STATISTICAL METHODS IN DATA MINING OF BRAIN DYNAMICS

By

ALLA R. KAMMERDINER

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

To my wonderful family.

ACKNOWLEDGMENTS

I would like to show my heartfelt appreciation to my advisor Dr. Panos M. Pardalos for his support and mentoring. Working with him has helped me grow not only professionally, but also as a person. I am also very grateful to other members who served on my supervisory committee, J. Cole Smith, William W. Hager, Vladimir L. Boginski, and H. Edwin Romeijn, for their valuable comments on my research for this dissertation. Last but not least, I thank my husband Jason, my parents Olga and Aleksandr, my brother Mikhail, and all the rest of my great family for their unconditional love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, NOMENCLATURE, OR ABBREVIATIONS

Most of the notation is unambiguously defined in the text where it is introduced. To provide some general guidelines, we include the following list of commonly used symbols.

ARIMA(p, d, q)	autoregressive integrated moving average process
AR(p)	autoregressive process of order p
CPV	Cauchy principal value
$\overline{X}$	complex conjugate of X
$\rightarrow$	converges to
$\Rightarrow$	converges in distribution to
$*$	convolution of functions
det	determinant
$\in$	element of
$=$	equals
$:=$	equals by definition
$\equiv$	equivalent to
$\widehat{x}$	estimator of x
$\exists$	exists
E	expectation
exp	exponential function
$\forall$	for all
I	identity matrix
I_K	$(K \times K)$ identity matrix
$\Im$	imaginary part of a complex value
∞	infinity
$I(d)$	integrated process of order d
$\sim$	is distributed as
lim	limit

ln	natural logarithm
Pr	probability of a random event
rank	rank of a matrix
$\Re$	real part of a complex value
$\sum$	sum of terms
$\circ$	superposition
trace	trace
X'	transpose of X
Var	variance
VAR(p)	vector autoregressive process of order p

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

STATISTICAL METHODS IN DATA MINING OF BRAIN DYNAMICS

By

Alla R. Kammerdiner

May 2008

Chair: Panos M. Pardalos
Major: Industrial and Systems Engineering

This study discusses statistical approaches used for data mining of multichannel electroencephalogram recordings. Such recordings represent massive data sets that contain hidden patterns of complex dynamical processes in the brain. Formally, multichannel EEG can be viewed as a multiple time series, and therefore, a natural idea for summarizing such data is to utilize autoregressive modeling of multivariate stochastic processes.

In particular, we thoroughly discuss various concepts and approaches related vector autoregressive processes, including stable stationary VAR models of order p and nonstationary systems with integrated and cointegrated variables, as well as procedures for estimating parameters of the systems (e.g. order, lag, or cointegration rank).

The work highlights some stability issues that may arise in the application of vector autoregression to mining EEG data, and questions the applicability of Granger causality in the frequency domain to multichannel EEG.

Synchronization has been found to be an important characteristic of the abnormal brain dynamics manifested by epilepsy and Parkinson's disease. We review two approaches for extracting the instantaneous phase from time series. In this study, we generalize the concept of the phase synchronization, and propose a novel approach based on multivariate analysis via modeling cointegrated VAR(p) processes.

CHAPTER 1
INTRODUCTION

1.1 Statistical Methods for Data Mining

With significant improvements in obtaining, processing and storing the information electronically in the last decades, it became routine to accumulate large amounts of data in various fields of research and business. However, any information is only useful if it can be analyzed to draw some meaningful conclusions. A drastic increase in information loads makes the task of interpreting collected data especially challenging. Not only visual inspection and analysis of such massive data become extremely time-consuming and often ineffective, but also the limitations of the traditional numerical data analysis techniques in application to the massive collections of data necessitate the development of new approaches. As a result of continuous attempts of scientific community to extract useful information from large data sets, a multi-discipline field of data mining has been developed.

Data mining together with data preprocessing constitute the central part of a more general process of knowledge discovery in databases (KDD). The KKD process can be described as a sequence of actions, which selects the raw data in data warehouses and transforms the selected data in order to discover valid, understandable, novel and potentially useful knowledge from the data. Data preprocessing that is applied to raw data to improve the quality of the data often influences the selection and facilitates the application of data mining techniques. Proper preprocessing of raw data leads to a decrease in the time needed to mine the data, and boosts the overall mining efficiency.

The techniques used for data preprocessing can be roughly subdivided into data cleaning, data integration and data reduction [64]. Data cleaning techniques handle the problem of incomplete, inconsistent and erroneous data, remove the noise inherently present in the raw data, minimize redundancy in the data, etc. Data integration is concerned with combining heterogeneous data collected from different sources to form a

consistent data set. The process of data reduction amounts to identifying useful features, which are capable of adequately representing the data, and it is usually performed using dimensionality reduction and feature extraction methods.

The two fundamental tasks assigned to data mining are a descriptive task of discovering hidden patterns and relationships in given data, and a predictive task of forecasting or classifying the model's behavior from available data. Data mining includes regression, classification, clustering, image restoration, learning association rules and extracting functional dependencies, data summarization, etc. Data mining is closely connected to other research areas such as statistics, machine learning and artificial intelligence, optimization, visualization and databases. Data mining utilizes many important results from the related fields, while keeping the main focus on the algorithms and architectures, scalability of the number of features and instances, and automated managing of massive quantities of diverse data.

Many areas of data mining employ various approaches developed in the field of optimization. In particular, it is shown in [7] that many basic problems in data mining, including classification and clustering, can be formulated as mathematical programming problems and solved using optimization techniques.

In fact, Bradley at el. [7] demonstrated that a problem of minimizing the number of misclassified points in two-class classification can be viewed as a linear program with equilibrium constraints (LPEC). LPEC is a linear program (LP) with a single complementarity constraint. Such constraint imposes a condition of orthogonality between two linear functions. LPEC formulation arises in the instances when the constraints of the problem include another LP problem.

In addition, the problem of feature selection in two-class classification by finding a separating plane that utilizes minimum number of features can be given a mathematical programming formulation as a parametric problem. Furthermore, the classification via support vector machines (SVMs that find the separating plane maximizing the

margin between two classes while minimizing misclassification errors) can be stated as a quadratic programming problem [6]. Moreover, as indicated in [7], the above mathematical programming formulations have been extended to be effectively employed by other data mining approaches, including neural networks training, calculation of nonlinear discriminants, and building decision trees. The clustering problem has a complex formulation as a minimization problem with the objective given by a sum of the minimums of a set of convex functions [5]. In general, this objective function is neither convex nor concave. See review by Bradley at el. [7] for additional information about mathematical programming formulations for various problems in data mining, the application of optimization techniques, as well as the challenges that the field of data mining offers to optimization.

Many data mining approaches, such as classification, clustering, indexing and segmentation, have been applied to time series analysis [45]. Many traditional statistical approaches are also applied to mining time series. For instance, regression is one of the most commonly used techniques for modeling and forecasting time series. Among the statistical models applied to regression in time series are linear autoregression (AR), autoregressive moving average process (ARMA), autoregressive integrated moving average process (ARIMA), as well as their multivariate analogs (i.e. vector autoregression, etc.)

Time series arise in various applied areas, including economics and finance, meteorology, biomedicine, etc. For instance, the study of seismic activity related to earthquakes produces two-dimensional time series, where each measurement consists of the time and the magnitude of a registered seismic event. Many biomedical signals, such as electrocardiogram (ECG), electroencephalogram (EEG) and electrooculogram (EOG) represent time series that can be interpreted via application of regression, segmentation, neural networks, and other data mining methodologies. Sound signals are another example of time series that are effectively analyzed using different data mining techniques.

Time series, which originate in different fields, are generated by diverse underlying processes, and as a result they are often characterized by very distinct properties. Indeed, as indicated in [14], although the normal time scale is a very natural choice of parameter for the time series describing physical processes, the regular time looses its natural meaning when dealing with many financial time series. Because most of financial time series are irregularly spaced in physical time, the concept of "business time" or "intrinsic time" is introduced to represent a new time parameter with respect to which time series are regularly scaled. This procedure of time deformation allows the relabeled time series to be viewed as stationary on a new time scale. Financial time series also often exhibit clear seasonal trends, which obviously cannot be found when examining time series produced by speech.

Statistical testing of several multivariate time series determined that the time series from AUSLAN and BCI data sets can be considered stationary, whereas BCI MPI and EEG contained non-stationary time series [108]. As a result of inherent differences in time series data from diverse sources, some data mining methods that are successfully applied to time series in one research area may not necessarily be applicable to analysis of time series that stem from another applied field.

Time series obtained from electroencephalogram (EEG) recordings have several interesting properties that distinguish them from other time series. Although some studies apply one-dimensional modeling by considering one channel in EEG recording at a time, in general, EEG data should be treated as multivariate time series. The multivariate approach becomes especially important in view of its ability to investigate spatio-temporal dependencies in the EEG data in contrast to being limited to only temporal relations in a one-dimensional case. There is a disagreement among researchers studying EEG data on whether the series should be modeled by a non-linear stochastic process or they can be better described by a deterministic chaotic dynamical system.

Storage of continuous EEG recordings sampled from multiple channels at high frequency during several hours (or even days) from multiple subjects, wether it is for a sleep study or diagnosis of neurological disorder, may take gigabytes of memory. The application of data mining approaches to EEG time series allows automatic handling and analysis of such large data sets. With new technological advances in collecting EEG data, it becomes particularly important to develop new efficient data mining methods designed specifically for mining EEG data.

1.2 Electroencephalographic Recordings

Electroencephalography is one of the most commonly applied methods of extracting neurophysiological signals. It originated in 1875, when an English physician Richard Caton measured the electrical activity from the exposed brains of monkeys and rabbits [100].

Generally speaking, EEG represents a digital or a graphic record of the electrical activity in the brain, and can be measured by either non-invasive or invasive methods. EEG (obtained during a non-invasive procedure) is defined as a record of electrical activity of an alternating type measured from the scalp surface after being picked up by metal electrodes and conductive media [67]. There are two types of EEG produced by invasive procedures, the electrocorticogram, which measures the brain's electrical activity directly from the cortical surface, and the electrogram, which is an EEG obtained using deep probes.

EEG estimates and records the relative change in electric potentials produced by a large number of electric dipoles during a period of neural excitations. The activation of neurons (brain cells) generates local current flows in the brain. EEG records mostly the electrical currents that flow during synaptic excitations of the dendrites of numerous pyramidal neurons in the cerebral cortex. EEG recorded from the scalp surface can only detect the electrical activity produced by massive populations of active neurons. On the other hand, EEG recorded using deep probe electrodes implanted into the brain can pick up a signal from a small group of neurons, which can be further filtered out to obtain the

electric potentials generated by individual neurons. EEG has become an effective device in the area of neurological research as well as clinical neurology, because of its capacity to reveal both abnormal and normal brain activity. It is believed that by birth, human brain has already developed the full number of neural cells, which is approximately 10^{11} neurons [68]. This gives an average density of about 10^4 neural cells per cubic millimeter of the brain. Neural cells are interconnected through synaptic connections in the brain into neural nets. The brain of an average adult contains approximately 500 trillion synapses. The total number of neurons decreases with age. As a result the total number of synaptic connections declines with aging, even though the number of synapses per one neuron increases with age.

To ensure the consistency in referencing locations of electrodes in EEG experiments, the *International 10-20 system for EEG electrode placement* was developed [37]. The 10-20 EEG system is used to describe the respective locations of scalp electrodes during EEG recording in relation to the underlying area of cerebral cortex.

According to the 10-20 system, anatomical landmarks of a skull, such nasion, inion and preauricular points, are identified for consecutive placement of the electrodes at fixed distances from these points in steps of either 10 or 20 percent. This approach is devised to take into account possible variations of head size. In addition, the method is easily applicable in practical use. As a result, the 10-20 EEG system became very widely used for positioning electrodes.

In the 10-20 system, the points are denoted with one or two letters, and can be also followed by a number (as shown on Figure 1-1). The letters roughly represent the lobe location (with exception of letters C and Z), whereas the numbers serve for identifying the corresponding hemisphere. More specifically, the points located on the left hemisphere of the brain are represented by odd numbers (1, 3, 5, and 7), and the sites on the right hemisphere are marked with even numbers (2, 4, 6, and 8). The sites located on the frontal, temporal, parietal and occipital lobes are denoted by the corresponding initials

F, T, P, and O. The letters C and Z refer to the points placed in the central area. In particular, Z represents a point on the midline, and C refers to the line parallel to the midline. Note that the central area is not a lobe.

EEG signal resembles a collection of sinusoids of various amplitude and frequency. Power spectrum is extracted from the raw EEG data using Fourier transform to obtain the information about the contribution of sinusoidal waves of different frequency. The power spectrum of EEG is continuous, ranging from 0 Hz up to a half of the sampling frequency. Depending on the state of the brain, certain frequencies appear to be more prevalent. There are four major frequency bands, alpha, beta, delta and theta, which presence in EEG during various states of consciousness has been extensively studied. These bands represent sine waves of relatively low frequency, with delta ranging from 0.5 to 4 Hz, theta 4 – 8 Hz, alpha 8 – 13Hz, and beta over 13 Hz. Alpha waves discovered by Adrian and Matthews in 1934 are the best-known and the most studied among the four frequency bands [100]. They are induced by closing eyes and by relaxation, and terminated with eyes opening or due to thinking, calculating, and other analytical activities. In particular, in most people, eye closing produces rapid changes in brain activity manifesting themselves in EEG as an adjustment of the dominant frequency band from beta to alpha. EEG is capable of discriminating between different states, such as resting, alertness, stress state, various sleep stages, hypnosis, etc. Presence of beta band is dominant during the state of alertness with eyes open. Drowsiness or the resting are usually characterized by the rise in alpha activity. During the sleep, presence of lower frequency waves becomes more apparent. A higher proportion of delta band frequencies is observed during stages III and IV of the non-rapid eye movement sleep (NREM). EEG recorded from distinct regions in the brain exhibits different spectrum of wave frequencies. In addition, the brain patterns are unique for every individual.

Practical applications of EEG include epilepsy research and localization of the focus of epileptic seizures, testing of epilepsy drug effects; determining areas of damage due

to stroke, head injury, etc.; monitoring alertness, coma and brain death; testing afferent pathways by evoked potentials; research in sleep physiology and sleep disorder; controlling anaesthesia depth, etc. [3].

1.3 Feature extraction

Formally, EEG signal can be described as a deterministic multidimensional nonlinear non-stationary time series [94]. In order to properly reflect the spatio-temporal properties of brain dynamics, the analysis of EEG data must involve a simultaneous investigation of the dependencies across channels with respect to time.

Different features have been proposed for analysis of EEG time series, including Fourier transform, wavelets, cross-correlation, coherence, Granger causality and partial directed coherence, mutual information and transfer entropy, global and phase synchronization, Lyapunov exponents and correlation dimension, etc.

Since EEG can be viewed as a collection of sine waves, EEG series are often analyzed in a frequency domain. In addition, some frequency bands have shown to play specific roles in various states of consciousness, and so the frequency information in EEG can be particularly important. Subsequently, the Fourier transform with a running time window, also known as short time Fourier transform (STFT), became one of the most widely used methods for extracting features from EEG. STFT is obtained from Fourier transform by applying a time window function g with a time shift τ. Mathematically, STFT is given by the following formula:

$$S(\tau, f) = \int_{-\infty}^{+\infty} x(t) g(t-\tau) \exp\{-2i\pi \cdot f \cdot t\} dt, \tag{1–1}$$

where $S(\tau, f)$ denotes the STFT with time window g located at time τ, corresponding to frequency f; and $x(t)$ is a signal at time t. In other words, STFT $S(\tau, f)$ represents the power spectrum of the signal estimated around time τ. The drawback of STFT is that there is a trade off between time accuracy and frequency precision. By making the window

g smaller, the resolution of the time parameter is improved, unfortunately, at the expense of the resolution in frequency.

An alternative to Fourier transform is wavelet transform (WT), which is a transformation of the signal based on a special function, called mother wavelet (MW). The mother wavelet is shifted in time by a location parameter τ, and then adjusted by a scale parameter a. More precisely, the wavelet transform is defined by the following formula:

$$W(\tau, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi(\frac{t-\tau}{a})dt, \tag{1–2}$$

where ψ is a mother wavelet, a is scale parameter, τ is a time location parameter, and x is a signal.

The scale parameter a in WT is analogous to the frequency parameter f in STFT. In particular, the large values of parameter a ($a \gg 1$) stretch the wavelet, and so they represent low frequencies, whereas the small values of a ($a < 1$) shrink the wavelet function, which corresponds to higher frequencies. An advantage of using wavelets is that the high frequency components can be analyzed with a higher time accuracy than the lower frequency components of the signal.

As follows from (1–2), $W(\tau, a)$ can be interpreted as the projection of the signal onto the appropriately shifted and scaled wavelet ψ, i.e. $W(\tau, a)$ is a contribution of the wavelet to the signal $x(t)$.

While Fourier and wavelet transforms are usually applied to study each channel of EEG signal individually, the cross-correlation, coherence and Granger causality measure the interdependency between different channels. The cross-correlation function quantifies the linear correlation between two processes. Given two normalized signals $x(t)$ and $y(t)$ with zero means and unit variances, the cross-correlation between these signals is estimated as:

$$C_{xy}(\tau) = \frac{1}{N-\tau} \sum_{t=1}^{N-\tau} x(\tau + t)y(t), \tag{1–3}$$

where N denotes the total number of sampled points, and τ is a time delay parameter between two signals. The cross-correlation estimate ranges between -1 and 1. The positive values of cross-correlation indicate the direct correlation between the signals (i.e. x and y tend to be similar in both their absolute value and have the same sign), and the negative values correspond to inversely correlated signals (i.e. signals have similar absolute values, but different signs). Although zero cross-correlation value shows that the signals are not linearly correlated, it does not necessarily imply that two signals are not interrelated in a non-linear fashion.

The coherence function is a frequency domain analog of the cross-correlation measure. Coherence is obtained from cross-correlation by applying Fourier transform to (1–3). The estimate of the coherence spectrum of two signals is called periodogram. The periodogram is calculated by subdividing the signals into a number of epochs of the same length, and then applying the following formula:

$$P_{xy}^2(f) = \frac{|\overline{S_{xy}}(f)|^2}{|\overline{S_{xx}}(f)||\overline{S_{yy}}(f)|} \tag{1–4}$$

where $\overline{S_{xy}}(f)$, $\overline{S_{xx}}(f)$, $\overline{S_{yy}}(f)$ denote the average values of the cross-spectral density between x and y, and the individual auto-spectral densities of x and y, respectively. The periodogram values range between 0 and 1, where zero coherence indicates that the signals are linearly independent at a given frequency, and the maximum periodogram value of 1 shows that two signals are completely linearly dependent at a chosen frequency.

While the cross-correlation and coherence are features, which reflect the linear dependency between two channels in the data, the concept of Granger causality is capable of not only establishing the linear dependency, but also specifying the direction of such dependency. In other words, by applying Granger causality, it becomes possible to identify causal relationship among the channels of EEG. Granger causality is based on the multivariate autoregressive modeling of time series. It has also received an alternative reformulation in the frequency domain via spectral decomposition for stochastic processes.

Several studies have shown that EEG can be effectively modeled using chaos theory [35, 89]. In chaotic systems, trajectories originating from very close initial conditions diverge exponentially. The system dynamics are characterized by the rate of the divergence of the trajectories, which is measured by Lyapunov exponents and dynamical phase.

Short term largest Lyapunov exponent (denoted $STL_{\max}$), which is an estimate of the maximum Lyapunov exponent for non-stationary data, is a dynamical measure of the chaoticity in the brain. Next, the method for estimating $STL_{\max}$ is summarized.

First, using the method of delays [69], the embedding phase space is constructed from a data segment $x(t)$ with $t \in [0, T]$, so that the vector X_i of the phase space is given by

$$X_i = (x(t_i), x(t_i + \tau), \ldots, x(t_i + (p-1)\tau)), \tag{1–5}$$

where $t_i \in [1, T - (p-1)\tau]$, p is a chosen dimension of the embedding phase space, and τ denotes the time delay between the components of each phase space vector.

Next, the estimate L of the short term largest Lyapunov exponent $STL_{\max}$ is computed as follows:

$$L = \frac{1}{N_a \triangle t} \sum_{i=1}^{N_a} \log_2 \frac{X(t_i + \triangle t) - X(t_j + \triangle t)}{X(t_i) - X(t_j)}, \tag{1–6}$$

where N_a is the total number of local maximum Lyapunov exponents that are estimated during the time interval $[0, T]$; $\triangle t$ is the evolution time for the displacement vector $X(t_i) - X(t_j)$; $X(t_i)$ represents the point of the fiducial trajectory such that $t = t_i$, $X(t_0) = (x(t_0), x(t_0 + \tau), \ldots, x(t_0 + (p-1)\tau))$, and $X(t_j)$ is an appropriately selected vector that is adjacent to $X(t_i)$ in the embedding phase space. In [34], Iasemedis at el. suggested a method for estimating $STL_{\max}$ in the EEG data based on the Wolf's algorithm for time series [107].

The short term largest Lyapunov exponent $STL_{\max}$ is proved to be an especially useful EEG feature for studying the dynamics of the epileptic brain [35, 89]. In particular,

spatio-temporal transitions during interictal, preictal, ictal, and postictal states can be characterized by the changes in $STL_{\max}$ profiles [90].

1.4 Contribution Summary

1.4.1 Testing Applicability of Frequency Domain Estimates of Granger Causality for EEG time series

The rules of interaction between various parts of the brain are one of the key problems in theoretical and applied neuroscience. One of the most commonly used ways to record the neural information in multiple areas of the brain is via a multichannel electroencephalogram recording. Quantitatively, such recording is a multivariate time series. In the last decade, the investigation of causal relationships exhibited by the electroencephalographic time series became a very active area of research in neuroscience and related fields. Several approaches for studying causality in EEG are proposed, most of which are based on the definition of Granger causality via the spectral representation of time series (or processes), which was introduced by Geweke [25] for analysis of econometric series. Because frequency domain contains valuable information about the brain processes, Geweke's definition of causality seems to be particularly useful for analysis of EEG data.

Both definitions of Granger causality, the original one given by Granger and the frequency domain definition, are introduced via vector autoregressive modeling of multiple time series. Precisely, the linear vector autoregression is used to fit the data. Based on this model, two competing hypothesis about the data are considered and tested statistically to determine, which of these two assumptions is supported by the data. In other words, to test the causality, the hypothesis of data being modeled as linear autoregression (that does not include another series) is compared to the alternative of the data best described by including information from the other series.

Since Granger causality is defined based on linear vector autoregressive model (VAR), the applicability of Granger causality depends upon the underlying assumptions of the model. The fundamental assumptions of the VAR are stability, stationarity, and gaussian

distribution of the error term. If such conditions are violated, the model is not suitable for the data. In this case, one needs to find a model that is more appropriate for the given data. For example, many financial time series data exhibit periodic or seasonal behavior. To account for such cases, various modifications of vector autoregression (including those relaxing the stability condition) are constructed for analysis of econometric time series.

Although Bernasconi and König [2] tested the stationarity of different EEG data, and concluded that a time interval of 1 second is the interval on which the EEG time series can be considered stationary, the underlying assumptions of the VAR modeling such as stationarity and stability are rarely statistically tested in applications to EEG data.

In particular, the stability assumption means that the reverse characteristic polynomial of the model does not have roots inside a unit circle. In order to highlight the importance of the stability condition for VAR, it is necessary to point out that when stability is violated, the model may simply follow a random walk, or it may even exhibit explosive behavior.

To the best of our knowledge, the stability condition of the VAR estimated from the EEG time series was not investigated before our study [43]. In many studies that utilize the vector autoregression to examine Granger causality among series, the verification of the conditions assumed by the VAR model is often omitted.

The results of our numerical experiments indicate that the stability condition of vector autoregressive model is often violated in application to the EEG data. More specifically, we found that the stability assumption imposed on the linear VAR models may be violated even in the case when the sample size parameter T is much larger than the lag parameter p of the estimated model. In addition, we showed that despite the fact that it is common in practice to filter the EEG data within a certain frequency band, filtering the EEG time series within some restricted frequency band often results in significant reduction of the (T, p) domain, where the estimated VAR(p) models remain stable.

Based on our numerical studies, we concluded that suitable extensions of multivariate autoregression to unstable processes may fit the data better, and so, be more appropriate for the EEG time series analysis than linear vector autoregressive modeling [43].

Comprehensive statistical testing is necessary in order to make conclusions on what multivariate models are the most appropriate for extracting the directional dependencies between channels in a frequency domain from multichannel EEG data.

1.4.2 Generalization of Phase Synchronization via Cointegrated VAR

The temporal integration of various functional areas in different parts of the brain is believed to be essential for normal cognitive processes. This results in constant interaction among the brain regions. Many studies highlight the importance of neural synchrony in such large-scale integration [15, 101, 104, 105]. Actually, it was found that oscillation of various neuronal groups in given frequency bands leads to temporary phase-locking between such groups of neurons. This observation has stimulated the development of robust approaches that allow one to measure the phase-synchrony in a given frequency band from experimentally recorded biomedical signals such as EEG.

In particular, the importance of synchronization of neuronal discharges has been shown by a variety of animal studies using microelectrode recordings of brain activity [83, 96], and even at coarser levels of resolution by other studies in animals and humans [21].

The phase synchronization in the brain extracted from EEG data using Hilbert or wavelet transforms has recently been shown to be an especially promising tool in analysis of EEG data recorded from patients with various types of epilepsy [86].

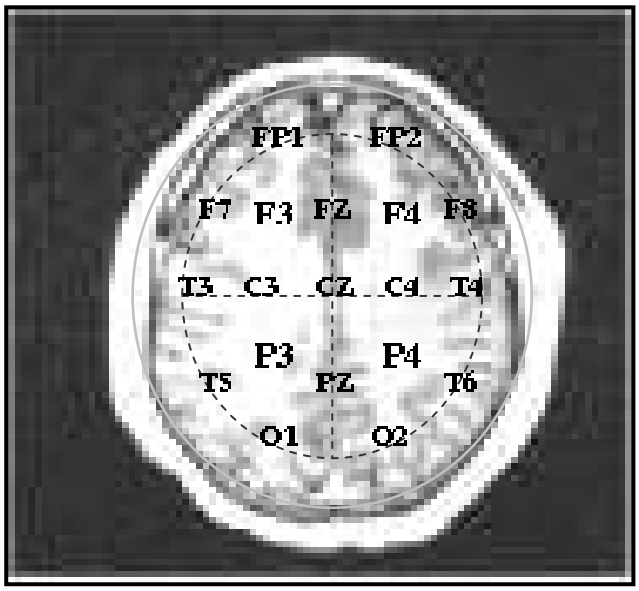
In our recent study [44], we introduce a novel concept of generalized phase synchronization, which is based on vector autoregressive modeling. This new notion of phase synchronization is constructed as an extension of the classical definition of phase synchronization between two systems. In fact, the phase synchronization is usually defined as the condition that some integer combination of the instantaneous phases of two signals is constant. Often this condition is relaxed by allowing for a bounded linear combination of two phases, in order

to account for noise in the measurements. This classical approach is clearly bivariate. But what if we are interested in studying a synchrony among several parts of a system? Is there such a notion?

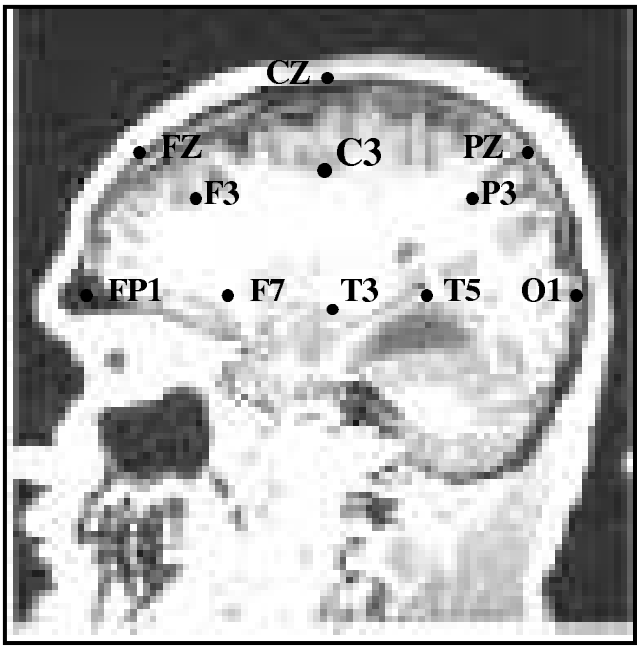
To construct a more general multivariate concept of phase synchronization, we extended the classical definition by considering a linear combination of phases for a finite number of signals that represents a stationary process. All the individual signals together form a common system described by some multivariate process. We note that a vector process, such that a linear combination of its individual components is a stationary process, can be modeled as a cointegrated vector autoregressive time series.

Furthermore, it is easy to see (as shown in Section 4.2.1) that the cointegrated rank of the regression determines how restricted the behavior of such system is. This means that the rank r of cointegrated autoregressive model, estimated from the multiple time series of the instantaneous phases, measures how large the vector subspace, which generates the changes in the phase values, is.

This new measure of cointegration was applied to absence epilepsy EEG data in [44]. The data sets collected from the patients with other types of epilepsy are currently being investigated.



A top view



B profile view

Figure 1-1. The International 10-20 system for placement of EEG electrodes

CHAPTER 2
AUTOREGRESSIVE MODELING OF MULTIPLE TIME SERIES

2.1 Multivariate Autoregressive Modeling in EEG Data Mining

Several methods for joint spatio-temporal analysis of multichannel EEG recordings based on the idea of Granger causality were presented in the last decade. The concept of Granger causality was first introduced by Clive Granger [26, 27] for measuring linear dependence and feedback in economic time series. Later, this idea was further extended by John Geweke [25], who proposed an equivalent measure based on the spectral representation of time series. Both Granger's and Geweke's approaches employ the vector autoregressive modeling to derive estimates of underlying causal relations in the data. However, the latter approach is found particularly useful for analysis of EEG time series, since it investigates the causal relation in the frequency domain instead of the time domain as in the former approach. In particular, the spectral measure of Granger causality proposed by John Geweke was employed on intracortical local field potentials recorded from 8 electrodes during go/no-go trials of cat's visual responses [2]. Another study [57] utilized a similar method of directed transfer function (which is equivalent to the spectral measure of Granger causality) to examine causal influences in the primate visual cortex during the task of visual pattern recognition. The direct transfer function approach to Granger causality was also applied to analyzing brain connectivity patterns on human EEG data recorded during stage 2 sleep [42].

Michael Eichler proposed a graphical approach for modeling Granger-causal relationships in multivariate time series [17] and later applied this method to studying connectivity in neural systems [18, 19]. Luiz Baccalá and Koichi Sameshima introduced a concept of the partial directed coherence for inference of Granger causality in the frequency domain based on the linear vector autoregressive modeling, and applied it to investigating the functional interactions among different brain structures [1, 92].

Vector autoregressive modeling is a common basis in various approaches to estimating Granger causality. Next, we review the details of the multivariate autoregressive modeling and investigate some limitations in its application to the EEG data.

2.2 Tests of Granger Causality

Many different tests of Granger Causality are developed. Some versions of the test are based on a vector autoregressive model, others are based on a multivariate moving average representation. For simplicity, we present the alternative definition based on the bivariate regression. For a detailed review of Granger causality tests, see the book on time series by Hamilton [30].

Let $X(t)$ and $Y(t)$, $t \in \mathbb{Z}$ denote two time series (or discrete time stochastic processes) with the corresponding realizations x_t and y_t, $t \in \mathbb{Z}$. Suppose that $\Omega_{X,t}$ and $\Omega_{Y,t}$ denote all the information about the realizations of processes X and Y, respectively, up to time t. Then, the relationship of Granger causality between such series can be formally defined as follows:

A time series $X(t)$ is said to *Granger-cause* $Y(t)$ if there exists $p = 1, 2, \ldots$ such that the mean squared error (MSE) of the p-step forecast of $Y(t)$ based on the information $\Omega_{X,t}$ and $\Omega_{Y,t}$ is smaller than the MSE of the p-step forecast of $Y(t)$ based on $\Omega_{Y,t}$ alone, i.e.

$$\exists p = 1, 2, \ldots \quad : \qquad \Sigma_Y(p|\Omega_{X,t}, \Omega_{Y,t}) < \Sigma_Y(p|\Omega_{Y,t}), \tag{2–1}$$

where $\Sigma_Y(p|\Omega)$ is the MSE of the p-step forecast of $Y(t)$ based on information Ω.

Using the above definition, we now present the test for Granger causality based on the bivariate autoregressive model. Suppose that for some integer lag parameter $p > 0$, the realizations of time series Y(t) are given by the model

$$y_t = \nu + \sum \alpha_i y_{t-i} + \sum \beta_i x_{t-i} + \varepsilon_t, \tag{2–2}$$

where ε_t is a *standard white noise* (or *innovation process*, i.e. ε_t has zero mean and zero autocorrelation).

We test the hypothesis

$$H_0: \quad \beta_i = 0 \quad \forall i = 1, 2, \ldots, p \tag{2–3}$$

against the alternative

$$H_1: \quad \exists j,\, 1 \leq j \leq p \quad \text{s.t.} \quad \beta_j \neq 0 \tag{2–4}$$

Note that if the null hypothesis is accepted, then a time series $Y(t)$ is believed to be NOT Granger-caused by $X(t)$. Meanwhile, rejecting the null hypothesis (i.e. accepting the alternative) means that $X(t)$ is believed to cause $Y(t)$ in Granger's sense.

Let T be the sample size parameter. The model parameters for the null hypothesis H_0 and the parameters for the alternative H_1 are estimated from the sample data using the ordinary least squares method (or other methods) to obtain the estimates of the forecast errors $\hat{\varepsilon}_{0t}$ and $\hat{\varepsilon}_{1t}$, respectively, $t = 1, 2, \ldots, T$. Then the sum of squared residuals RSS_0 under the assumption of null hypothesis H_0 is

$$RSS_0 = \sum_{t=1}^{T} \hat{\varepsilon}_{0t}^2, \tag{2–5}$$

and the sum of squared residuals RSS_1 under the alternative H_1 is

$$RSS_1 = \sum_{t=1}^{T} \hat{\varepsilon}_{1t}^2. \tag{2–6}$$

By conducting the F-test of the null hypothesis, one can find the test statistic

$$S_1 = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T - 2p - 1)} \sim F_{p,\, T-2p-1}. \tag{2–7}$$

If the test statistic S_1 exceeds the specified critical value, then the null hypothesis that $X(t)$ does not Granger-causes $Y(t)$ is rejected. Otherwise, H_0 is accepted.

An asymptotically equivalent test of Granger causality is given by the following statistic

$$S_2 = \frac{T(RSS_0 - RSS_1)}{RSS_1} \sim \chi^2(p). \tag{2–8}$$

It is noteworthy to point out that the tests of Granger causality are very sensitive to the choice of the lag length parameter p, and to the methods utilized for handling any non-stationarity in the time series.

The bivariate approach to testing Granger causality can be naturally extended to the multivariate case by partitioning the vector autoregressive process $Z(t)$ into two components $X(t)$ and $Y(t)$, so that $Z(t) = (X(t), Y(t))$, and then testing the suitable zero constraints on the coefficients of vector autoregression. For the derivation of the Wald statistic and the F-statistic for testing Granger causality in the multivariate case, see the book on multiple time series by Lütkepohl [59].

2.3 Vector Autoregressive Models (VAR)

2.3.1 Methods for VAR Parameter Estimation

The vector autoregressive (VAR) model of finite order serves as a foundation for establishing Granger causal relations in multidimensional time series.

Let p denote a positive integer, and let y_t denote the K-variate time series (i.e. realizations of K-dimensional process $Y(t)$). A vector autoregressive model of order p, denoted VAR(p), is formally defined as follows:

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{2–9}$$

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a $(K \times 1)$ random vector, $\nu = (\nu_1, \ldots, \nu_K)'$ is a fixed $(K \times 1)$ vector representing a non-zero mean $EY(t)$, the A_i, $i = 1, \ldots, p$ are fixed $(K \times K)$-dimensional coefficient matrices, and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})'$ is a K-dimensional *white noise* process (i.e. $E[\varepsilon_t] = 0$, $E[\varepsilon_s \varepsilon_t'] = 0$, for $s \neq t$, and $E[\varepsilon_s \varepsilon_t'] = \Sigma_\varepsilon$). It is assumed that the covariance matrix Σ_ε is nonsingular. In addition, three important conditions are usually imposed on the time series in the VAR model. The first condition is stability of the process $Y(t)$, the second is stationarity of $Y(t)$, while the third one supposes that the underlying white noise process ε_t is *Gaussian*.

Suppose that the the lag length parameter p is specified. Although, in the above definition of a VAR(p) model, the process mean ν, the coefficient matrices A_i, and the covariance matrix Σ_ε are assumed to be known, in practice, these parameters must be derived from the sample data. There are three main approaches to estimating the parameters of a VAR(p) time series, namely, the multivariate least squares method, the Yule-Walker estimator, and the maximum likelihood estimation [59]. Under the assumptions of stability and Gaussian distribution, these approaches lead to estimators with the same asymptotic properties. However, the asymptotic results should be used cautiously in inference from small samples. As a result, different approaches may sometimes lead to different results when estimating the model parameters using small samples.

Let us now briefly present the multivariate least squares estimation, which is a higher dimensional extension of the well-known method of ordinary least squares. For more detailed discussion, refer to [59].

Suppose that the available data include $(T + p)$ successive realizations of estimated multiple time series represented by K-dimensional vectors

$$y_{-p+1}, \ldots, y_0,\ y_1, \ldots, y_T$$

where p is the fixed lag length, and T is the sample size parameter. For convenience, we partition the data into the pre-sample $y_{-p+1}, \ldots, y_0$ and the sample $y_1, \ldots, y_T$ values. In addition, the following notation is introduced:

$$\begin{aligned} Y &:= (y_1, \ldots, y_T) && (K \times T), \\ Z_t &:= \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} && ((Kp+1) \times 1), \\ Z &:= (Z_0, \ldots, Z_{T-1}) && ((Kp+1) \times T), \end{aligned} \tag{2–10}$$

$$B := (\nu, A_1, \ldots, A_p) \qquad (K \times (Kp+1)),$$
$$\varepsilon := (\varepsilon_1, \ldots, \varepsilon_T) \qquad (K \times T),$$

where $t = 1, \ldots, T$. Then using this notation in (2–9), the vector autoregressive model of order p can be represented in the compact form:

$$Y = BZ + \varepsilon, \tag{2–11}$$

and the coefficients B of the model are given by the least squares estimator:

$$\hat{B} = YZ'(ZZ')^{-1}. \tag{2–12}$$

The covariance matrix can be estimated in various ways. Since $\Sigma_\varepsilon = E\ [\varepsilon_t \varepsilon_t']$, the estimator

$$\tilde{\Sigma}_\varepsilon = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t' = \frac{1}{T}(Y - \hat{B}Z)(Y - \hat{B}Z)' \tag{2–13}$$

is consistent. However, this estimator of the covariance matrix Σ_ε is not unbiased. Therefore, it is often replaced by the following unbiased estimator

$$\hat{\Sigma}_\varepsilon = \frac{1}{T - Kp - 1}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t' = \frac{1}{T - Kp - 1}(Y - \hat{B}Z)(Y - \hat{B}Z)' \tag{2–14}$$

Obviously, both estimators are consistent estimators of the covariance matrix, and they are asymptotically equivalent.

When estimating the coefficients of the vector autoregressive model from data, we assumed the order p of the VAR(p) to be known. In practice, however, it is unknown, and therefore, needs to be derived from the data. Since zero coefficient matrices are allowed, one could simply set p to some upper bound on the VAR order. On the other hand, selecting an unnecessary large p would affect the forecast precision of the estimated model. Therefore, it is advantageous to apply some suitable criteria for optimal selection of the lag length parameter p.

2.3.2 VAR Order Selection Criteria

Various criteria for choosing the optimal model order are developed. Some of the most commonly used are the final prediction error (FPE) criterion, Akaike's information criterion (AIC), Hannan-Quinn criterion (HQ), and Schwarz or Bayesian information criterion (SC) [59].

Let $\tilde{\Sigma_u}(m)$ denote the maximum likelihood estimator of Σ_u computed by fitting the VAR model of order m. The FPE criterion proposed by Akaike in 1969 is based on the idea that minimizing the mean square error improves the forecast of the model. For a VAR(p) time series, the FPE criterion is defined as

$$FPE(m) = \left(\frac{T + Km + 1}{T - Km - 1}\right)^K \det\left(\tilde{\Sigma}_u(m)\right) \tag{2–15}$$

Using the FPE criterion, the estimate $\hat{p}_{FPE}$ of the model order p is selected so that

$$FPE\left(\hat{p}_{FPE}\right) = \min_{m=1,\ldots,M} FPE(m), \tag{2–16}$$

where M denotes some upper boundary on the model order. In other words, first, for each $m = 1, \ldots, M$, the vector autoregressive model of order m is estimated from the data, and the respective values of the $FPE(m)$ are calculated using (2–15); then the order producing the smallest value of $FPE(m)$ is chosen among the possible orders $m = 1, \ldots, M$.

AIC is another popular order selection criteria that was also introduced by Akaike. Given a VAR(m) model, the Akaike information criteria is defined as follows:

$$AIC(m) = \frac{2mK^2}{T} + \ln\left(\tilde{\Sigma}_u(m)\right) \tag{2–17}$$

Similarly to the FPE criterion, the VAR(m) models are estimated for different $m = 1, \ldots, M$ to obtain the corresponding $AIC(m)$ values for each order. Then the estimate $\hat{p}_{AIC}$ of the model order p with the smallest $AIC(m)$ is selected.

The Hannan-Quinn order selection criterion, HQ, is given by:

$$HQ(m) = \frac{2mK^2 \ln \ln T}{T} + \ln\left(\tilde{\Sigma}_u(m)\right) \tag{2–18}$$

As before, among the model parameters $m = 1, \ldots, M$, the parameter m having the smallest value of $HQ(m)$ is chosen as the estimator $\hat{p}_{HQ}$ of the true model order p.

Last, but not least, we present Schwarz criterion, which was derived using Bayesian arguments. The SC is formulated as:

$$SC(m) = \frac{mK^2 \ln T}{T} + \ln\left(\tilde{\Sigma}_u(m)\right), \tag{2–19}$$

and the order minimizing $SC(m)$ is chosen among $m = 1, \ldots, M$ as the estimator $\hat{p}_{SC}$ of the model order p.

Some interesting statistical properties of the above criteria are proved in [59]. In particular, it is shown that AIC and FPE criteria for VAR order selection are asymptotically equivalent, although these estimators of the model order are not consistent. On the other hand, the other two criteria provide consistent estimators of the order parameter p. More precisely, in the univariate case ($K = 1$), the Hannan-Quinn criterion is consistent (i.e. $\lim_{T\to+\infty} \Pr\{\hat{p} = p\} = 1$). In addition, the HQ criterion is strongly consistent for $K \geq 2$ (i.e. $\Pr\{\lim_{T\to+\infty} \hat{p} = p\} = 1$). The SC is shown to be strongly consistent for any dimension K.

It is important to keep in mind that even though FPE and AIC do not provide consistent estimators, they are not necessarily inferior to HQ and SC. Actually, in small samples, and even in larger samples, FPE and AIC may produce better forecast, although they may not estimate the model order correctly.

2.3.3 Stability Condition and Other Assumptions of VAR

As mentioned above, the conditions of stability, stationarity and Gaussian distribution are usually imposed on time series when dealing with the VAR models. Below we define these conditions, and discuss their role. A K-dimensional VAR(p) time series (2–9) are

called *stable*, if

$$\det(I_K - A_1 z - \dots A_p z^p) \neq 0 \qquad \text{for} \quad \text{complex } z \; : \quad |z| \leq 1. \tag{2–20}$$

In other words, the VAR(p) process (2–9) satisfies the *stability condition* when its *reverse characteristic polynomial* (given by $\det(I_K - A_1 z - \dots A_p z^p)$) has no roots on and inside the complex unit circle.

The stability condition guarantees that there exists a moving average (MA) representation for the VAR(p) process. Also stability ensures that the process is a well-defined stochastic process with the distributions of its univariate components and joint distribution of the process y_t uniquely determined by the innovation process ε_t. For a stable VAR(p)process, both the process mean and the autocovariance are time-invariant (which, according to the definition below, implies stationarity).

When the stability condition is violated, the process variance is increasing with time and unbounded. Specifically, if the reverse characteristic polynomial of the time series has a single unit root, and all the other roots are outside the complex unit circle, then the time series behavior is similar to a *random walk*. In this special case, the variance increases linearly with time, the correlation between y_t and y_{t+h} approaches 1, and the process mean $E[Y(t)]$ exhibits a linear trend for $\nu \neq 0$. In addition, if one of the roots of the reverse characteristic polynomial lies strictly inside the complex unit circle, then such process is explosive, i.e. the process variance grows exponentially. Various approaches are developed in the time series literature to address the time series with the unit roots. For example, the unit roots can be removed by taking differences. However, the explosive time series are not as well-studied, because it is believed that an exponential increase in the variance of the economic time series is not well founded. As one can see the stability assumption plays an important role in VAR(p).

A wide-sense stationarity for stochastic processes is imposed on the VAR time series as follows. A stochastic process $Y(t)$ is considered *stationary* if

1. $E\ [Y(t)] = \nu$ for all t;
2. $E\ \left[(Y(t+h) - \nu)\,(Y(t) - \nu)'\right] = \Gamma_y(h) = \Gamma_y(-h)$ for all t and $h = 0, 1, \ldots$.

In other words, the stationarity condition supposes that the first and the second moments are time invariant. Also note that the process mean ν and the autocovariance matrix $\Gamma_y(h)$ are finite. It is shown (see Proposition 2.1 in [59]) that

A stable VAR(p) time series y_t, $t = 0, \pm 1, \pm 2, \ldots$ is stationary.

Since stability of a time series implies that the series is stationary, the stability condition (2–20) is sometimes cited in the literature as the stationarity condition. However, it is important to remember that these two conditions are not equivalent. In fact, although a stable vector autoregressive series is always stationary, the converse is not true, i.e. an unstable time series is not necessary non-stationary.

The Gaussian distribution assumption is introduced into the VAR(p) model through ε_t. Specifically, given representation (2–9) of the VAR (p), the innovation process ε_t is assumed to be *Gaussian white noise*. This condition implies that y_t is a Gaussian process, i.e. any subcollection $y_t, \ldots, y_{t+h}$ follows a multivariate normal distribution for all possible values of t and h.

2.4 Inegrated and Cointegrated VAR

In previous sections of this chapter, we considered VAR processes, for which the stationarity and stability assumptions are satisfied. However, in practice, many time series data are fit better by unstable non-stationary processes. In this chapter, we introduce integrated and cointegrated processes, which are found especially useful in econometric studies, and for which the stability and stationarity conditions are violated.

Recall that the VAR(p) process (2–9) satisfies the *stability condition* when its *reverse characteristic polynomial* $\det(I_K - A_1 z - \ldots A_p z^p)$ has no roots on and inside a complex unit circle. If an unstable process has a single unit root and all the other roots outside of the complex unit circle, then such process exhibits a behavior similar to that of a random walk. In other words, the variance of such process increases linearly to infinity,

and the correlation between the variables $Y(t)$ and $Y(t \pm h)$ tends to 1 as $t \to \infty$. On the other hand, when the root of reverse characteristic polynomial lies inside the unit circle, the process becomes explosive, i.e. its variance increases exponentially. In real-life applications, the former case is of the most practical interest.

This renders the following definition of an integrated process.

A one-dimensional process with d roots on the unit circle is said to be *integrated of order* d (denoted as $I(d)$).

It can be shown [59] that the integrated $I(d)$ process $Y(t)$ of order d with all roots of its reverse characteristic polynomial being equal to 1 can be made stable by differencing the original process d times. For example, the integrated $I(1)$ process $Y(t)$ becomes stable after taking the first differences $(1 - L)Y(t) = Y(t) - Y(t - 1)$, where L represents the lag operator. More generally, for the $I(d)$ process $Y(t)$, its transformation $(1 - L)^d Y(t)$ is stable.

An example of an integrated $I(d)$ process in the univariate case is an *autoregressive integrated moving average* process ARIMA(p, d, q), which is sometimes called fractionally differenced autoregressive moving average process for $d \in (-0.5, 0.5)$. The one-dimensional process $Y(t)$ is said to be ARIMA(p, d, q), if $Z(t) := (1 - L)^d Y(t)$ is a stationary autoregressive moving average ARMA(p,q) process, i.e.

$$\sum_{i=0}^{p} a_i Z(t - i) = \sum_{j=0}^{q} b_j \varepsilon_{t-j}, \tag{2–21}$$

where ε_{t-j}'s are independent normally distributed random variables with mean 0 and variance σ^2, and L is the differencing operator introduced above.

It is noteworthy to point out that taking differences may distort the relationship among the variables (i.e. one-dimensional components) in some VAR(p) models. In particular, this is the case for systems with cointegrated variables. It turns out that fitting VAR(p) model after differencing the original *cointegrated* process produces inadequate results. Next, we discuss such processes.

Cointegrated processes were first introduced by Clive Granger in 1981, and gained a great deal of popularity in both theoretical and applied econometrics. Indeed, many economic variables are expected to be in equilibrium relationship, for example, household income and expenditures, or prices of a given commodity in different markets.

Suppose that sampled values y_{it} of K different variables of interest $Y_i(t)$ are combined into the K-dimensional vectors $y_t = (y_{1t}, \ldots, y_{Kt})'$. In addition, suppose that the variables are in a *long-run* equilibrium relation

$$c\, Y(t) := c_1 \cdot Y_1(t) + \ldots + c_K \cdot Y_K(t) = 0, \tag{2–22}$$

where $c = (c_1, \ldots, c_K)'$ is a K-dimensional real vector. During any given time interval, the relation (2–22) may not necessarily be satisfied precisely by the sample y_t, instead we may have:

$$c\, y_t := c_1 \cdot y_{1t} + \ldots + c_K \cdot y_{Kt} = \varepsilon_t, \tag{2–23}$$

where ε_t is a stochastic process that denotes the deviation from the equilibrium relation at time t. If our assumption about the long-run equilibrium among individual variables $Y_i(t)$, $i = 1, \ldots, K$ is valid then it is reasonable to expect that the variables $Y_i(t)$ move together, i.e. the stochastic process ε_t is stable. On the other hand, this does not contradict the possibility that the variables deviate substantially as a group. Therefore, it is possible that although each individual component $Y_i(t)$ is integrated, there is a linear combination of $Y_i(t)$, $i = 1, \ldots, K$, which is stationary. Integrated processes with such property are called *cointegrated*.

Without loss of generality, we assume that all individual one-dimensional components $Y_i(t)$ $(i = 1, \ldots, K)$ are either $I(1)$ or $I(0)$ processes. Then the combined K-dimensional VAR(p) process

$$Y(t) = \nu + A_1 Y(t-1) + \ldots + A_p Y(t-p) + \varepsilon_t \tag{2–24}$$

is said to be *cointegrated of rank* r, when the correspondent matrix

$$\Pi = I_K - A_1 - \ldots - A_p \tag{2–25}$$

has rank r.

Since some one-dimensional components of the cointegrated VAR(p) process are integrated processes, one may be interested in testing the presence of a unit root in the univariate series. In the following section, we present a commonly used unit root test, which was derived by Dickey and Fuller [16].

2.4.1 Augmented Dickey-Fuller Test for Testing the Null Hypothesis of the Presence of a Unit Root

The augmented Dickey-Fuller (or ADF) test is a widely used statistical test for detecting the existence of a unit-root of the reverse characteristic polynomial in a univariate time series. By fitting an autoregressive AR(k) model, this test investigates the null hypothesis of an autoregressive integrated moving average ARIMA(p, 1, 0) process against the alternative of a stationary ARIMA($p + 1$, 0, 0) process. The limiting distribution of the ADF test for $p \leq k - 1$ was derived by Dickey and Fuller [16], and it can be shown that this distribution is the same for $k > 1$ and for $k = 1$. Fuller tabulated the approximate critical values for the ADF test with $k \geq 1$ and $p \leq k - 1$ for *specific* sample sizes.

Finite-sample critical values for the ADF test for *any* sample size were obtained by means of response surface analysis by MacKinnon [60], who also showed that an approximate asymptotic distribution function for the test can be derived via response surface estimation of quantiles [61].

Although the asymptotic distribution of the ADF test statistic does not depend on the lag order, it is noted by Cheung et al. [13] that empirical applications must deal with finite samples, in which case the distribution of the ADF test statistic can be sensitive to

the lag order. Taking this into account, they closely examined the roles of the sample size and the lag order in finding the finite-sample critical values of the ADF test.

As we noted above, the limiting distribution of the ADF test statistic is the same for $k > 1$ and $k = 1$. Hence, for simplicity, we consider the case of $k = 1$. In fact, let Y denote the autoregressive AR(1) model

$$Y(t) = c\, Y(t-1) + \varepsilon_t, \qquad t = 1, 2, \ldots \tag{2–26}$$

where $Y(0) = 0$, c is a real number, and $\varepsilon_t \sim N(0, \sigma^2)$ (i.e. ε_t is normally distributed with zero mean and variance σ^2 for all $t = 1, 2, \ldots$).

Note that when $|c| < 1$, the process Y(t) converges to a stationary process as $t \to \infty$; whereas, in the case of $|c| = 1$, the process $Y(t)$ is not stationary with variance $t\sigma^2$. Furthermore, when $|c| > 1$, not only the process is not stationary, but the variance of $Y(t)$ grows exponentially with time t.

From the AR(1) model (2–26), one can see that in the case when $c = 1$, in order to make the process stationary, the series can be appropriately transformed by differencing. Furthermore, notice that the condition $c = 1$ in (2–26) is clearly equivalent to the requirement that the reverse characteristic polynomial $\det(1 - cz) = 1 - z$ of AR(1) has a unit root. In other words, to determine whether an autoregressive time series AR(1) has a unit root, we must test the null hypothesis $H_0 : c = 1$.

Let $y_1, y_2, \ldots, y_T$ denote a sample of T consecutive observations of the AR(1) process $Y(t)$, then the maximum likelihood estimator of c is the least squares estimator

$$\widehat{c} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \tag{2–27}$$

Note that $\widehat{c}$ is a consistent estimator of the regression coefficient c.

Since each y_t, $t = 1, \ldots, T$ is a realization of an AR(1) process, it follows from (2–26) that $y_t = c\, y_{t-1} + \varepsilon_t$ holds, and so by plugging this last condition into Equation (2–27), the

estimator $\widehat{c}$ of the regression coefficient can also be written as:

$$\begin{aligned} \widehat{c} &= \frac{\sum_{t=1}^{T}(c\, y_{t-1}+\varepsilon_t)\, y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \\ &= \frac{c\, \sum_{t=1}^{T} y_{t-1}^2 + \sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2} \\ &= c + \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}. \end{aligned} \tag{2–28}$$

Subtracting c from both sides of Equation 2–28 and multiplying each side by T lead to the ADF statistic

$$T(\widehat{c}-c) = \frac{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2}. \tag{2–29}$$

Dickey and Fuller [16] derived the following representation of the limiting distribution for statistic $T(\widehat{c}-c)$:

$$T(\widehat{c}-c) \Rightarrow \frac{1}{2}\Gamma^{-1}(W^2-1), \quad \text{as} \quad T\to\infty \tag{2–30}$$

where

$$\Gamma = \sum_{i=1}^{\infty} d_i^2 X_i^2, \tag{2–31}$$

$$W = \sum_{i=1}^{\infty} \sqrt{2}\, d_i X_i, \tag{2–32}$$

$$d_i = \frac{2(-1)^{i+1}}{\pi(2i-1)}, \tag{2–33}$$

and random variables X_i, $i = 1, 2, \ldots$, are independent and identically distributed according to the normal distribution with zero mean and variance σ^2. Note that a symbol $\Rightarrow$ denotes convergence in distribution.

In [16], Dickey and Fuller considered the following "Studentized" statistic based on the likelihood ratio test of the hypothesis $H_0 : c = 1$.

$$\widehat{\tau} = \frac{\widehat{c}-1}{S}\left(\sum_{t=2}^{T} y_{t-1}^2\right)^{\frac{1}{2}}, \tag{2–34}$$

where

$$S^2 = \frac{1}{T-2}\left(\sum_{t=2}^{T}\left(y_t - \widehat{c}y_{t-1}\right)^2\right), \tag{2–35}$$

and $\widehat{c}$ is computed from (2–27)

Tables of the critical values for the asymptotic distributions of the ADF test statistic $T(\widehat{c}-1)$ and the statistic $\widehat{\tau}$ can be found in Fuller [22]. We summarize some of the information in Table 2-1, which lists the p-values for asymptotic distributions of $T(\widehat{c}-1)$ and $\widehat{\tau}$ corresponding to percentiles of 90, 95, and 99 percent.

2.4.2 Phillips-Ouliaris Cointegration Test

The unit root tests based on analysis of residuals were introduced by Phillips [75]. In particular, in his study Phillips first considered two statistics Z_α and Z_t for testing the null of no cointegration in time series.

Because many unit root tests, constructed before 1987, were founded on the assumption that the errors in the regression are independent with common variance (which is rarely met in practice), Phillips wanted to relax the rather strict condition that the time series are driven by independent identically distributed innovations. In other words, he wanted to develop the testing procedures based on the least squares regression estimation and the associated regression t statistic, which would allow for rather general weakly dependent and heterogeneously distributed sequence of error terms.

The properties of asymptotic distributions of residual based tests for the presence of cointegration in multiple time series were thoroughly investigated by Phillips and Ouliaris [76]. The characteristic feature of these tests is that they utilize the residuals computed from regressions among the univariate components of multivariate series. The residual based procedures developed by Phillips and Ouliaris are designed to test the null of *no cointegration* by means of testing the null hypothesis of the unit root presence in the residuals against the alternative of a root that lies inside the complex unit circle. The hypothesis H_0 of the absence of cointegration is rejected, if the null of a unit root in the

residuals is rejected. In the nutshell, the procedures are simply residual based unit root tests.

As noted in [76], the residual based unit root tests are asymptotically similar, and can be represented via the standard Brownian motion. Moreover, the ADF and Z_t tests are proved to be asymptotically equivalent. However, these two tests are not as powerful as the test based on statistic Z_α, because it was shown by Phillips and Ouliaris [76] that the rate of divergence under cointegration assumption is slower for the ADF and Z_t than other tests, such as the Z_α-statistic test. The later test (i.e. the cointegration test based on Z_α) is also widely known as the *Phillips-Ouliaris cointegration test*.

It is noteworthy that the null hypothesis for the Phillips-Ouliaris test is that of *no cointegration* (instead of cointegration). This formulation is chosen because of some major pitfalls found in procedures that are designed to test the null of cointegration in multiple time series. These defects (discussed in more detail in [76]) are significant enough to be a strong argument against the indiscriminate use of the test formulations based on the null of cointegration, and to support the continuing use of residual based unit root tests.

Consider the K-dimensional vector autoregressive process $Y(t)$. Let us partition $Y(t) = (U_t, V_t')'$ into the univariate component $U_t = Y_1(t)$ and the $(K-1)$-dimensional $V_t = (Y_2(t), \ldots, Y_K(t))'$.

The residuals are determined by fitting linear cointegrating regression

$$U(t) = c\, V(t) + \xi_t, \qquad t = 1, 2, \ldots \tag{2–36}$$

Residual based tests are formulated to test the null hypothesis that the multiple time series $Y(t)$ are not cointegrated using the *scalar* unit root tests, such as the ADF test, which are applied to the residuals ξ_t, $t = 1, 2, \ldots$ in (2–36).

In [76], the ADF test as well as two additional tests Z_α and Z_t, developed earlier by Phillips [75], were applied to check for the presence of a unit root in the residuals ξ_t. In

order to perform the unit root test, we fit an AR(1) model to ξ_t, $t = 1, 2, \ldots$ according to

$$\xi_t = \widehat{\alpha}\, \xi_{t-1} + \varrho_t, \qquad t = 1, 2, \ldots \tag{2–37}$$

Then the statistic Z_α in Phillips-Ouliaris test is defined as follows:

$$Z_\alpha = T(\widehat{\alpha} - 1) \; - \; \frac{1}{2} \cdot \frac{s_{Tl}^2 - s_\varrho^2}{\frac{1}{T^2} \sum_{t=2}^{T} \xi_{t-1}^2}, \tag{2–38}$$

whereas the Z_t statistic is given by the following formula:

$$Z_t = \left(\sum_{t=2}^{T} \xi_{t-1}^2 \right)^{\frac{1}{2}} \cdot \frac{(\widehat{\alpha} - 1)}{s_{Tl}} \; - \; \frac{1}{2} \cdot \frac{s_{Tl}^2 - s_\varrho^2}{s_{Tl} \left(\frac{1}{T^2} \sum_{t=2}^{T} \xi_{t-1}^2 \right)^{\frac{1}{2}}}, \tag{2–39}$$

where

$$s_\varrho^2 \;=\; \frac{1}{T} \sum_{t=1}^{T} \varrho_t^2, \tag{2–40}$$

$$s_{Tl}^2 \;=\; \frac{1}{T} \sum_{t=1}^{T} \varrho_t^2 + \frac{2}{T} \sum_{s=1}^{T} w_{sl} \sum_{t=s+1}^{T} \varrho_t \varrho_{t-s}, \tag{2–41}$$

$$w_{sl} \;=\; 1 - \frac{s}{l+1}. \tag{2–42}$$

Note that s_ϱ^2 and s_{Tl} are consistent estimators for the variance σ_ϱ^2 of ϱ_t and the partial sum variance $\sigma^2 = \lim_{T \to \infty} E\left(\frac{1}{T} S_T^2\right)$, where $S_T = \sum_{t=1}^{T} \xi_t$ is the partial sum of the error terms in (2–37).

The critical values for Z_α and Z_t statistics can be found in [76] (Tables I and II). Phillips and Ouliaris tabulated the values for cointegrating regressions with at most 5 explanatory variables. Some estimates of the critical values for the Phillips-Ouliaris test (Z_α) are listed in Table 2-2.

2.4.3 Estimation of Cointegrated VAR(p) Processes

Several methods can be employed to estimate the parameters of a Cointegrated VAR(p) model, including modifications of the approaches used for estimation of the standard VAR(p) processes.

In this section we present the maximum likelihood approach to estimating a *Gaussian* cointegrated VAR(p) process. Suppose y_t is a realization of a K-dimensional VAR(p) process with cointegration rank r, such that $0 < r < K$. Without loss of generality, we assume that $Y(t)$ has zero mean, i.e. the intercept $\nu = 0$ in (2–24).

Given a realization y_t, $t = 1, 2, \ldots$, of $Y(t)$, one seeks to determine the coefficients of the following model:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t+p} + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{2–43}$$

subject to the constraint

$$\operatorname{rank}(\Pi) = \operatorname{rank}(I_K - A_1 - \ldots - A_p) = r. \tag{2–44}$$

Note that ε_t is assumed to be a Gaussian white noise with a nonsingular covariance matrix Σ_ε. Furthermore, the initial conditions $y_{-p+1}, \ldots, y_0$ are supposed to be fixed.

In order to impose the cointegration constraint, the model (2–43) is reparameterized in the following fashion [59]:

$$\Delta y_t = D_1 \Delta y_{t-1} + \ldots + D_{p-1} \Delta y_{t-p+1} + \Pi y_{t-p} + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{2–45}$$

where $\Delta y_t = y_t - y_{t-1}$, and matrix Π can be represented as a product $\Pi = HC$ of matrices of rank r, i.e. H is $(K \times r)$ and C is $(r \times K)$.

Consider

$$\begin{aligned} \Delta Y &:= [\Delta y_1, \ldots, \Delta y_T], \\ \Delta X_t &:= \begin{bmatrix} \Delta y_t \\ \vdots \\ \Delta y_{t-p+2} \end{bmatrix}, \\ \Delta X &:= [\Delta X_0, \ldots, \Delta X_{T-1}], \\ D &:= [D_1, \ldots, D_{p-1}], \\ Y_{-p} &:= [y_{1-p}, \ldots, y_{T-p}]. \end{aligned} \tag{2–46}$$

Then the log-likelihood function for a sample of size T can be written as:

$$\begin{aligned}\ln l &= -\frac{KT}{2}\ln[2\pi] - \frac{T}{2}\ln\left[\det \Sigma_\varepsilon\right] \\ &\quad -\frac{1}{2}\text{trace}\left(\left(\Delta Y - D\Delta X + HCY_{-p}\right)' \Sigma_\varepsilon^{-1}\left(\Delta Y - D\Delta X + HCY_{-p}\right)\right). \quad (2\text{–}47)\end{aligned}$$

The proof of the following theorem on the maximum likelihood estimators of a cointegrated VAR process can be found in [59] (Proposition 11.1).

Theorem 1. *(reproduced from [59])*

Define

$$\begin{aligned} M &:= I - \Delta X'(\Delta X \Delta X')^{-1}\Delta X, \\ R_0 &:= \Delta Y\, M, \\ R_1 &:= Y_{-p} M, \\ S_{ij} &:= \frac{1}{T} R_i R_j', \quad i = 0, 1. \end{aligned}$$

Let G be the lower triangular matrix with positive diagonal such that $GS_{11}G' = I_K$.

Denote $\lambda_1 \geq \ldots \geq \lambda_K$ to be the eigenvalues of $GS_{10}S_{00}^{-1}S_{01}G'$,

and

$v_1, \ldots, v_2$ be the corresponding orthonormal eigenvectors.

Then the log-likelihood function in (2–47) is maximized for

$$\begin{aligned} C &:= [v_1, \ldots, v_r]' G, \\ H &:= -\Delta Y M Y_{-p}' C' \left(C Y_{-p} M Y_{-p}' C'\right)^{-1} \\ &= -S_{01}C'\left(CS_{11}C'\right)^{-1}, \\ D &:= \left(\Delta Y + HCY_{-p}\right)\Delta X\left(\Delta X \Delta X'\right)^{-1}, \\ \Sigma &:= \frac{1}{T}\left(\Delta Y - D\Delta X + HCY_{-p}\right)\left(\Delta Y - D\Delta X + HCY_{-p}\right)'. \end{aligned}$$

The maximum is

$$\max[\ln l] = -\frac{KT}{2}\ln[2\pi] - \frac{T}{2}\left(\ln\left[\det S_{00}\right] + \sum_{i=1}^{r}\ln(1-\lambda_i)\right) - \frac{KT}{2}. \tag{2–48}$$

2.4.4 Testing for the Rank of Cointegration

Based on Theorem 1, one can easily derive the likelihood ratio statistic for testing a candidate value r_0 of the cointegration rank r of a VAR(p) process against a larger cointegration rank r_1.

Given a VAR(p) process $y(t)$ defined by (2–24), suppose we wish to test a hypothesis H_0 against an alternative H_1, where

$$H_0 : r = r_0 \qquad \text{against} \qquad H_1 : r_0 < r \leq r_1. \tag{2–49}$$

Under assumption that the noise ε_t is a Gaussian process, the maximum of the likelihood function for a cointegrated VAR(p) model with cointegration rank r is computed in Theorem 1. From that result, the value of the LR statistic for testing (2–49) can be determined in the following manner:

$$\begin{aligned}\lambda_{LR}(r_0, r_1) &= 2\left[\ln L_{\max}(r_1) - \ln L_{\max}(r_0)\right] \\ &= T\left[-\sum_{i=1}^{r_1}\ln(1-\lambda_i) + \sum_{i=1}^{r_0}\ln(1-\lambda_i)\right] \\ &= -T\sum_{i=r_0+1}^{r_1}\ln(1-\lambda_i),\end{aligned} \tag{2–50}$$

where $L_{\max}(r_i)$, $i = 0, 1$, denotes the maximum of the Gaussian likelihood function for cointegration rank r_i. The advantage of this test is in the simplicity with which the LR statistic can be computed. On the other hand, the asymptotic distribution of the LR statistic (2–50) is nonstandard. Specifically, the LR statistic is not asymptotically distributed according to χ^2-distribution. Nevertheless, the asymptotic distribution of the cointegration rank test statistic λ_{LR} depends only on two factors:

- the difference $K - r$ between the process dimensionality and the cointegration rank; and
- the alternative hypothesis.

As a result, the selected percentage points of the asymptotic distribution of the test statistic λ_{LR} were tabulated by Johansen and Juselius in [38]. The percentage points of asymptotic distribution of λ_{LR} are given in Tables 2-3 and 2-4.

Table 2-1. Critical values of the asymptotic distributions of the $T(\widehat{c}-1)$ and $\widehat{\tau}$ for performing unit root check by the ADF test (reproduced from [22])

Statistic	90%	95%	99%
$\widehat{c}$	0.93	1.28	2.03
$\widehat{\tau}$	0.89	1.28	2.01

Table 2-2. Critical values of the asymptotic distributions of the Z_α statistic for testing the null of no cointegration (Phillips-Ouliaris demeaned, reproduced from [76]). Parameter n ($n = K - 1$) represents the number of explanatory variables

n	90%	95%	99%
1	-17.0390	-20.4935	-28.3218
2	-22.1948	-26.0943	-34.1686
3	-27.5846	-32.0615	-41.1348
4	-32.7382	-37.1508	-47.5118
5	-37.0074	-41.9388	-52.1723

Table 2-3. Percentage points of the asymptotic distributions of the $\lambda_{LR}(r, K)$ for testing the cointegration rank (reproduced from [38])

$K - r$	90%	95%	99%
1	6.69	8.08	11.58
2	15.58	17.84	21.96
3	28.44	31.26	37.29

Table 2-4. Percentage points of the asymptotic distributions of the $\lambda_{LR}(r, r+1)$ for testing the cointegration rank (reproduced from [38])

$K-r$	90%	95%	99%
1	6.69	8.08	11.58
2	12.78	14.60	18.78
3	18.96	21.28	26.15

CHAPTER 3
PHASE SYNCHRONY IN BRAIN DYNAMICS

In this chapter we introduce a concept of phase synchronization, and consider two methods for estimating the phase of a signal, specifically using the Hilbert transform and via the wavelet transform.

3.1 The Role of Phase Synchronization in Neural Dynamics

The word "synchrony" originates from a combination of two Greek words $\sigma\upsilon\nu$ (syn, meaning common) and $\chi\rho o\nu o\varsigma$ (chronos, meaning time), and it can be translated as "happening at the same time". A concept of synchronization can be defined as a process of active adjustment between the rhythms of different oscillating systems due to some kind of interaction or coupling between them [78]. Synchronization phenomena were discovered in the seventeenth century by C. Huygens who first observed synchronization between two pendulum clocks hanging from a common support [33]. Since then, the study of synchronization between dynamical systems became an active field of research in many scientific and technical disciplines, including solid state physics [74], plasma physics [84], communication [11, 48], electronics [72, 77], laser dynamics [20, 87, 98], and control [80, 88].

Synchronization phenomena can also be found in physiological systems, such as heart and brain. Synchronization processes in physiological systems were discovered by B. van der Pol in the beginning of the twentieth century. In particular, van der Pol was the first to apply oscillation theory to the human heart [103].

One of the important research areas in neuroscience explores the role of synchronization in neural dynamics. Much effort is given to investigation of synchronization phenomena on all different levels of organization of brain tissue, starting with pairs of individual neurons to larger scales, such as within a given area of the brain or between distinct parts of the brain.

Many studies emphasize that normal cognitive processes call for the transient integration of numerous functional areas in various regions of the brain, and as a result, the brain regions are in constant interaction with each other [15, 101, 104, 105]. Neural synchrony plays a vital role in such large-scale integration. In fact, various neuronal groups oscillate in specific frequency bands and become phase-locked over a limited period of time. This observation has stimulated the development of robust approaches that allow one to measure the phase-synchrony in a given frequency band from experimentally recorded biomedical signals, such as EEG.

In particular, the importance of synchronization of neuronal discharges has been shown by a variety of animal studies using microelectrode recordings of brain activity [83, 96]. The findings in the microelectrode-recording studies are also supported at coarser levels of resolution by other studies in animals and humans [21].

An electrophysiological signal is recorded via a low impedance extracellular microelectrode by placing the microelectrode sufficiently far from individual local neurons in order to prevent any particular cell from dominating the signal. Next, to obtain the local field potential (LFP), the signal is low-pass filtered, with a cut off at approximately 300 Hz. Due to the low impedance and positioning of the micro electrode, the recorded signal is predominantly induced by the activity of a large number of neurons. The unfiltered signal reflects the sum of action potentials from cells within approximately 50–350 micrometers from the tip of the electrode [53] and slower ionic events from within 0.5–3 millimeters from the tip of the electrode [39]. The spike component of the signal is removed by low-pass filter, whereas the lower frequency signal, the LFP, is preserved in the signal. It is assumed that the local field potential characterizes the synchronized input into the observed area, in the contrast to the spike data, which represents the output from the area.

Local field potentials (LFPs) of various degrees of spatial resolution can be recorded by scalp EEG or MEG. In fact, studies have shown that the presence of gamma and beta

band responses can be detected during visual discrimination tasks on the human scalp [97] and in subdural electrocorticograms [50, 54]. In addition, some recent findings suggest that long-range synchronization analogous to the one found in microelectrode studies in animals can also be detected between surface recordings [82].

It has been shown that synchronization is a significant attribute of the signal recorded from the patients affected by several neurological disorders. In particular, researchers have found that epilepsy [65] and Parkinsons disease [99] manifest as a pathological form of the synchronization process.

It is noted in [56] that although the cross-correlograms between spike discharges may be adequate for microelectrode studies, the quantification of phase synchrony between meso- or macro-electrodes (i.e. EEG/MEG, intracranial recordings) calls for entirely different methods. Therefore, they emphasize an importance of clearly distinguishing between synchrony as an appropriate estimate of phase relation, and the classical measures of coherence or spectral covariance that have been extensively used in neuroscience [8, 10, 62]. Le Van Quyen et al. discuss two important limitations of coherence [56].

The first limitation arises because the standard approaches for measuring coherence [12] based on Fourier analysis are known to be highly dependent on the stationarity of the measured signal, whereas the signals recorded from the brain, such as EEG, appear to be clearly non-stationary. Applying the timefrequency estimation method, which is not founded on the assumption of stationarity, could improve this limitation towards estimating a stable, instantaneous coherence as well as synchrony between two concurrent brain signals.

The second limitation stems from the fact that classical coherence is a measure of spectral covariance. Hence, it is not able to separate the effects of amplitude and phase in the relations between two signals. Because we are concerned with examining the specific hypothesis that phase-locking synchrony is the pertinent biological mechanism of transient integration in the brain, coherence serves only as an indirect measure. In order

to investigate the phase relations between different areas in the brain directly, the phase component should be extracted separately from the amplitude component for a given frequency or frequency band, which can be quite unstable or even chaotic. In a nutshell, coherence gives only an indirect and approximate indication of phase synchrony. There has been a general increase of interest in understanding bivariate data by studying their phase synchronization over time not only in neuroscience, but also in other research fields [86]. In other words, even though our discussion of the phase synchrony is focused on EEG data, its applications can also be extended to the fields other than neuroscience.

Classical concept of the synchronization of two oscillators is described as an active adjustment of their rhythmicity that manifests in phase-locking between the synchronized oscillators. Specifically, given two signals $X_1(t)$ and $X_2(t)$, and their corresponding instantaneous phases $\phi_1(t)$ and $\phi_2(t)$, the basic definition of the *phase locking* states that

$$n\phi_1(t) - m\phi_2(t) = C \equiv const, \tag{3–1}$$

where integers n and m specify the phase locking ratio.

When investigating phase synchrony in neurophysiological signals, one must assume that the constant phase locking ratio is valid within a limited time interval T, which usually means a few hundreds of milliseconds. When examining neural signals, one must keep in mind that discovering the presence of the phase locking between EEG recordings from two distant parts of the brain is not straightforward. The detection of phase synchrony in neural signals is problematic because of several factors particularly when working not on the level of a single neuron, but rather with large neuronal populations, whose activity is picked up by macroscopic or meso- electrodes. As noted in [56], as a consequence of volume conduction effects in brain tissues, the activity of a single neuronal population can be recorded by two distant electrodes, which results in spurious phase-locking between their signals. Furthermore, in non-invasive EEG, the true synchronies are hidden in a significant background noise, and so, in the synchronous

state, the phase shifts back and forth around some constant value. Hence, the signals can be viewed as *synchronous* or *not synchronous* only in a statistical sense. Therefore, this necessitates the development of novel approaches that are capable of extracting the true phase synchronies from noisy data, and so, the condition (3–1) must be adjusted to account for the noise as follows:

$$|n\phi_1(t) - m\phi_2(t)| \leq C, \tag{3–2}$$

where C denotes a positive constant.

The following two key steps are instrumental in investigating the phase synchrony:

1. estimate instantaneous phase of each signal;
2. provide a statistical criteria to quantify the degree of phase-locking.

Two methods for detecting phase-locking applied to neuronal signals have recently been considered in the literature. Tass and colleagues [99] extracted the instantaneous phases from original signals by means of the Hilbert transform, and then applied to magnetoencephalographic (MEG) motor signals in patients affected by Parkinson's desease [99], and also to the synchronization between cardiovascular and respiratory rhythms [93]. On the other hand, Lachaux et al. [50] estimated the phases from the original signals by means of convolution with a complex wavelet, and then applied it to EEG and intracranial data during cognitive tasks [51, 82].

The first step in quantifying phase synchronization between two time series X and Y is the determination of their instantaneous phases $\phi_X(t)$ and $\phi_Y(t)$. This is achieved either via the Hilbert transform or via the wavelet transform. These approaches are presented in the next two sections.

3.2 Phase Estimation using Hilbert Transform

The first method used to extract the instantaneous phase from the time series is based on the *analytic signal approach*, which was first introduced by D. Gabor [23] and later extended for model systems and experimental data [70, 86].

The *Hilbert transform* of a given real-valued function $f(t)$ with domain T is defined as a real-valued function $\hat{f}(t)$ on T as follows:

$$\hat{f}(t) = CPV \int_{-\infty}^{+\infty} f(\tau)g(t-\tau)d\tau = CPV \int_{-\infty}^{+\infty} g(\tau)h(t-\tau)d\tau, \tag{3–3}$$

where

$$g(t) := \frac{1}{\pi t}, \qquad t \in T,$$

and symbol CPV signifies that the integral is taken in the sense of *Cauchy principal value*.

Notice that $\hat{f}(t)$ can be viewed as a convolution $g(t) * f(t)$ of the original function $f(t)$ with the function $g(t)$. This means that the Hilbert transform can be performed by applying an ideal filter, whose amplitude response equals to 1, and phase response is a constant $\pi/2$ lag at all frequencies.

Given an arbitrary continuous real-valued time series $X(t)$, the corresponding *analytic signal* is defined as the following complex-valued function:

$$\xi_X(t) = X(t) + \imath \cdot \hat{X}(t) = a_X(t) \cdot \exp\{\imath \cdot \phi_X(t)\}, \tag{3–4}$$

where t denotes time, $\imath$ is a unit on the complex axis, $\hat{X}(t)$ denotes the Hilbert transform of the time series $X(t)$, $a_X(t)$ is the corresponding instantaneous amplitude, and $\phi_X(t)$ represents the instantaneous phase of the signal via Hilbert convolution.

It follows from Equation 3–4 that the *instantaneous phase* $\phi_X(t)$ of $X(t)$ can be computed as:

$$\phi_X(t) = \arctan\left\{\frac{\hat{X}(t)}{X(t)}\right\}. \tag{3–5}$$

A key advantage of the analytic approach is that the phase can be easily computed for an arbitrary broad-band signal. On the other hand, instantaneous amplitude and phase have a clear physical meaning only if $X(t)$ is a narrow-band signal. Therefore, filtration is required in order to separate the frequency band of interest from the background brain activity.

3.3 Phase Estimation via Wavelet Transform

An alternative approach to determining the instantaneous phase of the signal is based on the wavelet transform. This method of phase estimation was proposed by Lachaux and colleagues [50, 52], and is similar to the Hilbert transform method presented above. In their approach, Lachaux et al. extract the instantaneous phase by applying the convolution of the original signal with a complex Morlet wavelet. They consider the *Morlet wavelet* (also known as *Gabor function*) at time t and frequency ω given by the following formula:

$$\psi_{t,\omega}(\tau) = \sqrt{\omega} \cdot \exp\left\{\imath \cdot 2\pi\omega \cdot (\tau - t)\right\} \cdot \exp\left\{-\frac{(\tau - t)^2}{2\sigma^2}\right\}. \tag{3–6}$$

Notice that $\psi_{t,\omega}(\tau)$ is the product of a sinusoidal wave at frequency ω, and a Gaussian function centered at time t, with a standard deviation σ proportional to the inverse of ω. It depends solely on σ, which sets the number of cycles of the wavelet to $6\omega\sigma$.

According to [56], given the time series $X(t)$, the coefficient of the Morlet transform as a function of time t and frequency ω is defined as follows:

$$W_X(t,\omega) = \int_{-\infty}^{+\infty} X(\tau) \cdot \overline{\psi_{t,\omega}}(\tau) d\tau, \tag{3–7}$$

where $\overline{\psi_{t,\omega}}(\tau)$ denotes the complex conjugate of the Morlet wavelet $\psi_{t,\omega}(\tau)$.

The following slight modification of the Morlet wavelet is introduced in [81]:

$$\psi(t) = \exp\left\{-\frac{t^2}{2\sigma^2}\right\}\left(\exp\left\{\imath \cdot \omega_0 t\right\} - \exp\left\{-\frac{\omega_0^2\sigma^2}{2}\right\}\right), \tag{3–8}$$

where parameters ω_0 and σ represent the center frequency and the rate of decay of the wavelet function, respectively. This is proportional to the number of cycles and related to the frequency span by the uncertainty principle.

Similarly to the above, a complex time series of wavelet coefficients is obtained via the convolution of $X(t)$ with $\psi(t)$:

$$W_X(t) = (\psi \circ X)(t) = \int_{-\infty}^{+\infty} \psi(\tau) X(t - \tau) d\tau = \widetilde{a}_X(t) \cdot \exp\left\{\imath \cdot \widetilde{\phi}_X(t)\right\}, \tag{3–9}$$

where $\widetilde{a}_X(t)$ and $\widetilde{\phi}_X(t)$, respectively, are the instantaneous amplitude and the phase of the signal $X(t)$ extracted via the Morlet wavelet.

As in the case of the Hilbert transform, the phases can be determined from Equation (3–9) as

$$\widetilde{\phi}_X(t) = \arctan\left\{\frac{\Im\left[W_X(t)\right]}{\Re\left[W_X(t)\right]}\right\}, \tag{3–10}$$

where $\Re\left[W_X(t)\right]$ and $\Im\left[W_X(t)\right]$ denote the real and imaginary parts of the complex transformed time series $W_X(t)$, respectively.

3.4 Comparison between Two Approaches to Phase Extraction.

The above two definitions of the instantaneous phase are closely related, despite the fact that they are based on very different transformations. In particular, the connection between the phases obtained via the Hilbert and wavelet transforms was demonstrated experimentally in [56] and also explained theoretically by Quian Quiroga and colleagues [81].

In a nutshell, the phase $\widetilde{\phi}_X(t)$ extracted from the signal using the wavelet transform corresponds approximately to the phase $\phi_X(t)$ determined via the Hilbert convolution, which would be performed after band pass filtering the time series. Furthermore, if the phase estimation based on wavelet transform were performed by a convolution with an analytic wavelet, and if this wavelet were applied to do the band pass filtering in the Hilbert approach, then such approaches would, in fact, be equivalent.

It is easy to see that in the method based on the wavelet convolution, the center frequency ω and the rate of decay σ of the wavelet can serve as parameters that allow us to modify the frequency range of interest. On the other hand, the actual phase extraction via the Hilbert transform is free of parameters, and so the correspondent phase preserves information from the entire power spectrum and not just the main frequency band as in the case of the wavelet convolution. As a result, it is possible to achieve a comparison of narrow band and broad band synchronization simply by using both methods of phase extraction without performing any additional filtering.

3.5 Measures of Phase Synchrony

Various measures of phase synchrony between two signal are proposed based on the phases extracted via the Hilbert and the wavelet transforms, including standard deviation, mutual information and Shannon entropy [49, 56]. However, most of the currently used measures of phase synchronization are based on *bivariate* indexes.

In Chapter 4, we propose a novel *multivariate* approach to detecting phase synchronization in the phases extracted from multiple time series, such as multichannel EEG.

CHAPTER 4
APPLICATION OF VECTOR AUTOREGRESSION TO MINING BRAIN DYNAMICS

4.1 Numerical Issues in Estimating Parameters of Vector Autoregression from EEG

In the applications of Granger causality and related measures to the EEG, the directional dependencies in neural data are analyzed based on autoregressive modeling. Although, Bernasconi and König [2] applied statistical testing to verify the stationarity of the data, and established the duration of the stationary interval for EEG to be approximately 1 second. The statistical testing of underlying assumptions of the VAR (which was thoroughly discussed in Chapter 4) is often omitted in the later studies. As shown above, the stability condition is a very important assumption of vector autoregression. In this study, we examine how the parameters of the model order and sample size influence the stability of the derived VAR model.

In order to estimate VAR(p) model from data and investigate the properties of the derived model, the rodent intracranial EEG data were used. The data set consisted of the electroencephalographic recordings from 6 electrodes (implanted in left frontal, right frontal, two left hippocampal and two right hippocampal parts of the rodent's brain) sampled at 200 Hz.

To examine the applicability of the vector autoregressive modeling to EEG data, we estimated the VAR(p) model parameters for different values of lag order p and different sample sizes T. The sample sizes $T \in [1, 300]$ were used, and the lag length parameter p varied between 1 and 30. In addition, we filtered data using a Rectangular band pass Hamming window with 100 coefficients into the frequency bands of 0 – 30 Hz, 30 – 60 Hz, 60 – 90 Hz, 90 – 120 Hz, and 120 – 150 Hz. The raw data and the five differently filtered data represented separate data sets. For each data set, we ran the model estimation procedure with $T = 1, 2, \ldots, 300$ and $p = 1, 2, \ldots, 30$. The procedure for estimating coefficients of the model was implemented in the MATLAB environment based on the multivariate least squares method presented above. For every VAR model derived from

the data, the stability condition was checked, and the number of the roots of the reverse characteristic polynomial (RCP) on and inside the complex unit circle was stored. For each data set, a 3-dimensional surface plot was produced by graphing (T, p, n), where n denotes the number of the RCP roots on and inside the unit circle, whereas T and p represent the sample size and the model order (or lag length), respectively. The obtained surface plots are displayed in Figures 4-1 and 4-2.

The surface plot in Figure 4-1 supports the fact [59] that the sample size parameter should significantly exceed the lag length p, i.e. $T \gg p$. On the other hand, it can be seen from Figure 4-1 that even for $T \gg p$, the stability assumptions may still be violated. Indeed, for $T = 132$ considerably larger than $p = 2$, the estimated VAR has two RCP roots on or inside the unit circle ($n = 2$).

Figure 4-2 shows that for the filtered data, the (T, p) region, where the stability condition of the estimated VAR(p) model is violated, covers almost the whole domain. Whereas the (T, p) region that corresponds to stable VAR(p) models is much smaller than the stable region in Figure 4-1, and characterized by large T and very small p. Very similar results were obtained for the number of the RCP roots inside the unit circle, when estimating parameters of VAR(p) with different p using filters in the 30–60 Hz, 60–90 Hz, 90–120 Hz, and 120–150 Hz bands.

The experiment was repeated with consistent results on various samples from the six data sets. The results of our experiments clearly show that the stability condition imposed on the linear VAR model is often violated even for the parameters $T \gg p$. Furthermore, filtering the data within some restricted frequency band often leads to reduction of the (T, p) domain where the estimated VAR(p) models remain stable. In practice, the EEG data are often filtered within a certain frequency band. In many studies that utilize the vector autoregression to investigate Granger causality in the biomedical time series, the verification of the conditions assumed by the VAR model is often omitted. From our point

of view, the relevant statistical tests justifying the suitability of the model should always be performed when estimating the model from data.

Various modifications of vector autoregression (which relax the stability condition of no roots inside unit circle) are developed for analysis of economic time series. Although Bernasconi and König [2] examined the stationarity of different EEG data, and concluded that a time interval of approximately one second is the interval during which the EEG time series can be considered stationary, the underlying assumptions of the VAR modeling such as stationarity and stability are rarely statistically tested in applications to EEG data. The results of our experiments indicate that the stability assumptions of vector autoregressive model are often violated in application to the EEG data. This observation suggests that suitable extensions of multivariate autoregression to unstable processes may fit the data better, and as a result, such extensions of VAR may be more appropriate for the EEG time series analysis than the standard linear vector autoregressive modeling.

Additional statistical testing is required in order to make conclusions on what multivariate models are the most suitable for extracting the directional dependencies between channels in a frequency domain from multichannel EEG data.

4.2 Multivariate Approach to Phase Synchrony via Cointegrated VAR

We propose a new approach to measuring the synchrony among the instantaneous phases extracted from multivariate time series. This approach is based on the Cointegrated VAR modeling of time series.

Given the signal represented formally as a multiple time series $X(t)$, one can extract the instantaneous phases $\phi_{X_i}(t)$ from each one-dimensional component $X_i(t)$ of the signal as shown in Chapter 3 (either via a convolution with the Morlet wavelet or by applying the Hilbert transform). The phase extraction procedure produces a new multiple time series $\phi_X(t)$ of the correspondent phases.

Next, we derive new measures of phase synchrony of the signal based on the concepts introduced in Chapter 3. Let us observe that the left-hand side of Equation (3–1)

represents the linear combination of the respective phases $\phi_{X_1}(t)$ and $\phi_{X_2}(t)$ with integer coefficients. Also recall that condition (3–1), which defines phase-locking between two signals $X_1(t)$ and $X_2(t)$, needs to be modified in practice to account for the noise in the signal. Taking into account presence of the stochastic noise in the phase series, let us introduce a *modified* concept of the phase synchrony between two signals *by relaxing the integrality condition on the coefficients in the linear combination* as follows.

Two signals $X_1(t)$ and $X_2(t)$ are considered to be *generally phase-synchronized*, if the correspondent instantaneous phases $\phi_{X_1}(t)$ and $\phi_{X_2}(t)$ satisfy the condition below:

$$\exists\, c_1, c_2 \quad : \quad c_1\phi_{X_1}(t) + c_2\phi_{X_2}(t) = z_t, \tag{4–1}$$

where $z_t \sim N(C, \sigma')$ is a stochastic variable that represents the deviation from the constant level C as a result of the noise. Notice that in the contrast to condition (3–1) in the classic definition of phase synchronization, the coefficients c_1 and c_2 in the definition (4–1) of generalized phase-synchrony do not need to be integer.

Furthermore, it is straightforward that the new condition (4–1) means that a two-dimensional process $X(t) = (X_1(t), X_2(t))'$ is cointegrated. Based on this observation, we can extend our modified concept of phase synchronization between two signals to the multivariate case in the following manner.

The multichannel signal $X(t) = (X_1(t), \ldots, X_K(t))$ is considered to be *phase-synchronized of rank* r, if the process $\phi_X(t)$ composed of the correspondent instantaneous phases $\phi_{X_i}(t)$, $i = 1, \ldots, K$ is cointegrated of rank r.

In the subsequent subsections, we first discuss the role of the cointegration rank in the framework of multivariate phase-synchronization, and then apply this approach to multichannel EEG data collected from the patients with absence epilepsy.

4.2.1 Cointegration Rank as a Measure of Synchronization among Different EEG Channels

Note that integrated autoregressive processes $I(d)$ are shown to exhibit behavior similar to that of a random walk. In a short paper [66], Michael Murray used an example of drunkard and her dog to illustrate the concept of the cointegration. To explain our reasoning behind the rank of cointegration as a measure of synchrony, we briefly summarize and then further extend his analogy.

The nonstationary processes (such as a random walk) are often introduced by teachers of statistics by comparing them (it) with the drunkard's walk. The drunkard wonders aimlessly, so that the direction of each step is random and completely independent of her previous steps. In other words, the meandering of the drunkard is described by a random walk:

$$x_t - x_{t-1} = \varepsilon_t, \quad t = 1, 2, \ldots, \tag{4–2}$$

where x_t represents the position of the drunk at time t, and ε_t is a stationary white-noise, which models the drunk's step at time t).

As Murray noticed [66], an unleashed puppy is another creature, whose behavior reminds a random walk. Indeed, each new scent that puppy's nose comes upon dictates a direction for the pup's next step so strongly that the last scent along with its direction is forgotten as soon as the new scent appears. Having shown that the puppies follow the random walk y_t, $t = 1, 2, \ldots$, let us represent the puppy's walk as:

$$y_t - y_{t-1} = \epsilon_t, \quad t = 1, 2, \ldots, \tag{4–3}$$

where ϵ_t is a stationary white-noise (i.e. puppy's step at time t).

The well-known feature of a random walk is that the best predictor of the future value is the most recently observed one. In other words, the longer it has been since we had seen the drunk, or the dog, the further away from the initial place, on average, they

are at the moment. As a result, even if the drunk and the dog crossed their walks at some location, as the time goes on, they tend to wander further away from each other.

However, if the puppy belongs to the drunkard, then they will remain relatively close to each other at all the time, similarly to the individual integrated processes that together form a cointegrated process. Indeed, the drunk would still wonder aimlessly in a random walk fashion, as would her puppy. However, from time to time she would remember about her dog and call for it, the puppy would recognize her voice, and bark. They would hear each other and make their next step in each other's direction.

The paths of the drunk and her dog are still nonstationary, but they are no longer independent from each other. As a matter of fact, at each time, the puppy and its master are likely to be found not far from each other. If this is true, then the distance between two paths is stationary, and the walks of the drunk x_t and her dog y_t are said to be *cointegrated*, i.e. x_t and y_t are integrated $I(1)$, and there is a *linear combination* of x_t and y_t (with non-zero weights) that is $I(0)$, i.e. stationary.

Mathematically, the *cointegrating relationship* between a lady and her puppy can be written as:

$$x_t - x_{t-1} = \varepsilon_t + c(y_{t-1} - x_{t-1}), \tag{4–4}$$

$$y_t - y_{t-1} = \epsilon_t + d(x_{t-1} - y_{t-1}), \tag{4–5}$$

at time $t = 1, 2, \ldots$. Note that ε_t and ϵ_t, as before, represent the stationary white-noise steps of the drunk and her dog.

Since Equation 4–4 can be easily rewritten in form of (2–45) as follows:

$$\Delta \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ \epsilon_t \end{bmatrix} - \begin{bmatrix} c & -c \\ -d & d \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix}, \tag{4–6}$$

then

$$\Pi = \begin{bmatrix} c & -c \\ -d & d \end{bmatrix},$$

and so, $\text{rank}(\Pi) = 1$. This shows that the cointegrating relationship between the drunk lady and her puppy has the cointegration rank 1.

Note that $\text{rank}(\Pi) = 0$, if and only if $c = d = 0$. In such case, (4–4) becomes simply a system of equations (4–2) and (4–3), which models *two* independent random walks driven by independent white noise processes ε and ϵ. On the other hand, when at least one of the coefficients c and d is non-zero, then by multiplying system (4–6) by a vector $[d, c]'$, we have:

$$d\Delta x_t + c\Delta y_t = d\varepsilon_t + c\epsilon_t, \quad t = 1, 2, \ldots, \tag{4–7}$$

which means that the model is driven by a *single* common stochastic trend $d\varepsilon_t + c\epsilon_t$.

Although the example described by Murray is clearly a bivariate cointegrated VAR(1), it can be extended to an illustration of the multivariate cointegrated process. Consider, for example, a heard of sheep guarded by two dogs, where the sheep wonder aimlessly in the field, while the dogs run around and bring the sheep that have strayed too far back into the flock. Say, for example, a faster dog guards sheep from the east, south, and west, whereas a slower dog - from the north, then the cointegrated process appear to have the cointegration rank of 2. Clearly, two dogs are able to keep a flock of sheep closer together, than a single dog can. In other words, the higher cointegration rank the more restrictive it is.

In fact, let us consider a K-dimensional cointegrated vector autoregressive process, and let r denote the cointegration rank of the process. Similarly to the bivariate example above, we can see that when the rank is zero ($r = 0$), the univariate components of the process are independent, and the model is driven by K independent white noise processes (i.e. there is no cointegration). In the case of $r = 1$, we can decompose the multivariate

process onto $K-2$ independent components, and two dependent components that form a common stochastic trend. Hence, in the case $r=1$, the cointegrated model is driven by $(K-2)+1=K-1$ independent stochastic processes. By induction, we can show that for a cointegrated VAR process with the cointegration rank r, $0<r<K-1$, the VAR model is generated by $K-r$ independent stochastic trends.

Therefore, the smaller is the cointegration rank r, the larger is the number $K-r$ of the underlying independent stochastic trends, and so (the larger) is the vector space in which our cointegrated model can travel. And the other way around, increasing the cointegration rank of the model shrinks the underlying domain of the process, i.e. makes it bounded to a smaller hyperplane. For $r=K$, the VAR(p) is a stable process, which clearly has the most constrained domain. For $r=0$, the VAR process is not cointegrated and unrestricted.

Thus, in the framework of generalized phase-synchronization introduced above, the cointegration rank represents a fundamental measure of synchrony in the multi-channel signal, such as EEG. In particular, we say that the signal is *completely asynchronous*, if the cointegration rank r is zero. On the other hand, when the multivariate process is stable (i.e. the rank coincides with the dimension of the process, $r=K$), the signal is said to be *perfectly synchronous*.

4.2.2 Absence Seizures

Absence seizures (or petit mal seizures) are known to occur in several forms of epilepsy, whereas absence epilepsy refers to a type of epilepsy in which only the absence seizures occur. Absence epilepsy is usually characterized by age of onset, and often affects teenage population. Absence seizures usually begin in childhood or adolescence, and often run in families, which may suggest a genetic predisposition. Absence seizures are marked by momentary lapses of consciousness. Absence seizures often have no visible symptoms, although some patients may have purposeless movements during a seizure, such as rapidly blinking eyes. Absence seizures often have a brief duration, and a person may resume

the previous activity immediately after the seizure [73]. These brief seizures can happen several times during a day, but in some patients, the frequency of absence seizures can be as high as hundred of times a day, which interferes with the daily activities of a child such as school. In some cases of childhood absence epilepsy, the seizures stop when a child reaches puberty. Absence seizures exhibit a characteristic spike-and-wave EEG pattern at a 3 Hz frequency [73].

Figure 4-3 displays a multichannel EEG recording that includes an absence seizure. The duration of the seizure is approximately 4 seconds. The figure vividly illustrates a characteristic spike-and-wave activity during the seizure.

4.2.3 Numerical Study of Synchrony in Multichannel EEG Recordings from Patients with Absence Epilepsy

The proposed approach to studying synchronization among multiple channels was applied to analysis of EEG data recorded from the patience with absence epilepsy.

First, the multiple time series of the instantaneous phases were extracted from the raw EEG data using the Hilbert transform approach as described in Section 3.2. In particular, we took advantage of the functions `hilbert` and `angle` readily available in the MATLAB R 2006a environment.

The VAR modeling and testing were implemented using the R 2.6.1 statistical software. In our analysis of the instantaneous phases, we incorporated `ar`, `adf.test`, `po.test`, `cajolst` and other functions found in packages *tseries* and *urca*.

Next, we illustrate our approach on the example of the EEG data file that includes three seizure intervals. The file contains a 16-channel recording of scalp EEG sampled at the 200 Hz frequency as well as two auxiliary channels, which were discarded. The instantaneous phase values were estimated from the EEG time series by means of Hilbert transform, and the resulting phase series were tested using the ADF test introduced in Section 2.4.1. Specifically, we applied the Augmented Dickey-Fuller procedure to test the

presence of a unit root in the individual univariate components of the multiple time series of estimated phases.

The results of our experiments for seizures 1, 2, and 3 are presented in Tables 4-1, 4-2, and 4-3, respectively. The channels, for which the ADF unit root test has detected a presence of a unit root at the significance level $\alpha = 0.01$, are listed as *integrated*. Whereas the channels, for which the null hypothesis of a unit root has been rejected by the ADF at the 1 percent level, are denoted by *stationary*. The channels for which the p-values of the ADF test exceed 2.5% are marked with *. Notice that all three seizure segments are considered stable, when the ADF is applied at a 0.025 significance level.

Next, we fit vector autoregression to the multiple time series of phase estimates, for each of three different segments (before, during, and after a seizure) in order to determine appropriate lag length parameter p. To find appropriate lags p, the Akaike Information Criteria (AIC) was used. This led us to choose several lag length for each segment and each seizure. Finally, Johansen cointegration rank procedure was applied to determine the values of cointegration rank r for each case. The results are summarized in Tables 4-4, 4-5, and 4-6.

Notice that during the seizure the system becomes stable, especially when modeled using a short estimate of the lag parameter. Since the durations of the seizure 1 and seizure 2 are rather short, and only include 440-500 sample points, the models estimated under a long lag parameter may not adequately represent the underlying processes in seizure 1 and 2. On the other hand, seizure 3 is estimated based on almost 1200 sample values, and therefore the long lag model of a longer seizure 3 may be more realistic, than the long lag models for shorter seizures 1 and 2. Overall, the models based on a short lag p for all three seizures provide an evidence of absolute synchronization among the channels. Whereas, the the pre-seizure and post-seizure models are more likely to be less restricted, and seem to exhibit a cointegration rank between 9 and 16.

Table 4-1. Results of the ADF unit root tests for each channel during three segments (2 seconds immediately before seizure, during seizure, and 2 seconds after seizure) for seizure 1. Note that the significance at 2.5% level is denoted by *

Seizure #	pre-seizure	seizure	post-seizure
stationary	3,4,7,9,15	1-3,5-14	1,3,5,6,8,10,12,15
integrated	1*,2*,5,6*,8*,10,11,12*,13*,14*,16*	4,15,16	2*,4*,7,9,13*,14,16*

Table 4-2. Results of the ADF unit root tests for each channel during three segments (2 seconds immediately before seizure, during seizure, and 2 seconds after seizure) for seizure 2. Note that the significance at 2.5% level is denoted by *

Seizure #	pre-seizure	seizure	post-seizure
stationary	3,4,7,9-11,13,16	1-16	3,7,11-16
integrated	1*,2*,5*,6,8*,12*,14*,15	none	1*,2,4*,5,6*,8,9*,10

Table 4-3. Results of the ADF unit root tests for each channel during three segments (2 seconds immediately before seizure, during seizure, and 2 seconds after seizure) for seizure 3. Note that the significance at 2.5% level is denoted by *

Seizure #	pre-seizure	seizure	post-seizure
stationary	7,11,13,14	1-16	2,4,5,11,13,15
integrated	1*,2*,3,4*,5*,6*,8*,9,10*,12,15,16*	none	1,3*,6*,7,8,9*,10*,12,14*,16

Table 4-4. Results of the Johansen cointegration rank procedure for the multiple series during three segments (2 seconds immediately before seizure, during seizure, and 2 seconds after seizure) for seizure 1. Significance level is 1%. Full rank is denoted by †

Segment	pre-seizure	seizure	post-seizure
long lag	$p = 22, r = 12$	$p = 23, r = 11$, $p = 20, r = 13$	$p = 20, r = 12$
short lag	$p = 2, r = 13$	$p = 2, r = 16$†	$p = 2, r = 10$

Table 4-5. Results of the Johansen cointegration rank procedure for the multiple series during three segments (2 seconds immediately before seizure, during seizure, and 2 seconds after seizure) for seizure 2. Significance level is 1%, full rank is denoted by †

Seizure #	pre-seizure	seizure	post-seizure
long lag	$p = 21, r = 14$	$p = 26, r = 12$, $p = 20, r = 9$	$p = 20, r = 10$
short lag	$p = 2, r = 16$†	$p = 3, r = 16$†	$p = 2, r = 13$

Table 4-6. Results of the Johansen cointegration rank procedure for the multiple series during three segments (2 seconds immediately before seizure, during seizure, and 2 seconds after seizure) for seizure 3. Significance level is 1%, full rank is denoted by †

Seizure #	pre-seizure	seizure	post-seizure
long lag	$p = 24, r = 13$	$p = 26, r = 16$†, $p = 20, r = 16$†	$p = 20, r = 13$
short lag	$p = 2, r = 9$	$p = 2, r = 16$†	$p = 2, r = 16$†

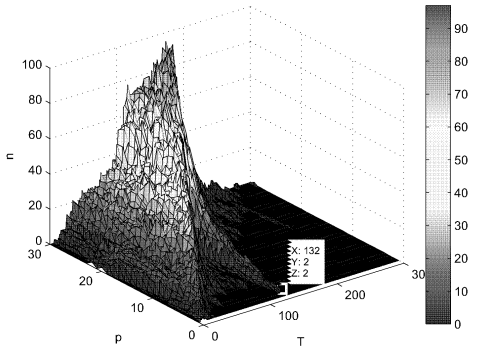


Figure 4-1. Numbers n of roots of the reverse characteristic polynomial (RCP) for VAR(p), which lie on and inside the complex unit circle, computed for different sample sizes T and for different model orders p using the raw data

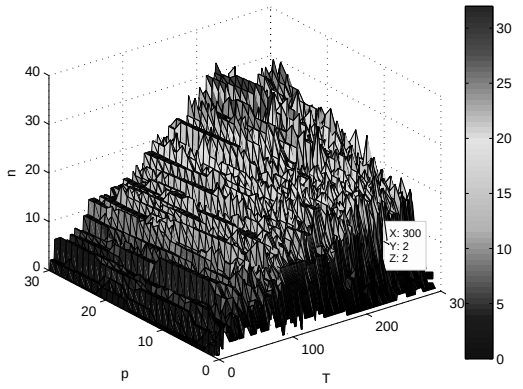


Figure 4-2. Numbers n of roots of the reverse characteristic polynomial (RCP) for VAR(p), which lie on and inside the complex unit circle, computed for different sample sizes T and for different model orders p using the 0–30 Hz band filtered data

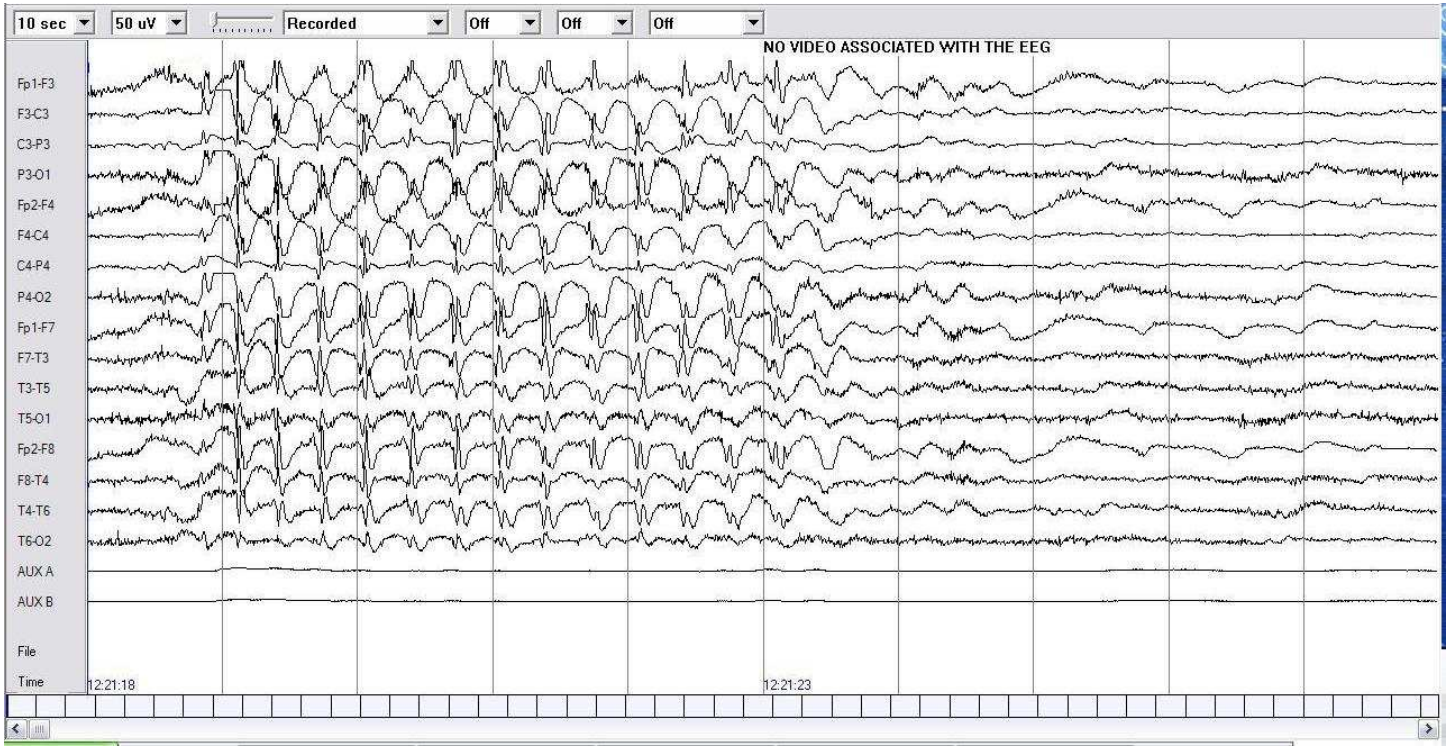


Figure 4-3. Segment of multichannel EEG recording that contains a four second long absence seizure

CHAPTER 5
CONCLUSION

Investigation of spatio-temporal properties of the EEG data by data mining and optimization approaches posts various challenges. Numerous features and methods have been proposed for studying the multivariate series that is EEG. The analysis of EEG time series is often approached from two different points of view, the one that treats EEG data as produced by a deterministic chaotic dynamical system, and the other more traditional approach of linear autoregressive modeling.

In this work, we investigated several statistical approaches that are recently introduced for data mining brain dynamics. In particular, we examined the application of vector autoregressive modeling and linear Granger causality to raw and filtered EEG data.

Motivated by recent success in application of phase synchronization to analysis of dynamic processes in epileptic brain, we developed a concept of *generalized synchronization* based on the novel idea of extending the classical synchronization condition of a bounded linear combination of two phases. This simple bivariate condition in the multivariate case is analogous to a cointegrating relationship in the multiple time series. Thus, we can analyze the synchrony among different parts of the common interrelated system (such as a human brain), by modeling the phases extracted from a finite number of signals in the systems by means of cointegrated vector autoregression. Moreover, we showed that the cointegration rank in the cointegrated VAR model of the phase time series can be viewed as a measure of synchrony among the phases of different components of the EEG signal.

Not only this new measure of multivariate phase synchrony can be tested on various biomedical data, such as multichannel EEG recorded from an epileptic brain, but also the new multiple phase synchronization can be employed in different areas of applied and theoretic research (including physics, communication, electronics, laser dynamics, and

control) for studying synchronization among several dynamical systems or a system that consists of several parts.

REFERENCES

[1] L. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biol. Cybern.*, vol. 84, pp. 463-74, 2001.

[2] C. Bernasconi and P. Koënig, "On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings," *Biol. Cybern.* vol. 81, pp. 199–210, 1999.

[3] R. D. Bickford, "Electroencephalography," in Adelman G. ed. *Encyclopedia of Neuroscience*. Cambridge, MA: Birkhauser, pp. 371–373, 1987.

[4] C. D. Binnie, "Long term EEG recording and its role in clinical practice," *IEE Colloquium on Data Logging of Physiological Signals*, pp.5/1-5/2, 1995.

[5] P. S. Bradley, O. L. Mangasarian, and W. N. Street, "Clustering via concave minimization," in M. C. Mozer, M. I. Jordan, and T. Petsche, eds.,*Advances in Neural Information Processing Systems*, vol. 9, pp. 368–374, 1997.

[6] P. S. Bradley and O. L. Mangasarian, "Feature selection via concave minimization and support vector machines," in J. Shavlik, *Machine Learning Proceedings of the Fifteenth International Conference*, pp. 82–90, 1998.

[7] P. S. Bradley, U. M. Fayyad, and O. L. Mangasarian, "Mathematical Programming for Data Mining: Formulations and Challenges," *INFORMS Journal on Computing*, vol. 11, no. 3, pp. 217–238, 1999.

[8] S. L. Bressler, R. Coppola, and R. Nakamura, "Episodic multiregional cortical coherence at multiple frequencies during visual task performance," *Nature*, vol. 366, pp. 153156, 1993.

[9] R. Brown and L. Kocarev, "A unifying definition of synchronization for dynamical systems," *Chaos*, vol. 10, pp. 344-349, 2000.

[10] T. H. Bullock and M. C. McClune, "Lateral coherence of the electrocorticogram: a new measure of brain synchrony," *Electroencephalogr Clin Neurophysiol*, vol. 73, pp. 479-498, 1989.

[11] T. L. Carroll and L. M. Pecora, "Cascading synchronized chaotic systems," *Physica D*, vol. 67, pp. 126, 1993.

[12] G. C. Carter, "Coherence and time delay estimation," *Proc IEEE*, vol. 75, pp. 236-255, 1987.

[13] Y.-W. Cheung, K. S. Lai, "Lag Order and Critical Values of the Augmented Dickey-Fuller Test," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 277–280, 1995.

[14] R. Cont, "Statistical Properties of Financial Time Series," in R. Cont and J. Yong editors, *Mathematical Finance: Theory and Practice, Lecture Series in Applied Mathematics*, vol. 1, 1999.

[15] A. R. Damasio, "Synchronous activation in multiple cortical regions: a mechanism for recall," *Semin Neurosci*, vol. 2, pp. 287-296, 1990.

[16] D. A. Dickey, and W. A. Fuller, "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, vol. 74, pp. 427–431, 1979.

[17] M. Eichler, *Graphical modelling of multivariate time series*, University of Heidelberg: Preprint, 2001.

[18] M. Eichler, R. Dahlhaus, and J. Sandkuhler, "Partial correlation analysis for the identification of synaptic connections," *Biol. Cybern.*, vol. 89, pp. 289-302, 2003.

[19] M. Eichler, "A graphical approach for evaluating effective connectivity in neural systems," *Philos. Transact. R. Soc. B* vol. 360, pp. 953-67, 2005.

[20] L. Fabiny, P. Colet, and R. Roy, "Coherence and phase dynamics of spatially coupled solid-state lasers," *Phys. Rev. A*, vol. 47, pp. 4287, 1993.

[21] W. J. Freeman, "Spatial properties of an EEG event in the olfactory bulb and cortex," *Electroencephalogr Clin Neurophysiol*, vol. 44, pp. 586-605, 1978.

[22] W. A. Fuller, *Introduction to Statistical Time Series*, 2nd ed. New York: John Wiley, 1996.

[23] D. Gabor. "Theory of communication," in *Proc. IEEE London*, vol. 93, pp. 429, 1946.

[24] P. Georgiev, A. Cichocki, and H. Bakardjian, "Optimization techniques for Independent Component Analysis with Applications to EEG data," In P. M. Pardalos, J. C. Sackellares, P. R. Carney, L. D. Iasemidis, eds., *Quantitative Neuroscience: Models, Algorithms, Diagnostics, and Theraputic Applications*, Kluwer Academic Publishers, pp. 53–68, 2004.

[25] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *J. American Statis. Association*, vol. 77, pp. 304–313, 1982.

[26] C. W. J. Granger, "Economic Processes Involving Feedback," *Information and Control*, vol. 6, pp. 28–48, 1963.

[27] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrics*, vol. 37, pp. 424–438, 1969.

[28] C. W. J. Granger, "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *J of Econometrics*, vol. 16, pp. 121–130, 1981.

[29] M. J. Griffiths, P. Grainger, M. V. Cox, and A. W. Preece, "Recent advances in EEG monitoring for general anaesthesia, altered states of consciousness and sports performance science," in *The 3rd IEE International Seminar on Medical Applications of Signal Processing*, pp. 1–5, 2005.

[30] J. D. Hamilton, *Time Series Analysis*. Princeton, 1994.

[31] M.-J. Hoeve, R. D. Jones, G. J. Carroll, and H. Goelz, "Automated detection of epileptic seizures in the EEG," in *Engineering in Medicine and Biology Society, Proceedings of the 23rd Annual International Conference of the IEEE*, vol. 1, pp.943–946, 2001.

[32] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, 2nd edition. Kluwer Academic Publishers, 2000.

[33] C. Huygens. *Horoloquium Oscilatorium.* Paris, 1673.

[34] L. D. Iasemidis, "On the dynamics of the human brain in temporal lobe epilepsy," Ph.D. Dissertation, University of Michigan, Ann Arbor, 1991.

[35] L. D. Iasemidis, and J. C. Sackellares, "The evolution with time of the spatial distribution of the largest Lyapunov exponent on the human epileptic cortex," in D.W. Duke, and W.S. Pritchard, eds., *Measuring chaos in the human brain*, pp. 49–82, World Scientific, 1991.

[36] L. D. Iasemidis, D.-S. Shiau, J. C. Sackellares, P. M. Pardalos, and A. Prasad, "Dynamical resetting of the human brain at epileptic seizures: application of nonlinear dynamics and global optimization tecniques," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 3, pp. 493–506, 2004.

[37] H. H. Jasper, "The ten-twenty electrode system of the International Federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 370–375, 1958.

[38] S. Johansen and K. Juselius, "Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, vol. 52, pp. 169–210, 1990.

[39] E. Juergens, A. Guettler, and R. Eckhorn. "Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical- and EEG-potentials, but not in human EEG," *Experimental Brain Research*, vol. 129, pp. 247-259, 1999.

[40] Y. Kaji, M. Akutagawa, F. Shichijo, H. Nagashino, Y. Kinouchi, and S. Nagahiro, "Analysis for Brain Activities during Operations Using Measured EEG," in *Conf Proc IEEE Eng Med Biol Soc.* vol. 6, pp. 6003–6006, 2005.

[41] T. Kalayci and O. Ozdamar, "Wavelet preprocessing for automated neural network detection of EEG spikes," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 14, no. 2, pp.160–166, 1995.

[42] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol. Cybern.*, vol. 85, pp. 145–57, 2001.

[43] **A. Kammerdiner**, P. Xanthopoulos, and P. M. Pardalos, "Numerical limitations in application of vector autoregressive modeling and Granger causality to analysis of EEG time series," In *AIP Conf. Proc.*, vol. 953, pages 232–245, 2007.

[44] **A. Kammerdiner**, and P. M. Pardalos, "Analysis of multichannel EEG recordings based on generalized phase synchronization and cointegrated VAR," In *Proc. of Conference on Computational Neuroscience 2008* 2008 (submitted).

[45] E. Keogh, and S. Kasetty, "On the need for time series data mining benchmarks: a survey and empirical demonstration," in *Proceedings of the Eighth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*, pp. 102–111, 2002.

[46] M.E. Kirlangic, D. Perez, S. Kudryavtseva, G. Griessbach, G. Henning, and G. Ivanova, "Fractal dimension as a feature for adaptive electroencephalogram segmentation in epilepsy," in *Engineering in Medicine and Biology Society. Proceedings of the 23rd Annual International Conference of the IEEE*, vol. 2, pp.1573–1576, 2001.

[47] W. A. Kittel, C. M. Epstein, and M. H. Hayes, "EEG monitoring based on fuzzy classification," *Circuits and Systems, Proceedings of the 35th Midwest Symposium,* vol. 1, pp.699–702, 1992.

[48] L. Kocarev and U. Parlitz, "General approach for chaotic synchronization with applications to communication," *Phys. Rev. Lett.*, vol. 74, pp. 5028, 1995.

[49] T. Kreuz, *Measuring Synchronization in Model Systems and Electroencephalographic Time Series from Epilepsy Patients*, Dissertation. NIC Series, vol. 21, 2003.

[50] J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Hum. Brain Mapp.*, vol. 8, pp. 194, 1999.

[51] J. P. Lachaux, E. Rodriguez, J. Martinerie, C. Adam, D. Hasboun and F. J. Varela, "Gamma-band activity in human intracortical recordings triggered by cognitive tasks," *Eur J Neurosci*, vol. 12, pp. 2608-2622, 2000.

[52] J. P. Lachaux, E. Rodriguez, M. Le Van Quyen, A. Lutz, J. Martinerie, and F. J. Varela, "Studying single-trials of phase-synchronous activity in brain," *Int. J. Bifurcation Chaos Appl. Sci. Eng.*, vol. 10, pp. 2429, 2000.

[53] A. D. Legatt, J. Arezzo, and H. G. Vaughan, "Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials," *J. of Neuroscience Methods*, vol. 2 no. 2,pp. 203–217, 1980.

[54] M. Le Van Quyen, C. Adam, J. P. Lachaux, J. Martinerie, M. Baulac, B. Renault, et al., "Temporal patterns in human epileptic activity are modulated by perceptual discriminations," *Neuroreport*, vol. 8, pp. 1703-1710, 1997.

[55] M. Le Van Quyen, J. Martinerie, C. Adam, and F. J. Varela, "Nonlinear spatio-temporal interdependencies of interictal intracranial EEG recordings from patients with temporal lobe epilepsy: Localizing of epileptogenic foci," *Physica D*, vol. 127, pp. 250-65, 1999.

[56] M. Le Van Quyen, J. Foucher, J.-P. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, and F. J. Varela, "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony," *Journal of Neuroscience Methods*, vol. 111, pp. 83-98, 2001.

[57] H. Liang, M. Ding, R. Nakamura, and S. L. Bressler, "Causal influences in primate cerebral cortex during visual pattern discrimination," *Neuroreport*, vol. 11, pp. 2875–80, 2000.

[58] Q. Liu, M. Sun, M. L. Scheuer, and R. J. Sclabassi, "Patient tracking for video/EEG monitoring based on change detection in DCT domain," in *Bioengineering Conference, 2005. Proceedings of the IEEE 31st Annual Northeast*, pp.114–115, 2005.

[59] H. Lütkepohl, *Intoduction to multiple time series analysis*. Berlin, Heidelberg: Springer-Verlag, 1991.

[60] J. G. MacKinnon, "Critical Values for Cointegration Tests," in R. F. Engle and C. W. J. Granger, eds.,*Long-Run Economic Relationships: Readings in Cointegration*,New York: Oxford University Press, pp. 266–276, 1991.

[61] J. G. MacKinnon, "Approximate Asymptotic Distribution Functions for Unit- Root and Cointegration Tests," *Journal of Business & Economic Statistics*, vol. 12, pp. 167–176, 1994.

[62] V. Menon, W. J. Freeman, B. A. Cutillo, J. E. Desmond, M. F. Ward, S.L. Bressler, et al., "Spatio-temporal correlations in human gamma band electrocorticograms," *Electroencephalogr Clin Neurophysiol*, vol. 98, no. 2, pp. 89-102, 1996.

[63] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond*. World Scientific, 1987.

[64] S. Mitra and T. Acharya, *Data mining: multimidea, soft computing, and bioinformatics*. Hoboken, NJ: John Wiley & Sons, Inc., 2003.

[65] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, "Mean phase coherence as a measure for phase synchronization and its application to the EEG of epileptic patients," *Physica D*, vol. 144, pp. 358-369, 2000.

[66] M. P. Murray "A Drunk and Her Dog: An Illustration of Cointegration and Error Correction," *The American Statistician*, vol. 48, no. 1 , pp. 37–39, 1994.

[67] E. Niedermeyer, F. H. Lopes da Silva, *Electroencephalography: Basic principles, clinical applications and related fields*, 3rd edition. Philadelphia: Lippincott, Williams & Wilkins, 1993.

[68] P. L. Nunez, *Neocortical Dynamics and Human EEG Rhythms.* New York: Oxford Univ. Press, 1995.

[69] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a Time Series," *Phys. Rev. Lett.*, vol. 45, no. 9, pp.712–716, 1980.

[70] P. Panter. *Modulation, noise, and spectral analysis.* New York: McGraw-Hill, 1965.

[71] H. S. Park, Y.H. Lee, S. Doo S, et al., "Detection of epileptiform activity using wavelet and neural network," *Engineering in Medicine and Biology Society*, vol.3, pp. 1194, 1997.

[72] U. Parlitz, L. Junge, W. Lauterborn, and L. Kocarev, "Experimental observation of phase synchronization," *Phys. Rev. E*, vol. 54, pp. 2115, 1996.

[73] J. B. Patten, *Neurological Differential Diagnosis*, 2nd ed., Springer-Verlag, 1996.

[74] D. W. Peterman, M. Ye, and P. E. Wigen, "High frequency synchronization of chaos," *Phys. Rev. Lett.*, vol. 74, pp. 1740, 1995.

[75] P. C. B. Phillips, "Time series regression witha unit root," *Econometrica*, vol. 55, pp. 277–301, 1987.

[76] P. C. B. Phillips and S. Ouliaris, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, vol. 58, pp. 165-193, 1990.

[77] A. S. Pikovsky, "Phase synchronization of chaotic oscillations by a periodic external field," *Sov J Commun Technol Electron*, vol. 30, pp. 85, 1985.

[78] A. S. Pikovsky, M. G. Rosenblum, and J. Kurths. *Synchronization. A universal concept in nonlinear sciences.* Cambridge, UK: Cambridge Univ. Press, 2001.

[79] O. A. Prokopyev, V. Boginski, W. Chaovalitwongse, P. M. Pardalos, J. C. Sackellares, and P. R. Carney, "Network-based Techniques in EEG Data Analysis and Epileptic Brain Modeling," in P. M. Pardalos, V. Boginski and A. Vazacopoulos, eds., *Data Mining in Biomedicine*, Springer, pp. 559–573, 2007.

[80] K. Pyragas, "Continuous control of chaos by self-controlling feedback," *Phys. Lett. A*, vol. 170, pp. 421, 1992.

[81] R. Quian Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, "Performance of different synchronization measures in real data: A case study on electroencephalographic signals," *Phys. Rev. E*, vol. 65, pp. 041903, 2002.

[82] E. Rodriguez, N. George, J. P. Lachaux, J. Martinerie, and F. J. Varela, "Perceptions shadow: Long-distance synchronization in the human brain," *Nature*, vol. 397, pp. 340-343, 1999.

[83] P. R. Roelfsema, A. K. Engel, P. König, and W. Singer, "Visuomotor integration is associated with zero time-lag synchronization among cortical areas," *Nature*, vol. 385, pp. 157-161, 1997.

[84] E. Rosa Jr., W. B. Pardo, C. M. Ticos, J. A. Walkenstein, and M. Monti, "Phase synchronization of chaos in a plasma discharge tube," *Int. J. Bifurc. Chaos*, vol. 10, pp. 2551, 2000.

[85] M. Rosenblum, A. Pikovsky, J. Kurths, "Phase synchronization of chaotic oscillators," *Phys Rev Lett*, vol. 76, pp. 1804-1807, 1996.

[86] M. G. Rosenblum, A. S. Pikovsky, C. Schäfer, P. Tass, and J. Kurths, "Phase synchronization: From theory to data analysis," in: S. Gielen, F. Moss F, eds. *Handbook of Biological Physics*, vol. 4. Elsevier Science, Neuro-informatics, 1999.

[87] R. Roy and K. S. Thornburg, "Experimental synchronization on chaotic lasers," *Phys. Rev. Lett.*, vol. 72, pp. 2009, 1994.

[88] N. F. Rulkov, L. S. Tsimring, and H. D. I. Abarbanel, "Tracking unstable orbits in chaos using dissipative feedback control," *Phys. Rev. E*, vol. 50, pp. 314, 1994.

[89] J. C. Sackellares, L. D. Iasemidis, R. L. Gilmore, and S. . Roper, "Epilepsy - when chaos fails," in K. Lehnertz, J. Arnold, P. Grassberger, and C. E. Elger, eds., *Chaos in the brain?*, World Scientific, 2002.

[90] J. C. Sackellares, L. D. Iasemidis, D.-S. Shiau, P. M. Pardalos, and P. R. Carney, "Spatio-temporal transitions in temporal lobe epilepsy," in P. M. Pardalos, J. C. Sackellares, P. R. Carney, L. D. Iasemidis, eds., *Quantitative Neuroscience: Models, Algorithms, Diagnostics, and Theraputic Applications*, Kluwer Academic Publishers, pp. 223–238, 2004.

[91] S. E. Said and D. A. Dickey, "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," Biometrika, vol. 71, pp. 599-607, 1984.

[92] K. Sameshima and L. Baccala, "Using partial directed coherence to describe neuronal ensemble interactions," *J. Neurosci. Methods*, vol. 94, pp. 93-103, 1999.

[93] C. Schäfer, M. G. Rosenblum, J. Kurths, and H. H. Abel, "Heartbeat synchronized with ventilation," *Nature*, vol. 392, pp. 239-240, 1998.

[94] F. Lopez de Silva, "EEG analisys:Theory and practice; Computer-assisted EEG diagnosis: Pattern recognition techniques," in E. Niedermeyer, F. H. Lopes da Silva, eds. *Electroencephalography: Basic principles, clinical applications and related fields*, pp. 871–919, 1987.

[95] C. A. Sims, "Money, Income and Causality," *American Economic Review*, vol. 62, pp. 540–552, 1972.

[96] W. Singer and C. M. Gray, "Visual feature integration and the temporal correlation hypothesis," *Annu Rev Neurosci*, vol. 18, pp. 555-586, 1995.

[97] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier, "Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in human," *J Neurosci*, vol. 17, pp. 722-734, 1997.

[98] D. Y. Tang, R. Dykstra, M. W. Hamilton, and N. R. Heckenberg, "Experimental evidence of frequency entrainment between coupled chaotic oscillations," *Phys. Rev. E*, vol. 57, no. 3, pp. 3649, 1998.

[99] P. Tass, M. G. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, et al., "Detection of n:m phase locking from noisy data: application to magnetoencephalography," *Phys Rev Lett*, vol. 81, pp. 3291-3294, 1998.

[100] M. Teplan, "Fundamentals of EEG measurement," *Measurement science review*, vol. 2, no. 2, pp. 112-23, 2002.

[101] G. Tononi and G.M. Edelman, "Consciousness and complexity," *Science*, vol. 282, pp. 1846-1851, 1998.

[102] R. D. Traub, M. A. Whittington, E. Buhl, J. G. Jefferys, and H. J. Faulkner, "On the mechanism of the gamma to beta frequency shift in neuronal oscillations induced in rat hippocampal slices by tetanic stimulation," *J Neurosci*, vol. 19, pp. 1088-1105, 1999.

[103] B. van der Pol and J. van der Mark, "The heartbeat considered as a relaxation oscillation, and an electrical model of the heart," *Phil. Mag.*, vol. 6, pp. 763, 1928.

[104] F. J. Varela, "Resonant cell assemblies: A new approach to cognitive functions and neuronal synchrony," *Biol Res*, vol. 28, pp. 81-95, 1995.

[105] F. J. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie, "The brain web: phase synchroniztion and large-scale integration," *Nature Rev Neurosci*, vol. 2, pp. 229-39, 2001.

[106] N. Wiener, "The Theory of Prediction," in E. F. Beckenbach, ed. *Modern Mathematics for the Engineer*, New York: McGraw-Hill, 1956.

[107] A. Wolf, J. B Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponent from a time series," *Physica D*, vol. 16, pp. 285–317, 1985.

[108] K. Yang, and C. Shahabi, "On the Stationarity of Multivariate Time Series for Correlation-Based Data Analysis," in *Proceedings of Fifth IEEE International Conference on Data Mining (ICDM'05)*, pp. 805–808, 2005.

BIOGRAPHICAL SKETCH

Alla Revenko Kammerdiner was born in Kiev, Ukraine. An older of two children, she grew up in Kiev, Ukraine, graduating from School #32 in 1994. She earned her B.S. in Probability Theory and Mathematical Statistics from the National Taras Shevchenko University of Kyiv in 1998.

In January 2001, Alla joined a graduate program in the Mathematics Department at the University of Florida. Upon graduating in May 2004 with her M.S. in mathematics, Alla entered the Ph.D. program in industrial and systems engineering at the University of Florida.

Alla has been happily married to Jason R. Kammerdiner for the last 3 years. On February 24, 2008 she completed her first marathon in 3:53:09.