, the rest of the methodology is the same as for EMD.

A re-definition of the signal mean envelope by means of new characteristic points is proposed in [33]. This method, called the Turning Tangent EMD, provides a new geometric definition of the mean envelope operator that interpolates the barycenters of particular oscillations. The method is shown to have decomposition performance comparable with EMD, and is computationally light, but requires signals to be in class  $C^1$ , or to have an appropriate numerical estimation of the derivative.

**Replacement of sifting process:** Given the empirical nature of the sifting process, variations of EMD that aim to replace the sifting process with a more mathematically tractable framework have also been proposed. The decomposition approach described in [94] makes use of a constrained optimization step to replace the sifting process, which seeks a trend of order  $k$  belonging to the class of spline functions. This approach is studied with respect to the sensitivity of the method to its parameters, and the accuracy of the method is found to depend on the estimation of the parameters. Also, the approach requires a pre-processing step to de-noise signals containing noise.

A similar approach which replaces the sifting process with a convex optimization procedure is presented in [95]. This method is based on proximal tools, and compared to the previously mentioned approach of [94], does not require the use of splines or a first estimate of the trend of order  $k$ . However, this method too is strongly dependent on the appropriate selection of its two parameters, and the sensitivity of the decomposition to the parameter values as well as robustness of the method to signals with noise is yet to be studied.

Approaches that seek to provide mathematically defined envelopes in the context of a sifting process implemented in a partial differential equation based framework are described in [96][97][98]. Although these models show good decomposition performance in certain scenarios, the models are defined only for a certain class of signals, or require model parameters

to be empirically chosen, or both.

### **3.3 Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS)**

In the previous Section, a number of different variations of EMD were highlighted, which address issues with different aspects of the EMD algorithm. While these variations enhance the algorithm, in general the enhancement comes at the cost of signal adaptivity of the algorithm, which is a unique strength of EMD. The same is the case with the mathematical models developed for different aspects of EMD, such as the sifting process. While the mathematical formulations provide capability for more precise analysis of different aspects of the algorithm, the models require estimation of their parameters for efficient and correct signal decomposition, which in turn has to be done empirically, or requires certain assumptions to be made. The model parameter estimation for one type of real life signals cannot, in general, be carried over to a different signal type, thereby requiring a different estimation for each signal type. Due to its data adaptive nature, the EMD algorithm is by nature empirical, and while mathematical models can be provided for certain aspects of the algorithm, the model parameters have to be estimated empirically for different signal types, thereby reducing the efficacy of the mathematical formulation.

On the other hand, the original EMD algorithm, and the different variations, many of which were referenced in the last Section, do not provide the facility for a time-scale based decomposition process, whereby signals can be decomposed according to a specified time-scale. In this Section, a novel variation of EMD, named Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS), will be described, which allows signal decomposition based on a particular time-scale, using intelligent peak selection for envelope formation during the sifting process.

The sifting process of EMD considers the oscillatory modes in the signal at the very local time-scale, and the EMD algorithm defines steps to extract these modes. The most

local time-scale is defined by two consecutive extrema, hence identification of all the extrema (peaks) in the signal is an important part of the sifting process. As described previously, the sifting process involves approximating the local average of the signal to be decomposed. This is done by forming upper and lower envelopes for the signal by interpolating through all the identified maxima and minima in the signal. The local average is then approximated as the mean of the envelopes. This also means that a change in the choice of extrema will result in limiting the time-scale over which the sifting process allows an oscillatory mode in the signal to pass un-decomposed.

EMD-MPS exploits this idea of selective extrema selection. In the EMD-MPS method, the sifting process uses intelligent peak selection in short-time windows of length  $\tau$ , where  $\tau$  represents a length of time from within which only the extrema with the maximum/minimum value are selected. Based on different values of  $\tau$ , different decompositions of a signal into modes termed as  $\tau$ -functions are possible. Therefore the short-time window acts as an operator which allows separation of different frequency components in a signal into  $\tau$ -functions, as determined by the length  $\tau$  of the short-time window. A relation between the decomposed frequency components and the value of  $\tau$  will be established and empirically verified.

In the next sections, first the EMD-MPS method will be described, and then the behaviour of EMD-MPS analyzed by applying it to decompose fractional Gaussian noise (fGn) with different values of the Hurst exponent. The spectral analysis of the obtained  $\tau$ -functions will be used to establish decomposition properties of EMD-MPS, as has been done previously in the case of EMD [37][99]. Importantly, the relationship between EMD and EMD-MPS will be pointed out, and it will be shown how EMD-MPS provides novel insight into EMD-based decomposition.

### 3.3.1 EMD-MPS Method

The sifting process in EMD-MPS uses a criterion for choosing the extrema based on short-time windows of length  $\tau$ , instead of using a time-scale based on successive extrema, as is done in the case of EMD. Let an operator  $W_i^\tau(\cdot)$  be defined,  $W_i^\tau(\cdot)$ ,  $i = 1...k$ ,  $i \in \mathbb{Z}$ ,

$0 < \tau < L$ ,  $L \in \mathbb{R}$ , which, given a signal  $x[n]$  of length  $L$ , produces the  $i$ -th  $\tau$ -function  $T_i$ , such that  $T_i[n] = W_i^\tau(x[n])$ . This happens as defined in Algorithm 2, and can be explained as follows:

1. For a given signal  $x[n]$ , a short-time window length denoted by  $\tau$  is chosen, and for each interval  $\tau$  over the whole signal length, the highest/lowest from among the maxima and minima within  $\tau$  are selected.
2. The maxima/minima thus identified (one maxima and minima each per  $\tau$ ) are connected using cubic splines to form the upper and lower envelopes.
3. The mean  $E_{n(mean)}$  of the upper and lower envelopes is calculated, and  $x[n]$  is updated by subtracting the mean from it  $x[n] \leftarrow x[n] - E_{n(mean)}$ .
4. These steps, which are similar to EMD-based sifting, are continued till a stopping criterion is met, at which point  $x[n]$  is reduced to the  $\tau$ -function  $T_1$ .
5. This  $\tau$ -function is subtracted from  $x[n]$  to get a residue, which is then taken as the starting point instead of  $x[n]$ , and previous steps of the algorithm are repeated to find all the  $\tau$ -functions  $T_i$  in the signal.

Unlike IMFs extracted by the EMD algorithm, the coarse-grained  $\tau$ -functions may contain different coexisting modes of oscillation, each superimposed on the other. This happens because the choice of extrema based on short-time window length  $\tau$  limits the scale over which the sifting process allows frequency component(s) to pass, resulting in  $\tau$  setting an upper limit on the periods of the oscillations that can be included in any given  $\tau$ -function obtained using the EMD-MPS method. This limit is determined by :

$$F = \frac{F_s}{\tau} \quad (3.1)$$

where  $F_s$  represents the sampling frequency. As an example for this relation, a value of  $\tau = 25$  (in samples) corresponds to a frequency value  $F = 40$  samples/second for  $F_s = 1000$

---

**Algorithm 2** EMD-MPS Algorithm

---

- 1: Choose a short-time window  $\tau$ .
  - 2: For each interval  $\tau$  over the whole signal length, identify the highest/lowest from among the maxima/minima within  $\tau$ .
  - 3: Find the upper and lower envelopes  $E_{n(U)}$  and  $E_{n(L)}$  by interpolating all maxima/minima identified (one maxima and minima each per  $\tau$ ).
  - 4: Calculate the local mean of the upper and lower envelopes  $E_{n(mean)} = \frac{E_{n(U)} + E_{n(L)}}{2}$ .
  - 5: Update  $x[n]$  by subtracting the mean from it  $x[n] \leftarrow x[n] - E_{n(mean)}$ . Continue the previous steps till a stopping criterion is met, at which point  $x[n]$  is reduced to a  $\tau$ -function  $T_1$ .
  - 6: Subtract  $T_1$  from  $x[n]$  to get a residue  $r[n]$ .
  - 7: Take  $r[n]$  as the starting point instead of  $x[n]$ , and repeat the previous steps of the algorithm till all  $\tau$ -functions  $T_i$  in the signal are found.
- 

samples/second. Using this value of  $\tau$ , only one peak (maxima and minima each) in each 25 sample interval will be used in the envelope formation, and the sifting process should then decompose all  $F \leq 40$  samples/second oscillatory components, and let all components with  $F > 40$  samples/second pass through un-decomposed in one  $\tau$ -function. In practice, the value of  $\tau$  is qualified by a scaling constant  $k$ , such that  $\hat{\tau} = k\tau$ , and  $0 < k \leq 1$ . The relation in Equation 3.1 and the scaling constant  $k$  are empirically validated in Sections 3.3.2.4 and 3.3.2.5.

Assuming Nyquist sampling, it can also be seen that the minimum value of  $\tau$  is given by  $\tau_{min} = \frac{F_s}{F_{max}}$ , where  $F_{max} = \frac{F_s}{2}$ . This gives the minimum value of  $\tau$  as  $\tau_{min} = 2$ . At this value of  $\tau$ , all the maxima and minima present in the signal are selected, and EMD-MPS is equivalent to the EMD algorithm. This means that all frequency components present in the signal are decomposed, and the  $\tau$ -functions are the same as IMFs. This relation of EMD-MPS with EMD allows shedding new light on the decomposition behaviour of EMD in terms of EMD-MPS, which is done in later sections. Similarly,  $\tau_{max}$  represents the maximum possible value of  $\tau$ , at which value the signal remains un-decomposed. Therefore the range

of possible values of  $\tau$  can be written as:

$$2 \leq \tau < \frac{F_s}{F_{min}} \quad (3.2)$$

where  $F_{min}$  is the lowest frequency present in the signal to be decomposed. Practically, however, the maximum value of  $\tau$  will be limited by the length of the signal (see also Section 3.3.2.3 for relation to EMD algorithm termination).

### 3.3.2 Important Properties of EMD-MPS

In this Section, some important properties of EMD-MPS will be established and discussed. As with EMD, the properties of EMD-MPS will be established empirically using numerical means. The main tool here will be the decomposition of fractional Gaussian noise using EMD-MPS. Fractional Gaussian noise (fGn) represents a versatile model for a full-spectrum process which is not dominated by any particular frequency band, and has been previously used in studies to establish properties of EMD [99]. The fGn of parameter  $H$ , which is the Hurst exponent, can be defined as the zero-mean stationary Gaussian process with autocorrelation sequence  $r_H[k] := E\{x_H[n]x_H[n+k]\}$ ,

$$r_H[k] = \frac{\sigma^2}{2}(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}) \quad (3.3)$$

For  $H = 0.5$ , fGn reduces to discrete white noise, whereas other values of  $H$  correspond to non-zero correlations (negative for:  $0 < H < 0.5$ ; positive for:  $0.5 < H < 1$ ).

For this work, extensive experiments were carried out with decomposition of fGn processes using EMD-MPS with different values of  $\tau$  in the operator  $W_i^\tau(\cdot)$ . For each of the different values of  $H$  used (0.2, 0.5, 0.8, and also 0.1 and 0.9, but for fewer values of  $\tau$ ), 2500 independent sample paths of fGn were generated, each of length 2048 samples. EMD-MPS was applied to each of the 2500 sample paths, using the following values of  $\tau$  and  $i$  in the operator  $W_i^\tau(\cdot)$ :  $i = 1, 2, 3, 4, 5$ , and  $\tau = 1 - 50$ . This means that for each value of  $\tau$  for each sample path, five  $\tau$ -functions ( $T_1$  to  $T_5$ ) were obtained. The number of  $\tau$ -functions was fixed



to have a uniform reference for different values of  $\tau$  for each path of fGn, and also because for larger values of  $\tau$ , the higher index  $\tau$ -functions have insignificant temporal or spectral content.

### 3.3.2.1 Filter bank behavior of EMD-MPS

Filter banks represent a collection of bandpass filters which can isolate different frequency bands in the input signal. Spectral analysis of IMFs obtained by application of EMD on white noise or fGn has shown frequency responses similar to that of a dyadic filter bank, as discussed later in Appendix A. In order to test the behaviour of  $\tau$ -functions  $T_i$  obtained by decomposing fGn with EMD-MPS, using different values of  $\tau$  in the operator  $W_i^\tau(\cdot)$ , the power spectrum of each  $\tau$ -function ( $T_1$  to  $T_5$ ) was estimated by computing its autocorrelation function for each fGn sequence, which was then ensemble averaged over all sequences and then Fourier transformed. Figure 3.2 shows the spectra of the  $\tau$ -functions for three different values of  $\tau$  for fGn sequences with  $H = 0.5$ .

It can be observed from Figure 3.2 that when  $\tau$ -functions display a band-pass behavior, the pass-bands overlap in a way such that the lower half-band of  $\tau$ -function  $T_i$  represents a frequency range which is roughly covered by the upper half-band of  $\tau$ -function  $T_{i+1}$ , indicating a quasi-dyadic filter-bank structure. In the case of EMD, the dyadic filter-bank structure for IMFs has previously been quantified by establishing an exponentially decreasing relationship between the number of zero-crossings in IMFs (as an indication of the mean frequency in an IMF) and the IMF index number [99][33]. Since  $\tau$ -functions may contain multiple modes of oscillation, the number of zero-crossings is not a good estimate of the mean frequency. Instead, the center-frequency  $F_i$  of the  $\tau$ -functions  $T_i$  with band-pass frequency response was used, and it was found that  $F_i$  is a decreasing exponential of the  $\tau$ -function index  $i$  related by:

$$F_i \propto \beta^{-i} \quad (3.4)$$

where  $\beta$  is very close to 2. It is interesting to note that the value of  $\beta$  remains close to 2 even as  $\tau$  increases, i.e. the dyadic filter-bank behaviour is demonstrated as long as  $\tau$ -functions

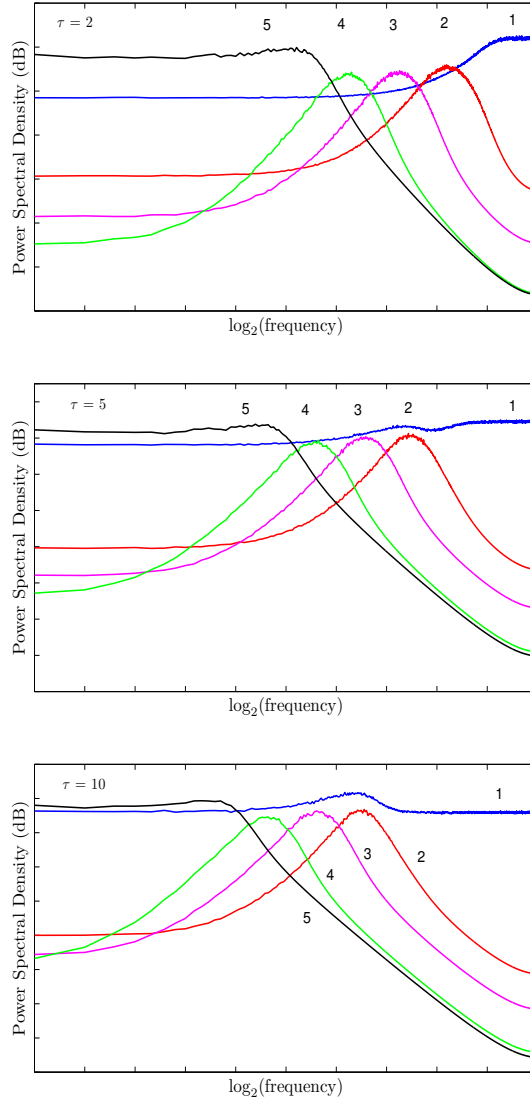


Figure 3.2: Mean spectra (power spectral density in  $dB$  on y-axis, and  $\log_2(\text{frequency})$  on x-axis) of  $\tau$ -functions  $T_1$  to  $T_5$ , obtained using three different values of  $\tau$  for 2500 fGn sequences with value of  $H = 0.5$ . Unlabeled numbers in the figures represent the indices of  $\tau$ -functions.

have band-pass like characteristics. This is illustrated in Figure 3.3, which shows, for the three values of  $H$  used, the values of  $\log_2(F_i)$  plotted against the  $\tau$ -function index  $i$ . As expected by Equation 3.4, the slopes of the least-square fits to the points are very close to -1 even as the value of  $\tau$  increases.

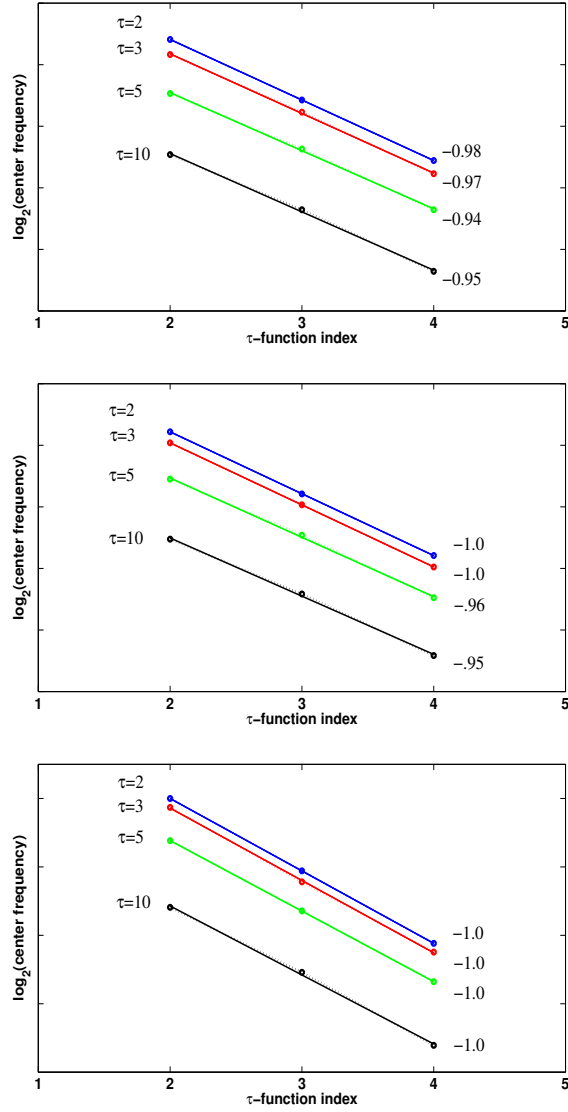


Figure 3.3: Center frequency of  $\tau$ -functions  $T_2$ ,  $T_3$ ,  $T_4$  obtained using different values of  $\tau$  plotted against the  $\tau$ -function index  $i$ , using different values of  $H$  (top:  $H = 0.2$ , center:  $H = 0.5$ , bottom:  $H = 0.8$ ). The slopes of the least-square fits and the values of  $\tau$  are indicated on the plots.

Self-similarity of the spectra of band-pass IMFs is an important property of EMD filter-banks [99]. For  $\tau$ -functions with band-pass frequency response, approximate self-similarity of spectra holds as well. For  $\tau = 2$ , where EMD-MPS is equivalent to EMD, self-similarity

holds according to the following relation:

$$S_i(f) = CS_j(f)(\beta^{(i-j)} f) \quad (3.5)$$

where  $S_i(f)$  represents the power spectrum of the  $i$ -th  $\tau$ -function  $T_i$  with band-pass spectrum,  $i > j \geq 2$ ,  $\beta \approx 2$ , and  $C = \beta^\alpha$ , with  $\alpha = 2(2H - 1)$ . When normalized according to Equation 3.5, the power spectra of the  $\tau$ -functions obtained using EMD-MPS collapse to a single curve, as shown in Figure 3.4. However, the value of  $C$  in Equation 3.5 holds only for values of  $\tau$  close to 2. As the value of  $\tau$  increases, the value of  $\alpha \rightarrow 0$ , leading to  $C \rightarrow 1$ . This change with increasing values of  $\tau$  is not uniform for different values of  $H$ , as is illustrated for  $H = 0.2$  in Figure 3.4, where a lower value of  $C$  given by  $C = \beta^{\frac{\alpha}{2}}$  leads to  $\tau$ -functions collapsing onto a single curve for  $\tau = 5$ .

One other aspect of EMD-MPS that may be noted from the filter-bank behaviour is that the spectrum of  $\tau$ -functions starts changing from band-pass to low-pass as  $\tau$  increases. For sufficiently large value of  $\tau$ , all  $\tau$ -functions  $T_i$ ,  $i > 1$  demonstrate low-pass behaviour. This is to be expected, since as the value of  $\tau$  increases, most of the frequency components pass un-decomposed in the first  $\tau$ -function, and only low frequency components present in the fGn sequence are decomposed into  $\tau$ -functions with index  $i > 1$ . Similarly, for a large enough increase in  $\tau$ ,  $\tau$ -function  $T_1$  assumes approximately all-pass characteristics. Additionally, the behaviour of  $\tau$ -function  $T_1$  is different from other  $\tau$ -functions. For lower values of  $\tau$ ,  $T_1$  displays high-pass behaviour, as in the case of EMD (corresponding to  $\tau_{min} = 2$ ). However, as  $\tau$  is increased,  $T_1$  displays band-pass characteristics, whereas for large values of  $\tau$ ,  $T_1$  assumes characteristics akin to low-pass behaviour. It can however be noticed from Figure 3.2 that for  $\tau$ -function  $T_1$ , the attenuation in the "stop-band" is much less than for other  $\tau$ -functions (except when  $T_3$ ,  $T_4$  and  $T_5$  assume low-pass characteristics with increasing values of  $\tau$ ). Importantly, though,  $T_1$  and  $T_2$  share approximately the same frequency bands but with different levels of attenuation. This can explain the efficacy of EMD-MPS in de-noising of signals, as described in [100]. When  $T_2$  assumes low-pass behaviour for large values of  $\tau$ ,  $T_1$  displays approximately all-pass characteristics.

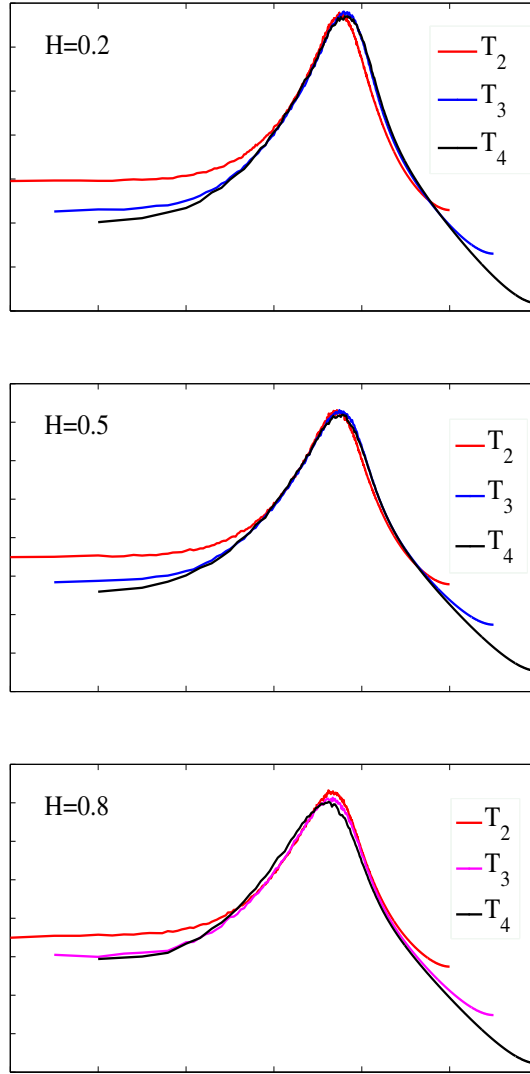


Figure 3.4: Normalized power spectra (with power spectral density in  $dB$  on y-axis, and  $\log_2(frequency)$  on x-axis) of 3  $\tau$ -functions according to Equation 3.5. The  $\tau$ -functions were obtained using a value of  $\tau = 5$ . The spectra for  $H = 0.2$  were obtained using  $C = \beta^{\alpha/2}$  in Equation 3.5.

### 3.3.2.2 Horizontal and vertical behavior

Given that the decomposition of EMD-MPS and EMD is driven by the sifting process, the similarity in behaviour is not surprising. However, according to Equation 3.1, we can relate the dyadic behaviour across modes ( $\tau$ -functions or IMFs) to a change in the value of  $\tau$ : a

change in the value of frequency by half implies an increase in value of  $\tau$  by a factor of 2. This means that after the extraction of the first mode (whether IMF or  $\tau$ -function), the sifting process behaves as if the value of  $\tau$  has been doubled. To test this, the value of  $\tau$  was increased by a factor of 2 after each  $\tau$ -function had been extracted, with this change resulting in no difference in the spectra of the  $\tau$ -functions. However, increasing the value of  $\tau$  by a factor of 4 after each  $\tau$ -function extraction resulted in decrease of the spectra to half the previous value.

This suggests that the "horizontal" (i.e. across modes) dyadic filter-bank behaviour also exists "vertically" (i.e. for the same mode, but for specific values of  $\tau$ ) for EMD-MPS, in that the same  $\tau$ -function  $T_n^\tau$  displays a dyadic behaviour for different values  $\tau_j$ , where  $\tau_{j+1} = 2\tau_j$ . As for the horizontal case, the vertical behaviour can be quantified as:

$$F_n^\tau \propto \beta^{-\tau_j} \quad (3.6)$$

where  $\beta$  is very close to 2. This is shown in Figure 3.5, which demonstrates the vertical behaviour for values of  $\tau_j$  starting from 2 (upper plot) and 4 (lower plot) using  $F_2^\tau$ . The solid lines in both plots are the least-square fits with a slope nearly equal to -1.

At the same time, Figure 3.5 reveals a very interesting phenomenon for values of  $\tau$  starting from 2 (upper plot), which corresponds to normal EMD behaviour. There is a very clear deviation of the frequency value  $F_2^2$ , indicating a lower value than expected. This implies that for EMD decomposition, the oscillatory components in the first mode have higher frequencies than expected by the vertical model, and going from mode 1 to 2 is not the doubling of  $\tau$ , but instead amounts to a  $\tau_2$  which is almost three times  $\tau_1$ ,  $\tau_2 \approx 3\tau_1$ , (slight difference in values for different values of  $H$  is not discussed here). According to Equation 3.1, the frequencies to be separated and the value of  $\tau$  are inversely proportional, hence for good separation of components with frequencies  $f_1$  and  $f_2$ ,  $f_2 > f_1$ , the frequency ratio  $f = \frac{f_1}{f_2}$  should have a value  $f \approx \frac{1}{3}$ . Interestingly, the separation limit established for two-tone separation in [101] is given by  $f \approx \frac{2}{3}$ , above which the two components cannot be separated by EMD regardless of the amplitude ratio of the components. Using the analysis presented here, it can be said

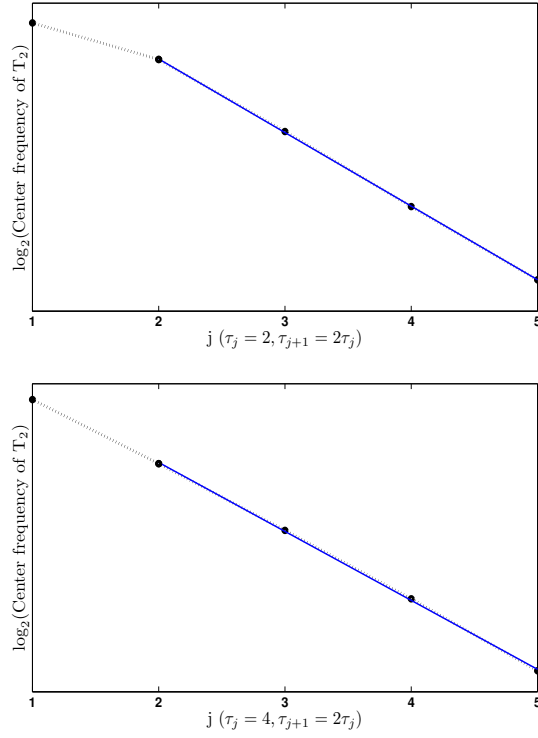


Figure 3.5: Center frequency of  $\tau$ -function  $T_2$  ( $\log_2$  values) obtained for different values of  $\tau$ , plotted against the index  $j$ .

that the separation limit for EMD decomposition of full spectrum random noise sequences is less than that for the two-tone model, and for all practical purposes, EMD can separate components according to a separation limit  $f_{sep} \ll \frac{2}{3}$ .

Similarly, the difference in behaviour of the first IMF has been mentioned previously in [102], but formulation of this behaviour in terms of the vertical behaviour presented by EMD-MPS, as done in the last paragraph, can provide better insight into the decomposition properties of EMD, and relate it to previous results as well. As an example, the band-pass to low-pass change in IMF behaviour [99] can be explained keeping in view IMF extraction in terms of an increase in  $\tau$  as mentioned at the end of Section 3.3.2.1.

Furthermore, the vertical behaviour can be used to demonstrate and explain the evolution of the probability density function (PDF) of the extracted modes, which has been shown to be bi-modal for the first IMF, and Gaussian-like for other IMFs in the case of EMD

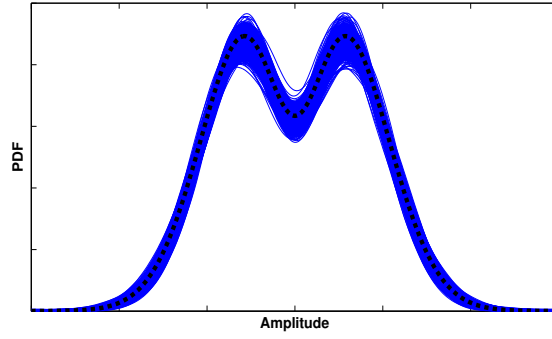
decomposition of white noise and fractional Gaussian noise [37][29]. The bi-modal nature of the first IMF is justified for the white noise case ( $H = 0.5$ ) in [29] as follows: the PDFs of Gaussian white noise maxima and minima are labeled by  $p_+(x)$  and  $p_-(x)$ . Then, since the first IMF has high-pass characteristics, and by construction EMD only maintains positive maxima and negative minima, the PDF of the first IMF can be expressed as  $p_*(x) := \max(p_-(x), p_+(x))$ , which reproduces the main structure of the bimodal PDF of the first IMF.

None of the previous works have studied the evolution of the PDF from bi-modal for IMF 1 to Gaussian-like for IMF 2 and above. However, it is possible to observe the evolution of the PDF of the extracted modes using the vertical behaviour in the context of EMD-MPS. This is shown in Figure 3.6, which shows the PDF of three  $\tau$ -functions  $T_i^\tau$ , where  $i$  is the index of the  $\tau$ -functions, and  $\tau$  is the value used to obtain the  $\tau$ -function  $T_i$ . The  $\tau$ -functions have been obtained from 2500 fractional Gaussian noise sequences with a value of the Hurst exponent  $H = 0.5$ , as described previously in Section 3.3.2. The PDF of each  $\tau$ -function was estimated using a Gaussian smoothing function estimate using 500 equally spaced points covering the whole range of data of the  $\tau$ -functions. The dashed lines in Figure 3.6 represent the average PDF of all the individual PDFs.

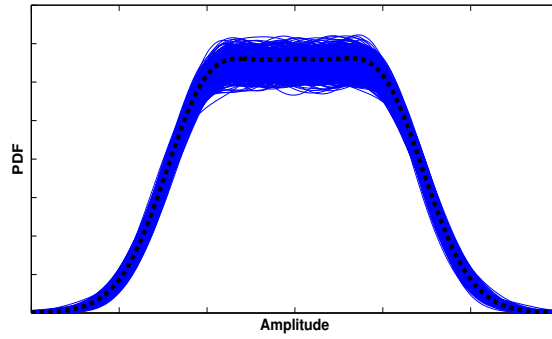
The bi-modal nature of the PDF of the  $\tau$ -function  $T_1^2$  is evident in Figure 3.6 (a). The  $\tau$ -function  $T_1^2$  has been obtained after application of EMD-MPS to fractional Gaussian noise sequences using a value of  $\tau = 2$ . At this value of  $\tau$ , EMD-MPS is the same as EMD, as mentioned in Section 3.3.1, and  $T_1^2$  is equivalent to IMF 1. Similarly, Figure 3.6 (c) shows the PDF of the  $\tau$ -function  $T_2^2$ , which is the second  $\tau$ -function extracted using a value of  $\tau = 2$ , and is therefore equivalent to IMF 2. It can be seen from Figure 3.6 (c) that the average PDF, shown in bold red line, appears approximately Gaussian, as expected based on previous results [37][29].

Interestingly, the PDF of  $\tau$ -function  $T_1^5$  shown in Figure 3.6 (b) has a curve shape intermediate between bi-modal and Gaussian-like. This intermediate curve shape represents the evolution of the PDF across modes, as explained by the vertical behaviour in the context of

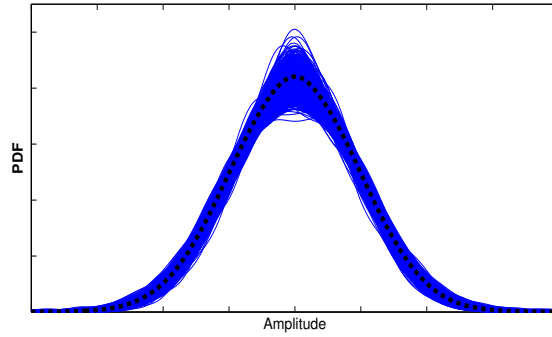




(a) PDF of  $\tau$ -function  $T_1^2$



(b) PDF of  $\tau$ -function  $T_1^5$



(c) PDF of  $\tau$ -function  $T_2^2$

Figure 3.6: Probability density function (PDF) estimates of 2500 realizations of three  $\tau$ -functions of fractional Gaussian noise with  $H = 0.5$  obtained using a Gaussian smoothing function estimate. The dashed lines represent the average PDF of all the PDFs of individual  $\tau$ -functions.

EMD-MPS. The  $\tau$ -function  $T_1^5$  represents the first mode extracted from the noise sequences using a value of  $\tau = 5$ . This allows us to use the vertical behaviour (for the same mode  $n$  represented as  $T_n^\tau$ , but for different values of  $\tau$ ) to explain the horizontal behaviour (across modes  $i$  represented as  $T_i^{\tau_n}$ , but for the same  $\tau$  value given by  $\tau_n$ ), since the extraction of modes represents an increase in the value of  $\tau$ , as discussed earlier in this Section. Therefore it can be seen that the change from the bi-modal PDF to a Gaussian-line PDF is gradual, and goes through intermediate stages. This can be explained in terms of the gradual change of the filtering behaviour of  $\tau$ -functions from band-pass to low-pass as  $\tau$  increases.

The work in [103] presents a discussion on the PDF of IMFs obtained by application of EMD on Gaussian white noise, and also considers the influence of data length on the PDF. One finding of the study in [103] relates to the shape of the PDF of IMFs with index  $i > 1$  becoming bi-modal and even multi-modal as the length of the data increases. Again using the vertical behaviour in the context of EMD-MPS, the evolution of the PDF due to an increase in the signal length can be demonstrated using the PDF of  $\tau$ -function  $T_1^5$ . This  $\tau$ -function represents the first mode extracted from the noise sequences using a value of  $\tau = 5$ , and the PDF is shown in Figure 3.7. It can be seen from Figure 3.7 that the PDF curve starts changing from a flat shape (compare to Figure 3.6 (b)) to multi-modal as the length of the noise sequence being decomposed increases.

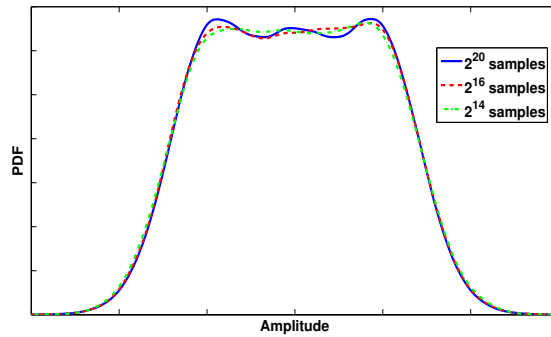


Figure 3.7: Probability density function (PDF)  $\tau$ -function  $T_1^5$  for different lengths of the fractional Gaussian noise signal with  $H = 0.5$ . As the length of the noise sequence increases, the PDF curve starts changing to multi-modal.

Another aspect of the study reported in [103] relates to testing the null hypothesis that the PDFs obtained from IMFs have a Gaussian distribution. This has been done using the Jarque-Bera and Lilliefors normality tests in [103], and the null hypothesis was rejected for all except a few IMFs obtained by using different parameters (sifting number, signal length) for the decomposition. However, in this work, the application of the same two normality tests to all the  $\tau$ -functions obtained using the different values of  $\tau$  led to a rejection of the null hypothesis, even for the cases where the shape of the PDF curve was Gaussian-like.

### 3.3.2.3 Convergence of EMD algorithm

Since EMD does not admit an analytic formulation, it is not possible to verify the properties of the algorithm mathematically. Among these properties is the convergence of the algorithm [31]. However, convergence of the EMD algorithm is accepted based on the empirical observation that the IMF  $c_{i+1}$  has strictly fewer local extrema than IMF  $c_i$ , as pointed out in [7], among others.

Interestingly, EMD-MPS can be used to explain the convergence of the EMD algorithm in terms of an increase in  $\tau$ . Previous studies have established the limit for the number of IMFs that can result from applying EMD to a signal as  $J \leq \log_2 N$ , where  $J$  is the number of IMFs obtained from a signal of length  $N$  [104]. In terms of EMD-MPS behaviour, we can formulate the extraction of an IMF as the doubling of  $\tau$ . Therefore, for the  $\lceil \log_2 N \rceil$ -th IMF, the value of  $\tau \geq N$ , which means no further decomposition is possible, since at most one minima and maxima each will be selected, and the algorithm will terminate as no envelope formation is possible. In practice, the number of IMFs can be less than  $\log_2 N$ , because, depending on the signal, the number of extrema can drop to less than 3 before  $\tau \leq N$ , which is when most implementations of EMD are programmed to terminate the algorithm.

The concept in the preceding paragraph is illustrated in Table 3.1 for a signal with length  $N=1024$  samples, which shows the doubling of  $\tau$  after each IMF is extracted, where the first IMF is extracted when  $\tau = 2$ , which corresponds to normal EMD, as previously explained in Section 3.3.1. For a signal length  $N = 1024$  samples, the number of IMFs  $J$  expected after

Table 3.1: Illustration of the convergence of EMD algorithm due to doubling of  $\tau$  after each IMF extraction for a signal with length  $N = 1024$  samples

IMF No.	$\tau$
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024

application of EMD to the signal is given by  $J \leq \log_2 N$ , which for this case leads to  $J \leq 10$ . It can be seen from Table 3.1 that for the  $\lceil \log_2 N \rceil$ -th IMF, which is IMF 10, the value of  $\tau = N$ . At this value of  $\tau$ , no further decomposition is possible, hence the algorithm will terminate.

#### 3.3.2.4 Relation between decomposed frequencies and $\tau$

To study the relationship between the decomposed frequencies and  $\tau$ , the center-frequency  $F_2$  of the second  $\tau$ -function  $T_2$  is used, since  $T_2$  maintains a band-pass behaviour over a large range of  $\tau$  for the data length used in these experiments. Figure 3.8, which plots  $F_2$  against  $\tau$  for 3 values of  $H$  used in this experiment, demonstrates the existence of a non-linear relationship between center-frequency  $F_i$  and  $\tau$ , and can also be seen as characterizing the non-linear nature of the extrema-based decomposition. It can be observed from Figure 3.8 that the change in frequency per unit change in  $\tau$  decreases as the value of  $\tau$  increases. As the value of  $\tau$  increases, the number of extrema used in the decomposition process (Section 3.3.1) decreases, hence leading to smaller change in the frequency content of the  $\tau$ -functions obtained. Also, for lower values of  $\tau$ , Figure 3.8 shows difference in the values of the center-

frequency  $F_2$  for different values of  $H$ , which corresponds to the properties of fGn with these values of  $H$ . However, this difference in the frequency values decreases as the value of  $\tau$  increases, which can be seen as the consequence of selecting fewer and fewer peaks as  $\tau$  becomes large, thereby reducing any correlations present in the noise sequences.

It is possible to empirically validate Equation 3.1 by using the following relation for a  $\tau$ -function  $T_i$  having band-pass frequency response with center frequency  $F$ :

$$\frac{\tau_n}{\tau_m} = x \left( \frac{F_n}{F_m} \right)^{-1}, \quad \forall n, m \in \mathbb{Z}, n \neq m \quad (3.7)$$

Equation 3.1 suggests a value of  $x = 1$  (since  $F_s = \tau F$ ), however  $x \approx 1$  in Equation 3.7 holds only for values of  $\tau > 6$  (for all values of  $H$ ), whereas  $0.7 \leq x \leq 1$  for  $2 \leq \tau \leq 6$ , with  $x$  approaching 1 as  $\tau$  increases to a value of  $\tau = 6$ . The reason for this behaviour is related to the different vertical behaviour inherent in the decomposition using a value of  $\tau = 2$ , as mentioned in Section 3.3.2.2.

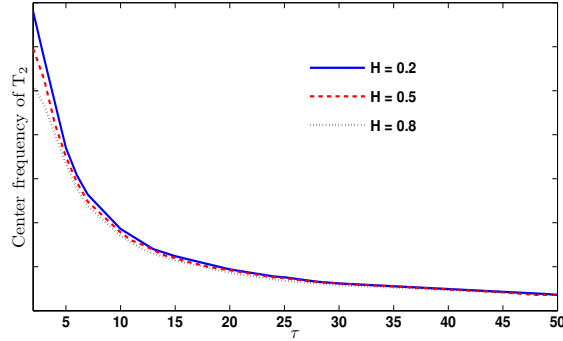


Figure 3.8: Relationship between decomposed frequencies and  $\tau$  for different values of  $H$ .

### 3.3.2.5 Scaling factor for $\tau$

It was previously mentioned in Section 3.3.1 that the value of  $\tau$  in Equation 3.1 is qualified by a scaling factor  $k$ . The scaling factor  $k$  was first estimated as  $0.5 < k < 0.63$ , based on experiments with different test signals and different values of  $\tau$  [100]. Although this range of  $k$  held well for simulated and real-world signals, the work in [100] reported a value

lower than  $k = 0.5$  giving much better results in terms of expected decomposition into  $\tau$ -functions. Using fGn analysis as described here, a better estimation of the scaling factor for  $\tau$  is available, which is now written as (valid for all values of  $H$ ):

$$0.25 < k < 0.44 \quad (3.8)$$

The range of the values of  $k$  was found by estimating the center frequency of  $\tau$ -function  $T_2$  resulting from a value of  $\hat{\tau}$  used for decomposition, and relating it to the value of  $\tau$  by  $\hat{\tau} = k\tau$ , where  $\tau$  is the value which would have resulted in the same frequency value according to Equation 3.1.

Importantly, the value of  $k$  is related to the difference in behaviour of  $\tau$  in the same way as previously mentioned, with the value of  $k = 0.44$  valid for values of  $\tau > 6$ , and  $0.25 \leq k < 0.44$  for  $2 \leq \tau < 6$ , with  $k$  increasing exponentially to 0.44 as  $\tau$  increases to a value of 6. The estimation of the scaling constant  $k$  also allows us to relate the change in frequency  $\Delta f$  to the change in  $\tau$  between values  $\tau_1$  and  $\tau_2$  as:

$$\Delta f = F_s \left( \frac{k_1}{\hat{\tau}_1} - \frac{k_2}{\hat{\tau}_2} \right) \quad (3.9)$$

For value of  $\tau > 6$ ,  $k_1 = k_2 = 0.44$ , whereas for other values of  $\tau$  the values of  $k_1$  and  $k_2$  depend on the value of  $\tau$  being used.

The scaling factor  $k$  has a lower value for lower values of  $\tau$ , and higher for higher values of  $\tau$ . This indicates that in order to separate higher frequency components from a signal using EMD-MPS, it is practical to use a value of  $\tau$  scaled by the lower limit in Equation 3.8. In simple terms, this can be explained as requiring more peaks to accurately separate higher frequency components.

### 3.3.2.6 Validation of the scaling factor $k$ using two-tone separation

The two tone separation problem using EMD has been studied in [101] using a performance measure to quantify the quality of separation of the two tones using the following signal

model:

$$x(t; f) = \cos 2\pi t + \cos(2\pi ft), t \in \mathbb{R}. \quad (3.10)$$

To validate the scaling factor described in the previous Section, EMD-MPS is used to test separation of two tones, using the same signal model, and the performance measure given by Equation 3.11. This performance measure gives a value close to zero for good separation, and a value approaching 1 in case of poor separation between components [101], as the high frequency tone (frequency  $f_1$ ) is expected to pass un-decomposed into  $\tau$ -function  $T_1$ , and the low-frequency tone (frequency  $f_2$ ) to be decomposed into  $\tau$ -function  $T_2$ .

$$c(f) = \frac{||T_1 - \cos 2\pi t||_2}{||\cos 2\pi ft||_2} \quad (3.11)$$

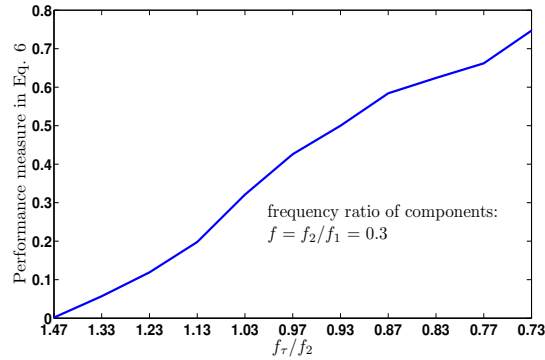


Figure 3.9: Value of the performance measure in Equation 3.11 plotted against  $f_\tau/f_2$ .

Figure 3.9 shows the results of using EMD-MPS to decompose the signal  $x(t)$ , where the frequency ratio  $f$  of the two components has a value 0.3. At this value of  $f$ , the performance measure  $c(f)$  should have a value close to zero [101]. The performance measure is plotted against  $f_\tau/f_2$ , where  $f_\tau = \frac{f_s}{\hat{\tau}} \cdot k$ , obtained using Equation 3.1 and the scaling relation  $\hat{\tau} = k\tau$ . Here  $\hat{\tau}$  represents the actual value of  $\tau$  used to decompose the signal, and  $k = 0.44$ . EMD-MPS does not decompose a frequency component with value greater than  $f_\tau$ , which passes un-decomposed in  $\tau$ -function  $T_1$ .

For this experiment, a value of the ratio  $f_\tau/f_2 \gg 1$  represents the case where  $f_\tau$  is greater than  $f_2$ , which means the tone with higher frequency  $f_1$  passes un-decomposed in  $\tau$ -function

$T_1$ , whereas the tone with frequency  $f_2 \ll f_\tau$  is decomposed into  $\tau$ -function  $T_2$ . However, the ratio  $f_\tau/f_2$  decreasing towards 1 and lower represents  $f_\tau$  becoming close to, and then less than  $f_2$ . In this case, a clean separation of the tones is not possible. This is depicted in Figure 3.9, where, for values of the ratio  $f_\tau/f_2$  close to and less than 1, the value of performance measure increases towards 1, showing an increasingly less effective separation of the tones.

### 3.3.2.7 Effect of modified peak selection on computational cost

The number of extrema used for the envelope formation in the sifting process of EMD-MPS depends on the value of  $\tau$  used for decomposition, as explained previously in Section 3.3.1. Using a fewer number of extrema for envelope formation, instead of all the possible extrema, also has implications for the computational cost of the sifting process, as discussed in this Section.

In the case of EMD, estimation of the computational cost of the algorithm depends on the particular implementation of the algorithm. Here we use the estimation of the computational cost of one sifting iteration in terms of the number of multiplications and divisions (number of flops) as given in [33]. Let  $F_{EMD}(c_i, k+1)$  be the number of flops for one sifting iteration required to obtain the  $i$ -th IMF  $c_{i,k+1}$  from  $c_{i,k}$ , where  $k$  denotes the number of sifting operations. Then  $F(c_i, k+1)$  is given by the following expression [33]:

$$F_{EMD}(c_i, k+1) = 18L + 15M_{EMD}(i, k) \quad (3.12)$$

where  $L$  represents the length of the signal being decomposed, and  $M_{EMD}$  is the number of extrema.

For EMD-MPS, the number of extrema  $M_{EMD-MPS}$  will depend on the the value of  $\tau$  used for decomposition. Given that only one maxima/minima is selected per  $\tau$ ,  $M_{EMD-MPS}$  will be given by:

$$M_{EMD-MPS} \leq \frac{L}{\tau} \quad (3.13)$$



It should be noted that Equation 3.13 has to be qualified with the value of  $\tau$  given by  $\tau > 2$ , since  $M_{EMD} = M_{EMD-MPS}$  when  $\tau = 2$ .

Using fewer extrema for envelope formulation, as is the case with EMD-MPS, leads to a decrease in the computational cost of the algorithm, with the number of flops for one sifting operation required to obtain the  $i$ -th  $\tau$ -function  $T_{i,k+1}$  from  $T_{i,k}$  given by the following expression:

$$F_{EMD-MPS}(T_i, k+1) \leq 18L + 15\left(\frac{L}{\tau}\right) \quad (3.14)$$

The total computational cost of the EMD-MPS method can be estimated by summing  $F_{EMD-MPS}(T_i, k)$  over both the number of iterations and the number of  $\tau$ -functions, as is done for the case of EMD [33]. It can be seen from Equation 3.14 that a larger value of  $\tau$  used for the decomposition will lead to a lesser overall computational cost.

### 3.4 Hierarchical Decomposition using Empirical Mode Decomposition-Modified Peak Selection

The ability of the EMD-MPS based method for a time-scale based decomposition allows the development of the concept termed here as hierarchical decomposition. Using hierarchical decomposition, it is possible to decompose the signals into frequency bands, which can be chosen based on application of Equation 3.1. This concept is illustrated in Figure 3.10, which shows decomposition of a signal into a three-level hierarchy, with the first, second and third levels denoted by  $a$ ,  $b$  and  $c$ .

The hierarchy of decomposition shown in Figure 3.10 is determined by selecting the values of  $\tau$  at each level of decomposition. The first decomposition level  $a$  is determined by selecting a value  $\tau_a$ , such that the signal with sampling frequency  $f_s$  is decomposed into two  $\tau$ -functions  $T_1^{\tau_a}$  and  $T_2^{\tau_a}$ . According to the frequency separation criterion given by Equation 3.1, the  $\tau$ -function  $T_1^{\tau_a}$  contains oscillatory components with frequency values  $f_{T_1^{\tau_a}} > f_{\tau_a}$ , where  $f_{\tau_a} = \frac{f_s}{\tau_a}$ . Similarly, the  $\tau$ -function  $T_2^{\tau_a}$  contains frequency components with frequency values  $f_{T_2^{\tau_a}} < f_{\tau_a}$ .

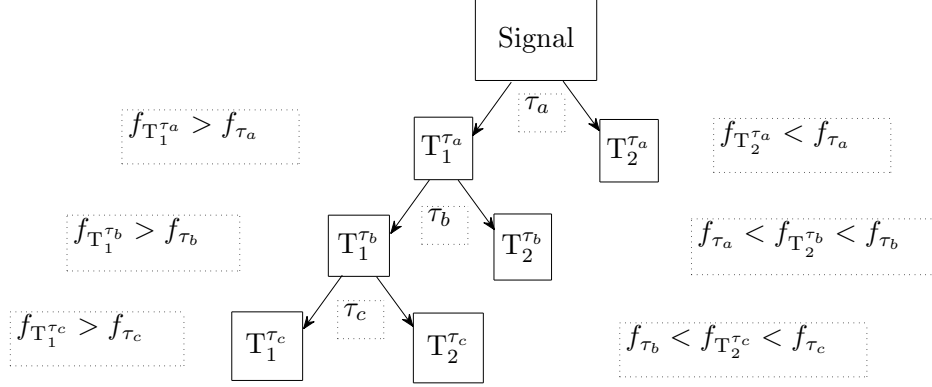


Figure 3.10: Hierarchical decomposition of a signal into a hierarchy of three levels a, b and c. The values of  $\tau$  given by  $\tau_a$ ,  $\tau_b$  and  $\tau_c$  determine the hierarchy of decomposition.

The second level  $b$  of the hierarchical decomposition is then obtained by using EMD-MPS to decompose the  $\tau$ -function  $T_1^{\tau_a}$  with a value of  $\tau$  given by  $\tau_b$ , where  $\tau_b < \tau_a$ . The  $\tau$ -function  $T_1^{\tau_a}$  is decomposed into two  $\tau$ -functions  $T_1^{\tau_b}$  and  $T_2^{\tau_b}$ , with the frequency separation value  $f_{\tau_b}$  determined by the value of  $\tau_b$  as  $f_{\tau_b} = \frac{f_s}{\tau_b}$ . In the second level of decomposition  $b$ , the  $\tau$ -function  $T_1^{\tau_b}$  contains oscillatory components with frequency values  $f_{T_1^{\tau_b}} > f_{\tau_b}$ . Importantly, though, the  $\tau$ -function  $T_2^{\tau_b}$  now contains oscillatory components within a range of frequencies given by  $f_{\tau_a} < f_{T_2^{\tau_b}} < f_{\tau_b}$ .

Similarly, the third level  $c$  of the hierarchical decomposition is then obtained by decomposing the  $\tau$ -function  $T_1^{\tau_b}$  using EMD-MPS with a value of  $\tau$  given by  $\tau_c$ , where  $\tau_c < \tau_b$ . As in the previous levels, the  $\tau$ -function  $T_1^{\tau_b}$  is decomposed into two  $\tau$ -functions, given in this case by  $T_1^{\tau_c}$  and  $T_2^{\tau_c}$ . Here, the  $\tau$ -function  $T_1^{\tau_c}$  contains oscillatory components with frequency values  $f_{T_1^{\tau_c}} > f_{\tau_c}$ , whereas the  $\tau$ -function  $T_2^{\tau_c}$  contains oscillatory components with frequencies in the range  $f_{\tau_b} < f_{T_2^{\tau_c}} < f_{\tau_c}$ .

It is important to note that the hierarchical decomposition may be started from a desired frequency separation level  $f_{\tau_a}$ , where  $\tau_a$  is the value used for EMD-MPS decomposition

obtained using Equation 3.1. This allows decomposition of a signal into desired frequency bands which can be obtained through  $\tau$ -functions  $T_2^{\tau_j}$ , where  $j$  represents a hierarchy level obtained through the use of a value of  $\tau$  given by  $\tau_j$ . The concept of hierarchical decomposition is next illustrated first by using a synthetic signal consisting of three sinusoids, and then by using a real-world EEG signal.

### 3.4.1 Illustration of Hierarchical Decomposition using Synthetic and Real Signals

The concept of hierarchical decomposition using EMD-MPS as described in Section 3.4 is illustrated using two examples in this Section. The first example uses a synthetic signal, whereas the second example uses a real-world EEG signal.

#### 3.4.1.1 Hierarchical decomposition of a synthetic signal

The synthetic signal used to illustrate hierarchical decomposition is shown in Figure 3.11. This signal has a sampling frequency  $f_s = 2048$  cycles/second, and consists of three sinusoids at frequency values 16, 64 and 256 cycles/second.

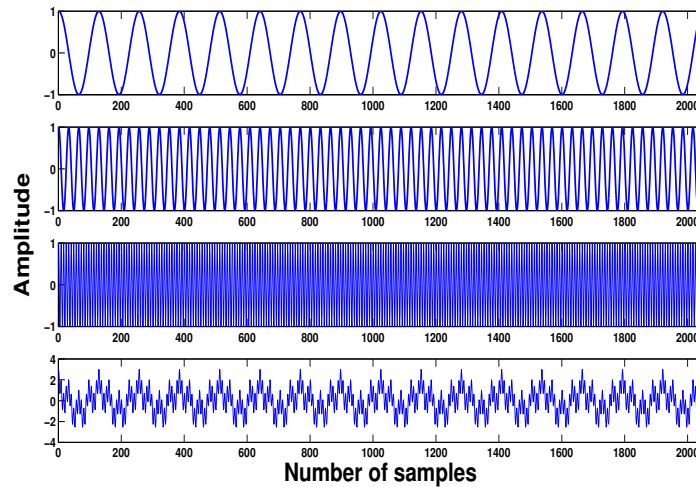


Figure 3.11: Synthetic signal (plot at the bottom) containing 3 sinusoids at frequency values 16, 64 and 256 cycles/second (top three plots).

Table 3.2: Separation of the frequency components of the synthetic signal shown in Figure 3.11 using hierarchical decomposition

Decomposition level	$\tau$ -function	$\tau$	Frequency component (cycles/second)
$a$	$T_1^{\tau_a}$	$\tau_a$	64 and 256
$a$	$T_2^{\tau_a}$	$\tau_a$	16
$b$	$T_1^{\tau_b}$	$\tau_b$	256
$b$	$T_2^{\tau_b}$	$\tau_b$	64

The power spectral density of the signal of Figure 3.11 is shown in Figure 3.12, which illustrates the three frequency components. For this signal, two levels of hierarchical decomposition can result in separating the three frequency components. The decomposition levels, the corresponding  $\tau$ -functions and the frequency components contained therein are shown in Table 3.2.

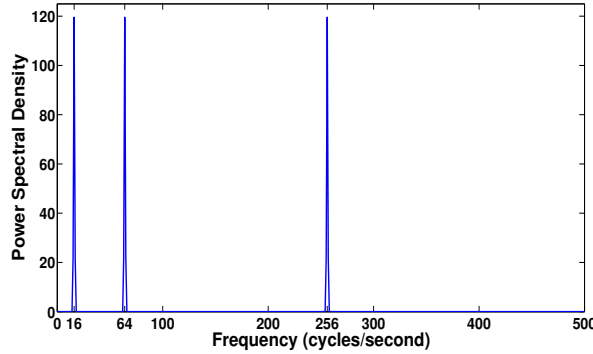


Figure 3.12: Power spectral density plot of the synthetic signal shown in Figure 3.11 which illustrates the three frequency components in the signal.

From Table 3.2, it can be observed that the three frequency components 16, 64 and 256 cycles/second are separated in the  $\tau$ -functions  $T_2^{\tau_a}$ ,  $T_2^{\tau_b}$  and  $T_1^{\tau_b}$  respectively. For frequency separation at levels  $a$  and  $b$ , the values of  $\tau_a$  and  $\tau_b$  are selected such that the frequency corresponding to  $\tau_a$  and  $\tau_b$ , given by  $f_{\tau_a}$  and  $f_{\tau_b}$  respectively, is such that  $16 < f_{\tau_a} < 64$  and  $64 < f_{\tau_b} < 256$ . For actual decomposition, the values of  $\tau_a$  and  $\tau_b$  were scaled by the scaling constant  $k$  as mentioned in Section 3.3.2.5, using a value of  $k = 0.44$ .

The power spectral density plots of the  $\tau$ -functions  $T_1^{\tau_a}$ ,  $T_2^{\tau_a}$ ,  $T_1^{\tau_b}$  and  $T_2^{\tau_b}$  are shown in Figure 3.13 to illustrate the separated frequency components in the  $\tau$ -functions.

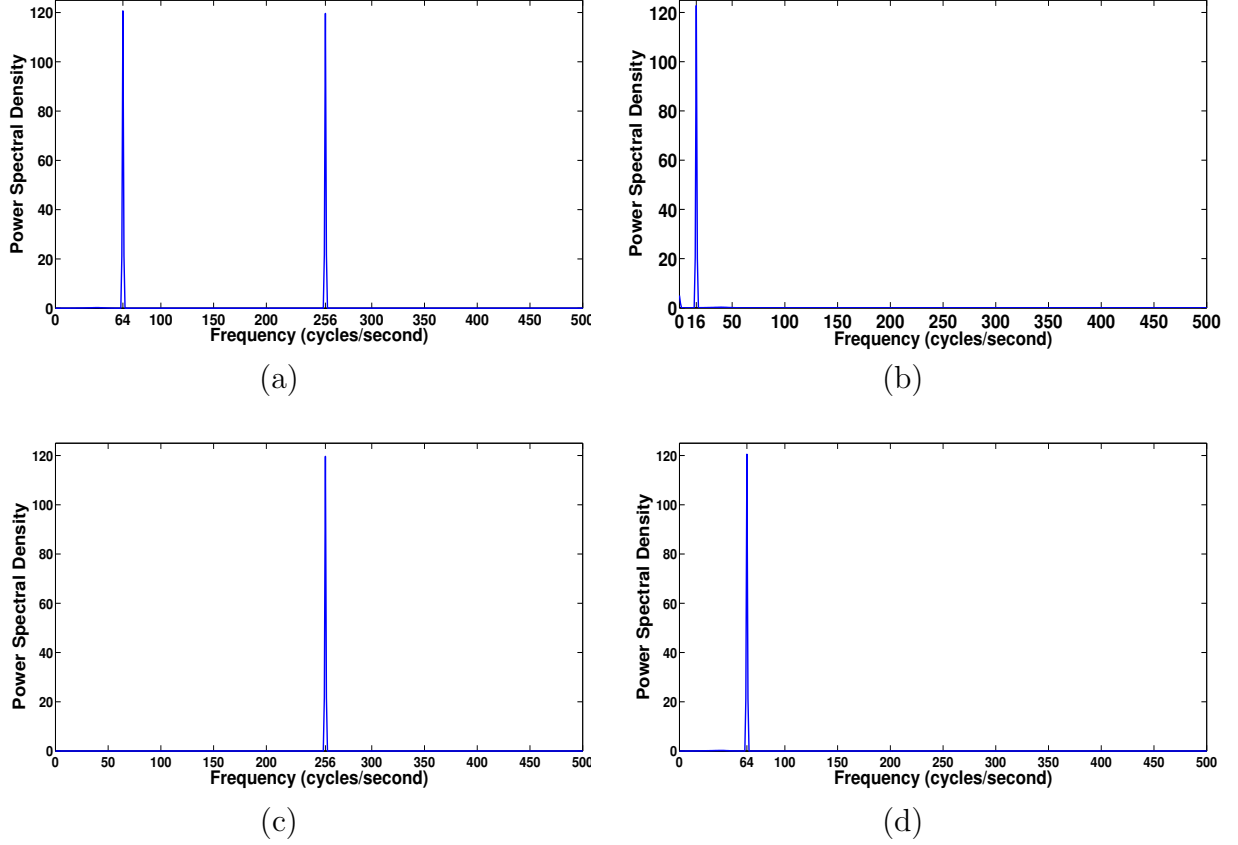


Figure 3.13: Power spectral density plots of the  $\tau$ -functions  $T_1^{\tau_a}$ ,  $T_2^{\tau_a}$ ,  $T_1^{\tau_b}$  and  $T_2^{\tau_b}$  obtained by hierarchical decomposition of the synthetic signal shown in Figure 3.11.

### 3.4.1.2 Hierarchical decomposition of an EEG signal

In this Section the hierarchical decomposition of a real-world scalp EEG signal segment will be described. For this purpose, and for later experiments also, the CHB-MIT Scalp EEG Database has been used, which is a publicly available database [105]. The signals in this database consist of multi-channel EEG signals sampled at a sampling frequency  $f_s = 256$  Hz.

To demonstrate hierarchical decomposition, a randomly chosen four second segment (1024 samples) was obtained from one channel. Further, this segment was taken from an interval marked as a seizure. A plot of the EEG segment is shown in Figure 3.14.

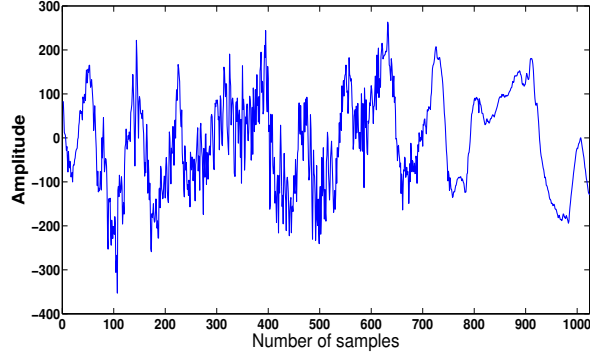


Figure 3.14: EEG signal segment of length four seconds (1024 samples) used to demonstrate hierarchical decomposition to separate frequency components in the frequency range  $3 \leq f \leq 29$ .

Although seizures can have a broad spectrum, it is accepted that seizures in recorded EEGs occur between 3 and 29 Hz [106]. Therefore the purpose of the hierarchical decomposition is to decompose the seizure segment into a frequency band given by the range  $3 \leq f \leq 29$  Hz. Here  $f$  represents the range of values of the frequency components present in the  $\tau$ -function containing the desired frequency band obtained after hierarchical decomposition.

EMD-MPS can be used for hierarchical decomposition of the EEG segment to obtain the desired frequency range  $3 \leq f \leq 29$  Hz in a  $\tau$ -function using a two-level hierarchy. This is illustrated in Table 3.3, which shows that the desired frequency range  $3 \leq f \leq 29$  is obtained in  $T_2^{\tau_b}$  in the second level of the hierarchical decomposition. For this hierarchical decomposition, the values of  $\tau_a$  and  $\tau_b$  were chosen according to the required frequency separation values  $f_{\tau_a}$  and  $f_{\tau_b}$  respectively, where  $f_{\tau_a} = 3$  Hz and  $f_{\tau_b} = 29$  Hz, and  $\tau_a = \frac{f_s}{f_{\tau_a}}$ , and  $\tau_b = \frac{f_s}{f_{\tau_b}}$ . For actual decomposition, the values of  $\tau_a$  and  $\tau_b$  were scaled by the scaling constant  $k$  as mentioned in Section 3.3.2.5, using a value of  $k = 0.44$ .

Table 3.3: Hierarchical decomposition of the EEG signal segment shown in Figure 3.14 using two-level hierarchical decomposition

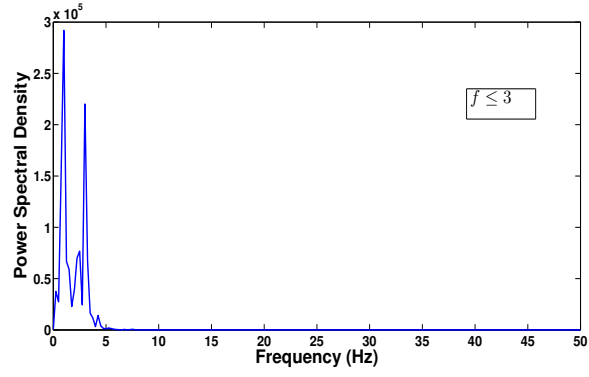
Decomposition level	$\tau$ -function	$\tau$	Frequency components (cycles/second)
$a$	$T_1^{\tau_a}$	$\tau_a$	$f > 3$
$a$	$T_2^{\tau_a}$	$\tau_a$	$f \leq 3$
$b$	$T_1^{\tau_b}$	$\tau_b$	$f > 29$
$b$	$T_2^{\tau_b}$	$\tau_b$	$3 \leq f \leq 29$

#### 3.4.1.3 Flexibility of equality signs in hierarchical decomposition of real signals

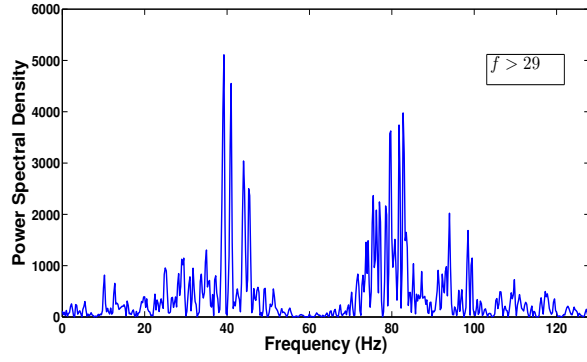
The power spectral density plots of  $\tau$ -functions  $T_1^{\tau_a}$ ,  $T_1^{\tau_b}$  and  $T_2^{\tau_b}$  mentioned in the previous Section are shown in Figure 3.15. The range of the frequency values of the oscillatory components contained in each frequency component are mentioned in each plot in the figure. There is evidence of frequency spillage beyond the ranges indicated in the figures. This happens due to mode-mixing, which may result from not having a sampling frequency with a value ten times higher than the frequency value of the highest frequency component present in the signal (Section A.1 in Appendix A), and also the ratio of frequency values of components being higher than two-thirds (Section A.2 in Appendix A). However, for real life signals it is not always possible to control these parameters. At the same time, the behaviour of extrema-based decomposition methods is well-understood, and as such the slight deviations from expected behaviour can easily be accounted for. Furthermore, it is also possible to eliminate or overcome mode-mixing by using noise-assisted decomposition [50], thereby minimizing the effect of deviation from expected behaviour.

#### 3.4.1.4 Depth of hierarchical decomposition

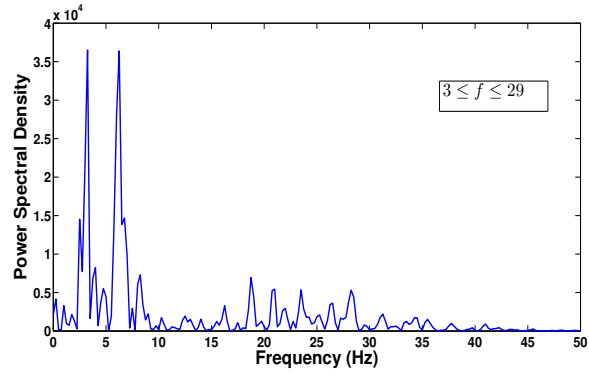
It was previously mentioned in Section 3.4 that the hierarchical decomposition may be started from a desired frequency separation level  $f_{\tau_a}$  using a value of  $\tau_a$ . From this level  $a$ , the hierarchical decomposition may be continued to further levels using appropriate values of  $\tau$ , as previously described. However, the scaling constant  $k$  and the minimum possible value



(a)



(b)



(c)

Figure 3.15: Power spectral density plots of the  $\tau$ -functions  $T_1^{\tau_a}$  (a),  $T_1^{\tau_b}$  (b) and  $T_2^{\tau_b}$  (c) obtained by hierarchical decomposition of the EEG signal shown in Figure 3.14.

of  $\tau$  as given by Equation 3.2, limit the depth of the hierarchical decomposition. This is explained next.



For a desired frequency separation value  $f$ , the value of  $\tau$  can be obtained as  $\tau = \frac{f_s}{f}$  using Equation 3.1, where  $f_s$  is the sampling frequency of the signal. However, the value of  $\tau$  has to be scaled by  $k$ , as explained in Section 3.3.2.5, so that the value of  $\tau$  used for decomposition is given by  $\hat{\tau}$ , where  $\hat{\tau} = k\tau$ . This leads to the following relationship between  $\hat{\tau}$  and  $k$ :

$$\hat{\tau} = \frac{f_s}{f} \cdot k \quad (3.15)$$

However,  $\hat{\tau}$  requires a value  $\hat{\tau} > 2$  in order to represent a time-scale based decomposition, since EMD-MPS at a value of  $\hat{\tau} = 2$  is reduced to the EMD algorithm, as explained in Section 3.3.1. Since EMD-MPS requires  $\hat{\tau} > 2$ , we have, from Equation 3.15,  $k \cdot \frac{f_s}{f} > 2$ , so that:

$$f < \frac{k}{2} \cdot f_s \quad (3.16)$$

where  $k = 0.44$ .

The depth of the hierarchical decomposition is determined by Equation 3.16, thereby limiting the number of levels in the decomposition hierarchy.

## 3.5 Chapter Summary

The focus of this Chapter was a novel modification of the EMD algorithm, namely EMD-MPS, which allows a time-scale based decomposition. The properties of EMD-MPS were studied empirically, and its use for a novel hierarchical decomposition methodology was presented. The next Chapter will present how EMD-MPS can be used for time-scale based de-noising and de-trending, and compare this to more involved and subjective approaches based on EMD.

# Chapter 4

## De-Noising and De-Trending with Empirical Mode Decomposition-Modified Peak Selection

### 4.1 Background

The efficacy of Empirical Mode Decomposition (EMD) for de-noising and de-trending of signals was discussed in Chapter 2. EMD decomposes signals into a set of intrinsic mode functions (IMFs), which contain oscillatory components of the signal at different time-scales. For cases in which the signal of interest is contaminated with noise, the IMFs resulting from decomposition of the signal using EMD can be quantified with respect to the information content or the noise, and a partial reconstruction of the signal by excluding the IMFs dominated by noise results in a de-noised signal. Similarly, partial reconstruction of the signal using only IMFs determined as containing the trend of the signal results in extraction of the signal trend. The most widely used methods used for de-noising and de-trending using EMD were also discussed previously in Chapter 2.

While EMD-based methods have proven to be successful for de-noising and de-trending of

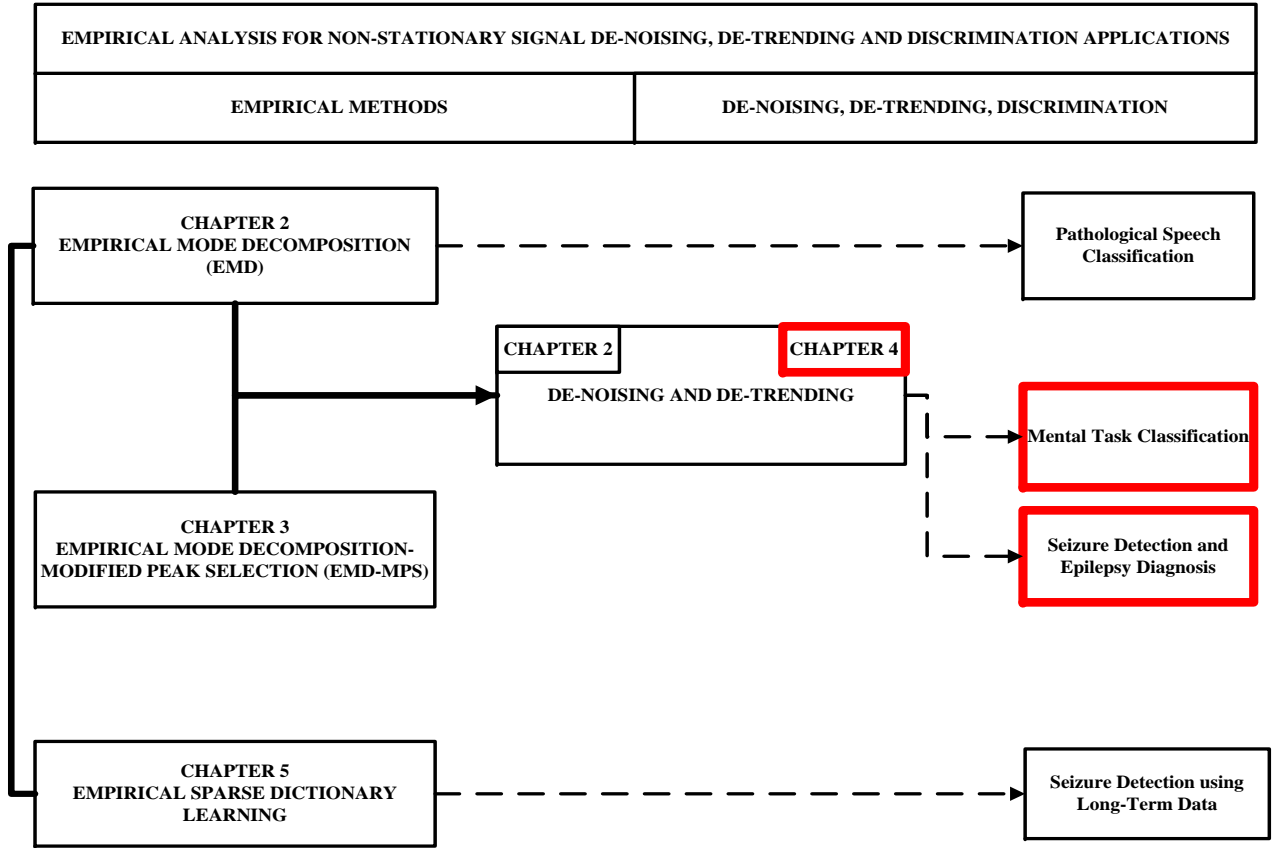


Figure 4.1: Chapter 4 focuses on de-noising and de-trending using the time-scale based decomposition made possible by the EMD-MPS method. Practical applications of time-scale base de-trending in terms of mental task classification and seizure detection are also presented.

a wide-variety of signals in diverse domains, there are a number of short-comings associated with EMD-based methods. EMD-based de-noising and de-trending methods were categorized into the following categories in Section 2.4 of Chapter 2:

1. Signal-specific methods
2. Statistical approaches
3. Energy and Energy-Ratio approaches

The signal-specific methods have been developed with particular signals and domains

in mind, e.g. speech signal enhancement and ground penetrating radar data de-noising. Therefore the analysis of IMFs obtained by decomposing these particular signals cannot be generalized to signals in diverse domains. On the other hand, the statistical approaches and energy and energy-ratio approaches are more general in application. However, these can suffer from mode-mixing, whereby oscillations at disparate time-scales may appear in one IMF at different time instants. Such issues can affect the accuracy of quantification of IMFs as consisting of noise or containing useful signal content. The effect of mode-mixing can be reduced using noise-assisted EMD, however such approaches are computationally prohibitive. These also require careful selection of the parameters of noise to be added to the signals before the decomposition process, thereby also making the overall methodology less adaptive. Furthermore, the statistical approaches have generally been formulated using white noise or fractional Gaussian noise models. For successful application of these methods, knowledge of statistical properties of the signals under analysis is required, or the statistical properties have to be accurately determined. This is not always possible, and if possible, requires an additional processing step. Also, a very important point related to all categories of EMD-based de-noising approaches relates to consideration of the first IMF as consisting mainly of noise, and exclusion of the first IMF in partial reconstruction for the de-noised signal. However, it has been argued in [28] that the first IMF contains important signal information, and hence removal of this IMF leads to important signal information being lost.

The most important aspect absent from the EMD-based approaches for de-noising and de-trending relates to incorporation of the time-scale information for de-noising and de-trending. This information is relevant as EMD decomposes signals into IMFs based on different time-scales. Also, in many cases the signal of interest is known to exist only within certain time-scales, and extraction of only that part of the signal can lead to simplified and accurate de-noising. A relevant example here is a speech signal contaminated with noise, where the speech of interest is bandlimited to a known frequency range. A decomposition approach which can be used to decompose non-stationary and non-linear signals can extract

information within the desired time-scale, if the decomposition can incorporate the time-scale information.

Incorporation of the time-scale information for de-trending is also extremely important, since the definition of trend may depend on the observation time-scale, as previously discussed in Section 2.4 of Chapter 2. Also, trends usually have time-scales associated with them, such as a 3-monthly or 6-monthly trend associated with changing values of a financial time-series, or the yearly or 5-yearly trend associated with climate data. A de-trending method should therefore allow trend extraction based on a required time-scale.

De-noising and de-trending methods based on EMD do not perform de-noising or de-trending based on time-scales, but instead on criteria requiring IMF analysis. De-noising and de-trending also requires partial reconstruction of the original signal by using only a subset of the IMFs resulting from the decomposition of the original signal. There is incorporation of the time-scale information in the work presented in [40], which studies application of EMD to de-trend non-stationary and non-linear time-series. This study uses climate data to demonstrate trends at different time-scales. The time-scales are found by a generalized zero-crossing method, which determines the time spans between various combinations of the critical points defined as the union of all of the zero-crossings and extrema of an IMF. A weighted mean of the different time spans is computed to obtain the time-scale. This method, however, requires complete decomposition of the signal into IMFs, and is also prone to the effects of mode-mixing, which can lead to inaccurate estimations of the time-scale. Noise-assisted EMD is also not likely to help, as the residual noise left in the IMFs can lead to inaccuracies in identification of zero-crossings and extrema. Furthermore, the extraction of trend by the method requires partial reconstruction of the signal.

It can therefore be argued that a decomposition methodology that allows the time-scale information to be incorporated into the decomposition process can allow time-scale based de-noising and de-trending. The de-noising and de-trending can then happen without requiring complete signal decomposition into IMFs followed by partial reconstruction using only a subset of the IMFs, which require identification using different criteria. Empirical

Mode Decomposition-Modified Peak Selection (EMD-MPS) is such a method, as presented in Chapter 3, which incorporates the time-scale information in the decomposition process. This is done by the following relation, also explained in Chapter 3.

$$F = \frac{F_s}{\tau} \quad (4.1)$$

where  $F_s$  represents the sampling frequency.

In the next sections, the use of EMD-MPS for time-scale based de-noising and de-trending of signals will be described, and illustrated using synthetic and real-world examples.

## 4.2 Time-scale based De-noising and De-trending

EMD-MPS allows a novel time-scale based de-noising and de-trending of signals, which does not require estimation of a trend model for model-based de-trending, or knowledge of the statistical properties of IMFs. In this context, with EMD-MPS, appropriate selection of  $\tau$  representing a required time-scale allows separation of the faster oscillations from the slowly-varying trend, where "slowly-varying" is controlled by the value of  $\tau$ , according to Equation 4.1. This in turn can be used for both de-noising and de-trending, depending on the time-scale used. EMD-MPS de-noising and de-trending can be related to EMD partial reconstruction, in the context of the vertical behaviour of EMD-MPS (Section 3.3.2.2 in Chapter 3), where this is equivalent to combination of lower order IMFs, which pass undecomposed in one  $\tau$ -function, and of higher-order IMFs, which combine to form the trend.

### 4.2.1 De-noising and De-trending of Synthetic Signals

As an example, we consider de-noising of a 10 Hz sinusoidal signal contaminated with additive white Gaussian noise (AWGN) at different signal-to-noise (SNR) ratios. This synthetic signal is similar to the examples considered in [84] for demonstrating the de-noising performance of noise-assisted EMD, and is also a realistic example in the context of electrical noise interference corrupting voltage signals [19]. Using an appropriate value of  $\tau$  selected using

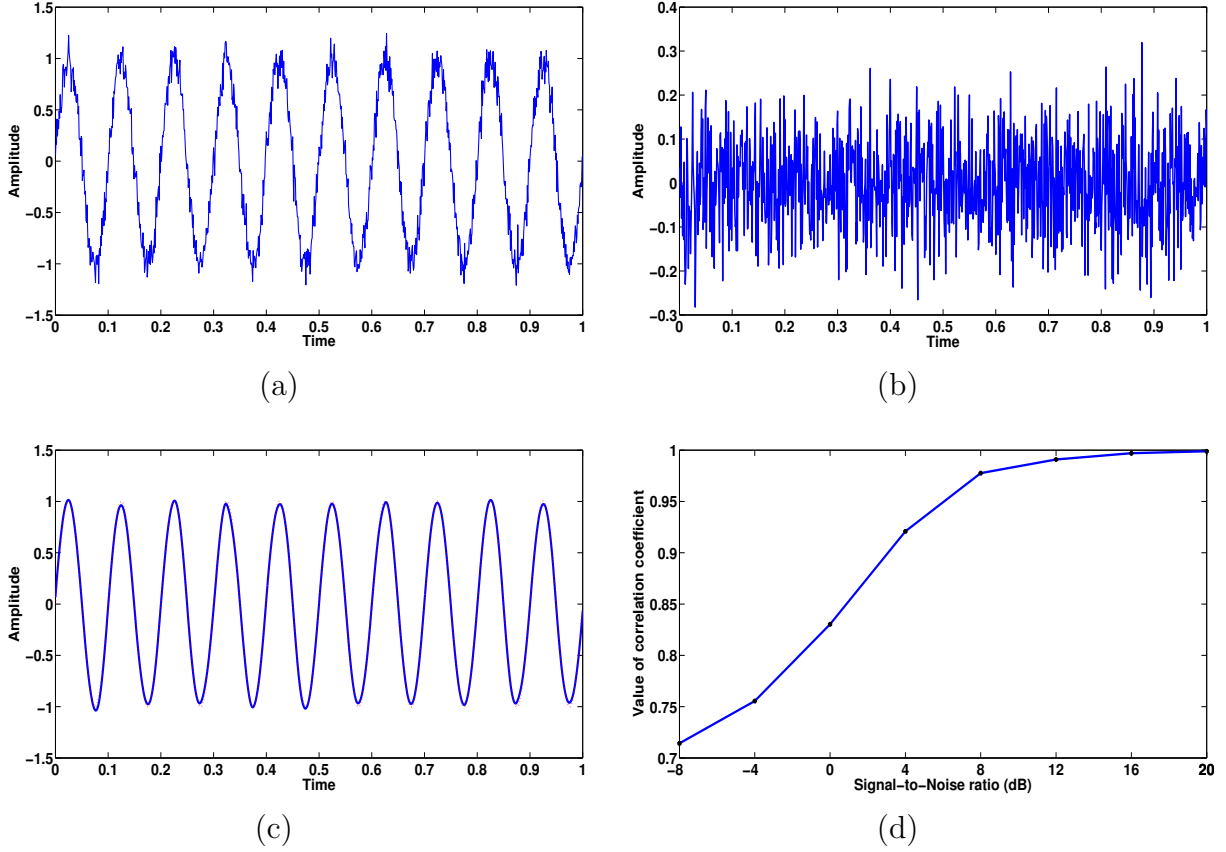


Figure 4.2: De-noising of a sinusoidal signal contaminated with additive white Gaussian noise (AWGN) using EMD-MPS. (a) Signal contaminated with AWGN: SNR = 20dB. (b) Extracted noise in  $\tau$ -function  $T_1$ . (c) Recovered signal in  $\tau$ -function  $T_2$  superimposed on the original signal. (d) Value of correlation coefficient ( $\tau$ -function  $T_2$  and original sinusoidal signal) plotted against SNR of test signal to quantify de-noising performance of EMD-MPS.

Equation 4.1, the noise is extracted in  $\tau$ -function  $T_1$ , whereas the low frequency sinusoid is extracted in  $\tau$ -function  $T_2$ . This is shown in Figure 4.2 for a sinusoid contaminated with AWGN with a SNR of 20 dB. However, as the SNR decreases, the degradation in the de-noised signal compared to the original signal is much less severe as compared to EMD, as demonstrated in Figure 4.2 (d), which plots the correlation coefficient as an objective measure of similarity between the de-noised and original signals for different values of the

SNR. On the other hand, in case of EMD, severe mode-mixing will render extraction of the original signal practically impossible even for relatively higher values of SNR, as shown in Figure 4.3.

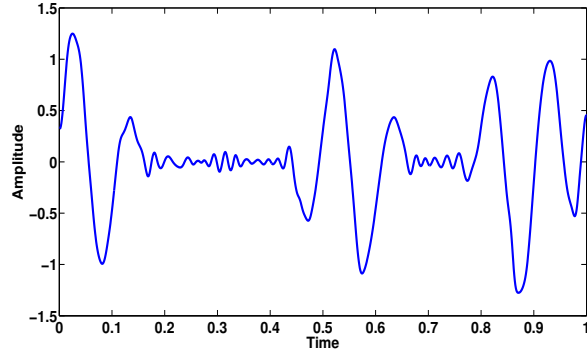


Figure 4.3: De-noising of a sinusoidal signal contaminated with AWGN using EMD. Severe mode-mixing results even at a relatively high SNR of 16 dB.

Also, since EMD-MPS does not attempt to decompose signals into mono-component IMFs, but instead into  $\tau$ -functions where different modes of oscillation may coexist, mode-mixing does not affect EMD-MPS as it does EMD. This is demonstrated using a large amplitude sinusoidal signal which has intermittently occurring small amplitude sinusoids of higher frequency values. Such signals will lead to mode-mixing when decomposed with EMD [8], and are also relevant practically [19]. An example of such a signal is shown in Figure 4.4, which also shows the IMFs resulting from decomposing the signal using EMD. As is clear from Figure 4.4 (b), the resulting IMFs demonstrate mode-mixing, and de-noising of the desired signal using partial reconstruction is not possible. On the other hand, Figure 4.5 shows the same signal, and the de-noised signal resulting from application of EMD-MPS with an appropriate value of  $\tau$  selected using Equation 4.1. As can be seen from Figure 4.5 (b), the de-noised signal is practically the same as the original signal, with some imperfections in the beginning samples due to algorithm implementation issues.

Numerous synthetic examples with different types of trend have been covered in literature (e.g. [51][43]). Some similar examples are utilized to demonstrate time-scale based de-noising and de-trending using EMD-MPS in Appendix B.



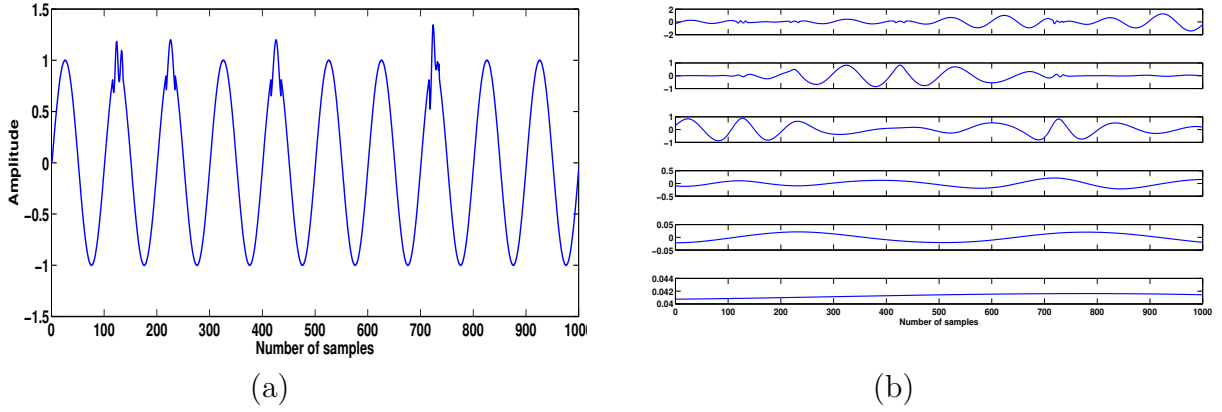


Figure 4.4: Demonstration of EMD mode-mixing, making removal of intermittently occurring low amplitude sinusoidal signals impossible. (a) Sinusoidal signal with intermittently occurring low amplitude components of higher frequency. (b) Mode-mixed IMFs resulting from EMD decomposition of signal in (a).

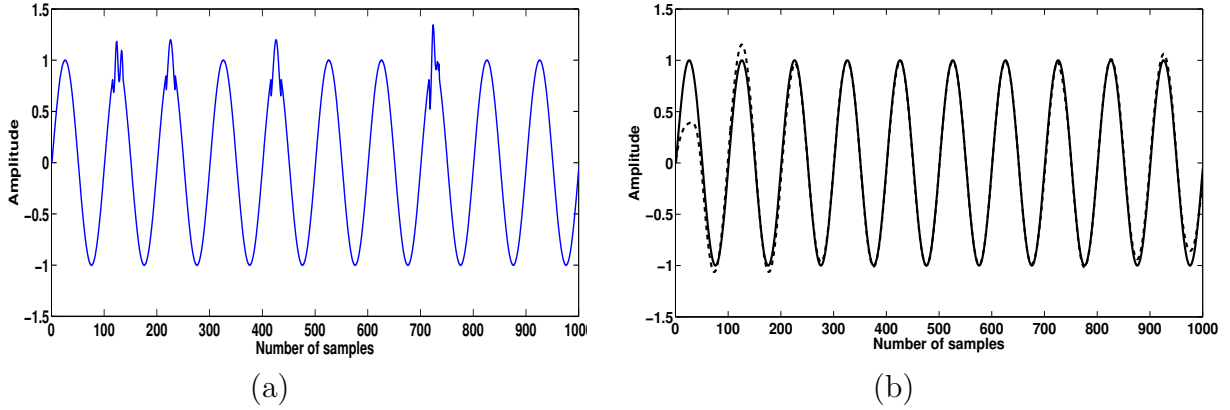
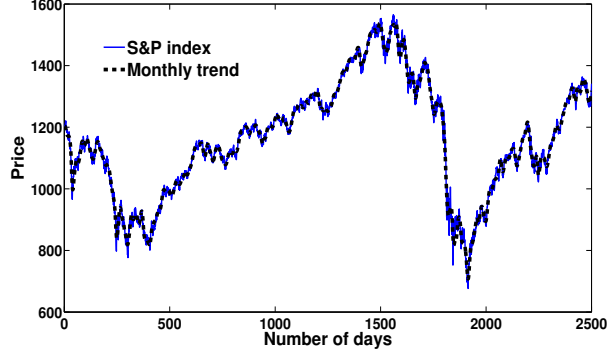


Figure 4.5: Removal of intermittently occurring low amplitude sinusoidal signals from the larger amplitude sinusoid using EMD-MPS. (a) Sinusoidal signal with intermittently occurring low amplitude components of higher frequency. (b) De-noised signal (dashed line) using EMD-MPS compared with the original signal (solid line).

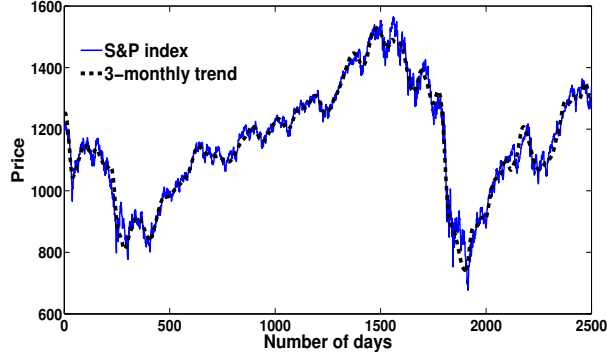
#### 4.2.2 De-noising and De-trending of Real-life Signals

In this Section, the use of EMD-MPS to de-noise and de-trend real life non-stationary signals based on required time-scales will be demonstrated. The signals used as examples in this Section consist of financial time-series, and biomedical signals consisting of EEG and ECG signals. Some further examples using pathological speech signals as well as ECG signals are

presented in Appendix B.



(a)



(b)

Figure 4.6: Time-scale based de-trending of S&P 500 index data from November 6, 2001 to October 11, 2011, showing the monthly trend in (a) and the 3-monthly trend in (b).

**Financial time-series:** The financial time-series used to demonstrate time-scale based de-trending using EMD-MPS is in the form of S&P 500 daily index from November 6th, 2001 to October 11th, 2011, which consists of 2500 data points. The time-scales for de-trending consist of the monthly and 3-monthly cycles. Therefore values of  $\tau$  representing monthly and 3-monthly cycles were calculated in the following way. Let  $F_s$  be equal to  $N$ , where  $N$  represents the number of days of the daily data being de-trended. For a 3-monthly cycle, let  $F$  be equal to 90, and the value of  $\tau$  is obtained from Equation 4.1. This way, the values of  $\tau$  representing monthly and 3-monthly cycles are calculated, and a different trend is extracted

from the S&P data for each of these two time-scales. The trends at different time-scales, superimposed with the original time-series, are shown in Figure 4.6, which illustrates the different levels of the slowly-varying detail captured by trends at these different time-scales.

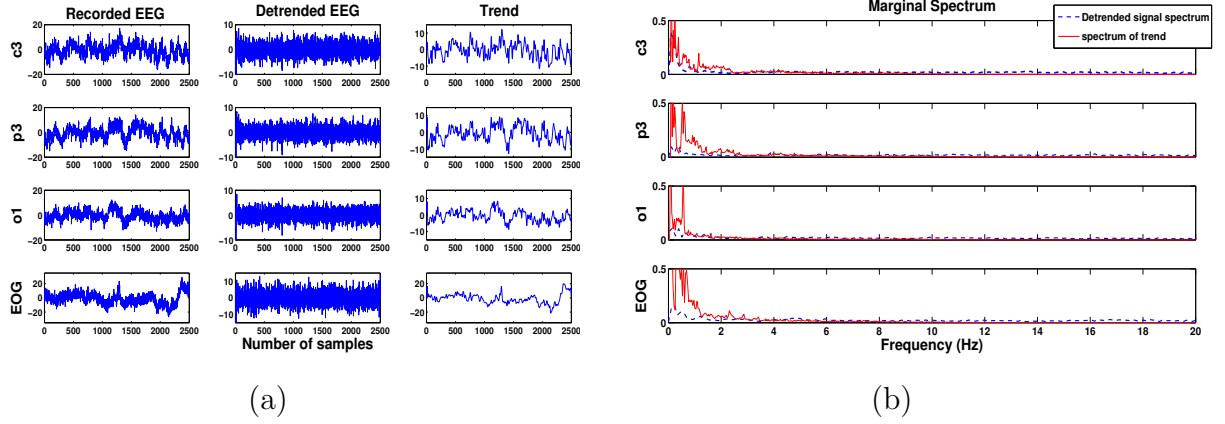


Figure 4.7: De-trending EEG signals using EMD-MPS. (a) De-trended EEG signals from 4 selected channels. (b) Marginal spectrum of the de-trended signal and the trend, showing clear separation between the spectra.

**EEG signals:** In EEG signals, the interference and the signal of interest are mixed together in a non-stationary and non-linear way, and an effective method to separate the signal of interest is required [107]. Here EEG signals from 4 selected channels are used to illustrate use of EMD-MPS to extract the low frequency trend according to a desired time-scale. The EEG signals have a length of 2500 samples (10 seconds duration), and have been sampled at 250 Hz. For these signals, the frequencies of interest are in the range of frequencies below 20 Hz. Traditional filtering methods would result in cutting off frequency content belonging to both the desired and undesired parts of the signal [107]. For a time-scale based de-trending corresponding to a frequency value of 20 Hz, a value of  $\tau = 12.5$  is used, as obtained using Equation 4.1. The  $\tau$ -functions, containing the de-trended signal and the trend, obtained after decomposing the EEG signals with EMD-MPS using this value of  $\tau$ , are shown in Figure 4.7 (a). The marginal spectrum of the  $\tau$ -functions obtained from all 4 channels are shown in Figure 4.7 (b). The marginal spectrum clearly shows the strength of the method in the low

frequency range (below 20 Hz), where overlapping of both spectra is common. EMD-MPS filtered out only the interference, leaving low-frequency activity undisturbed. This has also been achieved with minimal complexity of the method, compared to approaches based on EMD [107].

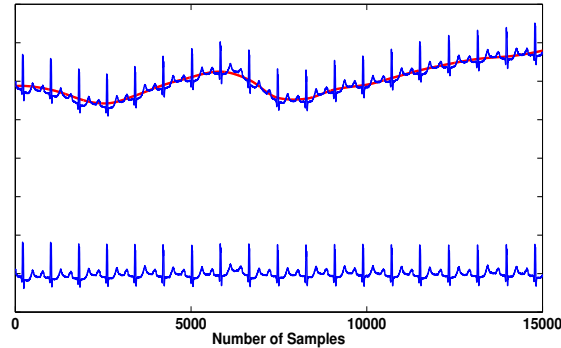


Figure 4.8: De-trending for baseline wander removal of a 15000 sample portion of Holter ECG signal. The upper graph shows the original signal along with the extracted trend (baseline wander), whereas the lower graph shows the de-trended ECG signal.

**ECG signals:** De-noising and de-trending of ECG signals is an important problem in signal processing, with baseline wander removal an important aspect of processing of ECG signals. As with other real-life signals, de-noising and de-trending methods for ECG signals based on non-adaptive filtering approaches have not shown to perform well in practice [108][109]. Numerous improved methods for ECG signal de-noising and de-trending have been proposed, which use data adaptive methods such as wavelets and EMD [110][111][109]. Wavelet-based methods however require careful selection of the wavelet basis function, and also an appropriate wavelet thresholding method.

Methods based on EMD are generally more adaptive in nature, as the basis is selected from the data, however partial reconstruction to obtain the desired signal may result in important QRS information being lost in the discarded IMFs, as some parts of the QRS complex might have been decomposed into IMFs judged to be noise. The method described in [111] first uses a windowing function to preserve the QRS complex in the IMFs, and then uses statistical properties of IMFs to identify the IMFs to be excluded as noise in the partial

reconstruction of the clean ECG signal. After this, a bank of low-pass filters is applied to the higher order IMFs to extract the baseline wander. The methodology presented in [112] combines EMD with wavelet shrinkage, by first using a windowing function on IMFs to preserve the QRS complex while removing as much noise as possible. Then these IMFs are transformed into the discrete wavelet domain, where soft thresholding based noise reduction is applied. The EMD-based technique presented in [109] for baseline wander removal is based on selective IMF reconstruction based on a slope minimization criteria for IMFs. The procedure requires taking the fast Fourier transform of the original signal, segmentation of the signal and calculation of the slopes for all segments, addition of individual slopes to form the global slope, and then minimization of the global slope by exclusion of higher order IMFs one by one. After a number of further steps, baseline wander removal is achieved.

As can be observed from the discussion just presented, the de-noising and de-trending methods used for ECG signals represent complicated procedures. Also, for EMD-based methods using partial reconstruction, the possibility of losing important signal information is very high, and complicated procedures have to be adopted to minimize this loss. On the other hand, EMD-MPS allows formulation of the baseline wander removal as a de-trending problem. By selection of an appropriate time-scale, the baseline wander can be removed from the ECG signal in just one decomposition step. This is illustrated in Figure 4.8, which shows baseline wander removal from a 15000 samples segment of a Holter ECG signal (Holter ECG signals are ambulatory in nature). The ECG signal has been sampled at 1024 samples per second, and a value of  $\tau = 1000$  was used for decomposition using EMD-MPS. This decomposition resulted in the trend representing the baseline wander being decomposed into  $\tau$ -function  $T_2$ , and the de-trended signal being contained in  $\tau$ -function  $T_1$ .

### 4.2.3 Discussion

In the previous Sections the effectiveness of EMD-MPS for time-scale based decomposition was demonstrated using different synthetic and real signals. These examples also illustrate the power of a data-adaptive decomposition approach, which is suitable for non-stationary

and non-linear signals. The use of a time-scale allows conceptually and practically simple decomposition, and does not require a complete decomposition followed by partial reconstruction based on some criteria. This is especially advantageous, since partial reconstruction done by excluding certain decomposed components may lead to a distorted signal, or a reconstructed signal with incomplete information, as useful signal information is spread in different IMFs.

#### 4.2.4 Experiment: Mental Task Classification using EEG Signals

This Section presents an experiment that uses EMD-MPS as the central part of the methodology for mental task classification. This experiment has the same objective as the experiment described in Section 2.4.1 of Chapter 2, where a methodology based on EMD was described. However, instead of using an EMD-based de-trending based on partial reconstruction of IMFs, a time-scale based de-trending based on EMD-MPS will be used. The next sections will describe the EMD-MPS based methodology for mental task classification in more detail, and also make comparison with the EMD based method previously described in Chapter 2.

##### 4.2.4.1 Background

For details about the background of this experiment and the data used, reference may be made to Sections 2.4.1.1 and 2.4.1.2 in Chapter 2, in order to avoid duplication of the details here.

This experiment presents a novel method for mental task classification based on application of EMD-MPS and Teager energy operator (TEO) on EEG signals. EMD-MPS is used to decompose the EEG signals into two  $\tau$ -functions representing the trend and a de-trended component. TEO is applied to both  $\tau$ -functions to obtain values of the average Teager energy, which are used to form the feature vectors. One-versus-one classification of mental tasks using these feature vectors is performed with a 1-NN classifier [81]. The overall methodology of this EMD-MPS based experiment is depicted in Figure 4.9. This method achieves an average correct classification rate of 87% for a one-versus-one classification scheme, improv-

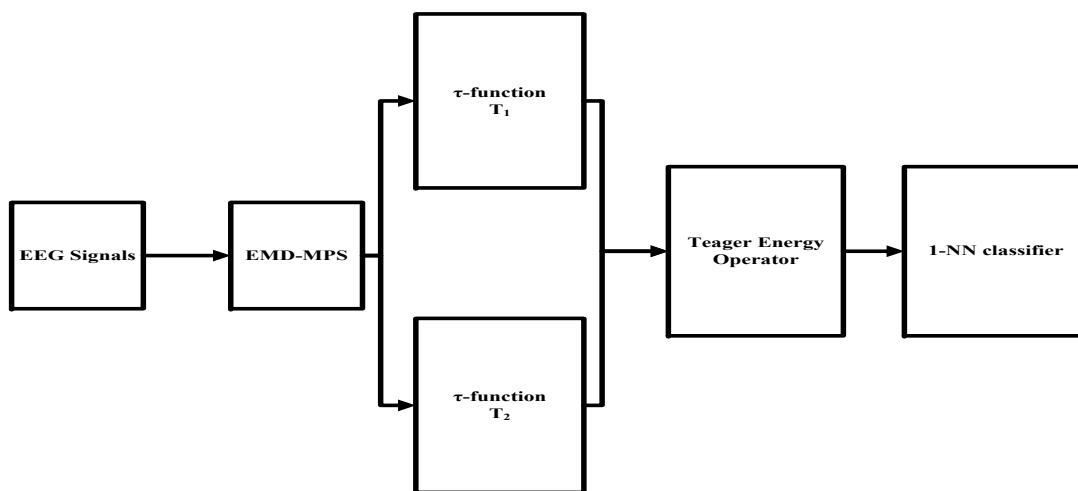


Figure 4.9: Schematic representation of the methodology described in this experiment.

ing on previous results using EMD-based methodologies on the same EEG signals [58][113]. At the same time, this EMD-MPS based method is characterized by the simplicity of the decomposition, feature extraction, as well as classification.

#### 4.2.4.2 Decomposition using EMD-MPS

EMD-MPS is used for a time-scale based de-trending of EEG signals, such that each EEG signals is decomposed into two  $\tau$ -functions, one representing the de-trended signal containing the higher frequency components, and the other representing the low frequency trend of the signal. This is done by appropriate selection of a value of  $\tau$  according to the relation presented in Equation 4.1. For de-trending of EEG signals used in this study, two different values of  $\tau$  were used to obtain two different sets of  $\tau$ -functions, containing frequency components separated according to Equation 4.1. The first value of  $\tau$  corresponds to a frequency value  $F = 8$  Hz, such that the  $\tau$ -function  $T_1$  contains frequency components with frequency values greater than 8 Hz, and the  $\tau$ -function  $T_2$  represents the trend containing frequencies lower than 8 Hz. Using Equation 4.1 and sampling frequency value  $F_s = 250$  Hz, the value of  $\tau$  obtained is given by  $\tau = 31.25$ . However, for decomposition, the value  $\hat{\tau}$  is used, which is  $\tau$  scaled by a constant  $k$  as described Section 3.3.2.5 of Chapter 3. A good estimate for the

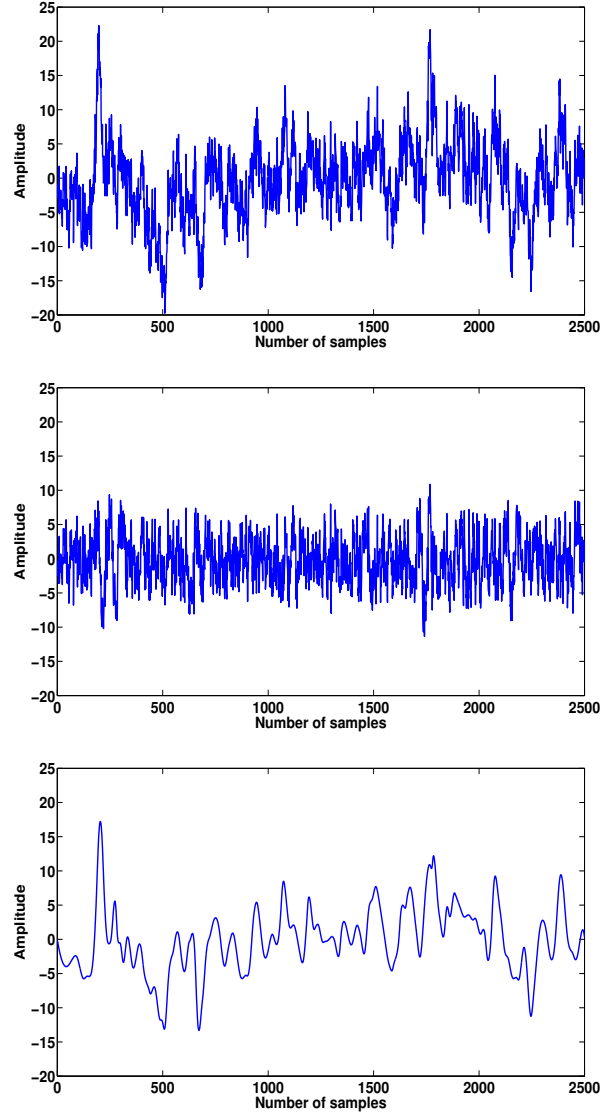


Figure 4.10: EEG signal from Subject 1 performing Task 2 (top).  $\tau$ -function  $T_1$ , representing de-trended signal (middle).  $\tau$ -function  $T_2$ , representing the trend (bottom). These  $\tau$ -functions have been obtained using a value of  $\tau = 31.25$  ( $\hat{\tau} = 14$ ), corresponding to a frequency separation value of  $F = 8$  Hz.

value of  $k$  is given by  $k \approx 0.44$  [114], such that  $\tau = 31.25$  corresponds to  $\hat{\tau} = 14$ . Similarly, the other set of  $\tau$ -functions were obtained corresponding to  $F = 4$  Hz, such that  $\tau = 62.5$ , with the corresponding  $\hat{\tau} = 28$ .

All EEG signals used in this study were decomposed using EMD-MPS using these two



values of  $\hat{\tau}$ . In this regard, Figure 4.13 shows an example EEG signal, and the  $\tau$ -functions  $T_1$  and  $T_2$  obtained with a value of  $\hat{\tau} = 14$ .

#### 4.2.4.3 Feature analysis

After the decomposition of EEG signals (6 EEG signals per subject per task, with 5 trials of each task), the TEO was applied to each of the two  $\tau$ -functions obtained per EEG signal to estimate the average Teager energy of each  $\tau$ -function.

The TEO was developed from the point of view of the energy required to generate a signal. This non-linear energy-tracking operator  $\Psi$  is given in its discrete form [54] as:

$$\Psi(x[n]) = x^2[n] - x[n+1] \cdot x[n-1] \quad (4.2)$$

The TEO is nearly instantaneous, given that only three samples are required for computing the energy at a given time instant, as can be seen from Equation 4.2. This also makes the operator easy to implement efficiently. The average Teager energy,  $e_i$ , for a  $\tau$ -function  $T_i$  is calculated as:

$$e_i = \frac{1}{N} \sum_{n=1}^N | \Psi[T_i(n)] | \quad (4.3)$$

where  $N$  is equal to the number of samples in the  $\tau$ -functions,  $\Psi[\cdot]$  is the discrete-time TEO and  $i = 1, 2$ .

The average Teager energy  $e_i$  is calculated for each  $\tau$ -function according to Equation 4.3. This way, for each EEG signal, there are two values of  $e_i$  corresponding to each  $\tau$ -function  $T_i$ . Given that 6 EEG signals per task are used, the feature vector contains 12 elements consisting of the average Teager energy values. This is a significant reduction in feature vector dimension compared to the previous EMD-based work [113], and other related works [58][15].

The feature vectors obtained for the tasks were used in a one-versus-one classification scheme using a linear classifier, namely the 1-NN classifier. The classification accuracy

for each task combination was estimated using the tenfold cross-validation method. The classification results are discussed in the next Section.

#### 4.2.4.4 Results

The classification results obtained with the methodology presented in this experiment are shown in Table 4.1. The classification accuracy for different task combinations listed in Table 4.1 has been obtained using  $\tau$ -functions extracted with both values of  $\hat{\tau}$  used in the analysis. It can be seen from Table 4.1 that the classification accuracy for most of the task combinations is greater than 80%, with many task combinations having a classification accuracy of 100%. This is true for both values of  $\hat{\tau}$  used for decomposition using EMD-MPS. However, for subjects 3 and 6, the average classification accuracy for all task combinations is considerably higher for a value of  $\hat{\tau} = 14$ , compared to a value of  $\hat{\tau} = 28$ . For other subjects, the difference in average classification accuracy over all tasks using both values of  $\hat{\tau}$  is not large, as can be seen in Figure 4.11. Overall, the average classification accuracy obtained with  $\hat{\tau} = 14$  is 86.8%, which is about 3% better than that obtained with  $\hat{\tau} = 28$ , and is higher than previously reported works [58][113][15].

Table 4.1: Classification Accuracy for Classification of Mental Tasks For All Subjects using 1-NN Classifier and 10-fold Cross-Validation.

Task Combination	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Mean	
	$\hat{\tau}=14$	$\hat{\tau}=28$	$\hat{\tau}=14$	$\hat{\tau}=28$	$\hat{\tau}=14$	$\hat{\tau}=28$	$\hat{\tau}=14$	$\hat{\tau}=28$	$\hat{\tau}=14$	$\hat{\tau}=28$	$\hat{\tau}=14$	$\hat{\tau}=28$	$\hat{\tau}=14$	$\hat{\tau}=28$	$\hat{\tau}=14$	$\hat{\tau}=28$
base-count	100%	100%	90%	90%	80%	60 %	100%	100%	70%	70%	90%	80%	90%	90%	<b>88.6%</b>	<b>84.3%</b>
base-letter	90%	90%	100%	100%	90%	90%	90%	100%	90%	70%	90%	50%	90%	70%	<b>91.4%</b>	<b>81.4%</b>
base-math	90%	100%	80%	90%	90%	70%	80%	100%	70%	80%	90%	90%	80%	90%	<b>82.8%</b>	<b>88.6%</b>
base-rot	100%	100%	90%	100%	60%	60%	80%	80%	80%	80%	90%	100%	100%	100%	<b>85.7%</b>	<b>88.6%</b>
letter-count	100%	100%	60%	90%	50%	50%	100%	100%	70%	70%	90%	80%	100%	80%	<b>81.4%</b>	<b>81.4%</b>
letter-rot	100%	100%	100%	100%	60%	70%	80%	60%	90%	90%	100%	90%	100%	100%	<b>90%</b>	<b>87.1%</b>
math-count	100%	100%	70%	60%	50%	50%	80%	70%	60%	60%	100%	100%	100%	100%	<b>80%</b>	<b>77.1%</b>
math-letter	100%	100%	100%	100%	80%	80%	100%	90%	100%	100%	100%	100%	100%	100%	<b>97.1%</b>	<b>95.7%</b>
math-rot	100%	100%	90%	90%	90%	60%	80%	60%	80%	90%	100%	70%	80%	90%	<b>88.6%</b>	<b>80%</b>
rot-count	80%	70%	100%	80%	50%	40%	80%	90%	70%	70%	100%	100%	100%	100%	<b>82.8%</b>	<b>78.6%</b>
Average	<b>96%</b>	<b>96%</b>	<b>88%</b>	<b>90%</b>	<b>70%</b>	<b>63%</b>	<b>87%</b>	<b>85%</b>	<b>78%</b>	<b>78%</b>	<b>95%</b>	<b>86%</b>	<b>94%</b>	<b>92%</b>	<b>86.8%</b>	<b>84.3%</b>

Tasks: base=baseline; math=multiplication; rot=rotation; count=counting;

The average classification accuracy for subject 3 is relatively low, having a value of 70%. This is due to three task-combinations having a low classification accuracy of 50%. For these task combinations, the following approach was used as a remedial measure. For both values of  $\hat{\tau}$ , the feature vectors obtained from  $\tau$ -functions  $T_1$  and  $T_2$  were used independently for classification. For the *math-count* task combination, the classification accuracy increased from 50% to 80% when only the feature vector obtained from  $T_2$  extracted with  $\hat{\tau} = 14$  was used. There was no change in the classification accuracy for the other two tasks, using either  $\tau$ -function obtained with either value of  $\hat{\tau}$ . Therefore, decomposition was performed using a lower value of  $\hat{\tau} = 7$ , which corresponds to a frequency separation value of  $F = 16$  Hz. For this value of  $\hat{\tau}$ , there was no change in classification accuracy for the task combinations *rot-count* and *letter-count* using feature vectors from both  $T_1$  and  $T_2$  together, or the feature vector from  $T_2$  individually. However, using only the feature vector obtained from  $T_1$ , the classification accuracy increased to 70% and 60% for the *rot-count* and *letter-count* task combinations respectively.

This shows that for *rot-count* and *letter-count* task combinations in case of subject 3, the de-trended component  $T_1$  containing the higher frequency components of the signals leads to better classification accuracy, whereas the lower frequency signal trend contained in  $T_2$  is more relevant for the *math-count* task combination. Importantly, the used  $\tau$ -functions in this case were obtained with different values of  $\hat{\tau}$ . This also demonstrates the flexibility of the method, whereby decomposition can be adapted to different time-scales, and either of the de-trended component, or the trend, can be used to extract discriminatory features for classification.

The average classification accuracy of nearly 87% obtained in this experiment improves on the classification accuracy of 85% reported in Section 2.4.1.6 of Chapter 2. However, the methodology presented in this experiment (denoted by **Experiment A**) improves on the previous EMD-based approach (denoted by **Experiment B**, and described in Section 2.4.1 of Chapter 2) in the following ways:

1. **Experiment B** uses EMD to decompose the EEG signals into at most  $\log_2(N)$  IMFs

[6], where  $N$  is the length of the signals. On the other hand, **Experiment A** uses EMD-MPS to decompose the signals into two  $\tau$ -functions, which makes the decomposition computationally more efficient.

2. **Experiment B** obtains the de-trended part of the signal and the signal trend by partial reconstruction based on checking a criterion. This criterion needs to be checked for every set of signals, and will change for different sets, e.g. if signal lengths are different. **Experiment A**, however, uses time-scale based de-trending of signals, which separates the trend according to a frequency separation criterion. Not needing to check a criterion, and not having to perform partial reconstruction after decomposition, makes **Experiment A** more computationally efficient and, importantly, portable to different signal types.
3. **Experiment B** uses two feature vectors of dimensions 30 and 18, whereas both feature vectors used in **Experiment A** have dimension 12. The improvement in classification accuracy in **Experiment A** is obtained with a reduced dimension feature vector.
4. Importantly, **Experiment A** allows flexibility in feature extraction, by allowing different time-scale based decompositions, to deal with difficult classification cases, as demonstrated earlier in this Section. This is a unique feature of the methodology.

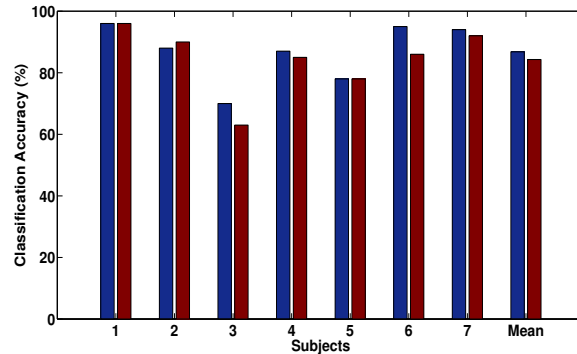


Figure 4.11: Classification accuracy for all subjects corresponding to  $\hat{\tau} = 14$  (left bars) and  $\hat{\tau} = 28$  (right bars).

#### **4.2.4.5 Experiment summary**

This Section presented a simple and effective method for mental task classification using EMD-MPS in conjunction with the TEO. This method is distinguished by the simplicity and flexibility of the decomposition, as well as the low dimensions of the feature vector. The efficacy of the feature vectors obtained is demonstrated by the high classification accuracy for one-versus-one mental task classification using a simple classifier. The average classification accuracy obtained is higher than previously reported results using similar approaches on the same data.

### **4.2.5 Experiment: EEG Seizure Detection and Epilepsy Diagnosis**

#### **4.2.5.1 Background**

Epilepsy is a neurological disorder affecting a very large number of people worldwide [59]. There is considerable research concerned with computer-based methods for seizure detection and epilepsy diagnosis using EEG signals [115]. Given the non-stationary and non-linear nature of EEG signals, signal processing methods for non-stationary signal analysis, such as EMD, time-frequency analysis, and wavelets, have been frequently used for automated seizure detection using EEG signals, e.g. [59][115][116][117]. The adaptive nature of EMD-based decomposition methods makes them particularly suitable for non-linear signal analysis [15].

In this experiment, a novel method for epilepsy diagnosis and seizure detection using EEG signals based on EMD-MPS is presented. The first part of the presented methodology consists of using EMD-MPS to decompose EEG signals into a de-trended component, and a trend, using a frequency separation criterion. Features are then obtained from the two decomposed components, which form feature vectors used for classification using a linear classifier, namely the 1-NN classifier. A schematic representation of the proposed methodology is shown in Figure 4.12.

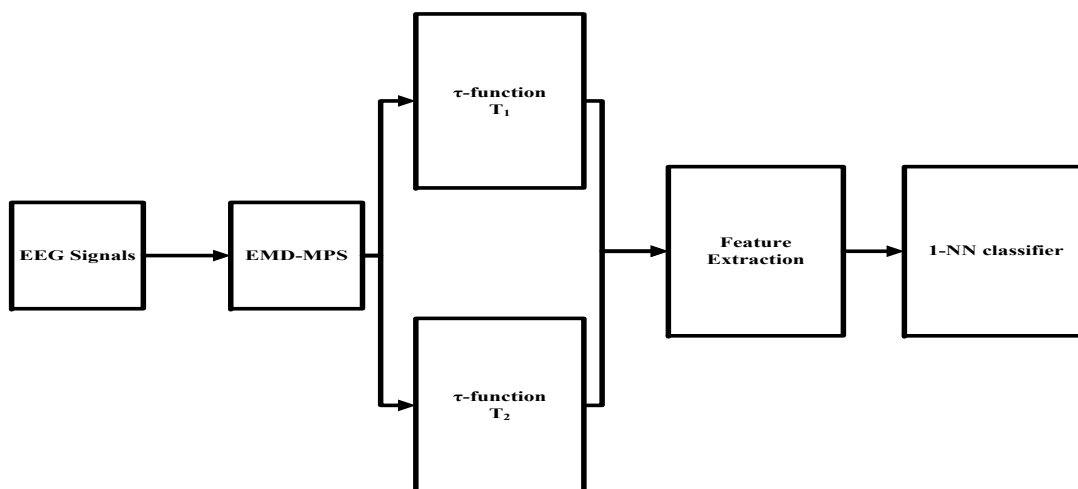


Figure 4.12: Diagram representing the methodology described in this experiment.

#### 4.2.5.2 Data

The EEG signals used in this experiment come from the data made public by the University of Bonn [118]. The data consists of EEG signals in five sets named A, B, C, D and E. Signals in sets A and B are from epilepsy free volunteers, whereas signals in sets C, D and E come from epilepsy patients. The signals in sets C and D have been recorded during epilepsy-free intervals, with set E containing only seizure signals. In each set there are 100 scalp EEG signals of 23.6 seconds duration sampled at 173.61 Hz, with each signal having 4097 samples, and a spectral bandwidth ranging from 0.5 Hz to 85 Hz. This data has been used in numerous previous studies (e.g. [26][115][116][117][119]), but in general many of the previous studies have tested classification only between sets A and E, and between sets ABCD and E. For this experiment, the proposed methodology is applied for testing binary classification in the following three scenarios:

1. Classification between sets A and E. This is the most commonly used scenario in previous works, and has been used to test the efficacy of proposed methodologies for seizure detection.
2. Classification between sets AB and CD. This is used to test performance of the method

in classifying signals mixed with different observational conditions or recordings at different spatial locations. Also, this scenario is relevant for the case of epilepsy diagnosis, as sets AB contain normal signals, and sets CD epileptic signals.

3. Classification between sets ABCD and E. This is relevant in terms of a seizure detection scenario, and also has relevance for clinical applications [116].

#### 4.2.5.3 Decomposition using EMD-MPS

EMD-MPS is used for a time-scale based de-trending of EEG signals, such that each EEG signal is decomposed into one  $\tau$ -function  $T_1$  representing the de-trended signal containing the higher frequency components, and another  $\tau$ -function  $T_2$  representing the trend of the signal. As in the experiment presented in Section 4.2.4, this is done by appropriate selection of a value of  $\tau$  according to Equation 4.1.

A value of  $\tau$  for decomposition of EEG signals using EMD-MPS was found as follows. First, a frequency separation value of  $F = 8$  Hz was chosen, such that  $T_1$  contains frequency components with frequency values greater than 8 Hz, and  $T_2$  represents the trend containing frequencies lower than 8 Hz. The value of  $F = 8$  Hz was objectively selected based on the highest classification accuracy obtained compared to other values of frequency separation evaluated.

Using Equation 4.1 and the signals' sampling frequency value of  $F_s = 173.16$  Hz, the value of  $\tau$  obtained for  $F = 8$  Hz is given by  $\tau = 21.6$ . However, for decomposition, the value  $\hat{\tau}$ , which is  $\tau$  scaled by a constant  $k$  is used, as described in the previous experiment. A good estimate for the value of  $k$  is given by  $k \approx 0.44$  [114], such that  $\tau = 21.6$  corresponds to value of  $\hat{\tau} = 9.5$ . Therefore, all EEG signals were decomposed using  $\hat{\tau} = 9.5$ . In this regard, Figure 4.13 shows an example EEG signal from set C, and the  $\tau$ -functions  $T_1$  and  $T_2$  obtained with a value of  $\hat{\tau} = 9.5$ .



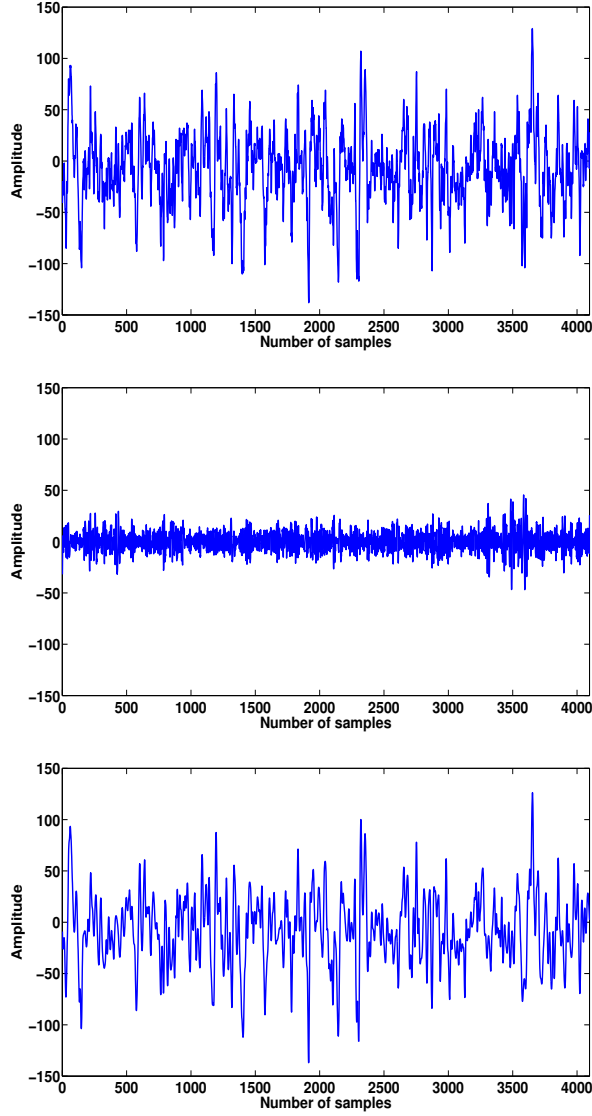


Figure 4.13: EEG signal from an epilepsy patient from set C (top).  $\tau$ -function  $T_1$ , representing de-trended signal (middle).  $\tau$ -function  $T_2$ , representing the trend (bottom). These  $\tau$ -functions have been obtained using a value of  $\tau = 21.6$  ( $\hat{\tau} = 9.5$ ) corresponding to a frequency separation value  $F = 8$  Hz.

#### 4.2.5.4 Feature analysis

After decomposition of the EEG signals into two  $\tau$ -functions each, a total of four features are obtained. One feature is extracted from the  $\tau$ -functions, and the remaining three from the frequency-domain representation of the  $\tau$ -functions obtained using the discrete Fourier

transform (DFT). Computationally fast implementations of the DFT algorithm are available in different software packages, hence this approach is expected to be computationally more efficient than time-frequency and wavelet decomposition based approaches, e.g. [115][117].

The DFT of both  $\tau$ -functions  $T_i$  results in a complex-valued function  $\hat{\mathbf{F}}_i$ . The real-valued single-sided amplitude spectrum of the  $\tau$ -functions, given by  $\hat{f}_i$ , is then obtained by taking the absolute value of  $\hat{\mathbf{F}}_i$ . Three features are subsequently extracted from  $\hat{f}_i$ .

The four extracted features are listed below:

1. The energy  $E_i$  of the  $\tau$ -functions  $T_i$ , given by:

$$E_i = \sum_{n=1}^N T_i^2[n], \quad i = 1, 2 \quad (4.4)$$

where  $N$  is the length of  $T_i$ .

2. The sum of the amplitude spectrum,  $S_{\hat{f}_i}$ , calculated as:

$$S_{\hat{f}_i} = \sum_{n=1}^M \hat{f}_i[n], \quad i = 1, 2 \quad (4.5)$$

3. The sparsity of the amplitude spectrum,  $SP_{\hat{f}_i}$ , calculated as:

$$SP_{\hat{f}_i} = \frac{\sqrt{M} - (\sum_{n=1}^M \hat{f}_i[n]) / \sqrt{\sum_{n=1}^M \hat{f}_i^2[n]}}{\sqrt{M} - 1}, \quad i = 2 \quad (4.6)$$

4. The sum of derivative of the amplitude spectrum,  $D_{\hat{f}_i}$ , calculated as:

$$D_{\hat{f}_i} = \sum_{n=1}^{M-1} \hat{f}_i[n]^2, \quad i = 1 \quad (4.7)$$

where  $\hat{f}_i[n] = \hat{f}_i[n+1] - \hat{f}_i[n]$ , for  $n = 1, \dots, M-1$ , and  $M$  is the length of  $\hat{f}_i$ .

The features  $SP_{\hat{f}_i}$  and  $D_{\hat{f}_i}$  have been successfully used previously for pathological speech classification [72]. In general, onset of seizure results in development of rhythmic activity

Table 4.2: Classification accuracy for the 3 scenarios (Section 4.2.5.2) using 1-NN classifier and 10-fold cross-validation

Sets:	A & E	AB & CD	ABCD & E
Classification Accuracy:	100%	99%	98.2%

typically containing multiple frequency components [120]. Therefore EEG signals containing seizure activity are expected to have lower values for the sparsity feature  $SP_{f_i}$  compared to values obtained from EEG signals without seizures. In order to reduce the number of feature vectors, this feature is extracted only from  $\tau$ -function  $T_2$ , which represents the low frequency trend of the EEG signals. This way, the difference in rhythmic activity occurring in the lower frequency range contained in the trend will be captured as a discriminative feature.

Similarly, the feature  $D_{f_i}$  is a good measure of abrupt changes and discontinuities in the signal representation in the frequency domain. This feature is extracted from the  $\tau$ -function  $T_1$ , which represents the de-trended part of the signals, and contains higher frequency components, where abrupt changes and discontinuities are more likely. In total, six feature vectors are obtained from the four features.

#### 4.2.5.5 Results

The four features described in the previous Section were used to form feature vectors in order to test classification between the sets in the three scenarios. In order to keep the overall methodology simple, a 1-NN classifier [81] was chosen, and classification results were estimated using the ten-fold cross validation method. The classification results thus obtained for the three scenarios are shown in Table 4.2, and discussed in the next sections.

**Classification between sets A and E:** The classification accuracy obtained for classification between sets A and E was 100%, thereby matching the results in recent works [115][116], and improving on previous results (e.g. as listed in [115]). Importantly, however, the 100% classification accuracy in this experiment has been obtained using just a single

feature, namely  $S_{\hat{f}_i}$ , the sum of the amplitude spectrum. This demonstrates the efficacy and simplicity of the proposed approach, as well as the utility of approaches based on adaptive signal decomposition.

**Classification between sets AB and CD:** The classification between EEG signals from healthy subjects (sets AB) and epileptic patients (sets CD) is relevant in terms of epilepsy diagnosis, and a 99% classification accuracy is obtained for this case using all four features. This is comparable to the 100% classification accuracy presented in [115], which is one work where this classification case is considered, using wavelet variances as features in conjunction with a 1-NN classifier. However, the method presented here achieves a high classification accuracy using a computationally simpler and adaptive decomposition approach compared to wavelet decomposition. Furthermore, it was found that removing the feature vector of feature  $E_i$  obtained from  $\tau$ -function  $T_2$  did not affect the classification accuracy. Hence only 5 feature vectors were used to obtain the classification accuracy of 99%.

**Classification between sets ABCD and E:** Classification between sets ABCD and E represents the case of seizure detection. For this scenario, and using the same five feature vectors as described in the previous Section, a classification accuracy of 98.2% was obtained. This result improves on the classification accuracy of 97.7% presented in [117], and is comparable to the classification accuracy of 98.3% presented in [116]. Both of these results, however, have been obtained with more involved signal analysis methods used with complicated classifiers, namely time-frequency analysis with feed-forward artificial neural network in the case of former, and multi-wavelet transform and entropy feature with multi-layer perceptron neural network (MLPNN) for the latter. The result of this experiment is also comparable to the classification accuracy of 100% presented in [115], which has been obtained with wavelet variance features extracted after wavelet decomposition of the signals and using a 1-NN classifier. In comparison with the methodology of [115], the signal analysis method of this experiment is computationally more simple and adaptive, as the decomposition does not require finding an appropriate basis.

#### 4.2.5.6 Experiment summary

In this experiment a novel EEG signal analysis method was presented. This method can be used for epilepsy diagnosis and seizure detection using a simple classification scheme. The signal analysis method is based on a novel decomposition scheme, which is characterized by its computational simplicity and adaptivity. Features are calculated from the decomposed components of the signal, and a 1-NN classifier is used to obtain high classification accuracy for different classification tasks. The classification results obtained are better than or comparable to other approaches using more involved signal analysis methods and complicated classifiers. A very important advantage of this method is the flexibility of decomposition and feature extraction, since decomposition is based on a frequency separation criterion, and different features may be extracted from the  $\tau$ -functions. This allows the method to be ported to a different set of biomedical signals in a different domain, as has been demonstrated in the context of mental task classification in Section 4.2.4.

### 4.3 Chapter Summary

This Chapter focussed on the de-noising and de-trending capabilities of EMD-MPS, and explained how a time-scale based de-noising and de-trending is made possible by EMD-MPS. The time-scale based de-noising and de-trending was demonstrated using synthetic signals, as well as different real life signals. Finally, the efficacy of time-scale based de-trending was illustrated using experiments for mental task classification and EEG seizure detection and diagnosis, both using EEG signals which are known to possess non-stationary and non-linear characteristics.

# Chapter 5

## Empirical Sparse Dictionary Learning

### 5.1 Background

Dictionary learning, or training, using a signal or a set of signals, as opposed to selecting a pre-existing basis (such as Fourier or Wavelet, among others), is a more recent approach to dictionary design that has been shown to better capture the structure and features specific to the signals being analyzed [121][122]. Most of the dictionary learning approaches have been designed in terms of sparse and redundant signal representation using over-complete dictionaries, with diverse applications such as signal and image compression, de-noising and face recognition, e.g. [123][124]. There is also considerable research for dictionary learning in the context of signal and image classification [125][126][127][128]. In such works, the problem of signal classification is also formulated as finding a sparse representation of the signal in a given over-complete dictionary. However, the difference from dictionary learning for sparse signal representation lies in the approach whereby redundant non-parametric dictionaries are first trained, and second, sparse signal representations are learned with an explicit discriminative goal. This makes such models different from reconstructive ones. For example, dictionary learning algorithms first learn dictionaries from the labeled data set, in either a supervised or weakly supervised setting, and use features of the sparse decomposition of the test signals for classification. These features can take the form of the reconstruction error or coefficients of orthogonal projections over atoms of a dictionary.

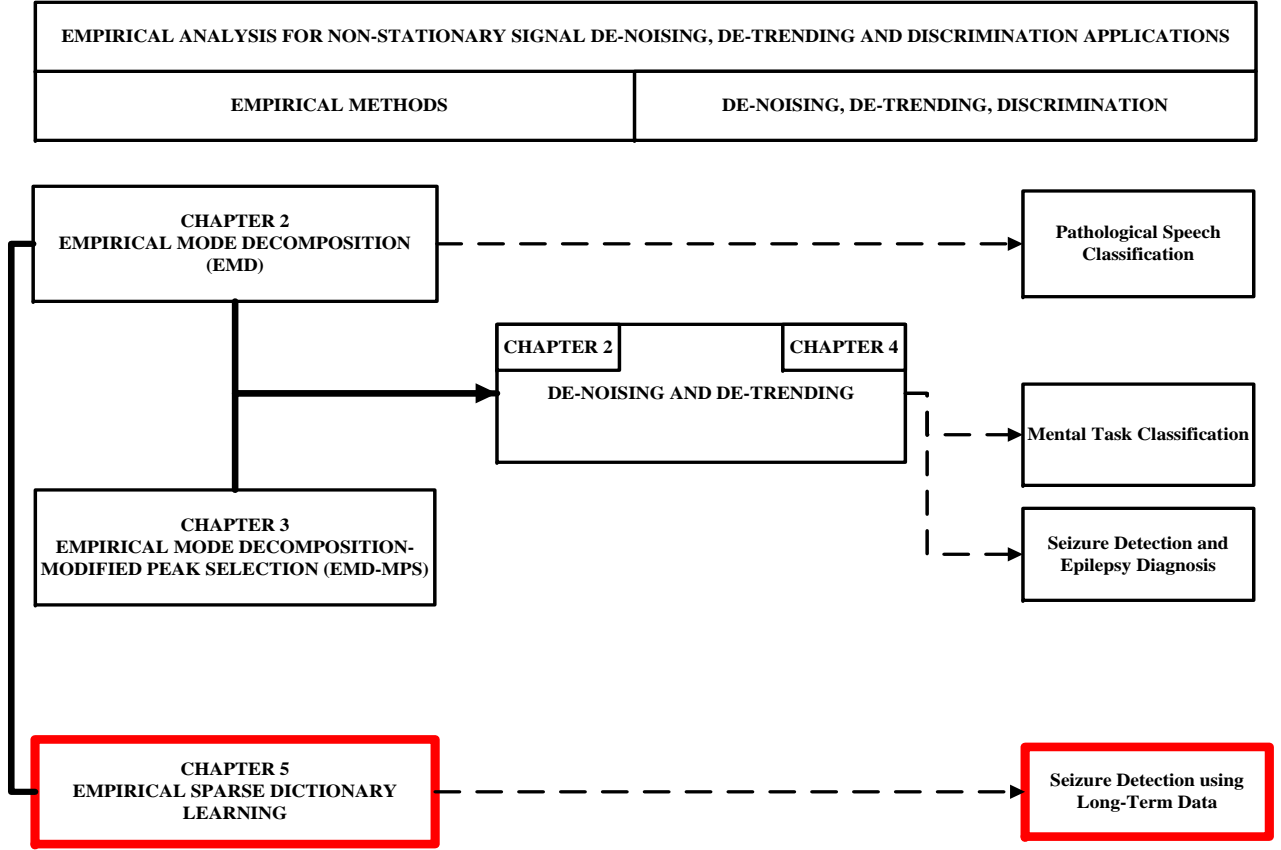


Figure 5.1: Chapter 5 presents details of the novel empirical sparse dictionary learning framework, which is based on signal decomposition using EMD. The use of the framework for signal classification is also described, including seizure detection using long-term data.

The reconstruction error, (given by  $\|\mathbf{x} - \mathbf{D}\alpha\|_2^2$ , where  $\mathbf{x}$ ,  $\mathbf{D}$  and  $\alpha$  are the signal, learned dictionary and coefficient vector, respectively) is a popular discriminant known to be effective for classification [127]. For example, given  $N$  signals belonging to  $N$  different classes, a dictionary could be learned for each class, which allows approximation of each signal with a constant sparsity using each dictionary to give  $N$  different reconstruction errors, which can then be used as classification features. In this context, the work in [127] proposes a measurement for the quality of sparse representation that caters to both the reconstruction error and the sparseness of the representation with respect to a learned dictionary, and shows this measurement to have discrimination power. Similarly, the reconstruction error is replaced

with a function representing the discriminative power in the objective function of the sparse representation in the work presented in [125], which makes the extracted feature (coefficient vector  $\alpha$ , given by  $\mathbf{x} = \mathbf{D}\alpha$ ) as discriminative as possible. In contrast to approaches where dictionary learning and classifier training are considered separate processes, more recent approaches utilize a mixed reconstructive and discriminative formulation, thereby simultaneously learning over-complete reconstructive and discriminative dictionaries [128].

Empirical Mode Decomposition (EMD) is a recent signal processing algorithm that was discussed in Chapter 2. EMD does not assume a pre-defined basis, rather the decomposition is data-adaptive in nature, and has been shown to be well-suited for analysis of non-stationary and non-linear signals. EMD decomposes a signal  $x(n)$  into a number of intrinsic mode functions (IMFs)  $a_q(n)$ ,  $q \in \{1, \dots, J\}$ , such that  $x(n) = \sum_{q=1}^J a_q(n)$  (ignoring the residue). Application of EMD results in separation of the temporal and spectral content of a signal into different IMFs. The lower index IMFs contain the faster oscillations present in the signal, whereas higher index IMFs contain the slower oscillations.

In this Chapter, a novel empirical framework for dictionary learning based on EMD is presented, with an application to signal classification. In this framework, training signals belonging to different classes are decomposed into IMFs using EMD, which form atoms of what is termed as a raw dictionary. This raw dictionary is then trained with the same training signals using a matching-pursuit like algorithm [129], which leads to a sparse learned dictionary. The sparsity of the dictionary is meant in terms of the dictionary dimensions, as the number of atoms of the raw dictionary is much smaller than the dimensions of the signals. The dictionary learning algorithm then further reduces the number of dictionary atoms. It should be mentioned here that the pursuit method is not for sparse coding, as generally used in dictionary learning approaches, but for learning a sparse dictionary. The under-complete trained dictionary so obtained can then be used for signal classification, whereby test signals of different classes are projected against the trained dictionary, and the projection coefficients thus obtained are used to classify the test signals.

Although the framework presented in this Chapter learns dictionaries empirically, the



use of the coefficient vector  $\alpha$  as a classification feature vector is also demonstrated. The coefficient vector is obtained using the trained dictionary. Further, even though the dictionary learning framework is not defined as reconstructive, signal reconstruction is demonstrated using properties of the learned dictionary. The potential of using the reconstruction error  $\epsilon$  as a feature for signal classification is also pointed out.

The use of EMD and IMFs to form dictionaries is scarce in literature. The work in [130] presents a decomposition methodology as an improvement over the original EMD method whereby an over-complete dictionary of analytically formulated IMFs is used for signal decomposition, with the approach formulated as a non-linear minimization problem. As has been checked in literature, there is no other work which uses IMFs for dictionary construction, specifically in the context of dictionary construction for signal classification. The framework presented in this Chapter does not seek to learn over-complete dictionaries which are reconstructive [131], or which combine reconstruction and discrimination [128]. Instead, the objective of this framework is to learn a sparse dictionary, which can be applied to signal discrimination tasks using traditional features (e.g. projection coefficients and coefficients vector). In this regard, EMD plays an important role in imparting discriminative capability to the learned dictionary, as using IMFs for dictionary formation, and learning the dictionary to select only the most relevant IMFs for signals of each class, allows a better capture of the discriminatory content of the signals in sparse dictionaries.

The utility of the proposed framework for signal classification tasks is tested using real-life EEG signals, and is described in Section 5.3. The classification accuracy obtained for these signals using different features validates the rationale behind the proposed empirical dictionary learning framework. Also, classification of biomedical signals in the context of dictionary learning approaches is an emerging area of research, and the proposed framework demonstrates the applicability of empirical approaches in this area. The results of investigation of application of the empirical dictionary learning framework for automatic seizure detection using long-term scalp EEG data will also be presented in Section 5.5. The next sections will present more details of the different aspects of the dictionary learning

framework.

## 5.2 Discriminative Dictionary Learning Algorithm

One starts with a training matrix  $\mathbf{X}_{Train}^c \in \mathbb{R}^{n \times k^c}$ , whose columns consist of  $k^c$  training signals  $x^c \in \mathbb{R}^n$  associated with class  $c$ , ( $c \in \mathbf{C}$ ,  $\mathbf{C} \supset \{c_1, \dots, c_K\}$ ), assuming there to be  $K$  classes. The first step of the methodology is to form a raw dictionary  $\mathbf{D}_{raw}^c = \{\psi^m\}_{m=1}^M$ , where the dictionary atoms  $\psi$  consist of IMFs, and  $M < n$ . The task is to learn a trained dictionary  $\mathbf{D}_{Train}^c = \{\psi^p\}_{p=1}^P$ , where  $P < M$ .

The dictionary formation and learning algorithm, called Discriminative Dictionary Learning Algorithm, consists of the following steps:

1. The signals  $x^c$  in  $\mathbf{X}_{Train}^c$  are decomposed into IMFs  $a_q$  using EMD, such that  $x^c = \sum_{q=1}^J a_q^c(n)$ . It should be noted that  $J$  is dependent on the implementation of the EMD algorithm, and might not be the same number for all signals  $x^c$ .
2. The IMFs thus obtained are used to form class-specific raw dictionaries  $\mathbf{D}_{raw}^c = [\mathbf{d}_1 | \mathbf{d}_2, \dots, | \mathbf{d}_L] \in \mathbb{R}^{n \times L}$ ,  $L = k^c \times J$ . The atoms arranged in columns are constrained to have  $l_2$ -norm less than or equal to 1, such that  $\mathbf{d}_r = \hat{\mathbf{a}}_q^c$ ,  $r \in \{1, \dots, L\}$ ,  $q \in \{1, \dots, J\}$ , and  $\hat{\mathbf{a}}_q^c = \mathbf{a}_q^c / \|\mathbf{a}_q^c\|_2$ .
3. The class specific dictionaries are merged together to form a combined raw dictionary  $\mathbf{D}_{raw}^c = [\mathbf{D}_{raw}^{c_1} | \mathbf{D}_{raw}^{c_2}, \dots, | \mathbf{D}_{raw}^{c_K}] \in \mathbb{R}^{n \times M}$ .

Starting from the raw dictionary  $\mathbf{D}_{raw}^c$ , the trained dictionary  $\mathbf{D}_{Train}^c$  is learned using the dictionary learning algorithm, which is described next.

### 5.2.1 Learning the Trained Dictionary

The learning algorithm uses a matching pursuit like algorithm to learn the trained dictionary  $\mathbf{D}_{Train}^c$ , which is shown in Algorithm 3. The termination of the algorithm is described next.

---

**Algorithm 3** Trained Dictionary Learning Algorithm

---

- 1: Initialize  $\mathbf{X}_{Train}^c$  {all training matrices},  $\mathbf{D}_{raw}^c = \{\psi^m\}_{m=1}^M$  {merged raw dictionary},  
 $\mathbf{D}_{Train}^c = []$  {empty matrix}
  - 2: **repeat** for each signal  $x^c$ :
  - 3:   Find the projection coefficient:  $\alpha_m = \langle x^c, \psi^m \rangle$
  - 4:   Select the dictionary atom  $\widetilde{\psi^m}$  with the highest value of  $|\alpha_m|$
  - 5:   **if**  $\widetilde{\psi^m}$  not in  $\mathbf{D}_{Train}^c$  **then** add  $\widetilde{\psi^m}$  to  $\mathbf{D}_{Train}^c$ :  $\mathbf{D}_{Train}^c = [\mathbf{D}_{Train}^c | \widetilde{\psi^m}]$
  - 6:   **end if**
  - 7:   Calculate the residue  $r^x = x^c - \widetilde{\psi^m} \langle x^c, \widetilde{\psi^m} \rangle$
  - 8:   Set the atom  $\widetilde{\psi^m} = 0$  in  $\mathbf{D}_{raw}^c$
  - 9:   Set:  $x^c = r^x$
  - 10: **until** Termination (see Section 5.2.1.1)
  - 11: Merge the class specific trained dictionaries to form a combined trained dictionary  $\mathbf{D}_{Train}^c$   
 $= [\mathbf{D}_{Train}^{c_1} | \mathbf{D}_{Train}^{c_2}, \dots, \mathbf{D}_{Train}^{c_K}] \in \mathbb{R}^{n \times P}$ .
- 

#### 5.2.1.1 Termination rules

There are two possible ways to terminate the Trained Dictionary Learning Algorithm:

1. The procedure shown in Algorithm 3 stops after a fixed number of iterations  $S$ . The number of iterations determines the size of the dictionary, as well as the "depth" of the residue. More iterations lead to a larger number of atoms in the dictionary (Section 5.2.2), and a "deeper" residue means that more higher index IMFs will also be added as dictionary atoms. This is due to Step 7 of Algorithm 3, which is similar to the decomposition step of EMD, given the subtraction of an IMF from the residue  $r^x$  after iteration  $x$ .
2. The second way of terminating the learning algorithm is to use a validation scheme to establish the number of iterations. The validation scheme described here is based on maximizing the distance between the vectors  $\Gamma_{c_i}$  and  $\Gamma_{c_j}$ , containing, respectively, the projection coefficients obtained using training signals  $\hat{x}^c$  from different classes  $c_i$  and  $c_j$ . Importantly, the signals  $\hat{x}^c$  have not been used for the initial dictionary formation or learning. For each signal  $\hat{x}^c$ , the projection coefficient  $\alpha_m = \langle \hat{x}^c, \psi^m \rangle$  is calculated, and added to the appropriate coefficients vector  $\Gamma_c$ . Letting  $S$  represent the number of iterations,  $\Gamma_c^S$  represents the vector of projection coefficients obtained after  $S$  iterations.

The number of iterations for termination of the algorithm is then decided according to the following criterion:

$$\max_S ||\Gamma_{c_i}^S - \Gamma_{c_j}^S||_2^2 \quad (5.1)$$

It should be mentioned here that other validation schemes are possible here as well, for example a scheme which maximizes the classification accuracy using a classification method.

## 5.2.2 Some Characteristics of the Learned Dictionary

### 5.2.2.1 Change in dictionary size

The raw dictionary  $\mathbf{D}_{raw}^C \in \mathbf{R}^{n \times M}$  has  $M$  atoms consisting of IMFs obtained from  $k$  training signals, each of which has  $n$  samples. Although the number of IMFs obtained through EMD is dependent on the EMD algorithm implementation, it has previously been established that the number of IMFs  $J$  obtained from a signal  $x(n)$  is such that  $J \leq \log_2(n)$  [104]. Therefore  $M \leq k \times \log_2(n)$ , where  $M < n$ . The trained dictionary  $\mathbf{D}_{Train}^C$ , learned using Algorithm 3, has  $P < M$  atoms. Since the same number  $k$  of training signals is used for learning the trained dictionary, a maximum of  $k$  atoms can be added to the trained dictionary after each iteration in the algorithm, as given by Step 4 of Algorithm 3. This means that the change in size of the dictionary, going from the raw dictionary to the trained dictionary, can be seen in the following ways:

1. The number of atoms  $M$  of the raw dictionary  $\mathbf{D}_{raw}^C$  can be at most  $k \times \log_2(n)$ . The number of atoms  $P$  of  $\mathbf{D}_{Train}^C$ , learned using Algorithm 3, can be at most  $k \times S$ , where  $S$  is the number of iterations after which Algorithm 3 terminates.
2.  $k \times S$  is an upper bound on the size of the trained dictionary  $\mathbf{D}_{Train}^C$ . This is due to Step 5 of Algorithm 3, whereby if the selected atom is already present in the dictionary, it is not added. Therefore the number of atoms added to the trained dictionary after each iteration can be less than  $k$ .
3. The change in size of the dictionary going from  $\mathbf{D}_{raw}^C$  to  $\mathbf{D}_{Train}^C$  can be represented as

the ratio:

$$\Delta(M \rightarrow P) = \frac{k \times \log_2(n)}{k \times S} \quad (5.2)$$

Equation 5.2 gives a good estimate of the magnitude of the decrease in dictionary size going from the raw to the trained dictionary. In general, the number of atoms  $M$  of the raw dictionary will be less than  $k \times \log_2(n)$ . Similarly, as mentioned in Point 2 previously, the number of atoms  $P$  in the trained dictionary will in general be less than  $k \times S$ . Therefore the actual magnitude of decrease may be more or less than the quantity in Equation 5.2. The maximum decrease in dictionary size occurs when Algorithm 3 terminates after  $S = 1$  iterations, since in this case  $P \ll M$ .

This is depicted in Figure 5.2, which compares the actual increase in dictionary size with the expected increase in size during the learning phase using examples presented in Section 5.3. It can also be seen from Figure 5.2 that the actual increase in dictionary size is nearly linear and matches the expected increase.

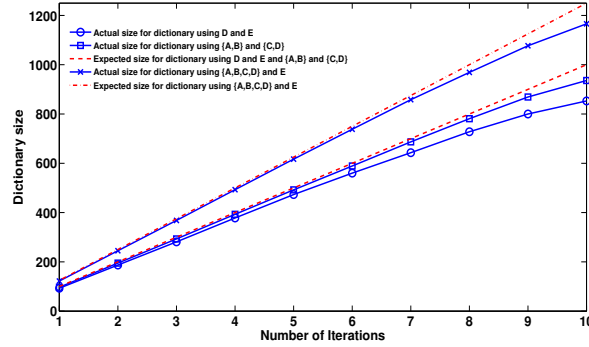


Figure 5.2: Increase in size of learned dictionary plotted against the number of iterations in Algorithm 3, showing expected and actual increase in dictionary size using examples described in Section 5.3.

### 5.2.2.2 Increase in dictionary learning time with number of iterations

The number of iterations  $S$  required to terminate Algorithm 3 has implications for the computational time it takes to train the learned dictionary. However, as explained in Section

5.2.2.1, the maximum number of atoms that can be added to the dictionary after each is given by  $k$ . Therefore one expects the computational time for dictionary learning to increase linearly with the increase in the number of iterations. This is depicted in Figure 5.3, which shows a nearly linear increase in computation time of Algorithm 3 with increase in the number of iterations using data described later Section 5.3.1.

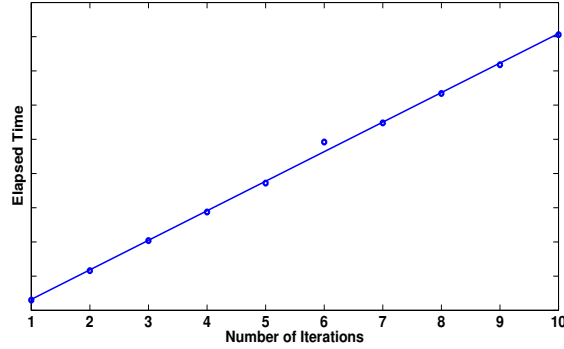


Figure 5.3: Increase in computational time of Algorithm 3 with increase in number of iterations. The solid line represents a least-squares fit. The values have been obtained using data described in Section 5.3.1.

### 5.2.2.3 Increase in computational time for signal projection with increase in learned dictionary size

The size of the trained dictionary  $\mathbf{D}_{Train}^C$  has implications for the computational time it requires to project test signals against the dictionary to obtain the projection coefficients. As with the nearly linear increase in dictionary size with increase in number of Algorithm 3 iterations presented in the previous Section, one can expect the computational time for signal projection against the trained dictionary to increase linearly with an increase in dictionary size for the same number of signals. This can be inferred from the linear increase in dictionary size presented in Section 5.2.2.1. This holds to be true in practice also, as is shown in Figure 5.4, which uses test signals and a trained dictionary created using data described in Section 5.3.1.

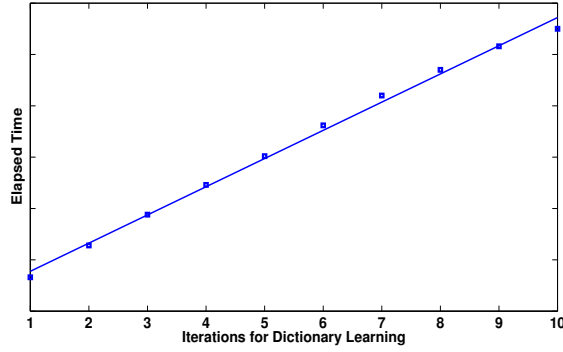


Figure 5.4: Increase in computational time for projecting a fixed number of signals against the trained dictionary  $\mathbf{D}_{Train}^C$ , with an increase in dictionary size, represented here by the number of iterations for Algorithm 3. The solid line represents a least-squares fit, and the values have been obtained using data described in Section 5.3.1.

#### 5.2.2.4 Semi-orthogonality of trained dictionary

Although the trained dictionary  $\mathbf{D}$  has been constructed empirically, it is nearly semi-orthogonal, with  $\mathbf{D}^T \mathbf{D} \approx I$ . Therefore it is possible to obtain the coefficients vector  $\alpha$  using the following relation:

$$\mathbf{x} = \mathbf{D}\alpha \quad (5.3)$$

In the context of the presented approach, it is possible to obtain the coefficients vector using the relation shown below:

$$\alpha = (\mathbf{D})^\dagger \mathbf{x} \quad (5.4)$$

where  $\mathbf{x}$  represents a testing signal, and  $(\mathbf{D})^\dagger$  is the left pseudo-inverse of the trained dictionary  $\mathbf{D}_{Train}^C$ .

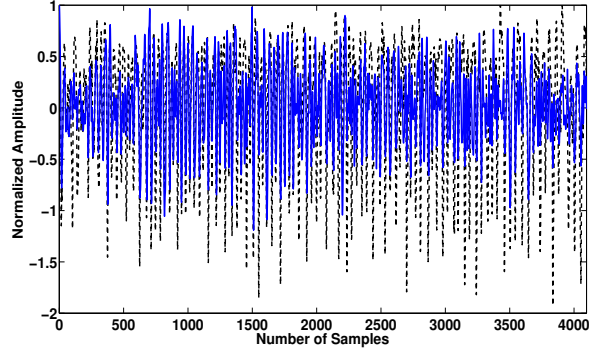
In addition, even though the learned dictionary  $\mathbf{D}$  is not designed as a reconstructive dictionary, it is possible to obtain a reconstruction of the signal  $\mathbf{x}$  using the relation:

$$\hat{\mathbf{x}} = \mathbf{D}(\mathbf{D})^\dagger \mathbf{x} \quad (5.5)$$

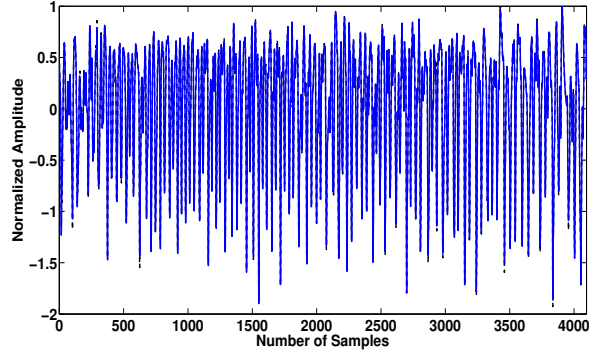
The reconstruction error, given below in Equation 5.6, is however expected to be large, especially when using a trained dictionary with a smaller number of atoms, which would

typically result if the dictionary learning algorithm terminates after a smaller number of iterations.

$$\epsilon = ||\hat{x} - x||_2 \quad (5.6)$$



(a)



(b)

Figure 5.5: Original signal in dashed line superimposed with the reconstructed signal obtained using the relation in Equation 5.5. For the plot in (a), the trained dictionary was obtained after one iteration of Algorithm 3, and after 10 iterations for the plot in (b).

Using a test signal from one class and a two-class trained dictionary constructed from the data described in Section 5.3.1, and using the relation in Equation 5.5, the reconstructed signal along with the original signal is shown in Figure 5.5 (a). As can be seen from this figure, the reconstruction error is large. However, the reconstruction error can be expected to decrease with a trained dictionary having a larger number of atoms. This is indeed the case, as can be seen in Figure 5.5 (b), where the reconstructed signal has been obtained



using a dictionary trained after 10 iterations of Algorithm 3.

Using the same signal as used in Figure 5.5, and a trained dictionary obtained after different number of iterations, the decrease in the reconstruction error with increase in dictionary size, as indicated by the iteration number, is shown in Figure 5.6.

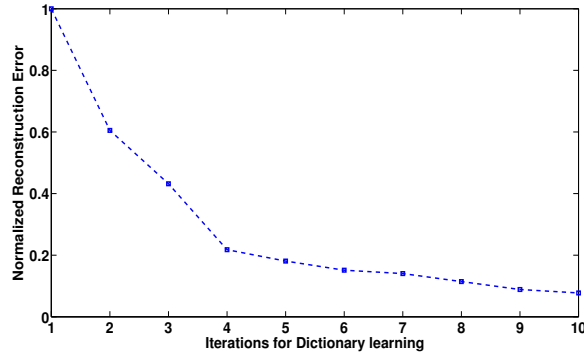
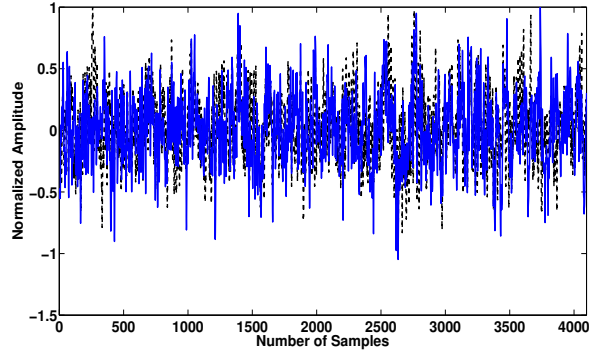


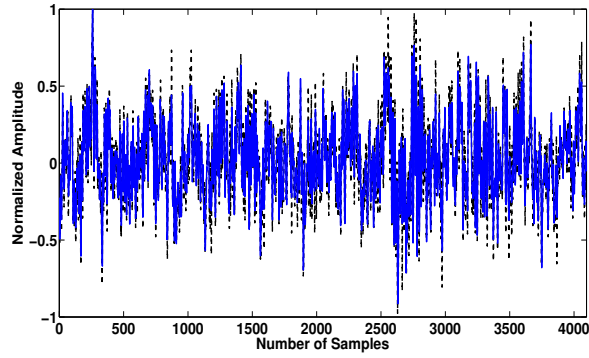
Figure 5.6: Decrease in reconstruction error with an increase in size of the trained dictionary, as indicated by the iteration number.

Using a test signal from the second class and the same two-class trained dictionary, the reconstructed signal along with the original signal is shown in Figure 5.7 (a). The reconstructed signal obtained using a dictionary trained after 10 iterations of Algorithm 3 is shown in Figure 5.7 (b). As before, the decrease in the reconstruction error is evident. Similarly, as in the previous case, the decrease in reconstruction error with an increase in size of the trained dictionary is indicated in Figure 5.8.

The graphs in Figures 5.6 and 5.8 are unique to the behaviour of the signals they represent, and also to the trained dictionary being used to demonstrate the reconstruction. However, decrease in the size of reconstruction error with an increase in trained dictionary size holds for all signals. This is to be expected, since, as explained in Section 5.2.1.1, more iterations of the procedure shown in Algorithm 3 result in a larger number of higher index IMFs being added to the trained dictionary. Furthermore, Step 7 of Algorithm 3 is similar to the decomposition step of EMD, therefore a larger number of iterations will result in a larger dictionary containing IMFs with indices covering a wide range over all possible index values. In this regard, Figure 5.9 shows four different atoms from the trained dictionary



(a)



(b)

Figure 5.7: Original signal in dashed line superimposed with the reconstructed signal obtained using the relation in Equation 5.5. For the plot in (a), the trained dictionary was obtained after one iteration of Algorithm 3, whereas for the plot in (b) the trained dictionary was obtained after 10 iterations.

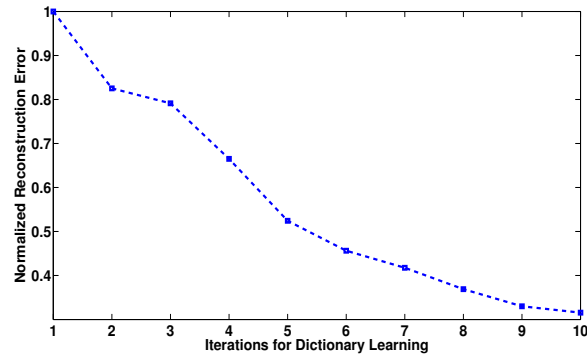


Figure 5.8: Decrease in reconstruction error with an increase in size of the trained dictionary, as indicated by the iteration number.

$\mathbf{D}_{Train}^C$  learned after 10 iterations of Algorithm 3. The plots (a) and (d) in Figure 5.9 represent IMFs of the first and last indices respectively, whereas plots (b) and (c) are IMFs of intermediate indices. In a trained dictionary learned after fewer iterations of Algorithm 3, most of the atoms in the trained dictionary would consist of IMFs of lower indices only, thereby leading to a large signal reconstruction error.

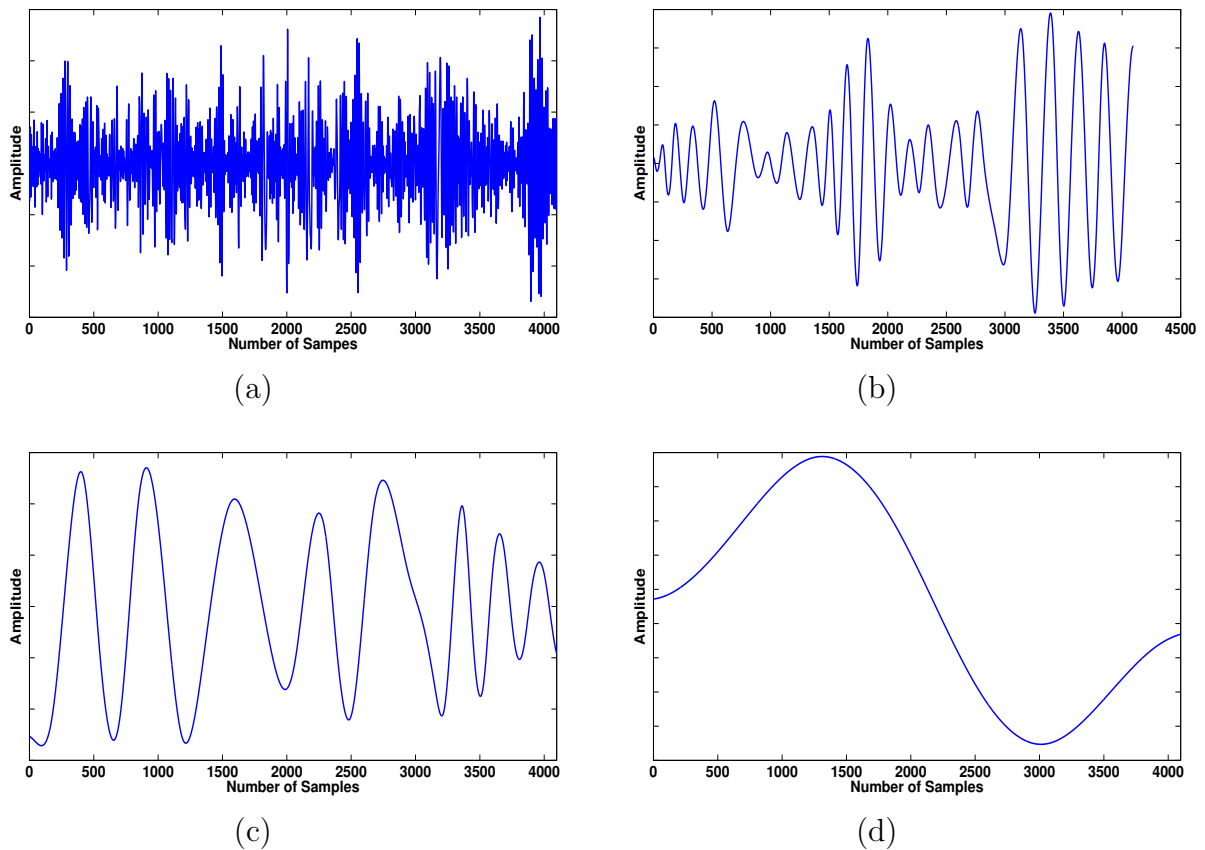


Figure 5.9: Examples of dictionary atoms from a trained dictionary  $\mathbf{D}_{Train}^C$  learned after 10 iterations of Algorithm 3.

### 5.2.2.5 Use of features other than the projection coefficient

Traditional approaches of classification using dictionary learning (e.g. [125]) use the relationship  $\mathbf{x} = \mathbf{D}\alpha$  to employ the coefficients vector  $\alpha$  as a feature vector obtained using a learned dictionary  $\mathbf{D}$  and a signal  $\mathbf{x}$ . As mentioned previously in Section 5.2.2.4, it is possible to obtain this feature vector with the proposed approach as well, using the relation in Equation 5.4. The use of  $\alpha$  as a feature vector will be described in Section 5.3.

Another feature that may be used for signal classification using the proposed approach is the reconstruction error given in Equation 5.6. As demonstrated in Figures 5.6 and 5.8, signals from different classes are expected to have different reconstruction behaviours using a trained dictionary according to the relation in Equation 5.5, which has the potential to make the reconstruction error a discriminant feature.

### 5.2.3 Signal Classification Using the Trained Dictionary

The trained dictionary can be used for signal classification by using coefficients of projections of test signals over the atoms of the trained dictionary as features to discriminate between different classes. The projection coefficients are expected to form distinct clusters in the feature space, and a classifier can then be used to classify between the projection coefficients obtained from signals of different classes. This expectation follows from the application of EMD to the training signals, which results in the different characteristics of signals of each class being decomposed into different IMFs. The dictionary learning algorithm then selects only the most relevant IMFs of each class [132], which not only results in a sparse dictionary, but one whose atoms contain the most relevant characteristics of the different classes being analyzed.

The projection coefficients are obtained in the same way as described in Step 3 of Algorithm 3. For each test signal  $x_t^c$  and all atoms  $\psi^m$  of the trained dictionary, the projection coefficient  $\alpha_m = \langle x_t^c, \psi^m \rangle$  with the maximum value is selected as the feature point. Similarly, it is possible to obtain the coefficients vectors using the test signals, as described in Section 5.2.2. Next the use of these features to classify test signals in different classification

scenarios is described.

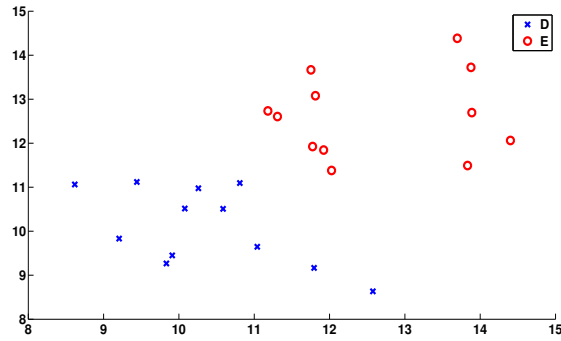


Figure 5.10: Scatter plot of projection coefficients of test signals from two classes, D (shown as crosses) and E (shown as circles). Clear separation of the signals in the feature space is visible.

## 5.3 Experiments

To demonstrate the methodology presented in the previous sections, real-life data in the form of EEG signals is used. The details of these EEG signals have been described previously in Section 4.2.5.2 of Chapter 4. This data consists of 5 sets of signals named A, B, C, D and E. Signals in sets A and B are from epilepsy free patients, whereas signals in sets C, D and E are from epilepsy patients. The signals in sets C and D have been recorded during epilepsy-free intervals, and signals in set E are seizure signals. For experiments using the proposed methodology, binary classification is tested in the following three scenarios:

1. Scenario I: Classification between signal sets D and E.
2. Scenario II: Classification between signal sets ABCD and E.
3. Scenario III: Classification between signal sets AB and CD.

The first two scenarios are relevant in the context of seizure detection, whereas the third scenario is meant to test the performance of the method for epilepsy diagnosis.

### 5.3.1 Scenario I

For classification between sets D and E, 50 signals from each set were decomposed into IMFs using EMD, and used to form the raw dictionary consisting of 1067 atoms. The dictionary was trained using these 50 signals as described in Algorithm 3, and the algorithm was terminated by the second method in Section 5.2.1.1. For this, 25 signals of each class were used in the validation step to estimate the maximum separation between the projection coefficient vectors, which was obtained after 1 iteration. The dictionary learning stage resulted in a 92% decrease in dictionary size, with the trained dictionary consisting of 83 atoms. Finally, 24 signals of each class were used as test signals, and projected against the trained dictionary to obtain the projection coefficients. Figure 5.10 shows a grouped scatter plot of the projection coefficients so obtained from test signals of sets D and E, clearly showing the separation between the classes in the feature space, thereby making the use of a classifier unnecessary.

### 5.3.2 Scenario II

To test the second scenario related to classification between sets ABCD and E, 25 signals from each individual set were decomposed into IMFs and used to form the raw dictionary consisting of 1303 atoms. The raw dictionary was trained using these 125 signals as per Algorithm 3, which terminated after 3 iterations according to Equation 5.1 using 5 validation signals from each of the individual sets in ABCD, and 20 validation signals from E. The dictionary learning step resulted in a trained dictionary containing 368 atoms, representing a reduction in dictionary size of 72%. Figure 5.11 shows examples of dictionary atoms taken from this trained dictionary. For testing, the remaining unused 70 signals in each individual signal set in ABCD, and the 55 unused signals in set E were used. These 335 test signals (280 belonging to set ABCD and 55 belonging to set E) were projected against the trained dictionary, and the projection coefficients obtained were used to test the discrimination performance between sets ABCD and E. For this case, the highest classification accuracy was obtained using a 3-NN classifier, with a correct classification rate of 95.8% obtained using 10-fold stratified cross-validation. The sensitivity and specificity values were 92.7%

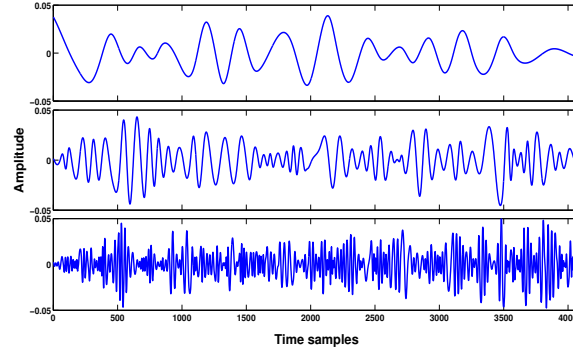


Figure 5.11: Example of dictionary atoms consisting of IMFs of decomposed signals taken from the trained dictionary described in Section 5.3.2

and 96.4% respectively.

### 5.3.3 Scenario III

The third scenario involves testing the proposed method for classification between classes AB and CD, representing normal and inter-ictal signals respectively. For this purpose, a raw dictionary was created using IMFs from 25 signals of each of the signal classes A, B, C and D, which consisted of 1050 atoms. A trained dictionary was learned using these 25 signals from each class, and 15 further validation signals from each class were used for terminating the algorithm, which, in this case, terminated after 1 iteration according to Equation 5.1. This resulted in a trained dictionary with 97 atoms, representing a decrease in dictionary size of 91%. Using the remaining 40 signals in each signal set as test signals, the coefficients of projections were used as features to test classification between classes AB and CD. For this scenario, the highest classification accuracy of 83.8% was obtained using linear discriminant analysis tested with 10-fold stratified cross-validation, giving a sensitivity of 87.5% and specificity of 80%.

In addition, to demonstrate the use of the coefficients vector  $\alpha$  as a feature vector, obtained using  $\alpha = (\mathbf{D})^\dagger \mathbf{x}$ , as described in Section 5.2.2, the classification performance using this feature vector was also tested. In this case the linear classifiers resulted in lower classification accuracies, close to 70% correct classification rate. However, the highest classification

accuracy of 89.6% was obtained using a Support Vector Machine (SVM) classifier using a quadratic kernel, having a sensitivity of 94.2% and a specificity of 85%.

### 5.3.4 Discussion

The classification accuracy for the three test scenarios presented previously compares favourably to the results of recent studies using the same EEG data, e.g. [117][116][115][26], where classification accuracy and sensitivity and specificity values in the range of 90% to 100% have been demonstrated. These values have been obtained using methodologies based on wavelets, time-frequency analysis, and higher-order statistical analysis in the EMD domain, and using a variety of classifiers, such as k-NN classifier, artificial neural network, and multi-layer perceptron neural network.

The method presented here is classifier agnostic, and it is possible to choose a classifier empirically, as has been done with the three scenarios presented. In the first scenario, clear separation of the two classes was visible in the feature space. For the second and third scenarios, linear classifiers provided high classification accuracy, whereas the highest classification accuracy when using coefficients vector  $\alpha$  as a feature vector in the third scenario was obtained using a SVM classifier with a quadratic kernel.

The main advantage of the proposed method lies in being able to create a dictionary using a small number of signals. This is also relevant in the context of the computational time requirements, since EMD decomposition of signals is the most computationally expensive step of the overall methodology. The dictionary learning step then further reduces the size of the dictionary, retaining only the atoms which have strong correlation with the training signals of different classes. A small number of dictionary atoms makes projection of test signals computationally fast, which is advantageous for long-term data, as presented in [106]. In such a scenario, the dictionary could be formed and trained using a small segment of the available data. Features for classification could be extracted from the remaining data by projecting the data against the trained dictionary, without having to manipulate or transform the whole data into a different domain, as is done in almost all current approaches.



## 5.4 Conclusions

The method presented in the previous sections represents a novel approach to sparse dictionary learning for signal classification based on EMD, with sparsity understood in terms of the low number of dictionary atoms compared to the signal dimensions. The dictionary atoms consist of IMFs, obtained from signals using EMD, hence the dictionary is learned from the data, and not formed using a pre-defined basis, making it more adaptive to the data to be analyzed. The dictionary learning framework is also distinguished by its simplicity, as the aim of the framework is not to learn reconstructive, or jointly reconstructive and discriminative dictionaries. Instead, the framework focusses on learning a sparse, under-complete dictionary, whose small number of atoms contain representative characteristics of their respective classes. At the same time, the framework allows the use of the learned dictionary to extract traditional features, which can be used for signal classification using an appropriate classifier. The utility of the empirical framework presented here was demonstrated using real-world data, and it was shown that it is possible to extract features using the under-complete dictionary which can be used for signal classification.

## 5.5 Automatic Seizure Detection Based on Empirical Dictionary Learning Using Long-term EEG Data

Automatic seizure detection has important implications for long-term monitoring, diagnosis and rehabilitation in case of epilepsy [106]. The traditional method of seizure detection, which relies on visual analysis of EEG recordings by trained professionals, is costly and tedious. This is particularly relevant when EEG data to be analyzed may include multiple EEG recordings from a single patient, consisting of multi-channel data and having a duration of many hours. Earlier in this Chapter the Empirical Sparse Dictionary Learning Framework was presented, and the application of this framework for classification of real-life EEG data was also demonstrated, as described in Section 5.3.

In this Section application of the Empirical Sparse Dictionary Learning Framework for

Table 5.1: Gender and Ages of Patients.

Gender	Age range (years)	Average age (years)
10 females and 5 males	1.5 - 22	9.3

automatic seizure detection in long-term scalp EEG recordings will be presented. Scalp EEG is a non-invasive measure of the electrical activity of the brain, and using this measure for seizure detection is challenging because the numerous classes which the brain’s electrical activity is composed of have overlapping characteristics [105].

The next sections will describe the EEG data used, details of the experimental set-up, as well as conclusions drawn from the seizure detection results.

### 5.5.1 Data

The data used to demonstrate the Empirical Sparse Dictionary Learning Framework has been obtained from the CHB-MIT Scalp EEG database [105]. This database consists of EEG recordings from pediatric subjects with intractable seizures. It should be noted that pediatric scalp EEG displays greater variability in seizure and non-seizure activity, compared to adult scalp EEG [105].

The original database consists of EEG recordings from 24 patients. There are multiple EEG recordings for each patient, obtained using the International 10-20 system of electrode positions and nomenclature, and the EEG recordings are available in the standard **.edf** file format. All EEG signals in the database have been sampled at 256 samples per second with 16 bit resolution. For each patient there are between 9 and 42 continuous **.edf** files. For most of the patients, the **.edf** files contain one hour of digitized EEG signals, though for some patients the **.edf** files may contain two or four hour long EEG signals. The signal length in files with seizures is generally shorter. A list of **.edf** files containing seizures (one or more) is available, and the start and end times of seizures within these files is annotated. All the records contain a total of 198 seizures.

Most of the **.edf** files contain data from 23 EEG signals, whereas a few contain 24 and 26 signals. In some records, signals other than EEG are also recorded, e.g. the ECG signal in case of patient number 4, and the vagal nerve stimulus (VNS) signal in some files belonging to patient number 9, however these other signals were not used for this study.

For the experiment described next, EEG recordings from 15 patients were used, mainly due to the enormity of the data available. Details related to the gender of the patients and their ages are shown in Table 5.1. From within the EEG recordings of these 15 patients, data consisting of 87 seizures was used, with the number of seizures per patient listed in Table 5.2.

Also, given the length of the signals in case of long-term EEG recordings, EEG signals are usually segmented into smaller intervals for efficient processing. Segments of 2 or 4 seconds length are generally used [105][106], and for the experiments in this study, it was decided to segment the EEG signals into 4 second segments, consisting of 1024 samples.

The plots in Figure 5.12 and Figure 5.13 show 27 second segments of EEG signals from patient number 1 during seizure and non-seizure states respectively. These figures present adequate illustration of the following aspects of scalp EEG signals [105]:

1. There is considerable overlap in the EEG associated with seizure and non-seizure states for epilepsy patients.
2. A set of EEG channels develop rhythmic activity composed of multiple frequency components following the onset of most seizures. The channels on which the rhythmic activity develops most prominently, and the structure of the rhythmic activity differs from patient to patient.
3. Despite the variance of seizures across patients, the seizures for any one patient show consistency, provided these emerge from the same brain region.

The points presented above represent some of the challenges associated with automatic seizure detection using scalp EEG data, consisting of multiple recordings of multi-channel signals. Keeping these points and the details of the data presented in this Section in view, the

data processing used by the Empirical Sparse Dictionary Learning Framework is described next.

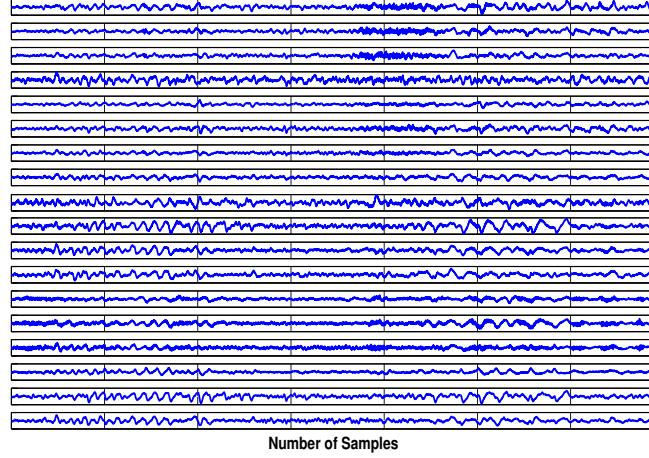


Figure 5.12: EEG signals from a 27 seconds seizure segment from a recording of patient number 1. The channels, from top to bottom, are, respectively: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ.

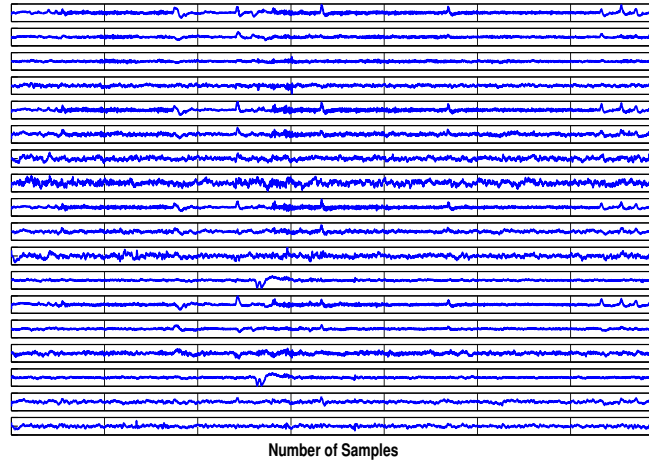


Figure 5.13: EEG signals from a 27 seconds non-seizure segment from a recording of patient number 1. The channels, from top to bottom, are, respectively: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ.

Table 5.2: Number of Seizures Used Per Patient.

Patient No.	Number of seizures used
1	6
2	3
3	7
4	4
5	5
6	7
8	5
9	4
10	6
11	3
12	11
13	8
14	7
15	8
Total Number of Seizures: 87	

### 5.5.2 Data Processing For Dictionary Creation and Learning

In this Section processing of the EEG data for dictionary creation and learning, and for subsequent testing of the EEG signals for automatic seizure detection will be described. First the processing of seizure signals from the 15 patients will be described, followed by description of the processing of non-seizure signals. The most important consideration for data processing was to use small amounts of the available data for dictionary creation, so as to create a sparse dictionary, which makes projection of a large number of test signals against the trained dictionary computationally fast. However, using fewer signals for dictionary creation and learning has the drawback of creating a dictionary which is not representative of the underlying characteristics of the data. Therefore the amount of data used for both seizure and non-seizure signals represented a compromise between these competing considerations. Furthermore, experiments with this amount of data allowed important conclusions to be drawn, which are presented in Section 5.5.5.

### 5.5.2.1 Data processing for seizure segments

The steps involved in the processing of the seizure segments are represented in the block diagram shown in Figure 5.14. These steps are as follows:

1. For each patient  $i$ , the data for  $k$  seizures from different recordings is loaded into the software from the **.edf** files. The number of seizures  $k$  is different for each patient, as listed in Table 5.2.
2. For each seizure segment, 3 channels were randomly selected, and the signals from these 3 channels were segmented into 4 seconds long lengths, consisting of 1024 samples.
3. The 4 second long segments from the 3 randomly selected channels were decomposed into IMFs using EMD for dictionary creation and learning, as described in Section 5.5.3.
4. The signals from rest of the channels were separated, and also segmented into 4 second long lengths. These 4 second long segments were kept separate for testing the methodology.

Using these steps, the total number of 4 second segments from the 3 randomly selected channels turned out to be 4635 from all 15 patients. The number of segments kept separate as test signals for projection against the trained dictionary were different for different patients, and their numbers are listed in Table 5.3.

### 5.5.2.2 Data processing for non-seizure segments

The block diagram representing the steps in the processing of non-seizure segments from all 15 patients is shown in Figure 5.15. These steps are similar to those presented for seizure segments in Section 5.5.2.1. However, there are also differences in the steps in order to incorporate the large amount of non-seizure signals involved.

The first step involved loading multiple **.edf** files representing non-seizure EEG signals representing recordings of more than 5 hours. The steps for raw dictionary creation are

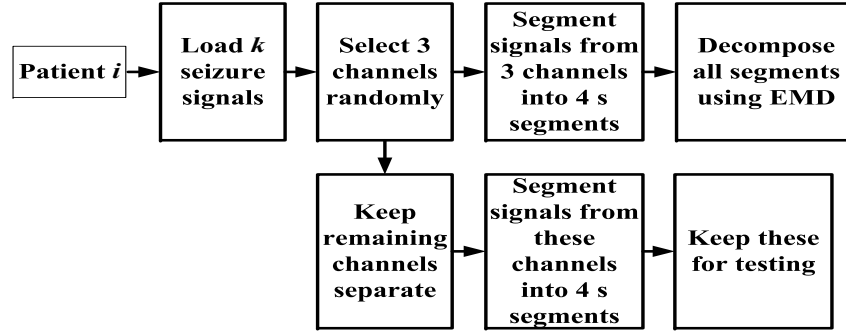


Figure 5.14: Block diagram representing the data processing for seizure signals from all 15 patients.

similar to those presented for seizure signals in Section 5.5.2.1, except that after segmentation of the signals into 4 second segments, the number of 4 second segments chosen for EMD decomposition is made the same as that for seizure segments, so as to avoid having a large difference between the number of IMFs resulting from both the seizure and non-seizure segments.

Once signals from 3 randomly channels were separated for dictionary creation and learning from the initially loaded data, the signals from rest of the channels were kept separate. In the case of non-seizure signals, the duration of these signals was at least one hour, with some patients having recordings of two or more hours as well. To reduce the amount of data, 10 channels were then randomly selected from these remaining channels, and these selected signals were then segmented into 4 second long segments. Further random selection was made from these 4 second long segments such that the selected segments represented EEG data of at least 5 hours. The actual number of non-seizure segments per patient used for testing is listed in Table 5.3.

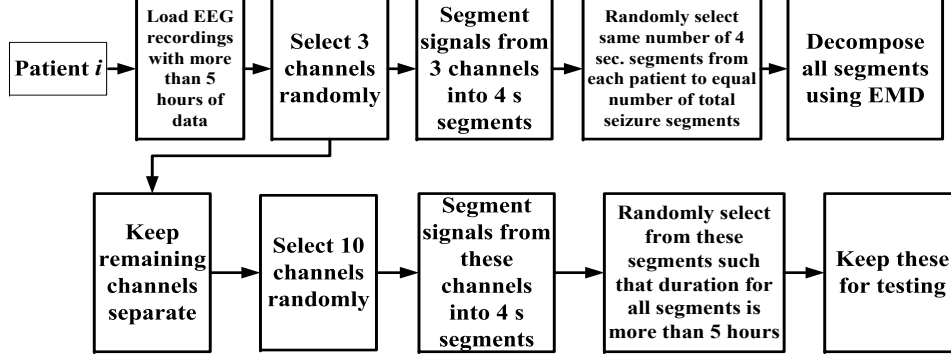


Figure 5.15: Block diagram representing the data processing for non-seizure signals from all 15 patients.

### 5.5.3 Dictionary Formation and Training

The number of 4 second segments selected for dictionary formation and training, as described in Section 5.5.2.1 and Section 5.5.2.2, was the same for both seizure and non-seizure segments at 4635 segments. These 4 second segments were decomposed into IMFs using EMD, resulting in 37863 IMFs from the seizure segments, and 38146 IMFs from non-seizure segments.. These IMFs formed the atoms of the class-specific raw dictionaries  $\mathbf{D}_{raw}^c$  representing the seizure and non-seizure signals. The class specific raw dictionaries were merged to form a combined raw dictionary  $\mathbf{D}_{raw}^C$  consisting of 76009 atoms, as described in Section 5.2. However, the combined raw dictionary is not sparse according to the definition of sparsity presented in Section 5.2.

Learning the trained dictionary  $\mathbf{D}_{Train}^C$  using Algorithm 3 resulted in the trained dictionary having 8830 atoms. This represents a magnitude decrease of almost 9 times, close to the maximum possible magnitude decrease of 10 according to Equation 5.2. Furthermore, due



to the large amount of computational time taken for dictionary learning due to the size of the raw dictionary and the number of training signals, Algorithm 3 was terminated after just one iteration. The plots in Figure 5.16 show four example atoms in the trained dictionary  $\mathbf{D}_{Train}^C$  which show a wide range of temporal and spectral characteristics.

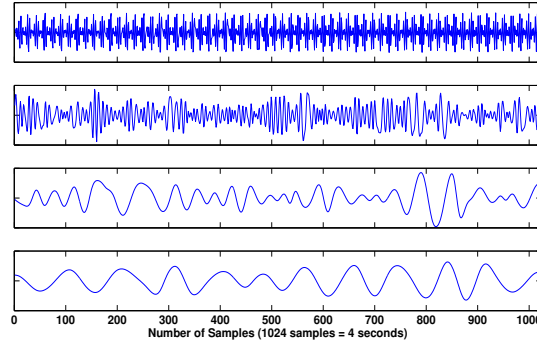


Figure 5.16: Examples of dictionary atoms in the trained dictionary  $\mathbf{D}_{Train}^C$ .

#### 5.5.4 Results

The results of patient-specific automatic seizure detection obtained by application of the Empirical Sparse Dictionary Learning Framework on the EEG data described in Section 5.5.1 are shown in Table 5.3. These results were obtained in the following way:

1. The test seizure and non-seizure signal segments from each patient were projected against the trained dictionary  $\mathbf{D}_{Train}^C$ , and the projection coefficients used as features for automatic seizure detection. The number of seizure and non-seizure segments for each patient listed in Table 5.3.
2. For classification, a SVM classifier with the Gaussian radial basis function (RBF) was used. An SVM classifier is a good choice, as seizure and non-seizure classes are often not linearly separable [105].
3. The SVM classifier was trained using a patient-specific methodology by using a specified fraction of the projection coefficients obtained by projecting the test signals against

the trained dictionary, as described in point 1. For this purpose, one-half of the seizure signal projection coefficients, and the same number of non-seizure signal projection coefficients were used to train the classifier. For example, there are 1590 seizure projection coefficients and 4780 non-seizure projection coefficients available for patient 1, as shown in Table 5.3; from these, 795 seizure and 795 non-seizure projection coefficients were used to train the classifier. The remaining 795 seizure and 3985 non-seizure projection coefficients were used to test automatic seizure detection. The same classifier training procedure was used for all patients.

The results shown in Table 5.3 show large variation in the values of classification accuracy, sensitivity and specificity. For some patients, e.g. patients 1, 2, 7 and 9, these values are close to or higher than 85%. However, for some patients, for example patient number 6, the classification accuracy, sensitivity and specificity values are close to 50%. The automatic seizure detection results presented in Table 5.3 allow important conclusions to be drawn about the whole experimental configuration as well as elucidate the implications for the application of the Empirical Sparse Dictionary Learning Framework to long-term data.

### 5.5.5 Discussion: From Global to Patient-Specific Dictionaries

A number of conclusions can be drawn from the wide variance in the automatic seizure detection accuracy rates, which have important implications for data processing, which was presented in Section 5.5.2, and dictionary formation and learning, presented in Section 5.5.3.

An important aspect of data processing for dictionary creation was to randomly select limited amount of EEG signals so as to keep the size of the raw dictionary  $\mathbf{D}_{raw}^c$  small. At the same time, the data selected for raw dictionary creation needs to be representative of the underlying characteristics and structures of the seizure and non-seizure states. With selection of only limited amount of data, it is possible to miss the various latencies, periodicities and fractal behaviours associated with the non-stationary and non-linear signals. This can be used to explain the variance in accuracy rates for automatic seizure detection across different patients, as not only do the underlying signal structures vary from one patient to the next,

Table 5.3: Seizure detection performance using data presented in Table 5.1.

Patient No.	Acc. (%)	Sens. (%)	Spec. (%)	No. of Seiz. Seg.	No. of Non-Seiz Seg.
1	90.3	91.3	90.1	1590	4780
2	86.1	87.8	86.0	770	4780
3	81.1	91.6	78.5	1910	4780
4	71.8	75.4	71.0	1790	4790
5	77.2	69.6	80.2	2690	4780
6	48.3	62.9	47.6	470	4790
7	84.2	85.4	83.9	1530	4790
8	79.8	80.7	79.1	4490	4780
9	86.6	94.0	85.4	1270	4790
10	77.1	72.2	78.2	1790	4790
11	81.7	86.3	78.5	3950	4780
12	61.3	56.2	62.4	1650	4780
13	64.5	54.6	66.3	1474	4780
14	78.1	57.1	79.6	630	4780
15	58.9	54.2	55.8	3514	4780

Acc. = Accuracy; Sens. = Sensitivity; Spec. = Specificity; Seiz. = Seizure; Seg. = Segments (4 second long test segments having 1024 samples).

there is variation in the signal patterns across different EEG channels in recordings from the same patient, as explained earlier in Section 5.5.1.

Although only a limited amount of EEG data per patient was selected for dictionary formation and learning, this data was selected randomly from the different EEG channels available in the signal recordings. In a previous iteration of the experiment, a limited amount of data was selected only from particular channels of the EEG data from all patients. This configuration resulted in very low automatic seizure detection accuracy rates for most of the patients under study. With the experimental configuration described here, random selection of EEG signals has helped better capture the patient-specific characteristics of seizure and non-seizure states, resulting in higher seizure detection accuracy rates for many patients.

However, the random selection of limited amounts of data from EEG recordings of all patients represents a global approach to dictionary formation and learning, whereby a single

raw dictionary is formed, trained into a learned dictionary, and then used for signal classification. The experimental results presented in Table 5.3 provide evidence of important drawbacks of the global approach to dictionary formation and learning, which are listed below:

1. A global dictionary formed and trained with limited amount of data does not fully capture the underlying characteristics of the data, specially when the data is non-stationary and non-linear in nature.
2. Although dictionary learning results in a considerable decrease in the size of the raw dictionary, the initial raw dictionary will be large in size for the case where the data to be analyzed has a large volume. Any attempt to limit the size of the dictionary by using a limited amount of data leads to the problem mentioned in the previous point. Furthermore, the raw dictionary, and the subsequently obtained learned dictionary, will not be sparse according to the definition of sparsity given in Section 5.2.

Therefore, for cases where a large amount of data from multiple sources is available, which for the current example under study consists of long-term EEG recordings from multiple patients, a source-specific dictionary formation and training strategy is required. In the context of the current study, this strategy would entail construction of patient-specific dictionaries.

#### **5.5.5.1 Patient-specific dictionaries**

A patient-specific dictionary in the context of automatic seizure detection using long-term EEG data would entail dictionary formation and training using EEG data from individual patients only. This has several important consequences for dictionary formation and learning, such as:

- Incorporation of larger amounts of patient data for dictionary formation and learning, thereby allowing the dictionary to be representative of underlying characteristics of data specific to each individual patient. For example, in the context of the current

study, instead of using only few channels from the seizure and non-seizure signals, data from all the available channels could be used, which will allow the dictionary to capture seizure and non-seizure characteristics spread over different channels.

- Even when using larger amounts of data for dictionary creation and learning, a patient-specific dictionary will allow faster computational time for dictionary formation and learning, as well as for projection of test signals against the trained dictionary, since the patient-specific dictionary will be more sparse than a global dictionary.

An important aspect of the patient-specific dictionary is that such a dictionary can be formed and trained once, with the available data, and then used by large amounts of data, which may continually be acquired. For the current study using long-term EEG data, this is feasible since the characteristics of the EEG data show considerable consistency across different recordings [105]. Even for a case where the characteristics of the collected data might change, a source-specific dictionary can be formed in a computationally efficient way, while retaining representative characteristics of the different classes of data from which it may have been formed and trained.

In the context of EEG signal analysis using long-term data, patient-specific methods have shown much better results [105][133]. It has also been established that non-patient specific methods exhibit poor accuracy due to the cross-patient variability in seizure and non-seizure activity, and exhibit good performance in some cases when restricted to analysis of seizure types that do not vary much across patients [105]. The EEG data used in this study is pediatric scalp EEG, which exhibits greater variability in seizure and non-seizure activity. Therefore, even with a global dictionary, the automatic seizure detection performance of the presented method, as presented in Table 5.3, represents good results, and holds greater promise for the same method based on patient-specific dictionaries.

## 5.6 Chapter Summary

In this Chapter, a novel empirical sparse dictionary learning framework for signal classification based on EMD was presented. At the heart of the framework is the discriminative dictionary learning framework, which learns sparse dictionaries, where sparsity quantifies the under-completeness of the learned dictionary. The dictionary formulation and learning is not meant to be reconstructive, or combined reconstructive and discriminative, as in other state-of-the-art approaches. The use of the framework was demonstrated with different experimental scenarios. Furthermore, the empirical sparse dictionary learning framework was used for automatic seizure detection using long-term EEG data. Based on the experimental results, it was concluded that the framework presented here could be used for patient-specific dictionaries for more robust signal analysis.

# Chapter 6

## Conclusion

Automated signal analysis provides objective and practical means for better understanding the characteristics of signals, and thus obtaining an accurate insight into the underlying mechanisms and processes of the systems which produce the signals. Also important for automated signal processing is the knowledge that signals in the real world are in general non-stationary in nature, and often the underlying systems producing the signals have non-linear dynamics. The methods for automated signal analysis, therefore, should be able to take the signal non-stationarity and non-linearity into account. In this regard, signal decomposition methods are particularly useful, as these bring out representative signal information hidden at characteristic time and frequency scales. This dissertation focused on empirical methods for signal analysis based on signal decomposition techniques, with emphasis on analysis of non-stationary and non-linear signals using Empirical Mode Decomposition (EMD) and new algorithms based on EMD. Based on the work presented in the dissertation, the following conclusions can be drawn:

- *EMD is a viable alternative to existing methods for signal analysis. Its properties and behaviour are well-understood through empirical and numerical means, which allows confident application of the algorithm in a particular domain.* The efficacy of EMD comes mainly from its adaptive nature, whereby the signal decomposition does not depend on a pre-defined basis, but the basis is instead derived from the data. This

makes it particularly suitable for analysis of non-stationary and non-linear signals. Furthermore, EMD decomposes a signal into components which are the same length as the original signal, and thereby allows analysis in the time domain, and also in the frequency domain, since mechanism for a meaningful transformation of the decomposed components in the frequency domain is also defined. EMD was applied for pathological speech classification (Chapter 2), for which a novel set of features was defined and calculated from the components decomposed from speech signals. The robustness of the proposed method was also demonstrated by pathological speech classification using telephone quality speech signals.

- *Since EMD decomposes signals based on their local time scales, it is possible to effectively use the decomposition for de-noising and de-trending of signals.* In addition, new frameworks for signal analysis based on de-noising and de-trending of signals can be developed. The EMD-based de-noising and de-trending approaches were categorized in Chapter 2, followed by development of a novel framework for mental task classification using EEG signals.
- *EMD does not allow time-scale based decomposition, whereby signals could be decomposed according to a specified frequency separation criterion. However, such ability would add much flexibility to the decomposition process, and open up a host of possibilities for innovative applications.* With this in mind, a novel modification of the EMD algorithm was proposed, which is named Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS) (Chapter 3), and which allows a time-scale based decomposition. The properties of EMD-MPS were studied, and its use for a novel hierarchical decomposition, which allows signals to be decomposed into selected frequency bands, was demonstrated. A novel time-scaled based de-noising and de-trending based on EMD-MPS was also shown, along with practical applications in the form of mental task classification and seizure detection using EEG signals (Chapter 4).
- *Empirical learned dictionaries, which are neither reconstructive, nor reconstructive-*



*discriminative, can be used effectively for signal classification.* A novel sparse empirical dictionary learning framework was presented, and a discriminative dictionary learning algorithm developed as part of this framework (Chapter 5). The properties of the learned dictionary were studied, and the features that can be obtained for signal classification were highlighted. Also, the use of the framework for signal classification was demonstrated, along with an application for automatic seizure detection using long-term EEG signals. Implications of the aforementioned application for patient-specific dictionaries, as well as the potential advantages when using long-term data, were also discussed.

The following directions for future work can be identified:

- In the current formulation of EMD-MPS, the value of the short-time window  $\tau$  once selected, is fixed for the whole length of the signal. Further enhancement of the EMD-MPS framework is planned, in which the value of  $\tau$  will not remain fixed for the whole length of the signal, but will be adaptive to the local properties of the signals being analyzed.
- In Chapter 3, it was shown how the decomposition behaviour of EMD could be formulated in terms of EMD-MPS, providing novel insight into EMD. Here, it may be realized that the modified peak selection as defined in the EMD-MPS algorithm, whereby local extrema are selected conditionally to time-windows of variable length, is analogous to the methodology of Re-scaled Range (R/S) analysis [134]. Re-scaled range is a statistical analysis used for estimating the scaling exponents of self-similar processes. In this context, conventional EMD, due to its wavelet-like property of mimicking dyadic filter banks, is one way of estimating the scaling exponents. It will therefore be an interesting research direction to explore if R/S analysis can be rephrased through EMD-MPS in such a context in terms of EMD, and/or vice-versa.
- Hierarchical decomposition based on EMD-MPS was also described and demonstrated in Chapter 3. Further analysis of hierarchical decomposition to establish its proper-

ties is warranted. This will also allow comparison of the properties of hierarchical decomposition with the wavelet decomposition.

- Although the empirical dictionary learning framework defined in Chapter 5 does not learn reconstructive dictionaries, it was shown that the under-complete learned dictionary does allow signal reconstruction, although with a large reconstruction error. The usability of the reconstruction error as a discriminative feature will form a research question to be explored. Additionally, an automatic seizure detection system using long-term EEG data was described in Chapter 5, along with a discussion of extending the work in the direction of patient-specific dictionaries. Extension of the methodology with a definition and utilization of patient-specific dictionaries forms the future work in this context. Another line of future work in the context of empirical dictionary learning framework will relate to incorporation of compressive sensing concepts into dictionary learning, whereby training signals can be replaced by the compressive sensing measurements representing these signals [135]. In this regard, use of real world signals other than the ones used in this dissertation will also be explored. An interesting issue to explore here would be formation and learning of semantic dictionaries, which may be considered as dictionaries containing specific signal patterns and characteristics as identified by domain experts, and using such dictionaries for further signal analysis and classification.

## List of Refereed Publications

1. Kaleem, M.F., Sugavaneswaran, L., Guergachi, A., Krishnan, S., *Application of Empirical Mode Decomposition and Teager energy operator to EEG signals for mental task classification*, Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2010, pp: 4590-4593.
2. Kaleem, M.F., Ghoraani, B., Guergachi, A., Krishnan, S., *Telephone-quality pathological speech classification using empirical mode decomposition*, Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2011, pp: 7095-7098.
3. Kaleem, M.F., Cetin, A.E., Guergachi, A., Krishnan, S., *Using a variation of empirical mode decomposition to remove noise from signals*, Proceedings of the 21st International Conference on Noise and Fluctuations (ICNF), 2011, pp: 123-126.
4. Kaleem, M.F., Ghoraani, B., Guergachi, A., Krishnan, S., *Pathological speech signal analysis and classification using empirical mode decomposition*, Medical & Biological Engineering & Computing, Vol. 51, No. 7, pp: 811-821, July 2013.
5. Kaleem, M.F., Guergachi, A., Krishnan, S., *A Variation of Empirical Mode Decomposition with Intelligent Peak Selection in Short Time Windows*, Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, pp: 5627-5631.
6. Kaleem, M.F., Guergachi, A., Krishnan, S., *Application of a Variation of Empirical Mode Decomposition and Teager Energy Operator to EEG Signals for Mental Task Classification*, Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013. pp: 965-968.
7. Kaleem, M.F., Guergachi, A., Krishnan, S., *EEG Seizure Detection and Epilepsy Diagnosis using a Novel Variation of Empirical Mode Decomposition*, Proceedings of the

35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013. pp: 4314-4317.

8. Kaleem, M.F., Guergachi, A., Krishnan, S., *Empirical Mode Decomposition Based Sparse Dictionary Learning with Application to Signal Classification*, Proceedings of IEEE Digital Signal Processing and Signal Processing Education Workshop (DSP/SPE), 2013. pp:18-23.

## **Publications under Preparation for Submission to Refereed Journals**

1. Kaleem, M.F., Guergachi, A., Krishnan, S., *Automatic Seizure Detection in Long-Term EEG Data Using Patient-Specific Dictionaries Based on Empirical Dictionary Learning*, To IEEE Transactions on Neural Systems and Rehabilitation Engineering.
2. Kaleem, M.F., Guergachi, A., Krishnan, S., *Hierarchical Decomposition Based on a Variation of Empirical Mode Decomposition*, To Signal, Image and Video Processing.

# Appendix A

## EMD: Algorithm Implementation Issues and Characteristics

This appendix contains discussion related to important EMD algorithm implementation issues, as well as characteristics of EMD.

### A.1 Algorithm Implementation Issues

**Stopping criterion for sifting:** The sifting iterations are stopped when the oscillatory mode obtained by subtracting the local mean from the signal fulfils the conditions of an IMF (Step 5 of Algorithm 1: Section 2.3, Chapter 2). Important here is the second condition for an IMF, whereby the mean value of the upper and lower envelopes formed by the extrema of an IMF is either zero, or close to zero. However, a large number of sifting operations in order to obtain a zero or close-to-zero mean can remove the physically meaningful amplitude variations in an IMF, reducing it to a FM signal of constant amplitude. Indeed, it has been pointed out in [31] that too many sifting operations result in IMFs devoid of physical meaning. The criterion to stop the sifting iterations such that the IMFs do not have the amplitude or frequency modulation unduly removed, as developed in [6], is based on limiting the size of the standard deviation calculated from two consecutive sifting results. This value

of the standard deviation, SD, is calculated as:

$$SD = \sum_{n=0}^N \frac{|(c_{1(k-1)})[n] - c_{1k}[n]|^2}{c_{1(k-1)}^2[n]} \quad (\text{A.1})$$

The value of SD recommended in [6] is between 0.2 and 0.3, with a value of 0.2 used in many implementations of the EMD algorithm. However, a flaw with this approach, as pointed out in [136], is that this approach is unrelated to the definition of an IMF, and only requires the successive IMF candidates to be nearly equal. The stopping criterion suggested in [136] is designed to stop the sifting when the number of the zero crossings and extrema differ by at most one for  $S$  sifting iterations. The value of  $S$  has to be selected empirically, though typically a value  $3 \leq S \leq 5$  has proven successful as the stopping criterion compared to the SD-based criterion [136], and is also used in implementations of the EMD algorithm [83].

A different and improved stopping criterion for the sifting process is described in [137], which is the default stopping criterion of a popular and freely available implementation of the EMD algorithm [79]. This criterion is based on two thresholds  $\theta_1$  and  $\theta_2$ , which is meant to guarantee globally small fluctuations in the mean (of the upper and lower envelopes) while allowing locally large variations in the oscillatory mode amplitude. These thresholds are meant to compare the amplitude of the mean with the amplitude of the corresponding mode during each sifting step, so as to prevent a very low amplitude threshold for the amplitude of the mean, which can lead to larger number of iterations and hence over-decomposition. The thresholds  $\theta_1$  and  $\theta_2$  work with the mode amplitude defined as  $a[n] := (E_U - E_L)/2$  and the evaluation function  $\sigma[n] := |m[n]/a[n]|$ , where  $m[n]$  is the mean defined in Step 3 of Algorithm 1. The sifting iterations are continued till  $\sigma[n] < \theta_1$  for a prescribed fraction  $(1 - \alpha)$  of the total signal duration, and  $\sigma[n] < \theta_2$  for the remaining fraction. The values set as default in the EMD implementation [79] are  $\alpha = 0.05$ ,  $\theta_1 = 0.05$  and  $\theta_2 = 10\theta_1$ .

**Border effects:** The border effects, also known as boundary effects or end effects, are related to the EMD algorithm operating on signals of finite lengths. This has implications for the accuracy of the envelope formation based on extrema, specially at both ends of the

signals. If not dealt with, the errors due to incorrect envelope formation at the ends of the signal can propagate inwards with each sifting iteration, and corrupt the mode extraction process, specially affecting the low frequency components in the signal [6]. A window frame approach is mentioned in [31], which involves prediction, and extension of data beyond the ends. The most popular and successful approach to mitigate border effects, however, involves reflecting, or mirroring, the extrema close to the signal ends [137][35]. This is also the technique used in the popular EMD algorithm implementation available at [79].

**Envelope interpolation:** A very important part of the EMD algorithm is the envelope interpolation using the identified maxima and minima (Step 2 in Algorithm 1). The type of interpolation chosen to form the envelopes affects the extracted IMFs, and different interpolation schemes have been considered, such as the cubic spline interpolation [6], B-spline interpolation [31], and Akima interpolation [35]. However, it has been empirically found that cubic spline interpolation gives the most preferable decomposition results [137][35]. Also, most implementations of the EMD algorithm use the cubic spline interpolation for envelope formation [137][83].

**Sampling requirement:** Although most real life signals are generated by continuous time processes, these signals are digitized for using modern digital signal processing methods. Since the EMD algorithm operates on quantized discrete-time signals, it is important that the extrema of the signal are correctly identified. In many cases, the extrema of the actual continuous time signal will not match with sampling instants and be correctly localized [35]. The misalignment of the sampling instant and local extrema of the signal results in an erroneous decomposition [138]. To mitigate this error, a fair amount of over-sampling is suggested [137][35]. The effect of sampling has been studied in [138] using linear frequency modulated signals, and a sampling limit, which is five times the Nyquist rate ( $f_s/10$ , where  $f_s$  is the sampling frequency) is recommended. This sampling limit is also tested on a bat echolocation pulse, which is a real world non-stationary and non-linear signal, and shown to hold for this signal as well. Similarly, the work in [139] mentions the minimum requirement

for the sampling period to be at most one half of the minimum distance between extrema, and provides a detailed analysis of the influence of sampling on EMD.

## A.2 Characteristics of Empirical Mode Decomposition

Since EMD does not admit an analytic formulation, its decomposition behaviour has been studied empirically using numerical simulations. This has allowed a clear understanding of its decomposition behaviour, and pointed out the issues with the method. This can help researchers make an informed decision about the application of the method to different types of signals. In this Section some of the more important characteristics of EMD will be presented.

**Mode-Mixing:** Mode-mixing is a very common phenomenon associated with the EMD algorithm, and refers to the situation where a single IMF contains multiple oscillatory modes, or a single oscillatory mode resides in multiple IMFs, or both [83]. Mode-mixing is caused by intermittency occurring in any part of the signal, where intermittency refers to an oscillatory component that comes into existence or disappears from a signal completely at a particular time-scale [87]. Mode-mixing and intermittency have been known since the EMD algorithm was proposed [6], and is an important issue because it can compromise the physical meaning of the IMFs, and also lead to incorrect instantaneous amplitude and frequency calculation (Section 2.3.1, Chapter 2). In this regard, Figure A.1 (a) illustrates the phenomenon of mode-mixing, using a 10 Hz sinusoidal signal which has intermittently occurring 50 and 100 Hz components. Intermittently occurring oscillatory components are reflected through intermittency in extrema, such that the extrema detected during the sifting process belong to different oscillatory components, leading to an envelope interpolation reflecting the mode-mixing, as shown in Figure A.1 (b). This in turn results in a mode-mixed IMF, which consists of oscillatory components at different time-scales, which is shown in Figure A.1 (c). Furthermore, as Figure A.1 (d) illustrates, after extraction of the mode-mixed IMF, the mode-mixing propagates to the rest of the extracted IMFs through intermittency in the



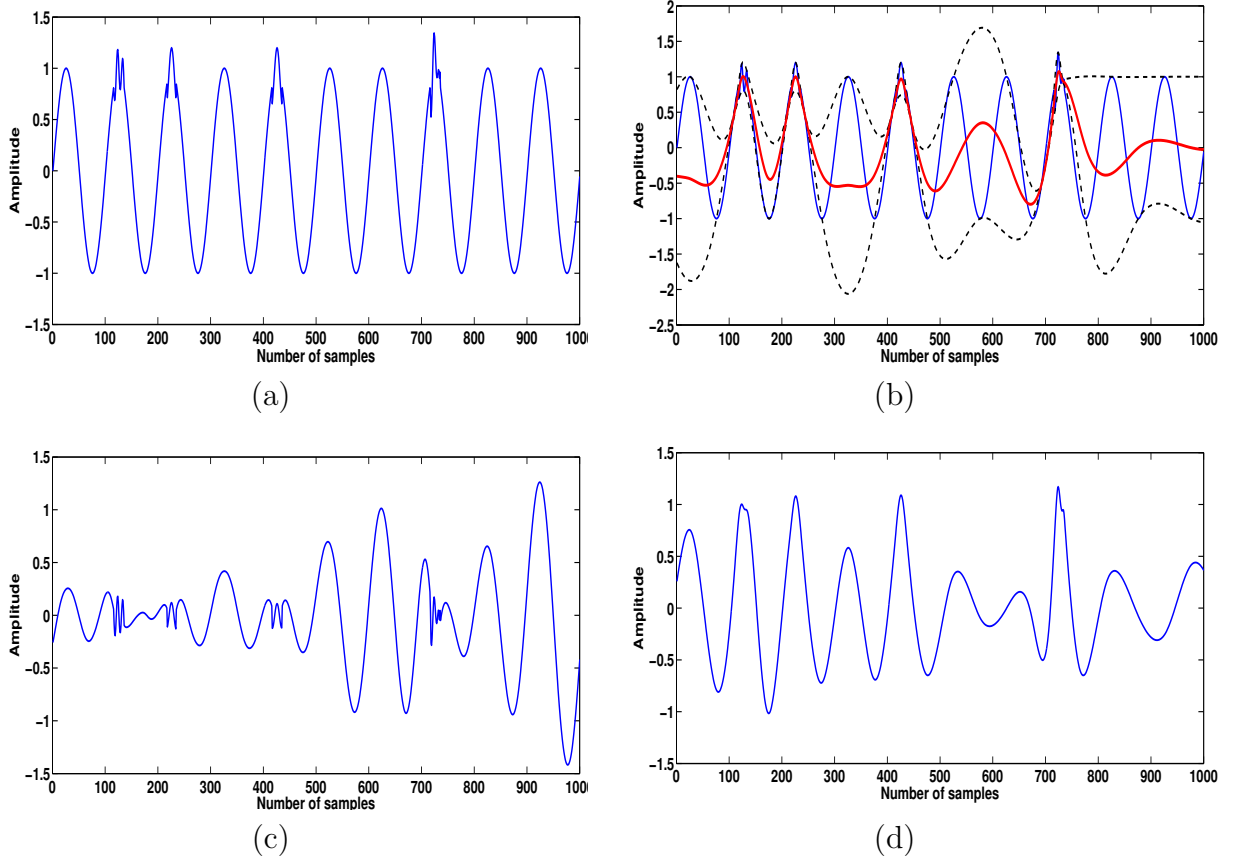


Figure A.1: Illustration of mode-mixing: (a) A 10 Hz sinusoidal signal containing intermittently occurring 50 and 100 Hz components; (b) Envelope formation using cubic spline interpolation. The dashed black lines represent the upper and lower envelopes, and the red line represents the average of the envelopes; (c) The first IMF obtained after the sifting process, showing mode-mixing, as it contains multiple oscillatory modes; (d) The residue obtained after the first IMF has been extracted, from which further IMFs will be extracted by the sifting process. The residue illustrates that mode-mixing is going to propagate to other IMFs as well.

residue. Many variations of the EMD algorithm have been proposed to overcome mode-mixing, some of which were discussed in Section 3.2 of Chapter 3.

**EMD as a Filter Bank:** Based on decomposition of white noise sequences using EMD, it was reported in [37] that EMD behaves as a dyadic filter which separates white noise into IMFs having a mean period twice that of the previous IMF. This was found to hold for both

uniformly distributed and normally distributed white noise. A similar study to ascertain the behaviour of EMD in case of stochastic conditions involving broadband noise not dominated by any particular frequency band is described in [82]. This study uses fractional Gaussian noise (fGn) with values of the Hurst parameter  $H$  varying from 0.1 to 0.9. With a value of  $H = 1/2$ , fGn reduces to white noise, whereas lower values of  $H$  ( $0 < H < 1/2$ ) refer to negative correlation, and higher values of  $H$  ( $1/2 < H < 1$ ) refer to positive correlation.

Using the power spectrum of the IMFs obtained after decomposition, it was found that the frequency profile of all the IMFs tends to arrange in a filter bank structure reminiscent of results of wavelet decompositions in similar situations. The filter associated with the first IMF (having index 1) tended to be essentially high-pass with some non-negligible low-pass characteristic in the lower half-band, whereas the IMFs of higher indices were organized in a set of overlapping band-pass filters. Also, each IMF having index  $(k + 1), k \geq 2$  had a frequency range which roughly occupied the upper half-band of the spectrum of the previous IMF with index  $k$ , representing a dyadic filter-bank structure. This was quantified using the number of zero-crossing in the IMFs,  $z_H[k]$ , which provides a rough indication of the mean frequency of each IMF  $k$ , with  $z_H[k]$  being a decreasing exponential function of the mode number  $k$  given by:

$$z_H[k] \propto \rho_H^{-k} \quad (\text{A.2})$$

where  $\rho_H$  has a value very close to 2 for all values of  $H$ .

Further analysis of fGn decomposition using EMD is described in [99], where self-similarity in the filter bank structure is quantified in terms of the following relation:

$$S_{k'}(f) = \rho_H^{\alpha_H(k' - k)} S_k(\rho_H^{k' - k} f) \quad (\text{A.3})$$

where  $S_k(f)$  is the power spectrum of the  $k$ -th IMF, and  $k' > k \geq 2$ . Self-similarity of the filter bank implies that the power spectra of the IMFs can be collapsed onto a single curve by appropriate re-normalization. This re-normalization is achieved with a value of  $\alpha_H = 2H - 1$

and  $\rho_H = 2$ . Based on this re-normalization, it was found that EMD acts on fGn as a dyadic filter bank of constant- $Q$  band-pass filters for  $H > 0.5$ . For the case of  $H < 0.5$ , EMD still acts as a dyadic filter bank, though the associated IMF filters are not band-pass and the re-normalization of the IMF spectra is not as well-behaved due to amplitude discrepancies. Furthermore, it is also possible to estimate the value of the Hurst exponent  $H$  through the progression of variance across the IMFs. Based on the re-normalization relation given in Eq. A.3, the variance  $V_H[k]$  of an IMF with index  $k$ ,  $k > 1$ , is such that:

$$V_H[k] \propto \rho_H^{(\alpha_H - 1)} k \quad (\text{A.4})$$

where  $\alpha = 2H - 1$ . According to this model, the IMF variance is an exponentially decaying function of the IMF index  $k$ ,  $k > 1$ , and the decay rate is a linear function of  $H$ . However, the model holds reasonably well for  $H \geq 0.3$ , with more discrepancy for smaller values of  $H$  [102][99]. Based on these techniques, the work in [99] also describes EMD based estimators of the Hurst exponent  $H$ , and compares these to Discrete Wavelet Transform based estimators. The performance of both types of estimators was found to be comparable.

It should also be mentioned here that the filter-bank behaviour is considered an essential property of EMD, and proposed variations of the EMD algorithm are also tested for conformity to this behaviour (e.g. [33][83][114]).

**Non-linearity of Decomposition:** EMD is a non-linear decomposition technique, such that decomposition of a sum of two signals may not produce IMFs equal to the sum of IMFs obtained after decomposition of each signal individually. The question of EMD linearity is posed in [7] in terms of whether the following expression holds:

$$\mathcal{E}(s_1 f_1 + s_2 f_2) = s_1 \mathcal{E}(f_1) + s_2 \mathcal{E}(f_2), \quad f_1, f_2 \in \mathcal{D}, s_1, s_2 \in \mathbb{R} \quad (\text{A.5})$$

where  $\mathcal{E}$  is the EMD operator resulting in IMFs, and  $\mathcal{D}$  represents the domain of  $f_1$  and  $f_2$ . This question is answered in the negative in [7], and a proof is also provided in terms of the following signal model:

$$f(x) = \cos(x) + \epsilon \cos(3x) \quad (\text{A.6})$$

The two components in the signal in Eq. A.6 fulfil the conditions of an IMF, and it is also shown in [7] that  $f$  also is an IMF under certain conditions, and hence will not be decomposed by EMD. The reason of EMD non-linearity is explained in terms of the reliance of the envelope formation process on the location and values of the local extrema, whose dependence on the signal is fundamentally non-linear in nature [7].

EMD non-linearity as shown in terms of a signal model consisting of tones also leads to the conclusion that separation of tones is not always possible using EMD. The tone separation aspect of EMD has been studied using mathematical and numerical analysis in [101] using a two tone model. According to the analysis presented in [101], it is possible to distinguish three domains with different behaviours depending on the frequency ratio and the amplitude ratio of the tones. The three domains correspond to the two components correctly identified and separated by EMD, considered as a single waveform with separation not possible, or a situation where the behaviour is either half-way between the first two domains, or results in incorrect decomposition with spurious IMFs. Important, however, is the critical value of the frequency ratio  $f$  of the two components, beyond which it is not possible to separate the two components, regardless of their amplitude ratio. The value of the frequency ratio for the two-tone model was established as  $f \approx \frac{2}{3}$  in [101]. Similarly, the frequency ratio for separation of components using EMD decomposition of full spectrum noise sequences was established as  $f \approx \frac{1}{3}$  in Section 3.3.2.2 of Chapter 3. Therefore, for all practical purposes, EMD requires a frequency ratio  $f \ll \frac{2}{3}$  for proper separation of components during the decomposition process.

# Appendix B

## EMD-MPS: De-noising and De-trending

This appendix contains examples based on synthetic and real-world signals to demonstrate time-scale based de-noising and de-trending using EMD-MPS.

### B.1 De-noising and De-trending of Synthetic Signals

The first two examples presented here are similar to the ones described in [51]. The first example represents a non-seasonal time-series  $y^1(t)$ , and the second a seasonal (i.e. containing a periodic component) time-series  $y^2(t)$ , given by the following expressions:

$$\begin{aligned}y^1(t) &= T_1(t) + \epsilon(t) \\T_1(t) &= \beta_0 + e^{\alpha t} \\y^2(t) &= T_1(t) + S(t) \\S(t) &= \beta_1 \cos(\frac{2\pi t}{12}) + \beta_2 \sin(\frac{2\pi t}{12})\end{aligned}\tag{B.1}$$

where  $t = (1, 2, \dots, N)$ ,  $N = 300$ ,  $\beta_0 = 100$ ,  $\beta_1 = 24$ ,  $\beta_2 = 32$ ,  $\alpha = 0.018$  and  $\epsilon$  is a white noise sequence.

The synthetic signals  $y^1(t)$  and  $y^2(t)$  along with the trend  $T_1(t)$  are shown in Figure B.1 (a) and (c) respectively. The results of extracting the exponential trend  $T_1(t)$  from both  $y^1(t)$

and  $y^2(t)$  using EMD-MPS are shown in Figure B.1 (b) and (d) respectively.

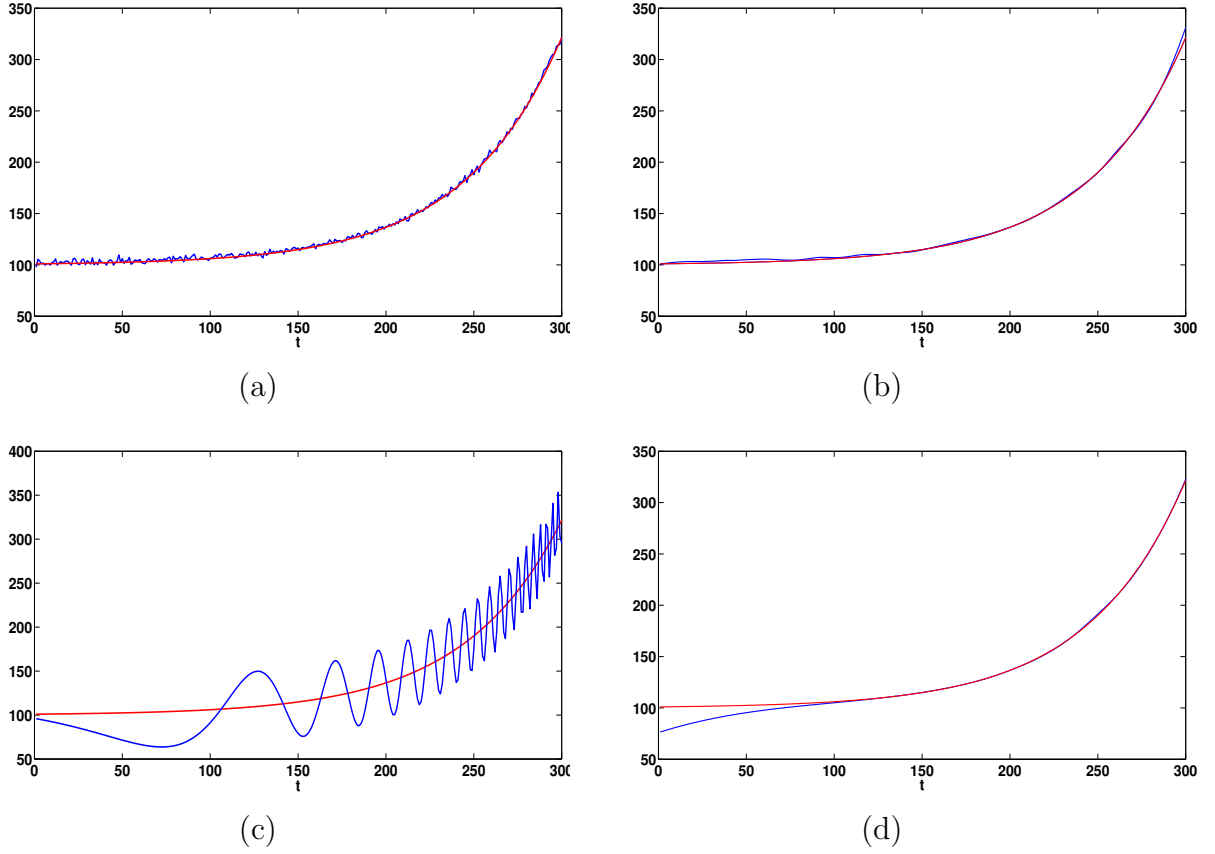


Figure B.1: Non-seasonal and seasonal time-series with exponential trends, represented by the expression in Eq. B.1. (a)  $y^1(t)$ . (b) Extracted exponential trend from  $y^1(t)$  compared with the original trend  $T_1(t)$ . (c)  $y^2(t)$ . (d) Extracted exponential trend from  $y^2(t)$  compared with the original trend  $T_1(t)$ .

A synthetic example of a process  $y^3(t)$  consisting of a piecewise linear trend  $T_2(t)$  with a superimposed AR(2) process  $Y(t)$ , inspired by [43], is given by the following expression:

$$\begin{aligned} y^3(t) &= T_2(t) + Y(t) \\ Y(t) &= 0.8Y(t-1) - 0.4Y(t-2) + \zeta(t) \end{aligned} \tag{B.2}$$

where  $t = (1, 2, \dots, N)$ ,  $N = 2000$  and  $\zeta(t)$  is a white noise process with variance  $10^4$ .

The time-series  $y^3(t)$  is shown in Figure B.2 (a), and the piecewise linear trend extracted by application of EMD-MPS to  $y^3(t)$  is shown in Figure B.2 (b) compared to the original trend  $T_2(t)$ . The extracted trend in Figure B.2 (b) shows slight amplitude variations, but these result from an implementation issue concerned with using spline interpolation for EMD-MPS envelope formation. A linear interpolation is expected to reduce these slight amplitude variations.

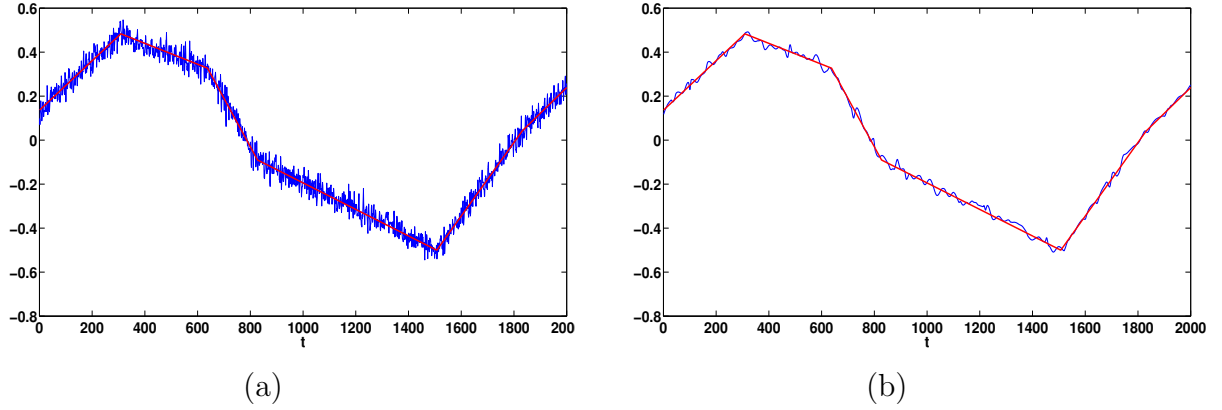


Figure B.2: Time-series consisting of an AR(2) process superimposed on a piecewise linear trend, represented by the expression in Eq. B.2. (a)  $y^3(t)$ . (b) Extracted piecewise linear trend from  $y^3(t)$  compared with the original trend  $T_2(t)$ .

As an objective measure of the quality of the extracted trend from synthetic signals  $y^1(t)$ ,  $y^2(t)$  and  $y^3(t)$ , the coefficient of correlation and a distance measure were calculated. The distance measure  $d$  is as defined in [51], and is given by the following expression:

$$d = \frac{\|T_i - \hat{T}_i\|_2}{\|T_i\|_2} \quad (\text{B.3})$$

where  $T_i, i = 1, 2$  is the original trend, and  $\hat{T}_i, i = 1, 2$  is the trend extracted using EMD-MPS. The values of these measures are given in Table B.1, and show a performance better than or comparable to what is demonstrated in [51][43], but using a much simpler and intuitive approach based on time-scale based de-trending.

Table B.1: Objective measures of the quality of trends extracted using EMD-MPS from synthetic signals  $y^1(t)$ ,  $y^2(t)$  and  $y^3(t)$ .

Measure	$T_1(y^1(t))$	$T_1(y^2(t))$	$T_2(y^3(t))$
Correlation coefficient	0.9997	0.9956	0.9990
$d$ (Eq.B.3)	0.0114	0.0430	0.0451

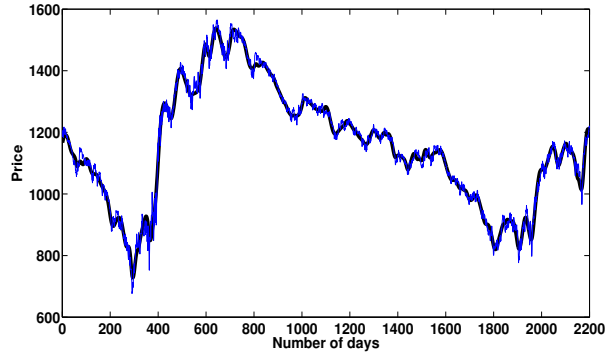
## B.2 De-noising and De-trending of Real-life Signals

**Financial time-series:** The first example in this section consists of a financial time-series in the form of S&P 500 daily index between August 1, 2001 to May 3, 2010. An example using the same time-series from a different date range was presented in Section 4.2.2 of Chapter 4, where calculation of the value of  $\tau$  for extracting trends at different time-scales was also described. In the same way, the 3, 6 and 9-monthly trends were extracted from the data being considered here using EMD-MPS with appropriate values of  $\tau$ . The results are shown in Figure B.3, and the variability of the trend based on the associated time-scale can be clearly seen.

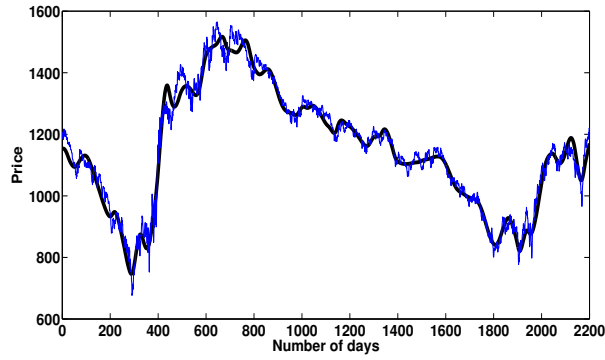
**Pathological speech:** Automated pathological speech detection using analysis of the speech signals is an active area of research [4]. Pre-processing of the speech signals is often used as part of such an analysis. Speech de-noising is also an important signal processing application [44][45].

Figure B.4 (a) shows a 0.5 second segment of a pathological speech signal. In order to limit the analysis of the speech signal to a spectrum of 2 kHz only, for example to remove unnecessary noise and interference, EMD-MPS can be applied to the signal to remove frequency components above the threshold of 2 kHz as  $\tau$ -function  $T_1$ , and to have the components with frequencies within the 2 kHz range in  $\tau$ -function  $T_2$ . The speech signal of Figure B.4 (a) is sampled at 25000 samples per second, and a value of  $\tau = 12.5$  corresponding to 2 kHz is needed for the purpose, according to the relation in Eq. 4.1. The  $\tau$ -function  $T_1$  containing the frequency components above the threshold is shown in Figure B.4 (b),

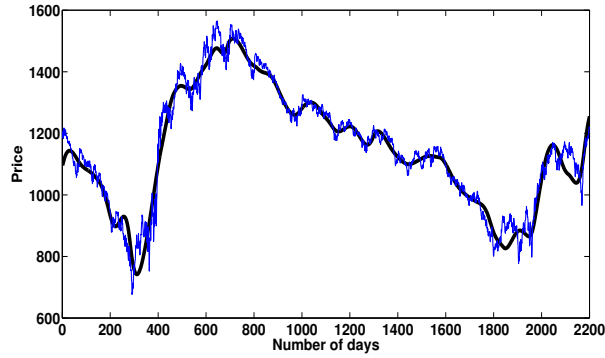




(a)



(b)



(a)

Figure B.3: Time-scale based de-trending of S&P 500 index data from August 1, 2001 to May 3, 2010, showing the 3-monthly trend in (a), 6-monthly trend in (b) and 9-monthly trend in (c).

whereas Figure B.4 (c) shows  $T_2$  containing the de-noised signal, which can be used for further analysis. The marginal spectrum of the two  $\tau$ -functions shown in Figure B.4 (d)

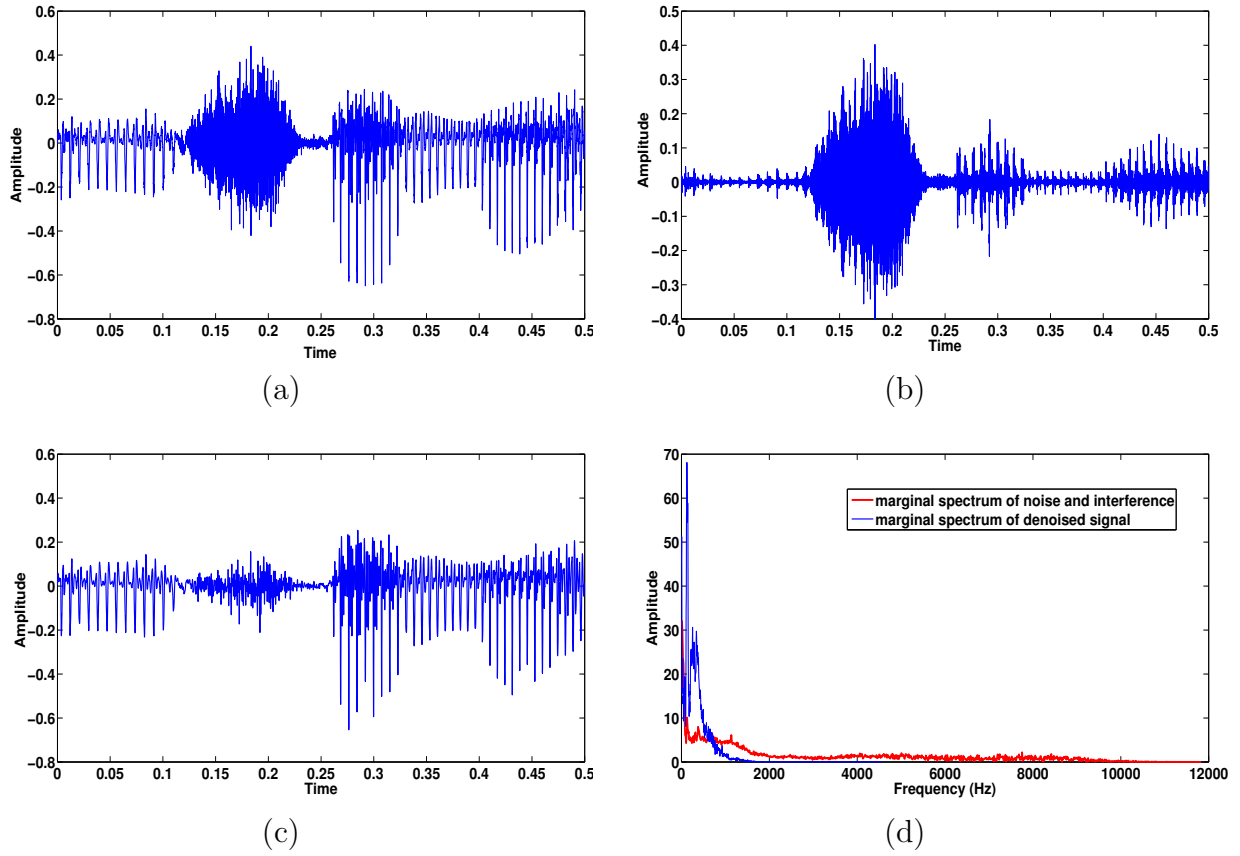


Figure B.4: Pathological speech signal de-noising using EMD-MPS. (a) A 0.5 second segment of a pathological speech signal. (b)  $\tau$ -function  $T_1$  containing oscillatory components above a frequency threshold of 2 kHz. (c)  $\tau$ -function  $T_2$  containing the de-noised signal. (d) Marginal spectrum of  $T_1$  and  $T_2$  showing separation of the spectra of the noise and de-noised signal.

clearly shows the separation of the spectra.

**ECG signals:** The use of EMD-MPS for baseline removal from ECG signals was discussed previously in Chapter 4. Another example demonstrating baseline wander removal from four different segments of ECG signals, using the same procedure as described in Section 4.2.2 of Chapter 4, is shown in Figure B.5.

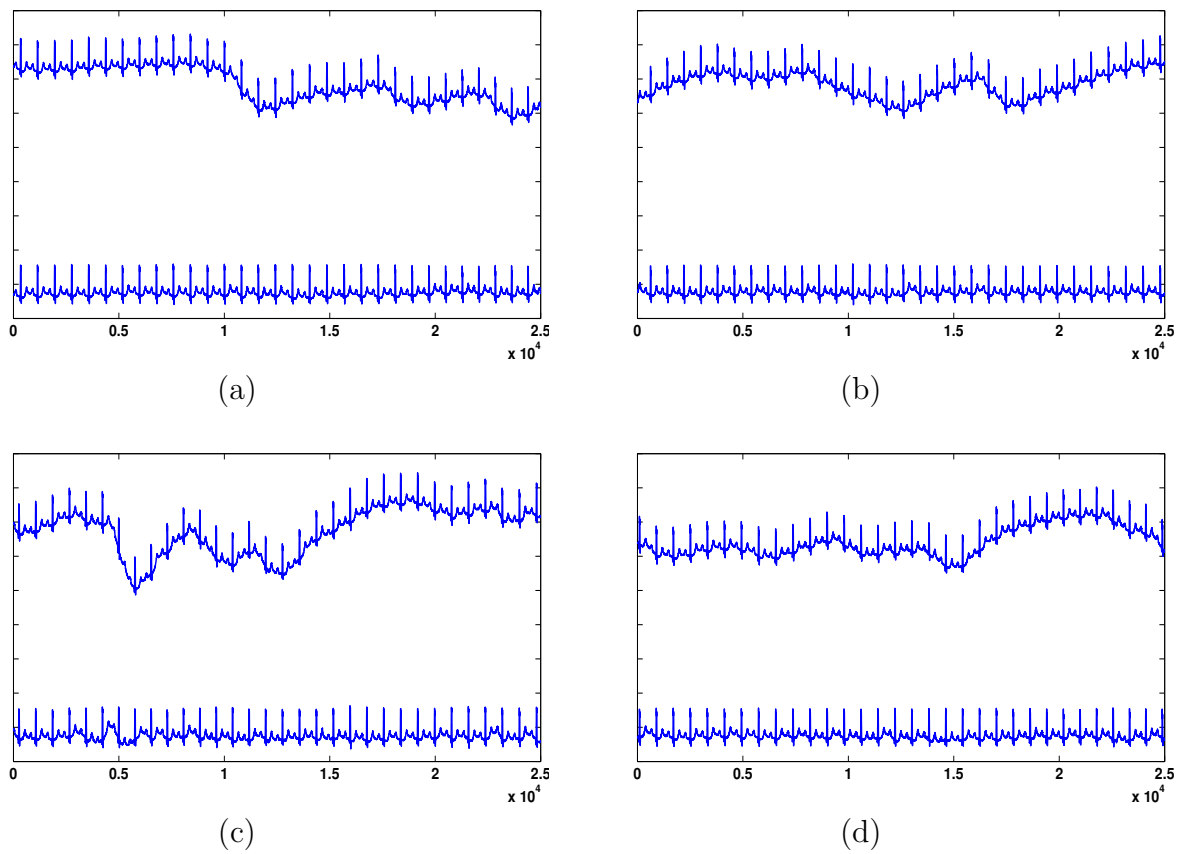


Figure B.5: Baseline wander removal from 25000 sample long segments of Holter ECG signals using EMD-MPS de-trending. The upper graphs in each figure represent the original ECG segments, whereas the lower graphs represent the de-trended ECG segments.

# References

- [1] “National Climate Data and Information Archive (Environment Canada),” [http://climate.weatheroffice.gc.ca/climateData/canada\\_e.html](http://climate.weatheroffice.gc.ca/climateData/canada_e.html). Accessed: 17 July 2013.
- [2] R. Moran, R. B. Reilly, P. de Chazal, and P. Lacy, “Telephony-based voice pathology assessment using automated speech analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 468–477, March 2006.
- [3] K. Drakakis, “Application of signal processing to the analysis of financial data [In the Spotlight],” *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 160–158, 2009.
- [4] M. Kaleem, B. Ghoraani, A. Guergachi, and S. Krishnan, “Pathological speech signal analysis and classification using empirical mode decomposition,” *Medical & Biological Engineering & Computing*, vol. 51, no. 7, pp. 811–821, July 2013.
- [5] B. Ghoraani, “Time-frequency Feature Analysis,” Ph.D. dissertation, Ryerson University, Toronto, Canada, September 2010.
- [6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

- [7] N. Tsakalozos, K. Drakakis, and S. Rickard, "A formal study of the nonlinearity and consistency of the Empirical Mode Decomposition," *Signal Processing*, vol. 92, no. 9, pp. 1961–1969, September 2012.
- [8] N. E. Huang, Z. Shen, and S. R. Long, "A New View of Nonlinear Water Waves: The Hilbert Spectrum," *Annual Review of Fluid Mechanics*, vol. 31, pp. 417–457, 1999.
- [9] F. He, S. A. Billings, H.-L. Wei, P. G. Sarrigiannis, and Y. Zhao, "Spectral Analysis for Nonstationary and Nonlinear Systems: A Discrete-Time-Model-Based Approach," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2233–2241, 2013.
- [10] Z. W. Danilo P. Mandic, Naveed ur Rehman and N. E. Huang, "Empirical Mode Decomposition-Based Time-Frequency Analysis of Multivariate Signals: The Power of Adaptive Data Analysis," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 74–86, October 2013.
- [11] B. M. Battista, C. Knapp, T. McGee, and V. Goebel, "Application of the Empirical Mode Decomposition and Hilbert-Huang transform to seismic reflection data," *Geophysics*, vol. 72, no. 2, pp. H29–H37, March-April 2007.
- [12] P. F. Pai, "Nonlinear vibration characterization by signal decomposition," *Journal of Sound and Vibration*, vol. 37, no. 2007, pp. 527–544, 2007.
- [13] G. G. Leisk, N. N. Hsu, and N. E. Huang, "Application of the Hilbert-Huang transform to machine tool condition/health monitoring," *AIP Conf. Proc.*, vol. 615, no. 1711 (2002), 2002. [Online]. Available: <http://dx.doi.org/10.1063/1.1472999>
- [14] N. E. Huang, M.-L. Wu, W. Qu, S. R. Long, and S. S. P. Shen, "Applications of Hilbert-Huang transform to non-stationary financial time series analysis," *Applied Stochastic Models in Business and Industry*, vol. 19, no. 3, pp. 245–268, 2003.
- [15] C. Park, D. Looney, P. Kidmose, M. Ungstrup, and D. Mandic, "Time-Frequency Analysis of EEG Asymmetry Using Bivariate Empirical Mode Decomposition," *IEEE*

- Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 4, pp. 366–373, 2011.
- [16] G. Fele-Zorz, G. Kavsek, Z. Noval-Antolic, and F. Jager, “A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups,” *Medical & Biological Engineering & Computing*, vol. 46, no. 9, pp. 911–922, September 2008.
  - [17] D.-J. Yua and W.-X. Rena, “EMD-based stochastic subspace identification of structures from operational vibration measurements,” *Engineering Structures*, vol. 27, no. 2005, pp. 1741–1751, 2005.
  - [18] R. Srinivasana, R. Rengaswamy, and R. Miller, “A modified empirical mode decomposition (EMD) process for oscillation characterization in control loops,” *Control Engineering Practice*, vol. 15, no. 2007, pp. 1135–1148, 2007.
  - [19] V. Agarwal and L. H. Tsoukalas, “Denoising Electrical Signal via Empirical Mode Decomposition,” in *Proceedings of iREP Symposium- Bulk Power System Dynamics and Control - VII, Revitalizing Operational Reliability, Charleston, SC, USA.*, 2007, pp. 1–6.
  - [20] K. Drakakis, “Empirical Mode Decomposition of Financial Data,” *International Mathematical Forum*, vol. 3, no. 25, pp. 1191–1202, 2008.
  - [21] B. M. Battista, A. D. Addison, and C. C. Knapp, “Empirical Mode Decomposition Operator for Dewowing GPR Data,” *Journal of Environmental and Engineering Geophysics*, vol. 14, no. 4, pp. 163–169, December 2009.
  - [22] T. B. J. Kuo, C. C. H. Yang, and N. E. Huang, “Quantification of Respiratory Sinus Arrhythmia using Hilbert-Huang Transform,” *Advances in Adaptive Data Analysis*, vol. 1, no. 2, pp. 295–307, 2009.

- [23] B. Mijovic, M. De Vos, G. I., J. Taelman, and S. Van Huffel, “Source Separation From Single-Channel Recordings by Combining Empirical-Mode Decomposition and Independent Component Analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, pp. 2188–2196, September 2010.
- [24] J. M. Hughes, D. Mao, D. N. Rockmore, Y. Wang, and Q. Wu, “Empirical Mode Decomposition Analysis for Visual Stylometry,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2147–2157, November 2012.
- [25] J. Lee, D. McManus, S. Merchant, and K. Chon, “Automatic Motion and Noise Artifact Detection in Holter ECG Data Using Empirical Mode Decomposition and Statistical Approaches,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1499–1506, June 2012.
- [26] S. Alam and M. Bhuiyan, “Detection of Seizure and Epilepsy Using Higher Order Statistics in the EMD Domain,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 2, pp. 312–318, 2013.
- [27] R. Fontugne, N. Tremblay, P. Borgnat, P. Flandrin, and H. Esaki, “Mining Anomalous Electricity Consumption Using Ensemble Empirical Mode Decomposition,” in *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, Canada*, 2013, pp. 5238–5242.
- [28] M. Chen, D. P. Mandic, P. Kidmose, and M. Ungstrup, “Qualitative assessment of intrinsic mode functions of empirical mode decomposition,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008, Las Vegas, NV, USA*, 2008, pp. 1905–1908.
- [29] P. Flandrin and P. Goncalves, “Empirical mode decompositions as data-driven wavelet-like expansions,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 4, pp. 477–496, 2004.

- [30] P. Flandrin, “Empirical Mode Decomposition vs. wavelets for the analysis of scaling processes (University Talk),” 2010, Accessed: July 29, 2013. [Online]. Available: <http://www.afaw.ulg.ac.be/scam/talks/flandrin.pdf>
- [31] Q. Chen, N. Huang, S. Riemenschneider, and Y. Xu, “A B-spline approach for empirical mode decompositions,” *Advances in Computational Mathematics*, vol. 2006, no. 24, pp. 171–195, May 2006.
- [32] T. Oberlin, S. Meignen, and V. Perrier, “An Alternative Formulation for the Empirical Mode Decomposition,” *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2236–2246, May 2012.
- [33] J. Fleureau, J.-C. Nunes, A. Kachenoura, L. Albera, and L. Senhadji, “Turning Tangent Empirical Mode Decomposition: A Framework for Mono- and Multivariate Signals,” *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1309–1316, March 2011.
- [34] P. Flandrin, P. Goncalves, and G. Rilling, “Detrending and Denoising with Empirical Mode Decompositions,” in *Proceedings of the 12th European Signal Processing Conference, Vienna, Austria*, 2004, pp. 1581–1584.
- [35] R. Rato, M. Ortigueira, and A. Batista, “On the HHT, its problems, and some solutions,” *Mechanical Systems and Signal Processing*, vol. 22, no. 2008, pp. 1374–1394, 2008.
- [36] O. Adam, “Advantages of the Hilbert Huang transform for marine mammals signal analysis,” *Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2965–2973, 2006.
- [37] Z. Wu and N. E. Huang, “A study of the characteristics of white noise using the empirical mode decomposition method,” *Royal Society of London Proceedings Series A*, vol. 460, no. 2046, pp. 1597–1611, March 2004.



- [38] J. Gao, H. Sultan, J. Hu, and W.-W. Tung, “Denoising Nonlinear Time Series by Adaptive Filtering and Wavelet Shrinkage: A Comparison,” *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 237–240, March 2010.
- [39] D. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [40] Z. Wu, N. E. Huang, S. R. Long, and C.-K. Peng, “On the trend, detrending, and variability of nonlinear and nonstationary time series,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 38, pp. 14 889–14 894, September 2007.
- [41] A. Moghtaderi, P. Borgnat, and P. Flandrin, “Trend Extraction for Seasonal Time Series Using Ensemble Empirical Mode Decomposition,” *Advances in Adaptive Data Analysis*, vol. 3, no. 1 & 2, pp. 41–61, 2011.
- [42] M. W. Watson, “Univariate Detrending Methods with Stochastic Trends,” *Journal of Monetary Economics*, vol. 18, no. 1986, pp. 49–75, 1986.
- [43] A. Moghtaderi, P. Borgnat, and P. Flandrin, “Trend Filtering: Empirical Mode Decompositions versus  $l_1$  and Hodrick-Prescott,” *Advances in Adaptive Data Analysis*, vol. 3, no. 1, pp. 41–61, March 2011.
- [44] D. Looney, L. Li, T. M. Rutkowski, D. P. Mandic, and A. Cichocki, “Speech Enhancement using Adaptive Empirical Mode Decomposition,” in *Proceedings of the 16th International Conference on Digital Signal Processing, Santorini-Hellas, Greece*, 2009, pp. 1–6.
- [45] T. Hasan and M. K. Hasan, “Suppression of Residual Noise From Speech Signals Using Empirical Mode Decomposition,” *IEEE Signal Processing Letters*, vol. 16, no. 1, pp. 2–5, January 2009.
- [46] D. Looney, L. Li, T. M. Rutkowski, D. P. Mandic, and A. Cichocki, “Ocular Artifacts Removal from EEG Using EMD,” in *Advances in Cognitive Neurodynamics ICCN*

- 2007, R. Wang, E. Shen, and F. Gu, Eds. Springer Netherlands, 2008, pp. 831–835.  
[Online]. Available: [http://dx.doi.org/10.1007/978-1-4020-8387-7\\_145](http://dx.doi.org/10.1007/978-1-4020-8387-7_145)
- [47] S. Baykut and T. Akgul, “MultiScale Zero-Crossing Statistics of Intrinsic Mode Functions for White Gaussian Noise,” in *Proceedings of the 18th European Signal Processing Conference (EUSIPCO-2010), Aalborg, Denmark*, 2010, pp. 135–138.
- [48] K. Khladi, A.-O. Boudraa, A. Bouchikhi, M. T.-H. Alouane, and E.-H. S. Diop, “Speech signal noise reduction by EMD,” in *Proceedings of 3rd International Symposium on Communications, Control and Signal Processing, ISCCSP 2008, Malta*, 2008, pp. 1155–1158.
- [49] D. Boutana, M. Benidir, and B. Barkat, “Denoising and characterization of heart sound signals using optimal intrinsic mode functions,” in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, Barcelona, Spain*, 2011, pp. 1–5.
- [50] Z. Wu and N. E. Huang, “Ensemble Empirical Mode Decomposition: A Noise Assisted Data Analysis Method,” *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [51] F. Mhamdi, J.-M. Poggio, and M. Jaidane, “Trend Extraction for Seasonal Time Series Using Ensemble Empirical Mode Decomposition,” *Advances in Adaptive Data Analysis*, vol. 3, no. 3, pp. 363–383, 2011.
- [52] C. Anderson and J. Bratman, “Translating Thoughts Into Actions by Finding Patterns in Brainwave,” in *Proceedings of 14th Yale Workshop on Adaptive and Learning Systems, Yale University, New Haven, CT, USA*, 2008, pp. 1–6.
- [53] Q. Shuren and J. Zhong, “Extraction of features in EEG signals with the non-stationary signal analysis technology,” in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), San Francisco, CA, USA*, 2004, pp. 349–352.

- [54] P. Maragos, J. Kaiser, and T. Quatieri, “On amplitude and frequency demodulation using energy operators,” *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1532–1550, 1993.
- [55] J.-C. Cexus and A.-O. Boudraa, “Non-stationary Signal Analysis by Teager-Huang Transform,” in *Proceedings of 14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy*, 2006, pp. 1–5.
- [56] Z. Zhao and C. Ma, “An Intelligent System for Noninvasive Diagnosis of Coronary Artery Disease with EMD-TEO and BP Neural Network,” in *Proceedings of International Workshop on Education Technology and Training and International Workshop on Geoscience and Remote Sensing, Shanghai, China*, 2008, pp. 631–635.
- [57] F. Jabloun, A. Cetin, and E. Erzin, “Teager Energy Based Feature Parameters for Speech Recognition in Car Noise,” *IEEE Signal Processing Letters*, vol. 6, no. 10, 1999.
- [58] P. Diez, V. Mut, E. Laciara, A. Torres, and E. Avila, “Application of the empirical mode decomposition to the extraction of features from EEG signals for mental task classification,” in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Minneapolis, Minnesota, USA*, 2009, pp. 2579–2582.
- [59] L. Orosco, E. Laciara, A. Correa, A. Torres, and J. Graffigna, “An Epileptic Seizures Detection Algorithm based on the Empirical Mode Decomposition of EEG,” in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Minneapolis, Minnesota, USA*, 2009, pp. 2651–2654.
- [60] A. Roy, C.-H. Wen, J. F. Doherty, and J. D. Mathews, “Signal Feature Extraction From Microbarograph Observations Using the Hilbert-Huang Transform,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1442–1447, 2008.

- [61] A. M. Bassiuny, X. Li, and R. Du, "Fault diagnosis of stamping process based on Empirical Mode Decomposition and learning vector quantization," *International Journal of Machine Tools and Manufacture*, vol. 47, no. 15, pp. 2298–2306, 2007.
- [62] V. Parsa and D. G. Jamieson, "Acoustic Discrimination of Pathological Voice: Sustained Vowels Versus Continuous Speech," *Journal of Speech, Language & Hearing Research*, vol. 4, no. 2, pp. 327–338, April 2001.
- [63] P. Henriquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. D. de Maria, "Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1186–1195, August 2009.
- [64] V. Parsa and D. G. Jamieson, "Identification of Pathological Voices Using Glottal Noise Measures," *Journal of Speech, Language & Hearing Research*, vol. 43, no. 2, pp. 469–485, April 2000.
- [65] N. Saenz-Lechona, J. I. Godino-Llorente, V. Osma-Ruiza, and P. Gomez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 120–128, April 2006.
- [66] A. Gelzinis, A. Verikas, and M. Bacauskiene, "Automated speech analysis applied to laryngeal disease categorization," *Computer methods and programs in biomedicine*, vol. 91, no. 1, pp. 36–47, July 2008.
- [67] G. Schlotthauer, M. E. Torres, and M. C. Jackson-Menaldi, "A Pattern Recognition Approach to Spasmodic Dysphonia and Muscle Tension Dysphonia Automatic Classification," *Journal of Voice*, vol. 24, no. 3, pp. 346–353, May 2010.
- [68] J. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 380–384, 2004.

- [69] K. Shama, A. Krishna, and N. U. Cholayya, “Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology,” *EURASIP Journal on Advances in Signal Processing*, vol. 2007, p. 9 pages, 2007.
- [70] M. Markaki, Y. Stylianou, J. Arias-Londono, and J. Godino-Llorente, “Dysphonia detection based on modulation spectral features and cepstral coefficients,” in *Proceedings of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, Texas, USA*, 2010, pp. 5162–5165.
- [71] K. Umapathy, S. Krishnan, V. Parsa, and D. G. Jamieson, “Discrimination of Pathological Voices Using a Time-Frequency Approach,” *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 3, pp. 421–430, March 2005.
- [72] B. Ghoraani and S. Krishnan, “A Joint Time-Frequency and Matrix Decomposition Feature Extraction Methodology for Pathological Voice Classification,” *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 11 pages, 2009.
- [73] G. Schlotthauer, M. E. Torres, and H. L. Rufiner, “Voice Fundamental Frequency Extraction Algorithm Based on Ensemble Empirical Mode Decomposition and Entropies,” in *Proceedings of 11th World Congress of IFMBE, Munich, Germany*, 2009, pp. 984–987.
- [74] —, “Pathological Voice Analysis and Classification Based on Empirical Mode Decomposition,” *Development of Multimodal Interfaces: Active Listening and Synchrony*, vol. 5967, pp. 364–381, 2010.
- [75] “Massachusetts Eye and Ear Infirmary Voice Disorders Database, Version 1.03 (CDROM),” Kay Elemetrics Corporation, Lincoln Park, NJ. (1994).
- [76] L. Sugavaneswaran, K. Umapathy, and S. Krishnan, “Exploiting the ambiguity domain for non-stationary biomedical signal classification,” in *Proceedings of the 31st Annual*

- International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Buenos Aires, Argentina, 2010, pp. 1934–1937.
- [77] N. Malyska, T. Quatieri, and D. Sturim, “Automatic Dysphonia Recognition using Biologically-Inspired Amplitude-Modulation Features,” in *Proceedings of 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, USA*, 2005, pp. 873–876.
  - [78] S. Furui, “On the role of spectral transition for speech perception,” *Journal of the Acoustical Society of America*, vol. 80, no. 4, pp. 1016–1025, October 1986.
  - [79] Matlab codes for Empirical Mode Decomposition algorithm. [Online]. Available: <http://perso.ens-lyon.fr/patrick.flandrin/emd.html>
  - [80] T. P. Hettmansperger and J. McKean, *Robust Nonparametric Statistical Models*, 2nd ed. CRC Press, 1998.
  - [81] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley and Sons, Inc., 2001.
  - [82] P. Flandrin, G. Rilling, and P. Goncalves, “Empirical Mode Decomposition as a Filter Bank,” *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, February 2004.
  - [83] N. U. Rehman and D. P. Mandic, “Filter Bank Property of Multivariate Empirical Mode Decomposition,” *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2421–2426, 2011.
  - [84] M. A. Colominas, G. Schlotthauer, M. E. Torres, and P. Flandrin, “Noise-Assisted EMD Methods in Action,” *Advances in Adaptive Data Analysis*, vol. 4, no. 4, pp. 1–11, 2012.
  - [85] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” in *Proceedings of the 2011*

- IEEE International Conference on Acoustics, Speech, and Signal Processing, Prague, Czech Republic*, 2011, pp. 4144–4147.
- [86] R. Deering and J. Kaiser, “The use of a masking signal to improve empirical mode decomposition,” in *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, USA*, 2005, pp. 485–488.
  - [87] X. Hu, S. Peng, and W.-L. Hwang, “EMD Revisited: A New Understanding of the Envelope and Resolving the Mode-Mixing Problem in AM-FM Signals,” *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1075–1086, 2012.
  - [88] B. Xuan, Q. Xie, and S. Peng, “EMD Sifting Based on Bandwidth,” *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 537–540, 2007.
  - [89] Y. Washizawa, T. Tanaka, D. Mandic, and A. Cichocki, “A Flexible Method for Envelope Estimation in Empirical Mode Decomposition,” in *Knowledge-Based Intelligent Information and Engineering Systems*, ser. Lecture Notes in Computer Science, B. Gabrys, R. Howlett, and L. Jain, Eds. Springer Berlin Heidelberg, 2006, vol. 4253, pp. 1248–1255. [Online]. Available: [http://dx.doi.org/10.1007/11893011\\_158](http://dx.doi.org/10.1007/11893011_158)
  - [90] Z. Xu, B. Huang, and K. Li, “An alternative envelope approach for empirical mode decomposition,” *Digital Signal Processing*, vol. 20, no. 2010, pp. 77–84, 2010.
  - [91] H. Hong, X. Wang, and Z. Tao, “Local Integral Mean-Based Sifting for Empirical Mode Decomposition,” *IEEE Signal Processing Letters*, vol. 16, no. 10, pp. 841–844, 2009.
  - [92] C. Li, X. Wang, Z. Tao, Q. Wang, and S. Du, “Extraction of time varying information from noisy signals: An approach based on the empirical mode decomposition,” *Mechanical Systems and Signal Processing*, vol. 25, no. 2011, pp. 812–820, 2011.
  - [93] A. Roy and J. Doherty, “Raised cosine filter-based empirical mode decomposition,” *IET Signal Processing*, vol. 5, no. 2, pp. 121–129, 2009.

- [94] T. Oberlin, S. Meignen, and V. Perrier, “An Alternative Formulation for the Empirical Mode Decomposition,” *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2236–2246, 2012.
- [95] N. Pustelnik, P. Borgnat, and P. Flandrin, “A Multicomponent Proximal Algorithm for Empirical Mode Decomposition,” in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania*, 2012, pp. 1880–1884.
- [96] E. Delechelle, J. Lemoine, and O. Niang, “Empirical Mode Decomposition: An Analytical Approach for Sifting Process,” *IEEE Signal Processing Letters*, vol. 12, no. 11, pp. 764–767, 2005.
- [97] E. H. S. Diop, R. Alexandre, and A. O. Boudraa, “Analysis of Intrinsic Mode Functions: A PDE Approach,” *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 398–401, 2010.
- [98] O. Niang, E. Delechelle, and J. Lemoine, “A Spectral Approach for Sifting Process in Empirical Mode Decomposition,” *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5612–5623, 2010.
- [99] G. Rilling, P. Flandrin, and P. Goncalves, “Empirical Mode Decomposition, Fractional Gaussian Noise and Hurst Exponent Estimation,” in *Proceedings of 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, USA*, 2005, pp. 489–492.
- [100] M. F. Kaleem, A. E. Cetin, A. Guergachi, and S. Krishnan, “Using a Variation of Empirical Mode Decomposition To Remove Noise From Signals,” in *Proceedings of the 21st International Conference on Noise and Fluctuations (ICNF), Toronto, Canada*, 2011, pp. 123–126.
- [101] G. Rilling and P. Flandrin, “One or Two Frequencies? The Empirical Mode Decomposition Answers,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 85–95, January 2008.



- [102] P. Flandrin and P. Goncalves, “Empirical Mode Decompositions as Data-driven Wavelet-like Expansions,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 4, pp. 1–20, March 2004.
- [103] G. Schlotthauer, M. E. Torres, H. L. Rufiner, and P. Flandrin, “EMD of Gaussian White Noise: Effects of Signal Length and Sifting Number on the Statistical Properties of Intrinsic Mode Functions,” *Advances in Adaptive Data Analysis*, vol. 1, no. 4, pp. 517–527, 2009.
- [104] Z. Wu and N. E. Huang, “On the Filtering Properties of the Empirical Mode Decomposition,” *Advances in Adaptive Data Analysis*, vol. 2, no. 4, pp. 397–414, 2010.
- [105] A. Shoeb, “Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment,” Ph.D. dissertation, Massachusetts Institute of Technology, September 2009.
- [106] Y. Liu, W. Zhou, Q. Yuan, and S. Chen, “Automatic Seizure Detection Using Wavelet Transform and SVM in Long-Term Intracranial EEG,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 6, pp. 749–755, 2012.
- [107] M. K. I. Molla, T. Tanaka, T. M. Rutkowski, and A. Cichocki, “Separation of EOG Artifacts from EEG Signals using Bivariate EMD,” in *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, USA*, 2010, pp. 562–565.
- [108] Suman, S. Devi, and M. Dutta, “Optimized Noise Canceller for ECG Signals,” in *Proceedings of International Conference on Intelligent Systems and Data Processing (ICISD) 2011, Gujarat, India*, 2011, pp. 10–17.
- [109] S. Pal and M. Mitra, “Empirical mode decomposition based ECG enhancement and QRS detection,” *Computers in Biology and Medicine*, vol. 42, no. 2012, pp. 83–92, 2012.

- [110] B. N. Singha and A. K. Tiwari, “Optimal selection of wavelet basis function applied to ECG signal denoising,” *Digital Signal Processing*, vol. 16, no. 2006, pp. 275–287, 2006.
- [111] M. Blanco-Velasco, B. Weng, and K. E. Barner, “ECG signal denoising and baseline wander correction based on the empirical mode decomposition,” *Computers in Biology and Medicine*, vol. 38, no. 2008, pp. 1–13, 2008.
- [112] M. A. Kabir and C. Shahnaz, “Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains,” *Biomedical Signal Processing and Control*, vol. 7, no. 5, pp. 481–489, September 2012.
- [113] M. Kaleem, L. Sugavaneswaran, A. Guergachi, and S. Krishnan, “Application of Empirical Mode Decomposition and Teager Energy Operator to EEG Signals for Mental Task Classification,” in *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Buenos Aires, Argentina, 2010*, pp. 4590–4593.
- [114] M. Kaleem, A. Guergachi, and S. Krishnan, “A Variation of Empirical Mode Decomposition with Intelligent Peak Selection in Short Time Windows,” in *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, Canada, 2013*, pp. 5627–5631.
- [115] S. Xie and S. Krishnan, “Wavelet-based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis,” *Medical & Biological Engineering & Computing*, vol. 51, no. 1-2, pp. 49–60, 2012.
- [116] L. Guo, D. Rivero, and A. Pazos, “Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks,” *Journal of Neuroscience Methods*, vol. 193, no. 1, pp. 156–163, 2010.
- [117] A. Tzallas, M. Tsipouras, and D. Fotiadis, “Automatic Seizure Detection Based on Time-Frequency Analysis and Artificial Neural Networks,” *Computational*

- Intelligence and Neuroscience*, vol. 2007, 2007. [Online]. Available: <http://dx.doi.org/10.1155/2007/80510>
- [118] R. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. Elger, “Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” *Physical Review E*, vol. 64, no. 6, 2001.
  - [119] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, “A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Subbands to Detect Seizure and Epilepsy,” *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 205–211, February 2007.
  - [120] A. Shoeb and J. Guttag, “Application of Machine Learning To Epileptic Seizure Detection,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel*, 2010, pp. 975–982.
  - [121] R. Rubinstein, A. Bruckstein, and M. Elad, “Dictionaries for Sparse Representation Modeling,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, June 2010.
  - [122] M. G. Jafari and M. D. Plumbley, “Fast Dictionary Learning for Sparse Representations of Speech Signals,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1025–1031, September 2011.
  - [123] M. Zibulevsky and M. Elad, “Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation,” *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1553–1564, March 2010.
  - [124] Q. Zhang and B. Li, “Discriminative K-SVD for Dictionary Learning in Face Recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, CA, USA*, 2010, pp. 2691–2698.
  - [125] K. Huang and S. Aviyente, “Sparse Representation for Signal Classification,” in *Pro-*

- ceedings of the Twentieth Annual Conference on Neural Information Processing Systems (NIPS), Vancouver, Canada, 2006, pp. 609–616.*
- [126] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, “Discriminative Learned Dictionaries for Local Image Analysis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, Alaska, USA, 2008, pp. 1–8.*
  - [127] I. Ramirez, P. Sprechmann, and G. Sapiro, “Classification and clustering via dictionary learning with structured incoherence and shared features,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, CA, USA, 2010, pp. 3501–3508.*
  - [128] Z. Jiang, Z. Lin, and L. Davis, “Learning a Discriminative Dictionary for Sparse Coding via Label Consistent K-SVD,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011), Colorado Springs, CO, USA, 2011, pp. 1697–1704.*
  - [129] S. Mallat and Z. Zhang, “Matching Pursuits With Time-Frequency Dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.
  - [130] T. Y. hou and Z. Shi, “Data-Driven Time-Frequency Analysis,” arXiv:1202.5621v1.
  - [131] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, November 2006.
  - [132] A. Ayenu-Prah and N. Attoh-Okine, “A Criterion for Selecting Relevant Intrinsic Mode Functions in Empirical Mode Decomposition,” *Advances in Adaptive Data Analysis*, vol. 2, no. 1, pp. 1–24, 2010.

- [133] A. S. Zandi, R. Tafreshi, M. Javidan, and G. A. Dumont, “Predicting Epileptic Seizures in Scalp EEG Based on a Variational Bayesian Gaussian Mixture Model of Zero-Crossing Intervals,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1401–1413, 2013.
- [134] J. Feder, *Fractals*. Plenum Publishing Corporation, New York., 1988.
- [135] F. P. Anaraki and S. M. Hughes, “Compressive K-SVD,” in *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, Canada*, 2013, pp. 5469–5473.
- [136] N. E. Huang, M.-L. C. Wu, S. R. Long, S. S. P. Shen, W. Qu, P. Gloersen, and K. L. Fan, “A confidence limit for the empirical mode decomposition and Hilbert spectral analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 459, no. 2037, pp. 2317–2345, 2003.
- [137] G. Rilling, P. Flandrin, and P. Goncalves, “On empirical mode decomposition and its algorithms,” in *Proceedings of the 6th IEEE/EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP '03), Trieste, Italy*, 2003.
- [138] N. Stevenson, M. Mesbah, and B. Boashash, “A sampling limit for the empirical mode decomposition,” in *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications (ISSPA), Sydney, Australia*, 2005, pp. 647–650.
- [139] G. Rilling and P. Flandrin, “Sampling Effects on the Empirical Mode Decomposition,” *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 43–59, 2009.