ON GEOMETRIC & TOPOLOGICAL METHODS FOR ANALYSIS OF BIOPHYSICAL TIME SERIES DATA

BY

IVAN T. ABRAHAM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Associate Professor Mohamed-Ali Belabbas, Co-Chair
Professor Yuliy Baryshnikov, Co-Chair
Professor Fatima Husain
Assistant Professor Zhizhen Zhao

Abstract

Time series analysis is a staple work horse in many fields, including climatology, econometrics, stock and derivatives markets, systems engineering, etc. Traditional analysis of time series data is focused on predicting or forecasting future values based on analysis and modelling of past values. In this work, we present methods of time series analysis that are based on geometric principles. First *cyclicity analysis*, a method of analysis of repeating but aperiodic signals, is introduced and treated in depth. This method is then used to examine two sets of brain imaging data, specifically functional magnetic resonance imaging data under both resting state and task paradigms. We present results that show our ability to fingerprint individuals using their resting state scans, detect slow cortical waves in the brain, and classify between groups in the dataset. Next we apply principles from *geometric diffusion process* in the context of manifold learning to show how synergy detection in eletcromyography data is best viewed as a nonlinear clustering problem rather than a factor analysis problem. We also present simple kinematic examples where linear methods fail to show that nonlinearity is inherent in even the simplest systems. Finally, we conclude this document with a review of results presented, some comments of a historical nature, ongoing trends in the fields that supplied the data, and what can be expected in the near future.

To my grandparents, eternal pillars who always wanted a doctor in the family[1]*, whose support always lives on, tangibly through space and intangibly through time, by way of prayers long forgotten.*

To my wife Sona, whose advent eclipsed any and all other attractors in my orbit; and whose kindness, patience, forbearence and support were a necessary condition for the conclusion of this document.

[1] To be fair, they did mean the other kind.

Acknowledgments

John Donne famously said "No man is an island entire of itself," and this work would not have been possible without support from a multitude of contributors. In no particular order, acknowledgements are due to my advisors, collaborators, colleagues, and friends.

- I am grateful to my advisors and committee members, Dr. Mohamed-Ali Belabbas, Dr. Yuliy Baryshnikov, Dr. Fatima Husain and Dr. Zhizhen Zhao for their patience, thoughtful inputs, advice and valuable mentoring.
- I am also indebted to my co-authors and collaborators, Dr. Somayeh Shahsavarani, Dr. Benjamin Zimmerman, Dr. Sara Schmidt, Dr. Citlali-Lopez Ortiz
- Some of the data analyzed in this work was provided by various collaborators. In particular,
  - Brain imaging data was obtained courtesy of Auditory & Cognitive Neuroscience Laboratory at the Department of Speech & Hearing Science at the University of Illinois, Urbana-Champaign.
  - Electromyography data was recorded by the Neuroscience of Dance in Health and Disability Laboratory at with the Department of Kinesiology at the University of Illinos, Urbana-Champaign.

  This work also used publicly available data from the Connectome Coordination Facility and the associated Human Connectome Project.
- Finally, I am thankful to my colleagues from the Coordinated Science Laboratory at the University of Illinos, Urbana-Champaign for many delightful and insightful conversations during my time there.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial intelligence
AR	Autoregressive
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
BOLD	Blood oxygen level dependent
CM	Correlation matrix
CNN	Convolutional neural network
DA	Discriminant analysis
DTM	Dynamic time warping matrix
DTW	Dynamic time warping
EMG	Electromyography
GAK	Global alignment kernel
GSR	Global signal regression
HCP	Human connectome project
HMM	Hidden Markov model
ICA	Independent component analysis
LCM	Lagged correlation matrix
LDA	Linear discriminant analysis
LM	Lead matrix
LOESS	Locally estimated scatterplot smoothing

LSVM	Linear support vector machine
ML	Machine learning
MNI	Montreal Neurological Institute
MPRAGE	Magnetization prepared rapid gradient-echo
MRI	Magnetic resonance imaging
NARX	Nonlinear autoregressive exogenous
NMF	Nonnegative matrix factorization
PCA	Principal component analysis
PLS	Partial least squares
PLS-DA	Partial least squares discriminant analysis
QDA	Quadratic discriminant analysis
QSVM	Quadratic support vector machine
ROI	Region of interest
SPM	Statistical parametric mapping
STC	Seasonal-trend-cyclic
STL	Seasona-trend decomposition by LOESS
SVM	Support vector machine
TR	Time resolution
VAE	Variational autoencoder
VAR	Vector autoregressive
VARIMA	Vector autoregressive integrated moving average
VARMA	Vector autoregressive moving average
fMRI	Functional magnetic resonance imaging
rs-fMRI	Resting-state functional magnetic resonance imaging

Chapter 1

Introduction

Data or information is all around us in its myriad forms, and we have evolved to constantly process massive amounts of it. For example, at a personal level from our surroundings, we unconsciously detect ambient temperature and humidity, wind speed and direction, objects in our periphery, sources of sounds and movements, etc. At another level, we also keep track of what time of day it is, what weather to expect in the near future, which season of the year it is, and so on. While this definition of what constitutes data and information may be too broad to be useful, there is a subset of such data that will be the main object of interest in this work, in particular data that systematically evolves with time. The analysis and study of such data can be traced all the way back to antiquity - for example, detailed maps of the heavens, including the motions of planets and stars, were charted by many ancient societies, all the way from the Indus valley to Mesoamerica. More recently, the study of *dynamical systems*, which give rise to such data, took off in earnest in the 18th century on the heels of the Englightenment and the advent of the Industrial Age. The theory of dynamical systems, and its numerous applications, has had tremendous success in the last couple of centuries, ushering in successively the machine age, the atomic age, the space age, and, finally now, the information age. In the broadest sense, a dynamical system is simply represented by the following *state space* equation:

$$\dot{x} = f\left(x, t, u\right) \tag{1.1}$$

where the variable x encodes the *state* of the system, which is necessarily dependent on natural time $t \in \mathbb{R}$, via some function f describing the dynamics, and another function u describing inputs into the system. Often, not all the states x will be observable or measurable, and, therefore, we tend to say

that the *output* y is described by:

$$y = h\left(x, t, v\right) \tag{1.2}$$

where the function h determines the manner by which the state x becomes the output y, subject to some external process v. While Eqs. 1.1 and 1.2 look tame and fairly innocuous, they are capable of describing almost any system that thought can conjure, precisely because we have placed no restrictions on what x, u, f, v or h may be. Thus, in a sense, Eqs. 1.1 and 1.2 are also too general to be useful.

For a fairly large class of systems, it is often possible to determine what f and h should be via modeling and experiments (e.g., multi-bar linkages, robots, coupled tanks, etc.) or approximate what they should be via simplifications (e.g., satellites, spacecraft, fighter jets, etc.). The former set tend to be systems where the dynamics can be described in closed form or the dimension of x and y is in the order of tens of states, while in the latter their dimensions may be in the order of thousands, and yet only a score or so physical effects dominate the dynamics. However, there is still another class of systems, often termed *complex systems*, where neither situation is true - the dynamics maybe intrinsically complicated and the dimension of the states also approaches thousands - for example the animal brain, human genomics, dynamics of the folding of proteins, etc. In such systems, it is nearly impossible to approximately describe f and h, let alone know what they should be; nor can they be adequately described by a smaller collection of variables more amenable to analysis. In such cases, what we may be left with are called the so called *time series* of *observables* y, and it is only by study of this quantity we can hope to infer something about the system that generated it.

The majority of this work is concerned with the analysis of collections of such observables y_k that arise from various sources. In the past decade, the ubiquity of data in almost every aspect of life has spawned "the age of Big Data" and, along with it, a proliferation of techniques to facilitate analysis of huge data sets. These advances have not just been in terms of new algorithms but also in terms of programming languages and paradigms, the applied mathematics that underpin them, as well as new platforms and tools that make it possible to analyze the titanic data sets available today. Yet working

with high dimensional data when there is a dearth of samples remains a challenging problem. The 'big' in big-data refers to the vast number of samples or data-points available for the investigator to analyze, that is, the *intrinsic* dimensionality or cardinality of the data set [1]. However, data can also be 'big' in another way - in terms of its *extrinsic* dimensionality, i.e. the number of variables of interest to the investigator far exceeds the number of observations he or she is able to make. Time series data are good examples of such data sets. Consider, for example, the thousand currently largest companies on the New York Stock Exchange and suppose the data set is the daily closing market price of each company's stock considered over a ten-year period. Let the objective of the analysis be to perform unsupervised clustering of the thousand companies - say to discover new or up and coming industry sectors - or to extract information about interrelations between them - say to better understand the market. Then, because each company's market price over the ten year period would be considered a single *sample* in this scenario, we would have a data set of intrinsic dimension a thousand; however, its extrinsic dimension would be approximately two and a half times that. Conversely, one can think of each day's closing price as an observation of a thousand variables. In any case, the sample sizes do not reach anywhere near the hundreds of thousands of samples required for training in contemporary machine learning techniques. Medical imaging datasets and/or biophysical signals are another source where the extrinsic dimensionality is naturally high - indeed a major part of this work focuses on functional Magnetic Resonance Imaging (fMRI) data where it costs roughly a couple of thousand dollars to generate a single data point.

When y has a systemic structure to it, for example periodicity in some of its elements, then tools such as the repertoire of Fourier analysis and related techniques often aid in characterization and study of them. On the other hand, one can examine relationships within the collection absent periodicity using concepts like correlation, auto-correlation, stationarity, cosine similarity, etc. In a similar vein, one of our tools here will be the concept of *cyclicity* [2]. *Cyclicity analysis* - which draws on ideas from differential geometry & topology and has connections with the *signature methods* introduced by Lyons et. al [3] - was presented as a tool to analyze the interrelationships between a collection of *aperiodic* yet repeating signals (see, for example, Figure 2.1) which were not easily amenable to analysis by traditional methods.

We expand this toolkit by way of addition of new algorithms & methods as well as new applications of related concepts. We then, demonstrate their use on real world datasets, specifically functional magnetic resonance imaging (fMRI) data both from healthy control participants and also participants suffering from the tinnitus condition.

While cyclicity and related methods seek to complement correlation and Fourier methods in explaining and finding interdependencies in the data, the classification, decomposition, and/or clustering of time series data is another important, if not parallel, objective in the pursuit of their analysis. Within this context of clustering, some of this work also analyzes electromyography (EMG) data with the help of nonlinear clustering tools to extract and/or delineate periods of *synergies*. A *synergy* can be thought of as the mechanism by which different muscles act in concert to achieve a task while eliminating *motor redundancy* inherent in the control of animal musculoskeletal structures. EMG data has long been used to attack this problem - that is, how does the human central nervous system choose from millions of possibilities for parameters associated with a motor task? More concretely [4],

> "... imagine a pointing task that requires a certain trajectory of a fingertip. How to select a combination of joint angles to produce the required trajectory? How to select muscle forces that would implement the joint rotations? How to select patterns of firing of individual motor units that would produce required muscle forces?"

Simply put, in terms of understanding motor control, there are clearly multiple redundant degrees of freedom, which is somehow naturally accounted for by the central nervous system (CNS). A long held view was one of a multilevel theory of movement control [5, 6, 7], one level of which was "synergy", in which coordinated actions by a set of effectors emerged. In this model, *synergies* are patterns of joint movements and muscle activations that are either hardwired or learned with practice over time, and help eliminate redundant degrees of freedom. Our objective in the second part of this work is to understand how one can relate the muscle activation potentials observed as EMG signals with the forces and moments that extensions or contractions of the muscle generates. We show that synergy detection and/or analysis is more appropriately viewed as a nonlinear clustering problem rather than the

signal decomposition problem commonly seen in the literature.

In light of the above, the rest of this document is structured as follows: In Chapter 2 we will review some classical methods of time series analysis and the background material required to introduce the methods used in this work. Chapter 3 will contain the bulk of exposition involving the mathematical treatment of our techniques. In Chapters 4 and 5, we will introduce and analyze the problems chosen for study by the application of methods in Chapter 3. Specifically, functional magnetic resonance imaging (fMRI) data and application of cyclicity analysis to it will be studied in Chapter 4. In the penultimate chapter we will examine the application of nonlinear clustering methods to electromyography (EMG) data. Finally, the last chapter will conclude this document with a cursory review of the history & context, a reiteration of open questions, and directions for future work.

Chapter 2

Preliminaries

Given the ubiquitousness of time series data, innumerable methods and techniques of analyzing them have been developed. Broadly speaking, traditional time series analysis methods can be chiefly divided into two camps - *time domain* methods and *spectral* methods. Within the former camp, the objective is frequently to perform forecasting, regression, modeling, etc., whereas in the latter case the objective is often to extract summary statistics and/or perform decompositions. While not meant to be exhaustive in any fashion, we briefly review some popular methods in the following.

2.1 A short review of time series analysis

2.1.1 Time domain methods

When the objective is to perform forecasting, modeling, or regression as is common in the fields of econometrics, statistics, meteorology, finance, etc., the methods most commonly in use tend to be based in the time domain. Potentially, the simplest such model arises when the time series' current value is hypothesized to depend on its past values in some fashion - the so called *autoregressive* (AR) model [8]. When the current value of a variable y depends on p past values, we write it as the AR(p) model

$$y_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t \tag{2.1}$$

where the subscript in y_t denotes that the value of the time series y at time t is under consideration, c is a constant, φ_i are model parameters, and ϵ_t is (often white) noise. Given y_i, various methods of estimating the φ_i exist, however the general sources of uncertainty in the model are: applicability of

the model to the given time series y, the correct value of p to be used, the noise term, and, paramountly, the uncertainty in autoregressive coefficients φ_i.

Using the AR model, we can also construct forecasts $\hat{y}_t$ of the process once the parameters φ_i have been estimated. The residuals $\epsilon_t := \hat{y}_t - y_t$ form their own time series, possibly with some structure. Applying the same autoregressive logic to the residuals, we can construct a model for the residual errors and predict the expected error. These error predictions can then be subtracted from the model prediction to provide an additional performance gain. This results in the very popular *autoregressive moving average* (ARMA) model. Here, we have that ARMA(p, q) is defined as:

$$y_t = c + \epsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{2.2}$$

where φ_i and θ_i are now the model parameters and q is the number of residuals used. Introducing differencing to account for nonstationarity results in the *autoregressive integrated moving average* (ARIMA) model, often written as ARIMA(p, d, q), where d now stands for the order of differencing applied. These models can also be extended to multidimensional times series resulting in *vector autoregressive* (VAR) models and their siblings VARMA and VARIMA models.

Other classical methods in the time-domain include *Hidden Markov Models* (HMM) and *nonlinear autoregressive exogenous* (NARX) models. NARX models are essentially derived from Eq. 1.1 and Eq. 1.2 by dropping the former and adopting $x := y$ in the latter allowing the equation to take the form [9],

$$y_t = h\left(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, v_t, v_{t-1}, v_{t-2}, v_{t-3}, \right) + \epsilon_t$$

where h may now represent a neural network, polynomials and/or other nonlinear functions. In contrast, HMMs do not drop the state x but rather "hide" it while constraining its relationship to y such that the output y_{t_0} only depends on the hidden state x_{t_0} at time t_0 and not any other past hidden states or outputs $(x_{t<t_0}, y_{t<t_0})$.

2.1.2 Spectral methods and decompositions

In contrast to regression and forecasting, the objective of most spectral analysis methods is usually to obtain decompositions of the given time series into simpler constituent components. For example, *Fourier analysis* aims to uncover the constituent frequencies present in a periodic signal by way of representing it as a sum of trigonometric functions; *principal component analysis* (PCA) applied to multivariate or multidimensional time series finds representations that encode most of the variability in the data with a few components, and so on. We call these *spectral* techniques because often at the heart of the method lies an eigenanalysis problem. For example, the PCA decomposition is obtained by examining the spectrum of the centered covariance matrix of the data, the Fourier components describe a periodic function in terms of an eigenbasis of complex exponentials and so on. Numerous algorithms and techniques fall under the grand umbrella of "spectral methods," so much so that to describe in detail even a handful of them would be beyond our scope. Nevertheless, the ideas applied and/or examined in this work frequently make comparison to the popular spectral methods which can be grouped together generically as factor-analysis methods which include non-negative matrix factorization (NMF), independent component analysis (ICA) and so on.

Yet another popular method which deserves a passing mention, the *seasonal-trend-cycle* (STC[1]) decomposes a time series into *seasonal*, *trend*, *noise*, and *cyclic* components, either additively or multiplicatively. That is, the time series is decomposed as

$$y_t = T_t + C_t + S_t + I_t \qquad \text{or} \qquad y_t = T_t \times C_t \times S_t \times I_t$$

where T_t stands for trend, C_t for cyclical, S_t for seasonal components and I_t denotes any residuals. In the STL method the trend and cyclical components are grouped into one [10]. While not strictly a spectral method, the decomposition that this technique yields is imminently interpretable and easy to understand.

[1]Also often called STL for *Seasonal-Trend decomposition using LOESS*[10] where LOESS itself is an acronym for *locally estimated scatterplot smoothing*.

2.1.3 Clustering and similarity measures

Classification or clustering of samples over time, based on their own past behaviors - both *auto* and *inter* relations - may be important for many purposes including inferring dynamics between them, detecting anomalies or outliers, gaining insights into their driving processes, etc. For an example, huge organizations may keep track of their external and intranet servers by monitoring relevant statistics over time. Any anomalies detected from such time series data could point to an imminent failure or an ongoing malicious attack. In such cases extracting or generating *features* that summarize statistics, or capture interpretable content from the multivariate time series, constitute a popular methodology for overcoming high extrinsic dimensionality. However, it may still be the case that the generated features themselves have high enough extrinsic dimensionality that the investigator may have to resort to further *dimension reduction* methods. Conversely, in the setting of machine learning, one may seek to *augment* the data set to increase the intrinsic dimension to be on par with the extrinsic dimension. For example, in image recognition or classification, flipping, rotating, or skewing the image is a common practice that can add additional data points to the set without changing its essential nature.

Clustering refers to techniques that aim to aggregate data that are 'similar' in nature together. Often, the metric used to define similarity is a parameter left to the choice of the user. Such techniques may be used to group time series behaving in a similar fashion together or may be used to group together periods of essentially the same activity within a single time series. Here, popular methods include the ubiquitous k-means adapted for time series (by way of an appropriate choice of similarity metric), kernel k-means, k-mediods, hierarchical and density based clustering to name a few.

Each technique presents its own advantages and disadvantages. For example, the Euclidean $k-$means algorithm requires the user to estimate a priori k, the number of clusters present in the data. Then it proceeds to initialize k centroids within the dataset and iteratively assigns data points so as to minimize within-cluster variance for each of the k clusters. More formally, the objective is to find amongst all k subsets $S = \{S_1, S_2, \ldots, S_k\}$ of the

data,

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 \tag{2.3}$$

where μ_i is the mean of the data points in S_i. Therefore, the naive algorithm at each step performs assignments, as mentioned above, and recalculates the cluster centroids. The iterations are said to be convergent when no more updates are required (or effect any change to the minimization). This algorithm is neither particularly fast nor guaranteed to find the optimum with local minima heavily influenced by choice of initial centroids. Kernel k-means extends k-means via the kernel trick to implicitly perform clustering in the embedding space associated with the chosen kernel. Kernel k-means is closely related to the spectral clustering methods [11], which is the main tool used in Chapter 5. For time series data, the norm in Eq. 2.3 is often replaced with an appropriate metric or similarity measure, e.g. the Dynamic Time Warping (DTW) distance for vanilla k-means or Global Alignment Kernel (GAK) for kernel k-means.

Recall that DTW method was introduced in the field of speech recognition to account for differences in the variation (in intonation, speed, timbre, etc.) of the same spoken content by different speakers [12]. Therefore, it was constructed to allow for a notion of "nearness" or distance between time series of unequal lengths, in contrast to the usual one where vectors are restricted to be of the same size. To wit, consider two given time series of not necessarily equal length, $X = (x_1, x_2, \ldots, x_k, \ldots, x_M)$ and $Y = (y_1, y_2, \ldots, y_j, \ldots, y_N)$, with the objective of finding the optimal alignment between the two. Here alignment means allowing for non-linear but monotonic warping of the time axis so that one series maybe transformed into the other. Specifically, an alignment between series of length M and N corresponds to a path through an $M \times N$ matrix that connects the $(1, 1)$ entry with the (M, N) entry using just $\rightarrow, \downarrow$ and $\searrow$ motions. Then the DTW algorithm begins by constructing a matrix C of size M by N where each element C_{kj} of the matrix corresponds to the optimal cost of the alignment between $X_k = (x_1, x_2, \ldots, x_k)$ and $Y_j = (y_1, y_2, \ldots, y_j)$, i.e., the first k points of X and the first j points of Y. This optimal cost is obtained as

$$C_{kj} = d_{kj} + \min(C_{(k-1)j}, C_{(k-1)(j-1)}, C_{k(j-1)}), \tag{2.4}$$

where d_{kj} is a distance metric, such as the Euclidean distance (with the specific choice being left to the practitioner) between the k-th and j-th symbols in the series. While a popular method in the analysis of time series data, the DTW distance is not a true metric, as it does not satisfy the triangle inequality nor is it amenable for use in the definition of a positive definite kernel. The global kernel alignment algorithm attempts to address this issue by changing the minimum in Eq. 2.4 with a soft-minimum so as to be able to define a kernel function appropriately [13, 14]. Many other similarity measures exist in literature, with survey papers describing them found as far back as forty-five years ago [15], as well as more recently [8, 16].

With this preliminary review of classical time series behind us, let us now proceed to describe the central problems of interest in this work.

2.2 Problem statement(s)

The main object of our analysis and study are multidimensional time series data. By dimension here, we mean that the dataset of interest, recorded over some $[0, T]$ interval or consisting of T steps, is comprised of multiple *channels* or *coordinates*. For example, in the case of fMRI recordings, this would be multiple regions of interest (ROIs) that provide changes in Blood Oxygen Level Dependent (BOLD) signal, or in the case of EMG data, this may correspond to different muscles reporting activity via various electrodes attached to them. The defining characteristic of the kinds of time series analyzed here is that they frequently exhibit repetition without a period. In this work, this is called *cyclic* but aperiodic behavior, an example of which is shown in Figure 2.1. The image of the collection of signals shown is constrained to be in $I := [-2, 2]$ for all t, i.e. they cycle through the values in I, albeit with no definite period. At this juncture, it is crucial to draw the distinction between periodic signals and cyclic signals to the forefront. A signal $x(t)$ is said to be periodic with period P if there exists a $P \in \mathbb{R}_+$ such that $x(t + P) = x(t)$. For an aperiodic signal there is no single P that can be found to be termed its period. A signal $y(t)$ is said to be *cyclic* if there exists a bounded monotonic function $\phi : t \to \phi(t)$ so that $y(\phi(t))$ is periodic. Such signals arise frequently in bio-physical contexts, for example the cardiac and respiratory rhythms which change frequently throughout the

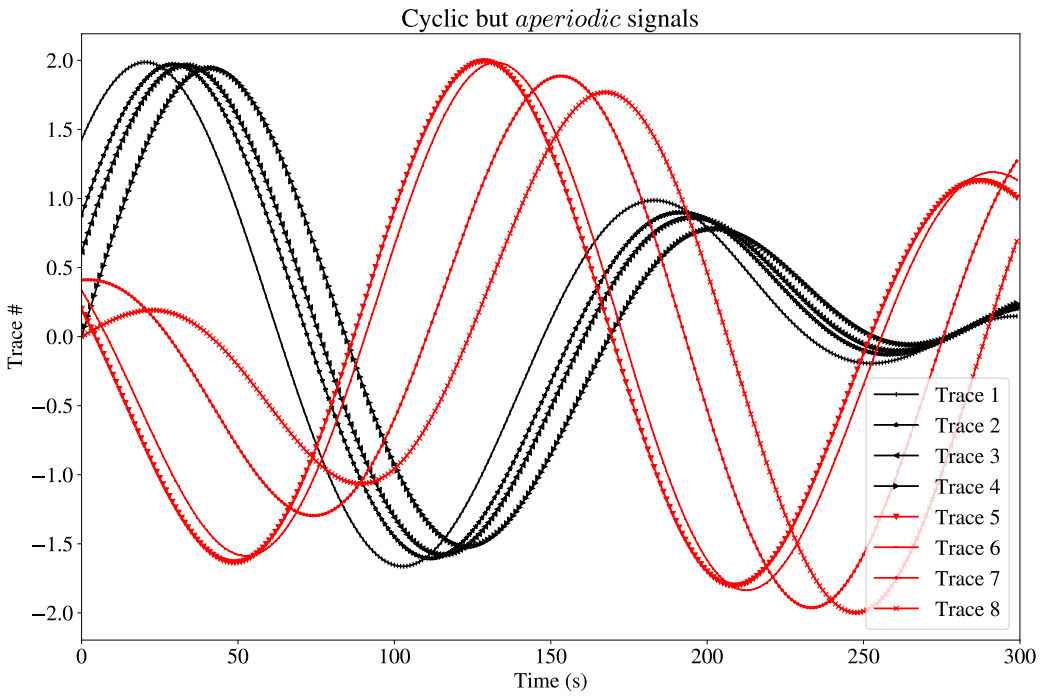


Figure 2.1: Signals that exhibit repeating but aperiodic behavior, termed *cyclic* that still exhibit structure. The order within the black sub-collections is $| \star \triangleright \triangleleft$ and that within the red sub-collection is $\square - \cdot \times$

day depending on various factors like arousal, rest, medication, exertion, etc. A function like $\phi(t)$ is essentially a re-parametrization of time; therefore, one interpretation of cyclic signals is to picture some internal processes with an unknown *private clock* and *internal state space* giving rise to the observable y via the function ϕ^{-1} (provided ϕ is strictly monotonic).

Yet within Figure 2.1, we see that the signals are not all random; they do exhibit a certain structure. Namely, we see that there are two sub-collections of signals, the red and black ones with the latter being a more cohesive bunch compared to the other. Even within the red sub-collection, it is clear that two of the signals coalesce around each other more than the others. In terms of order of appearance, we can see the black signals follow each other in the order: $| \star \triangleright \triangleleft$. Similarly, in the red sub-collection we have the order: $\square - \cdot \times$. Moreover, the structure just presented is preserved under monotonic transformations $\phi : t \to \phi(t)$. Then, naturally a few questions arise:

- Is it possible to recover the structure elucidated above from such a collection of signals?
- Is it possible to recover the fact there are two sub-collections in the

signal?

- Is this recovery preserved under reparametrization of time?

In certain cases, the analysis pipeline outlined in Chapter 3 will be able to answer the above questions affirmatively. In other cases, we may have to resort to different techniques. Before we proceed to describe our methods, we take a quick detour to present what questions we sought to answer with each of the datasets in use. Recall that the datasets used in this work are functional Magnetic Resonance Imaging (fMRI) data and electromyography (EMG) data.

2.2.1 Functional MRI datasets

We begin by briefly reviewing how the fMRI time series analyzed in this work are generated. Recall that at the heart of the fMRI machine is a Magnetic Resonance Imaging (MRI) scanner. MRI imaging works on the principle of nuclear magnetic resonance, namely, it utilizes the phenomenon that precessing atoms in a strong external magnetic field, give out characteristic energy signatures when returning to equilibrium. The said return being caused by intentional perturbation from equilibrium via radio-frequency (RF) pulses. These signatures can be studied and analyzed to glean information about the underlying tissue and body structure. Functional MRI of the brain, then, works by focusing on molecules within blood itself. Specifically, the hypothesis is that activity in any region of the brain requires energy, which necessitates the increased flow of oxygen rich blood to the said region. Since oxygen rich blood has an increased concentration of oxygenated hemoglobin (diamagentic), its characteristic signature can be mapped distinctly from that of deoxygenated hemoglobin (paramagnetic). This gives rise to what is called the Blood Oxygen Level Dependent (BOLD) signal. The fMRI time series studied in this work are all traces of various BOLD signals from different parts of the brain as recorded from individuals in a *resting* or *task modulated* state. An example of this BOLD signal is visualized in the left panel of Figure 2.2.
We examine two main sources of fMRI data in this work.

1. The first is fMRI data that was collected at the University of Illinois, Urbana-Champaign. In particular, data was recorded at the Auditory

& Cognitive Neuroscience Laboratory directed by Dr. Fatima Husain. This data was primarily collected in the resting-state and was done so with the objective of studying the neurological underpinnings of tinnitus as detailed in Chapter 4. The exact details of the fMRI recording pipeline is presented along with demographic information in Section 4.1.1.

2. The second is openly available fMRI data that was collected at the Washington University at St. Louis as part of the Human Connectome Project under the aegis of the Connectome Coordination Facility. In this dataset, fMRI scans were performed in both the resting state and task modulated paradigms. Whereas the previous cohort of subjects was a mixture of tinnitus afflicted and healthy individuals, all the data in the HCP dataset comprised of healthy individuals between the ages of eighteen and forty. More details regarding the HCP dataset can be found in Section 4.3 and in relevant literature [17, 18].

With the former dataset, we sought to utilize fMRI data for better understanding the neural underpinnings of *tinnitus* [19]. Tinnitus is a common neurological disorder that is characterized by a persistent perception of phantom noise often described as a ringing or buzzing sound. The tinnitus is typically associated with the presence of hearing loss and reportedly affects 1 in 6 individuals at some point of their lives. While some people habituate quite effectively with the presence of tinnitus, for many others it is a debilitating condition that can seriously affect their quality of life. The ringing or buzzing noise is characterized as phantom because no physical stimulus is actually present in the ambient environment. Analysis of the fMRI data collected at the Auditory & Cognitive Neuroscience Laboratory and presented in [19] further lead to [20] where we compared our proposed methods with other methods popular in the field. The overarching objective with this dataset was to gain insights into the brain processes, inter-region connectivity, and differences if any in the *functional* connectivity of the brain that could address the presence of tinnitus in the subject population. Such findings would allow for the development of diagnostic tools and/or intervention strategies for the management of the tinnitus condition.

In analysis of the dataset from Washington University at St. Loius, namely the Human Connectome Project (HCP) dataset, we sought to extend the

cyclicity analysis toolkit [21]. In particular, we showed how cyclicity analysis equipped with the so-called *chain of offsets model* can be used to recover an underlying network structure. For HCP data under the task paradigm, we showed that we are able to detect periods of directed activity between pairs of brain regions signifying the progression of a wave of activity.

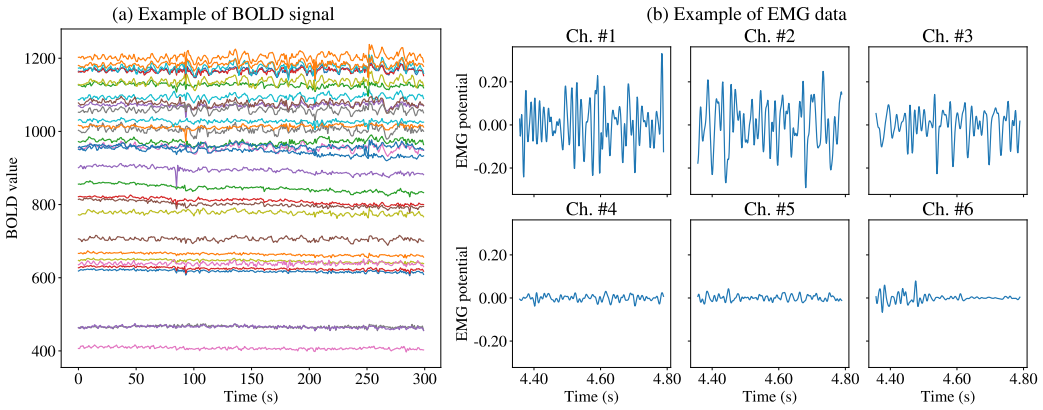


Figure 2.2: This figure visualizes exemplar data studied in this work. (a) The Blood Oxygen Level Dependent (BOLD) signal from 32 regions-of-interest (ROIs) is shown. (b) Here six channel EMG data sampled at 1000 Hz is shown for the duration of a sub-second task recording.

2.2.2 Electromyography data

The second source of biophysical time series data that is studied in this work is electromyograph (EMG) readings obtained from healthy individuals as they performed various tasks. EMG readings can be obtained from an individual in three principal ways:

1. **Intramuscular EMG**: Here EMG readings are obtained from the muscles directly by implanting small needle electrodes directly into the musculature. This is the oldest method of EMG recording and is still used in many applications where high degree of localization and selectivity is required (i.e. where individual motor unit activity needs to be recorded).
2. **Surface EMG**: In this method, EMG readings are obtained from the surface of the muscle via the skin, thereby covering a larger portion of the musculature and more muscle fibers. This method's distinct advan-

tage is its non-invasiveness, albeit at the cost of more signal processing to clean up the (possibly more) noisy signal and loss of selectivity.

3. **Multi-array EMG**: In this latest method, a multitude of electrodes are utilized in an arrayed fashion on the surface of the skin to provide localized information about the *depth* at which the muscle is activated apart from intensity of activation itself.

The dataset utilized in this work is of the second kind listed above, namely surface EMG recordings. The overarching aim with this latter dataset was to try and address the problem of *motor redundancy* by way of nonlinear techniques. In this context, i.e. within kinesiology, the *motor redundancy* problem is one of providing a mechanism by which the central nervous system (CNS) is able to coalesce the multitude of options available to it. For concreteness, it is not hard to see that a simple action like tracing an arc with one's finger on a blackboard involves multiple motor units, the hundreds of muscle fibers that comprise them, the tens of joints that link the musculature together, and so on. Thus, there are innumerably many degrees of freedom available to the CNS if one were to count from the level of individual muscle fibers, joint angles, etc. Yet, somehow the CNS is able to choose from amongst them reliably and repeatedly without fail. In addressing the problem of redundancy in motor control, researchers seek to provide a plausible mechanism by which the same is accomplished. With our work on the dataset collected at the Neuroscience of Dance in Health, Disease and Disability Laboratory under the supervision of Dr. Citlali Lopez-Ortiz, we show how nonlinear clustering techniques provide one way address this question. In particular, we are able to cluster the multichannel EMG data with the clusters being demarcated by changes in the force/moment generation in the task - data which is in fact absent as input into the clustering pipeline!

With this brief summary of the types of datasets that will be used in this work, we will now move on to the next chapter, where we will introduce the methods we will use to study the data.

Chapter 3

Methods

Now we turn to the task of describing the methods used in the analysis promised and previewed so far. We utilized methods related to *cyclicity analysis*, *nonlinear clustering*, and *machine learning techniques* along with some methods described in the previous section. We start with cyclicity analysis.

3.1 Cyclicity analysis

To describe the crux of our method in detail, let us formalize a few definitions we alluded to in Section 2.2.

Definition 3.1 (Path)**.** *is an alias for a multidimensional signal, i.e. a* $d-$*dimensional* ***path*** X *is simply a continuous mapping from a subset of the real line to the vector space* $\mathbb{R}^d$*,* $X : [s,t] \subset \mathbb{R} \to \mathbb{R}^d$*.*

The terms trajectory, path and signal are used interchangeably in this document and are all multidimensional unless specified in the context.

Definition 3.2 (Reparametrization)**.** *A* ***reparametrization*** *of a path over an interval* $[s,t]$ *is a change of variable* $X(t) \to X(\phi(t))$ *such that* $X(s) = X(\phi(s)), X(t) = X(\phi(t))$ *and* $\phi(\cdot)$ *is continuous, bounded and increasing over* $[s,t]$*.*

Note that some authors use the alternate spelling *reparameterization*. Recall that we said a periodic signal is necessarily cyclic but the converse need not be true. Indeed, we have that a signal or time series is periodic if shifting it in time ahead by a value P keeps the series unchanged. The smallest such P is called its *period*. On the other hand, a cyclic signal is one that repeats itself identically over and over again but perhaps with a variable speed. More precisely,

Definition 3.3 (Cyclic)**.** *A signal $g(t)$, is considered* ***cyclic*** *if there is a monotonically increasing bounded function, $\phi(t)$, so that $g(\phi(t))$ is periodic.*

Now to understand why we care about reparametrizations, let us consider a pulse and its shifts in time as shown in the left panel of Figure 3.1. In this figure, a Gaussian pulse and its progressive time shifts are shown so that the first (blue) and last (deep red) pulses have virtually no overlap. On the right panel in the same figure, we show the parametric plots of the first pulse against the rest of the shifted pulses. Here, the blue fixed pulse is plotted on the abscissa while the different red pulses are plotted on the ordinate.

Naturally, we see that when the shift $S = 0$ the parametric plot essentially becomes the straight line $y = x$ (c.f. the top row in the right panel of Figure 3.1 and imagine possibly, a Plot 0). Conversely, when there is no overlap at all, the parametric plots collapse onto the cardinal axes as shown tending to in Plot 9. However, in the middle, the parametric plots enclose areas of the plane by way of loops. It follows then that in the case of pair of sinusoids time shifted by S, the enclosed area would take on extreme values when the shift is a quarter period or three quarters of a period and a value of zero when the shift is a half or full period.

Given two time series x and y defined on $[a, b]$ such that $x(a) = x(b)$ and $y(a) = y(b)$, we can calculate the above mentioned signed area $A(x, y)$ enclosed by their parametric plot as

$$A(x, y) = \frac{1}{2} \cdot \int_a^b x(t) \cdot y'(t) - y(t) \cdot x'(t) dt \tag{3.1}$$

by way of Green's Theorem. Note that the above quantity is *reparametrization invariant* with respect to monotonic transformations $t \to \phi(t)$. In other words, if we were to think of a transformation $t \to \phi(t)$ as a warping of the time axis, it is clear that regardless of how quickly we traverse the image of $x(t)$ and $y(t)$, the curve in the $x - y$ plane remains unchanged and consequently, so should the area enclosed by it. This is easy to verify. Let $A(x, y)$

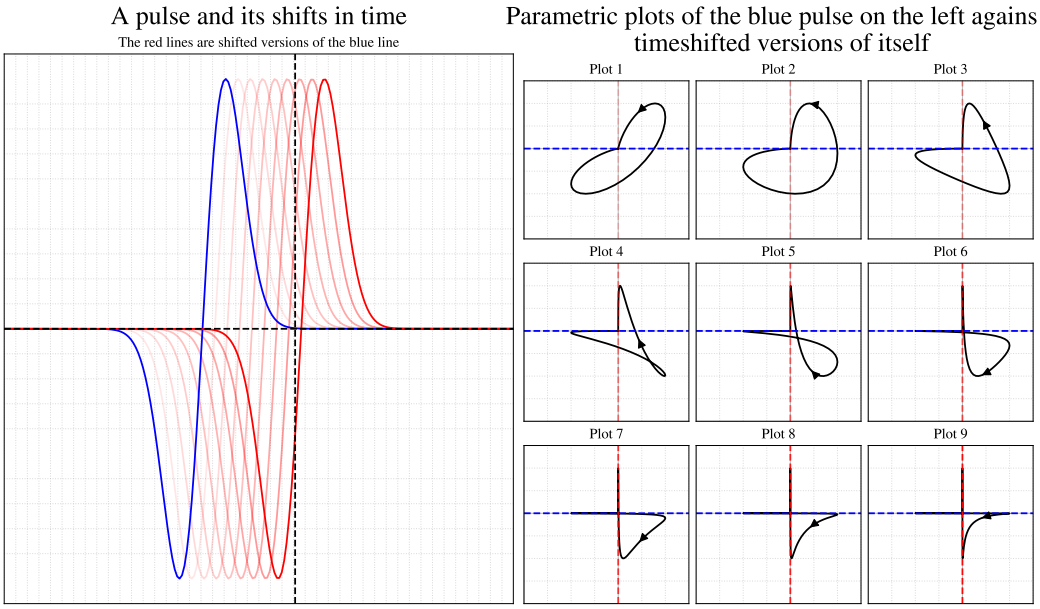


Figure 3.1: A pulse and its time shifts are shown on the left panel. The right panel shows parametric plots of the first (blue) pulse against its time shifts. We see that the parametric plots each enclose certain areas by forming loops. These areas are interpreted as signed areas because the parametric curve is considered to be oriented. A loop traversed in the counterclockwise direction contributes a positive area. We see that when the time shift $S = 0$ the parametric plot becomes a segment of $y = x$ and when $|S| >> 0$ the parametric plot collapses onto the cardinal axes.

be as defined in Eq. 3.1 and consider:

$$\begin{aligned} A\left(x\left(\phi(t)\right), y\left(\phi(t)\right)\right) &= \frac{1}{2}\int_a^b \left(x\left(\phi(t)\right) \cdot y'\left(\phi(t)\right)\phi'(t)\right) \\ &\quad - \left(y\left(\phi(t)\right) \cdot x'\left(\phi(t)\right)\phi'(t)\right) dt \\ &= \frac{1}{2}\int_a^b \left(x\left(\phi(t)\right) \cdot y'\left(\phi(t)\right) - y\left(\phi(t)\right) x'\left(\phi(t)\right)\right)\phi'(t) dt \end{aligned}$$

Let $z := \phi(t)$ so that $dz = \phi'(t)dt$. Then,

$$\begin{aligned} A\left(x\left(\phi(t)\right), y\left(\phi(t)\right)\right) &= \frac{1}{2}\int_a^b \left(x\left(z\right) \cdot y'\left(z\right) - y\left(z\right) x'\left(z\right)\right) dz \\ &=: A\left(x\left(z\right), y\left(z\right)\right) \end{aligned}$$

Moreover, we can in fact consider the enclosed areas to be signed areas, because the progression of time in the left plot of Figure 3.1 accords a natural orientation to the parametric curves in the 2-coordinate planes of the right panel of plots. By convention, a loop traversed in the counterclockwise direction is said to contribute a positively signed area. The determination of quantities related to multidimensional trajectories

$$X(t) = \left(x^1(t), x^2(t), \ldots, x^d(t)\right) \in \mathbb{R}^d$$

which remained invariant with respect to reparametrizations was studied by Kuo-Tsai Chen in the 1950s [22]. In particular, Chen established that it was their iterated integrals which comprised the reparametrization invariants of such signals. More concretely, let $X_t = X\left(t\right)$ be a d-dimensional path in $\mathbb{R}^d$ defined over an interval $[s, t]$. For any such path it is possible to define its n-th iterated integral [23] as,

$$\mathbf{X}_{s,t}^n = \int\limits_{s<u_1<\ldots<u_n<t} dX_{u_1} \otimes dX_{u_2} \otimes \ldots \otimes dX_{u_n} \tag{3.2}$$

where $\otimes$ denotes the tensor product[1]. Note that u_i in the above are simply integration variables. For example the first order iterated integral is simply

[1]For elements of a real valued finite dimensional vector space, this is simply the outer product.

the increment of the path over $[s,t]$:

$$\mathbf{X}^1_{s,t} = \int\limits_{s<u_1<t} dX_{u_1} = \int\limits_s^t dX\left(u_1\right) du_1 = X_t - X_s$$

where in the last equality we have dropped the argument to subscripts. Similarly the second order iterated integral is:

$$\begin{aligned}\mathbf{X}^2_{s,t} = \int\limits_{s<u_1<u_2<t} dX_{u_1} \otimes dX_{u_2} &= \int\limits_s^t \int\limits_s^{u_2} dX_{u_1} \otimes dX_{u_2} \\ &= \int\limits_s^t \mathbf{X}^1_{s,u_2} \otimes dX_{u_2}\end{aligned}$$

Example 3.1. As an example consider the path:

$$X_t := [0, 2\pi] \mapsto \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} \subset \mathbb{R}^2 \qquad \Rightarrow \qquad dX_t = \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}$$

Then we can compute,

$$\mathbf{X}^1_{s,t} = X_t - X_s = \begin{bmatrix} \cos t - \cos s \\ \sin t - \sin s \end{bmatrix}$$

and,

$$\mathbf{X}^2_{s,t} = \int\limits_s^t \int\limits_s^{u_2} \begin{bmatrix} \sin u_1 \sin u_2 & -\sin u_1 \cos u_2 \\ -\cos u_1 \sin u_2 & \cos u_1 \cos u_2 \end{bmatrix} du_1 du_2$$

where we have made use of the fact that $\otimes$ for finite dimensional vectors is the outer product. □

In general, Eq. 3.2 defines an n dimensional tensor so that an iterated integral of n-th order captures interrelationships between n-tuples of elements of X. By defining the zeroth order iterated integral to be unity, i.e. $\mathbf{X}^0_{s,t} := 1$, it is possible to group iterated integrals of all orders $n > 0$ together into a

single infinite dimensional entity called the *signature* of a path [24].

$$\mathbb{S}\left(\mathbf{X}_{\mathbf{s},\mathbf{t}}\right) := \left(1, \mathbf{X}^1_{s,t}, \mathbf{X}^2_{s,t}, \ldots\right) \tag{3.3}$$
$$\mathbf{X}^{n+1}_{s,t} := \int_s^t \mathbf{X}^n_{s,u} \otimes dX_u, \qquad n \geq 0$$

The *signature* of a path encodes many of its algebraic and geometric properties and can be shown to be invariant to translations and reparametrizations [25, 24]. The *signature method* promoted by the Lyons' school utilizes truncations of the signature as features in different machine learning applications. However, here we will focus on the second order component of the signature, in particular its antisymmetric part, i.e. *oriented areas*. When $n = 2$, we see that the antisymmetric part of Eq. 3.2 can be written as a matrix

$$\hat{\mathbf{A}} = \frac{1}{2} \cdot \int_{s<u_1<u_2<t} dX_{u_1} \otimes dX_{u_2} - dX_{u_2} \otimes dX_{u_1}$$
$$\Leftrightarrow \hat{A}^{ij} = \frac{1}{2} \cdot \int_{0<u_1<u_2<t} dx^i_{u_1} dx^j_{u_2} - dx^j_{u_1} dx^i_{u_2} \tag{3.4}$$

where each (ij)-element then captures signed area corresponding to the pair of curves (x^i, x^j).

The matrix in Eq. 3.4 is intimately related to the *lead matrix* of cyclicity analysis from [2]. Each (ij)-entry of the lead matrix is proportional to the sum of areas enclosed by the corresponding (ij)-pair of signals in their respective 2-coordinate planes weighted by their *winding numbers.* Recall that a *winding number* is the number of times a curve is traversed in either clockwise or counterclockwise around a point of interest. Figure 3.2 suffices to illustrate this point, and the reader may refer to the Appendix for a primer on the topic. In Figure 3.2 we show how the algebraic area for a pair of curves is defined. In the left panel, we see that the areas enclosed by the curves form loops which are assigned a sign depending on the orientation of traversal. Plot 1 has a positive area of π since counterclockwise traversal is by convention taken to contribute a positive area. Plot 2, however, has a net zero algebraic area because the two leaves in the figure contribute equal areas of opposite signs. The right loop is completed in a counterclockwise

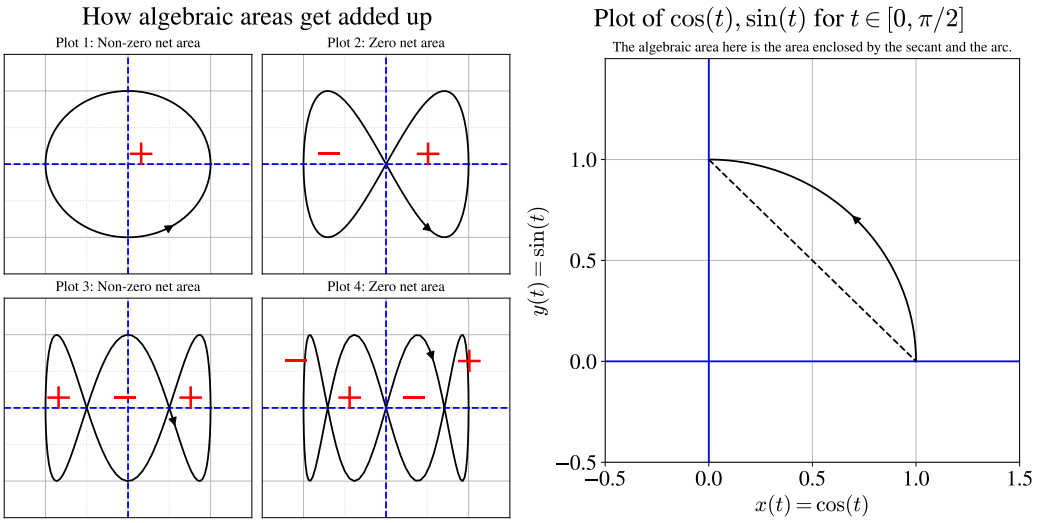


Figure 3.2: The algebraic area is the winding number weighted sum of the areas contributed by the loops. In the left panel, only the leftmost column of curves has a nonzero net area whereas for the right column, the net area is zero due to cancellations. These cancellations arise due the opposing nature of signs of the winding number for each of the *loops/leaves*. In the right panel we show how to interpret the algebraic area when the curve in the parametric plot does not form a closed loop. In this case, the area given by Eq. 3.4 for $i \neq j$ is the signed area between the curve and a chord connecting its two end points - a quantity called Levy area in some publications [26]. On the other hand Eq. 3.1 would return the area beneath the curve, i.e. area of the quarter sector of the unit circle.

fashion (hence a winding number of +1 around any point inside it) while the left loop is completed in a clockwise fashion (hence a winding number of −1 for points in the domain it encloses). Similarly, the algebraic area in Plot 3 is nonzero while that of Plot 4 is zero. The right panel shows how the area is interpreted when the pair of time series or trajectories do not result in closed parametric curves. In this case, the algebraic area - also called Levy area [26] - is the area enclosed by the curve and the chord connecting its end points. Indeed, one can verify that the pair $(\cos(t), \sin(t))$ generates the figure for $t \in [0, \pi/2]$ and that the area involved can be calculated from simple geometric intuition as:

$$\text{sector} - \text{triangle} = \frac{\pi}{4} - \frac{1}{2} = \frac{\pi - 2}{4}$$

whereas from Eq. 3.1 we see that the area A is:

$$\begin{aligned} A = \frac{1}{2} \cdot \int_0^{\pi/2} \cos(t) \left(\sin(t)\right)' - \sin(t) \left(\cos(t)\right)' dt &= \frac{1}{2} \cdot \int_0^{\pi/2} \cos^2(t) + \sin^2(t) dt \\ &= \pi/4 \end{aligned}$$

This illustrates the difference between the $i \neq j$ elements of $\hat{A}$ in Eq. 3.4 and area of Eq. 3.1. The former would read here as:

$$\begin{aligned} \hat{A}_{ij} &= A + \frac{1}{2} \sin(0) \left[\cos(\pi/2) - \cos(0)\right] - \frac{1}{2} \cos(0) \left[\sin(\pi/2) - \sin(0)\right] \\ &= A - \frac{1}{2} = \frac{\pi - 2}{4} \end{aligned}$$

since for a pair of trajectories (X, Y), the quantity in Eq. 3.4 is

$$\underbrace{\frac{1}{2} \int_s^t \left(X_u dY_u - Y_u dX_u\right) du}_{:=A} + \frac{1}{2} Y_s \left(X_t - X_s\right) - \frac{1}{2} X_s \left(Y_t - Y_s\right)$$

This motivates the following definition for the lead matrix.

Definition 3.4. *Given a collection of time series $X = (x_1(t), x_2(t), \ldots, x_n(t))$ with $t \in [a, b]$ we define the lead matrix M via Eq. 3.1 as*

$$M_{ij}(\tilde{X}) := -A\left(\tilde{x}_i, \tilde{x}_j\right) \tag{3.5}$$

where $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ *is the collection* X *linearly adjusted so that* $x_k(a) = x_k(b)$ *for all* k*.*

Note that the negative sign is present in this definition to be consistent with [2]. The primary use of the matrix M is to automatically extract *cyclic ordering* from the collection X via its eigen-analysis. To see why this would be possible, consider the matrix M when it arises from a collection X of single harmonics, $X_k = \sin(\alpha t + \phi_k), k = 1, \ldots, n$. In this case the lead matrix A defined via Eq. 3.5 by considering all pairs (x_i, x_j) is a rank 2 skew-symmetric $n \times n$. It is given by,

$$\left.\begin{aligned} a_{kl} &= C \sin(\phi_k - \phi_l) \\ a_{kl} &= -a_{lk} \end{aligned}\right\} \text{ for } \quad k, l = 1, \ldots, n, \quad k \leq l$$

where C is a constant function of α and the period of integration. Now consider the following result stated by Baryshnikov in [21] for which we offer a proof.

Lemma 3.1. *If a real skew-symmetric rank 2 operator* A *admits a decomposition* $A = z \otimes y - y \otimes z$ *where* z, y *are linearly independent then* (v, λ) *is an eigenvector-eigenvalue pair where*

$$v = -\frac{e^{-i\theta} z}{|z|} + \frac{y}{|y|}, \qquad \lambda = i \sin\theta \, |z| \, |y| \,, \qquad \theta = \arccos\left(\frac{z \cdot y}{|z| \, |z|}\right)$$

Proof: We have,

$$\begin{aligned} Av = \left(zy^T - yz^T\right) v &= zy^T \left(-\frac{e^{-i\theta} z}{|z|} + \frac{y}{|y|}\right) - yz^T \left(-\frac{e^{-i\theta} z}{|z|} + \frac{y}{|y|}\right) \\ &= \left(-\frac{e^{-i\theta} zy^T z}{|z|} + \frac{zy^T y}{|y|}\right) + \left(\frac{e^{-i\theta} yz^T z}{|z|} - \frac{yz^T y}{|y|}\right) \\ &= -e^{-i\theta} z \, |y| \cos\theta + z \, |y| + e^{-i\theta} y \, |z| - y \, |z| \cos\theta \\ &= \left(-e^{-i\theta} \, |z| \, |y| \cos\theta + |z| \, |y|\right) \frac{z}{|z|} + \left(e^{-i\theta} \, |y| \, |z| - |y| \, |z| \cos\theta\right) \frac{y}{|y|} \end{aligned}$$

The second term simplifies to

$$\left(e^{-i\theta} \, |y| \, |z| - |y| \, |z| \cos\theta\right) = -i \left(\sin\theta\right) |z| \, |y| = \lambda$$

establishing that

$$\left(e^{-i\theta}\,|y|\,|z| - |y|\,|z|\cos\theta\right)\frac{y}{|y|} = \lambda\frac{y}{|y|}$$

For the first term,

$$\begin{aligned}\left(-e^{-i\theta}\,|z|\,|y|\cos\theta + 1\cdot|z|\,|y|\right) &= \left(i\sin\theta\cos\theta + \sin^2\theta\right)|z|\,|y| \\ &= -i\sin\theta\,|z|\,|y|\,(-\cos\theta + i\sin\theta) = -\lambda e^{-i\theta}\end{aligned}$$

which gives

$$Av = \lambda\left(-e^{-i\theta}\right)\frac{z}{|z|} + \lambda\frac{y}{|y|} = \lambda v$$

□

Therefore if we write a component of v as $v_k = u_k + iw_k$ in terms of its real and imaginary parts then,

$$u_k = \frac{-\cos\theta}{|z|}z_k + \frac{y_k}{|y|}, \qquad w_k = \frac{\sin\theta}{|z|}z_k$$

which shows that if $(z_k, y_k) = (\cos\phi_k, \sin\phi_k)$ then the *order* of the collection of points on the unit circle $\{(\cos\phi_k, \sin\phi_k)\}$ is the same as that of the phase angles of $\{v_k\}$. Therefore, in the case of a signal consisting of a single harmonic we are able to determine the *cyclic order* of the phase shifts. Figure 3.3 shows the recovered cyclic ordering of the phase shifts in just such a case.

Further, supposing that a single harmonic dominates in the real signal, one hopes that rank two matrix PAP will closely approximate A in Frobenius norm, where P denotes the projection matrix onto the subspace spanned by the eigenvectors of the largest (in absolute value) eigenvalue [2]. Moreover, the rank two decomposition extends to the case when the collection contains harmonics that are integer multiples of t in different coordinates by considering any two harmonics $x_k(t) = \sin(\alpha_k t + \phi_k)$ and $x_j(t) = \sin(\alpha_j t + \phi_j)$. The area element corresponding to them in a lead matrix can be shown to be zero if $\alpha_k, \alpha_j \in \mathbb{Z}$ and $\alpha_k \neq \alpha_j$. Therefore, the lead matrix will then admit a decomposition

$$A = \sum_i^m z_i \otimes y_i - y_i \otimes z_i \tag{3.6}$$

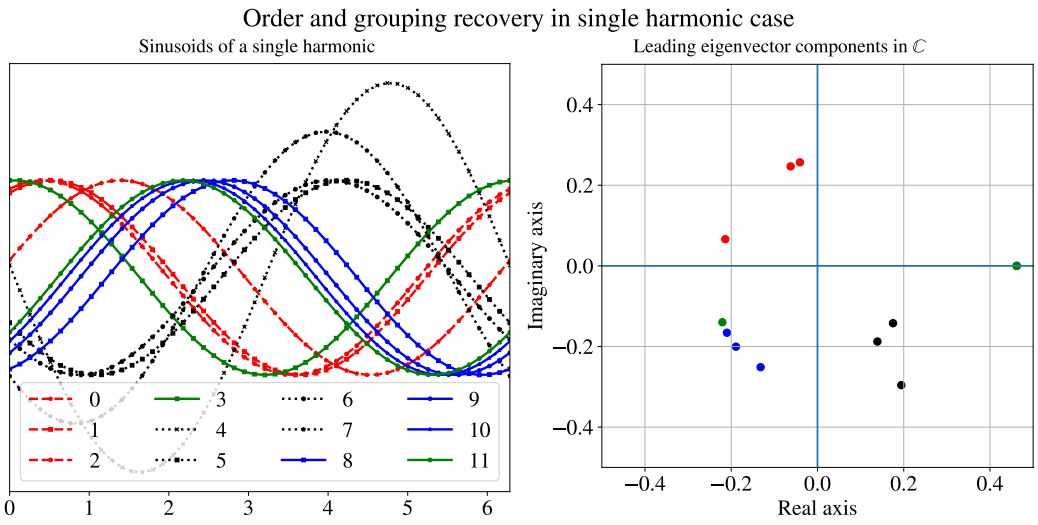


Figure 3.3: Here we see a collection of signals with a single harmonic. Two of them have higher magnitude than the others. They are also colored into groups of red, blue and green signals. The green signals are of opposite phase. On the right we see that the grouping structure, phase & amplitude difference are recovered in the leading eigenvector.

where m is the number of unique α_k and so it suffices to consider the m eigenvectors; each returning the cyclic ordering corresponding to one harmonic. In the absence of clean noiseless harmonics, we rely on the ratio of eigenvalues λ_1/λ_3 to assign a measure of quality to the recovered ordering (recall $\lambda_1 = \bar{\lambda}_2$) because $\lambda_1/\lambda_3 \to \infty$ as we approach the rank two case.

A collection $\{a_k f(t - \alpha_k)\}_k$ where the coordinates track essentially the same function f albeit with different offsets $\alpha_k \in S^1$ is called a *Chain-of-Offsets-Model* (COOM). Such a model is descriptive of systems where one process triggers another which yet again triggers another and so on, resulting in a self-sustained cycling behavior. Supposing that the observed Y_k arise from such a collection via a reparametrization as in Definition 3.3, then expanding the generative function in its internal state space via its Fourier coefficients[2]

$$c_m \sin\left[m\omega\left(t - \alpha_k\right) + \phi_m\right], \qquad m = 1, 2, \ldots \quad \text{and} \quad \omega = \frac{2\pi}{T}$$

one gets that the elements of the lead matrix are given by summation of the

[2]Ignoring the constant term.

terms:

$$A^m_{k,l} = 2\pi m |c_m|^2 a_k a_l \sin\left(m\left(\alpha_k - \alpha_l\right)\right), \quad m = 1, 2, \ldots \tag{3.7}$$

when integrated over $[0, T]$. Then, if one of the coefficients in the Fourier series dominates, we have that lead matrix is well approximated by the rank two skew-symmetric matrix of Eq. 3.7 in the Frobenius norm and by virtue of Lemma 3.1, the cyclic ordering of the phase shifts be approximately recovered. We now exhibit an example of the utility of this method in a network model.

3.1.1 An example in a gossip network

Consider a network model where the information propagates along nodes in a peer-to-peer fashion with propagation delay proportional to edge weights. Such node-to-node information exchange modulated by edge weights is prevalent in so called gossip networks and first passage percolation models. For our purposes, suppose a connected network G with $n = 12$ nodes as shown in Figure 3.4 has a signal $f(t)$ broadcast from a source node s. This graph was constructed by starting with the C_3 graph and adding nodes with $k = 3$ edges at a time. The edges are attached to vertices at random following a distribution proportional to the vertex degree. Suppose this signal,

$$f(t) := \sum_{k=1}^{s} r_k g(t, k), \tag{3.8}$$
$$\text{where} \quad g(t, k) = \exp\left(-\left(t - k\pi + cR_k\right)^2\right)$$

arrives at other nodes affected by a small amount of noise and the above mentioned propagation delay as shown in Figure 3.5. In Eq. 3.8 the amplitudes (r_k) and displacements (R_k) were drawn uniformly from the appropriate distributions so that $f(t)$ is a convex combination. Note that though the signal can arrive at a node in many different ways, only the first arrival is counted and the rest are ignored. Moreover, it is also possible that the shortest path might not be the single-hop distance.

The question then we seek to answer is whether the delayed arrival of the source signal (in essence, the phase shifts) can provide some information to us about the structure of the network. Running the observed collection depicted

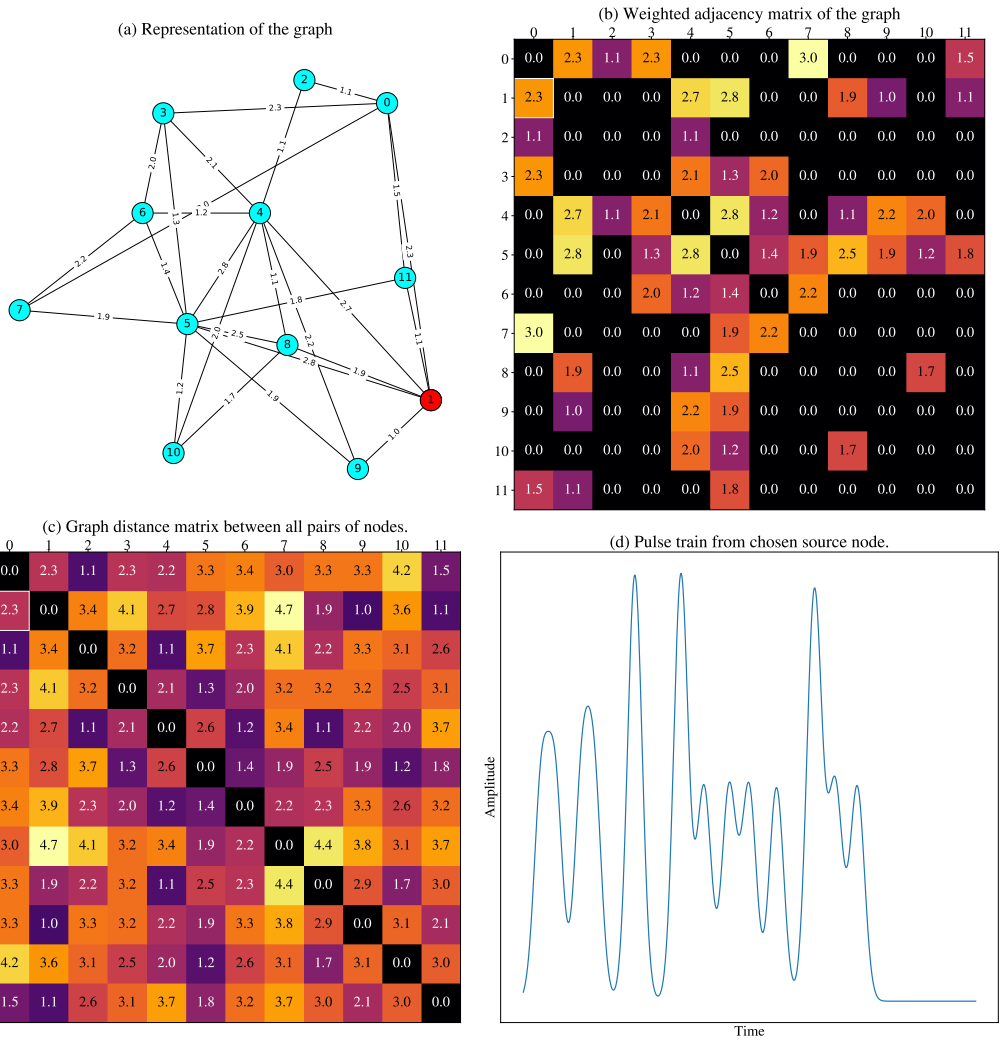


Figure 3.4: This figure shows a simulated example of information propagation along a network. The source of the signal is the red colored node in (a) with each node connected to it receiving the signal with a delay proportional to the connecting edge weights. These delays are represented in (b) where a 0 denotes a lack of connectivity between the nodes corresponding to that row/column. However, information can propagate throughout the graph as long as it is connected with the delays adding up. This is visualized in (c) where each entry is the graph distance between a pair of nodes. Finally (d) shows the noiseless waveform of the signal $f(t)$ of Eq. 3.8 emanating from node 5.

in Figure 3.5 through cyclicity yields results that hint in the affirmative. In this case, the obtained lead matrix shows that it is dominated by the leading eigenvalue with a $\lambda_1/\lambda_3 \approx 7$. Moreover, the plot of derived phase values versus the *known* sensor distance shows a nearly perfect linear dependence,

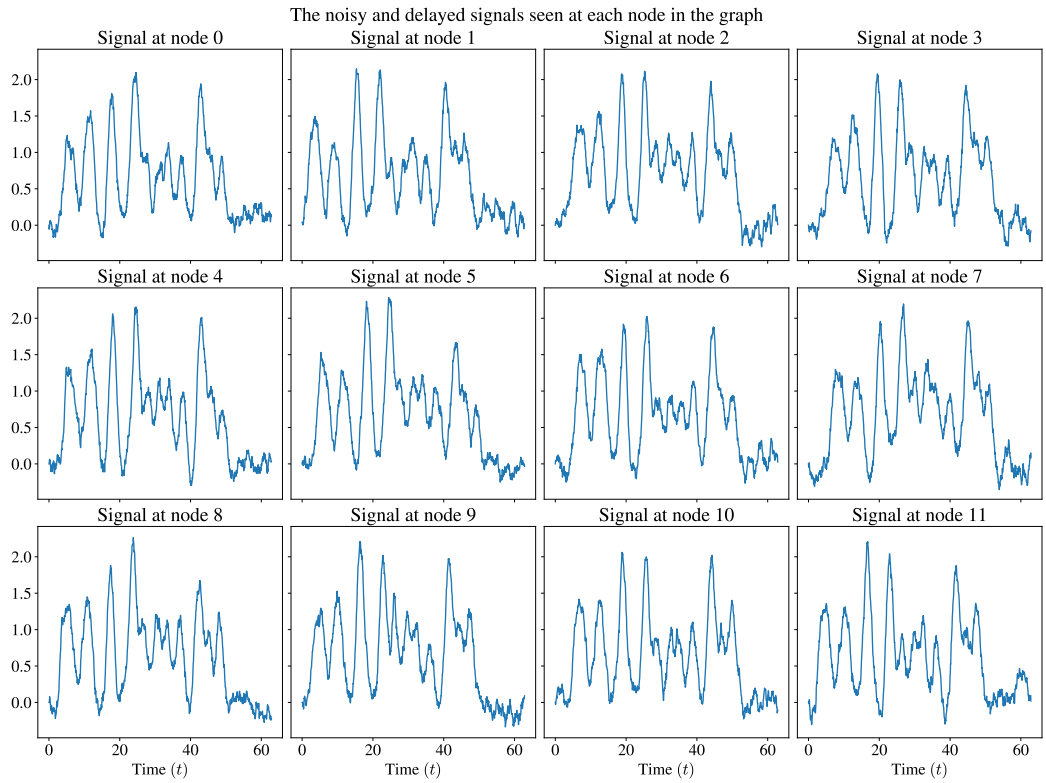


Figure 3.5: The signals arrive at each node with some corrupting noise and a delays. Note that even the source nodes sensor is assumed to have some noise and the delay is defined by the shortest path between a pair of nodes.

i.e. the lag-structure is recoverable. Typically the distance of the node from the source will be unknown; however, repeated analysis using multiple source nodes over time can possibly allow these to be inferred. While the area integral in Eq. 3.1 calculates as its value a quantity signifying the average leader-follower relationship over the entire interval of definition $[a, b]$, it is possible to leave the upper limit of integration as a variable,

$$A_{xy}(t) = \frac{1}{2} \cdot \int_a^t x(s) \cdot y'(s) - y(s) \cdot x'(s) \; ds \tag{3.9}$$

thereby creating a new time series that captures the time evolution of the elements of the lead matrix. This creates $n(n-1)/2$ new times series out of the original n ones, which we can eyeball collectively to infer interrelationships. See Figure A.1 for such pairs of area dynamics that correspond to the network we simulated above.

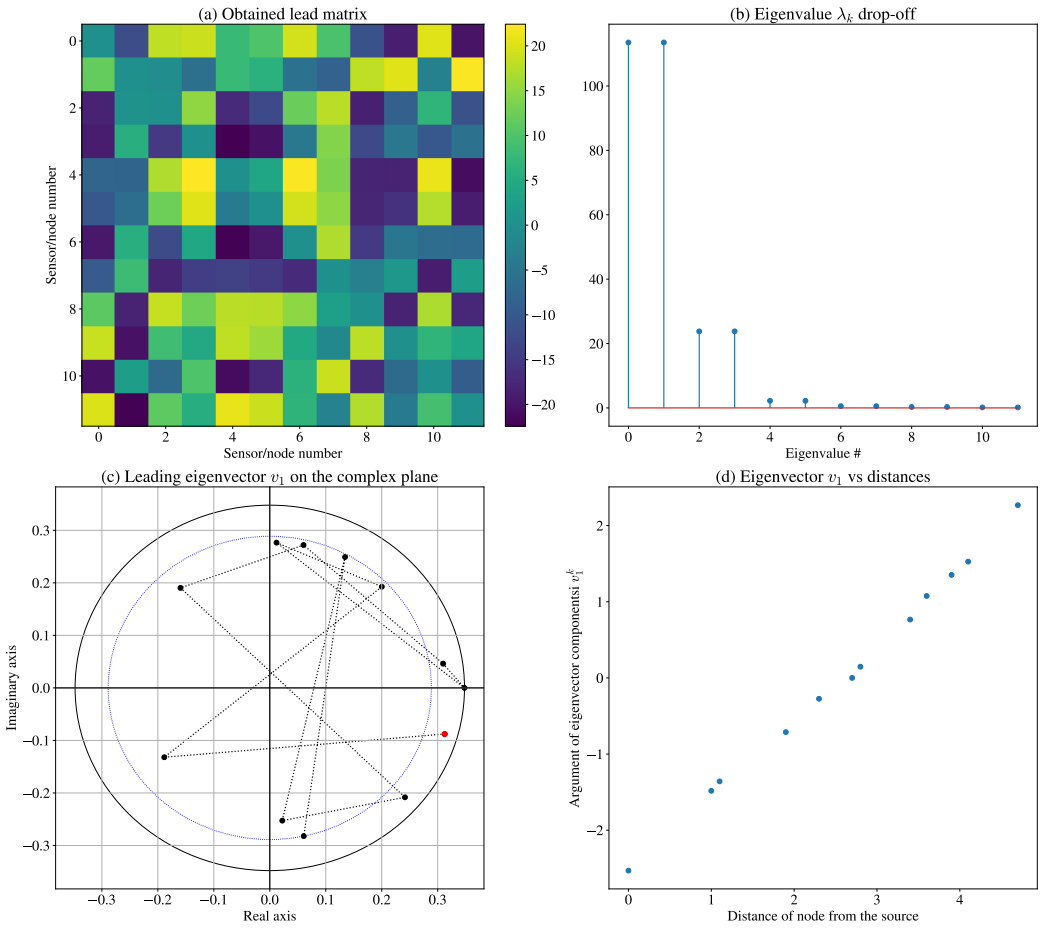
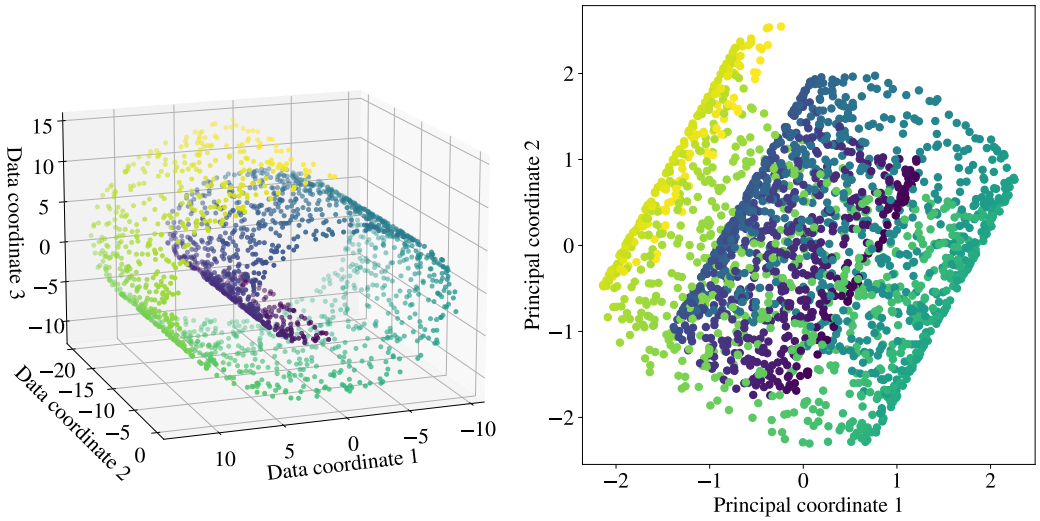


Figure 3.6: Results from cyclicity analysis on the collection of Figure 3.5. (a) The computed lead matrix. The interpretation of the entries shows for example that the nodes 3 and 4 are not too distant because of the nearly zero area between them. (b) The eigenvalue drop off shows that λ_1/λ_3 ratio to be quite high. (c) Shows the eigenvalue components plotted on the complex plane. (d) The plot of the derived phase values vs. the known distance of the node from the source shows a nearly linear dependence implying the lag stucture is recoverable.

In the next chapter we will see the application of cyclicity analysis in the study of fMRI data.

3.2 Nonlinear clustering

In this section, we will describe in detail the nonlinear clustering techniques that were applied to the EMG data mentioned in Section 2.2.2 and discussed in greater detail in Chapter 5. However, we first motivated the need for nonlinear methods by first describing some simple situations where linear methods can fail to account for the underlying structure of the data (cf. Figure 3.7).



(a) The data cloud resembling a 'swiss roll' (b) The data set after dimension reduction

Figure 3.7: The *swissroll* dataset and the result of application of PCA to it fails to unroll the data manifold.

The key here is to note that linear methods intrinsically consider features in the data at a *global* scale. Therefore, in situations where first approximations fail to capture adequately the intricacies happening at a *local* scale, linear methods are set up to fail; contrast the linearization of a nonlinear system about a stable equilibrium vs. unstable equilibrium. The domain of validity of the linearization is larger for the stable one. This would necessitate chaining together multiple linearizations to approximate the system in the unstable case. A similar situation may arise within data. To see how this may happen, consider Principal Component Analysis (PCA) and the now famous *swissroll* dataset in Figure 3.7(a). While commonly used as a dimensionality reduction technique, PCA can also be interpreted as an unsupervised learning method for clustering data. The hypothesis being that in the Principal

Coordinate (PC) space, data points that are *similar*, should ideally cluster together. However, by restricting oneself to a single, purely linear change of coordinates, we end up considering only the global structure in the data.

Application of PCA to this data set results in Figure 3.7(b), where in the PC space, points that would have been far apart ***if*** the local structure of the data had been considered, end up being mapped close to each other. In such datasets, methods that are inherently sensitive to the local structure of the data are required. In view of Figure 3.7(a), it is obvious that the point cloud being considered has intrinsic dimension of two. More formally, the points plotted lie in a two-dimensional *manifold* embedded in $\mathbb{R}^3$, the familiar three dimensional euclidean space.

Consider data points $x_i \in \mathbb{R}^n$ where $i = 1, \ldots, N$. We denote by X the *dataset*, represented as a $n \times N$ matrix and assume with no loss of generality that the dataset has zero mean. That is, each column of X is a datapoint/sample and the mean across the rows of the matrix X amounts to a zero vector. In PCA, one begins by constructing matrix $\Sigma = \dfrac{1}{N-1} XX^T$ and decomposing it as $\Sigma = UDU^T$ wherein the change of basis $Y = U^T X$ projects that dataset into PC space. If the data is not zero-mean then one achieves the same by first performing $X \leftarrow X - \mathbf{1}^T \cdot \mu\mathbf{1}$ where $\mathbf{1}$ is a n-long column vector of ones and μ is a vector holding the mean values. Observe that Σ is a positive semi-definite matrix and U is an orthogonal matrix, i.e, $U^T U = UU^T = I$ where I is the identity matrix. Diffusion maps and Laplacian eigenmaps generalize this fact by constructing positive semi-definite *kernels* whose eigenstructure provides a coordinate change that is essentially nonlinear. A kernel function $K : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ is one that satisfies, for two data points x_i and x_j,

1. Symmetry: $K(x_i, x_j) = K(x_j, x_i)$
2. Positive definiteness: $K(x, y) =\geq 0 \quad \forall x, y$

The choice of the kernel is informed usually by some prior knowledge of the statistics of the dataset or the process that generated the data point. Common ones include the radial basis function (RBF) kernel

$$k(x, y) = \exp\left(-\gamma \|x - y\|^r\right)$$

for $\gamma, r > 0$ and the polynomial kernel $K(x, y) = (x \cdot y + c)^r$ for $c, r > 0$.

When $r = 2, \gamma = \dfrac{1}{2\sigma^2}$ the former is called the Gaussian kernel and for $\gamma = \dfrac{1}{\sigma}\mathrm{i}$ and $r = 1$ the Laplace kernel. The Gaussian kernel is popularly used when the prior distribution of the data is unknown and the sigmoid kernel is often used as a proxy for neural networks where $\kappa = \dfrac{1}{n}$ is a popular parameter choice (n being the dimensionality of the data).

For each dataset X and kernel function $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, we can generate a $N \times N$ positive semi-definite kernel matrix K_X whose entries k_{ij} are given by $k_{ij} := K(x_i, x_j)$ for $i, j = 1, \ldots N$. In the following we make no distinction between the kernel function K and the matrix K_X formed by considering all pairs of data points x_i, x_j under said kernel function. Note that this kernel K (dropping the subscript X) is distinct from Σ in that dimension of K depends on the *size* or *cardinality* of the data set whereas the dimension of Σ depended on the *dimension* of the data set (i.e. the n number of features being observed per sample). Previously and in [1], this is referred to as a distinction between the *intrinsic* and *extrinsic* properties of our dataset.

3.2.1 Diffusion maps

Diffusion maps proceed then by furthering modifying the kernel matrix K to have some rather appealing properties we now describe. Consider for example the kernel function defining:

$$K = k_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma}\right) \tag{3.10}$$

Suppose the kernel function satisfied for each x_i the condition

$$\frac{1}{d_X}\sum_{x_j \in X} k(x_i, x_j) = 1$$

for some normalization constant d_X. One can then interpret its entries as representing transition properties of a random walk on the data set. Indeed, we can generate one-step Markov transition matrix $P = D^{-1}K$ where D is a diagonal matrix comprised of the row or column sums of K. Successive powers of P give then the transition probabilities after each step. Note that P is no longer symmetric but its spectral analysis $P = U\Lambda U^{-1}$ can be facilitated

by considering a related positive semi-definite kernel $D^{1/2}PD^{-1/2}$. See [1] for more details. This construction - of generating a transition matrix based on the diffusion kernel in Eq. 3.10 - allows one to see why the procedure extracts the underlying geometry of the sampled manifold. Note that if $x_i = x_j$ the corresponding entry in K is equal to one. Therefore, points x_i, x_j that are *locally* closer together generate higher probabilities - that is, the random walk is more likely to land on a data-point if it came from another point locally close to it. The eigenvectors U are called *diffusion coordinates* and the embedding to the *diffusion space* after t time steps then is given by $Y = U\Lambda^t$. Since the larger eigenvalues in Λ determine which coordinates are more important than the other, a dimension reduction is performed on the data by simply retaining only a few largest ones. The hyper-parameter σ then controls in effect the size of the *neighborhood* around each data-point that we examine to determine a local structure. The parameter t determines how much importance to give to the local structure since larger values of t increase the relative importance of some diffusion coordinates compared to each other.

3.2.2 Laplacian eigenmaps

Note that the kernel matrix K in the previous section is a dense matrix, i.e. nearly all of its entries are non-zero. The Laplacian eigenmap is essentially a *sparsified* version of the previous method. Instead of relying on the kernel matrix K directly, a sparsified K is constructed by zeroing out some entries. The entries to be zeroed out are determined by a k-nearest neighbor graph constructed from the data with vertices represented by the data points. Edges not in the graph correspond to row-column indices that need to be zeroed out. As opposed to a transition matrix P one instead considers the *graph Laplacian* $L := D - K$ where D is the same matrix from the previous section. Since we seek a mapping that respects the local structure, an embedding Y minimizing $\sum_{ij} \|y_i - y_j\|^2 K_{ij}$ will ensure that close points that end up being mapped far apart incur heavy penalties. A normalization constraint $[\|D^{1/2}y_i\| = 1$ can be imposed to avoid the arbitrary scaling factor in the embedding. As shown in [1], this problem can be cast as seeking an

embedding Y such that,

$$\underset{Y^T DY=I}{\arg\min}\, \mathrm{tr}\left(Y^T LY\right) \tag{3.11}$$

whose solution is provided by spectral analysis of the normalized Laplacian

$$D^{-1/2}LD^{-/2}$$

3.2.3 T-distributed stochastic neighbor embedding

Considering kernel in Eq. 3.10, it is somewhat obvious that the choice of the parameter σ makes a difference to the obtained lower-dimensional embedding. t-Distributed Stochastic Neighbor Embedding (t-SNE) is an award winning technique introduced which takes this idea one step further [27, 28]. In [27], the similarity measure between two data points x_i and x_j is defined as

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\|x_i - x_k\|^2/2\sigma_i^2\right)} \tag{3.12}$$

where σ_i is now the variance of a Gaussian centered at datapoint x_i. That is, $p_{j|i}$ is the conditional probability that x_i would consider x_j as its neighbor if neighbors were chosen in proportion to the probability density of a Gaussian distribution centered at x_i (with $p_{i|i}$ set to zero).

By allowing σ_i to vary between points, the algorithm is able to take into account local density differences along the data cloud in the high dimensional space. Given this distribution in the original space, SNE method seeks to create a lower dimensional embedding $\{y_i\}$ such that a similar conditional probability $q_{j|i}$ in the embedding space obtained via:

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum\exp\left(-\|y_i - y_k\|^2\right)} \tag{3.13}$$

remains "faithful". This is done by minimizing the Kullback-Leibler (KL)

divergences[3] between the original $p_{j|i}$ and $q_{j|i}$:

$$C = \sum_i KL\left(P_i||Q_i\right) = \sum_i \sum_j p_{j|i} \log\left(\frac{p_{j|i}}{q_{j|i}}\right) \tag{3.14}$$

in which P_i represents the conditional probability distribution over all other data points given data point x_i and similarly for Q_i. Note that in the lower dimensional space we set $\sigma = 1/\sqrt{2}$ for all data points so that the induced probability $q_{j|i}$ is a function of all points in the lower-dimensional embedding. Therefore, the SNE cost function is geared towards retaining the local structure of the data in the obtained mapping. For the sake of brevity we leave details of how to choose σ_i (it is a binary search over values of σ_i that produce a parameter called 'perplexity' which is a smooth measure for the effective number of neighbors) and other optimization/computational considerations to [27]. However the SNE methods suffers from the 'crowding problem' related to the curse of dimensionality. Roughly put, in higher dimensions it is possible for many more points to be equidistant from a given point (recall the surface area of the n-sphere is proportional to the n-th power of its radius). Therefore, in any mapping to lower dimensions, many points will try to 'occupy the same space'. The t-SNE method overcomes this problem by using distributions with much heavier tails than the Gaussian distribution in the embedding space, specifically the t-distribution so that,

$$q_{j|i} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|y_k - y_l\|^2\right)^{-1}} \tag{3.15}$$

3.3 Classification methods

Whereas clustering methods seek to obtain an arrangement of data-points that help understand its underlying geometry and features, and are generally called unsupervised learning methods, the *classification* task is one which is considered to a be a supervised learning technique. Here, the distinction is that we provide the machine with labeled data points so as to make inferences regarding its structure. Then, when presented with new information or data

[3]Recall that the KL divergence is a measure of similarity between probability distributions.

points, the classifier has to determine which of the previously seen labels apply to the new data. Support Vector Machines (SVMs) are a classical approach to this problem, formulated as early as 1963 [29]. While we use SVMs, here we discuss in detail only some related methods.

3.3.1 Sparse support vector machines

Recall that a traditional SVM is a maximum margin classifier developed based on the linear classification problem, involving the construction of a separating hyperplane that breaks up training observations based on the corresponding class labels. The margin being defined as a distance between any two hyperplanes that obtain separation in the linearly separable case, SVMs solve an optimization problem with some specific constraints on finding an optimal hyperplane that maximizes the margin. The standard risk function minimized by SVM includes two terms. The first one is the hinge loss function,

$$l(y) = \max(0, 1 - ty) = \max(0, 1 - t(w \cdot x + b)), \tag{3.16}$$

where t is the true label and y is the prediction, the elements of the vector w determining the optimal separating hyperplane. The second term is the squared l_2-norm of the weight vector w. In sparse linear SVM, the l_1-norm is used as the regularization term in lieu of l_2-norm allowing for a sparser solution of the weight vector. The non-zero elements of the weight vector then inherently correspond to variables that play more dominant roles in building the optimal separating hyperplane. This allows for interpreting the specific elements in a feature vector as substantially discriminative for the classification task [20].

3.3.2 Partial least squares discriminant analysis

PLS-DA is an iterative algorithm that makes use of the class labels in the training data to find transformation (or loadings) vectors a that has the following objective function:

$$\max_{(a_h, b_h)} \operatorname{cov}(X_h a_h, y_h b_h) \tag{3.17}$$

where a_h, b_h are the loading vectors for the data and the labels at iteration h and X_h is the residual. The initial iteration starts with $X_1 = X$ the data matrix. The obtained principal components can be shown to be equivalent to the eigenvectors of the non-singular portion of the covariance matrix given by $\frac{1}{(N-1)^2} X y^T y X^T$. PLS-DA can be thought of as a supervised version of Principal Component Analysis (PCA) because it attempts to achieve dimensionality reduction but with full awareness of the class labels. Aiming to maximize the covariance between the independent variables X and the corresponding dependent variable (class labels) of highly multi-dimensional data, PLS-DA finds a linear subspace of explanatory variables that allows for the prediction of y on new data with a reduced number of factors. Besides its use as for dimensionality-reduction and classification, it can be adapted to be used for feature selection [30].

3.3.3 Neural networks

Neural networks are a classical (going all the way back to the perceptron) method of learning. At the most basic level, the mimic the architecture of biological neurons. An *artificial* neuron is represented by a mathematical *node* that takes inputs via incoming weighted edge and combines them in some fashion to produce a single output. A column or array of neurons are said to form a *layer*. Typical models involve multiple layers starting with an input layer and ending with an output layer with varied models of interconnectivity between intervening layers depending on the type of network. A fully connected network where every node in a layer is connected to every node in successive layers are results in the classical multilayer perceptron of the 1980s. However, recently a regularized variation of them, termed convolutional neural networks (CNN), have become incredibly popular due to their huge success in dealing with image classification and related problems.

CNN is a class of artificial neural networks that uses convolution operation inspired by visual neuroscience [31]. Contrary to the conventional neural networks, CNN introduces sparse connectivity by sharing parameters leading to fewer number of parameters. Similar to conventional neural networks, CNN can use gradient-descent based algorithms to classify training observa-

tions according to their class labels. These general models can collectively called feedforward neural networks, i.e. the connections feed outputs of each layer forward (with "forward" taken to mean the direction from the input layer to the output layer). When the outputs can be fed back into the same layer or previous layer, we get a another general class of models called *recurrent* neural networks which have been shown to be Turing complete and have found applications in handwriting recognition, speech recognition, and related fields. Finally, one can set up far more general models by piecing together different types of layers, networks, and models. For example, by jointly training an *encoder* and *decoder* network where the encoder maps from the input space to a latent space, and the decoder maps from the latent space to the output space, with the objective to reduce reconstruction error, we have model popularly called *Variational Auto Encoder* (VAEs). VAEs are a popular method to augment data by synthesizing data by sampling from the latent space. Since we only minimally utilize neural networks in this work, a general treatment is skipped and specific details will be provided where necessary.

Chapter 4

Functional magnetic resonance imaging data

In this chapter, we deal with functional magnetic resonance imaging (fMRI) data. Apart from the problem of high extrinsic dimensionality, there are other reasons to motivate the development of new methods for such datasets like fMRI time series. In fMRI imaging studies, correlation based methods have long been one of the most popular. Correlation of pixel-time courses as a method of analysis was arguably first reported in 1993 by Bandettini & colleagues [32], and, by 1995, was used to define *functional connectivity* of the brain [33]. Regions of the brain may be structurally connected by white or gray matter; however BOLD signals from spatially distinct and unconnected regions of the brain can show high levels of correlation. Such regions are defined to be functionally connected and such connectivity between different regions of the brain is studied by generating a *correlation matrix* (CM) whose entries correspond to the correlation coefficient:

$$C_{i,j} = \frac{1}{\sigma\left(x_i\right)\sigma\left(x_j\right)} \cdot \frac{1}{T} \int_0^T x_i\left(t\right) x_j\left(t\right) dt \tag{4.1}$$

between each pair of rows in X where $\sigma(\cdot)$ is the standard deviation function and $[0, T]$ is the interval the time series is observed over. Study of more general correlation maps - seed based, graph based etc. - is one way of discovering so called large scale brain networks[34], see for example [35, 36, 37]. Some authors also study *lagged correlations* to account for the delay in propagation of neuronal activity in the brain [38, 39]. They do this by computing the value of τ at which

$$\hat{C}_{i,j}\left(\tau\right) = \frac{1}{\sigma\left(x_i\right)\sigma\left(x_j\right)} \cdot \frac{1}{T} \int_0^T x_i\left(t+\tau\right) x_j\left(t\right) dt \tag{4.2}$$

is extremized and call it the temporal lag between signals x_i and x_j. This gives rise to two matrices, a *lagged cross-correlation matrix* (LCM) comprised of entries in which (4.2) is extremized and the corresponding *time delay matrix* comprised of entries capturing the time lag between each pair of time series in X.

On the other hand, there has been an argument that cross-correlation based methods need to be carefully scrutinized [40]. For example, though correlation based methods have been long used in fMRI studies to study the connectivity structure in the brain they suffer from a few draw-backs [19, 20]. As noted in [20],

> ... it assumes the time series are not auto-correlated [40], but this assumption may not hold for resting state BOLD signals [41]. Second, the correlation analysis fails to capture the information expressed in lag structure [38, 42].

Even though other methods discussed in Chapter 3 may account for temporal lag and/or nonlinear relations between time series data, an essential problem still remains in the setting of machine learning algorithms. It is a common paradigm in fMRI based studies to treat the set of brain volumes over the entire scan duration as as a single sample and to use estimates of functional connectivity between brain regions' time courses as features by vectorizing covariance matrices. However, many classification and/or clustering algorithms implicitly assume that features used are generally uncorrelated. This assumption does not hold water in the case of correlation based methods because elements of the feature vector are intimately inter-related, since these matrices are sampled from a positive semi-definite cone [43]. This further necessitates the development of new methods to either generate different kinds of features from multi-variate time series data, or for algorithms and procedures to effectively de-correlate the data.

As mentioned in Section 2.2.1, we used data from two sources for analysis. Recall, the first was collected at the Auditory & Cognitive Neuroscience Laboratory with a primary objective of better understanding the neural underpinnings of *tinnitus*. Chronic tinnitus is a common and sometimes debilitating condition that lacks scientific consensus on physiological models of how the condition arises as well as any known cure. Tinnitus is classified into two types, objective and subjective. In objective tinnitus there is a clear

physical cause - often structures near the ear - for the perception of noise, and is much rarer than subjective tinnitus where no underlying cause can be found. Tinnitus is often accompanied by hearing loss and is more common among people who have had head or neck injuries, or exposure to concussive blasts. As such, it is a common condition among veterans. However, due to the internal nature of the condition, there exist no diagnostic methods or known cures. It is hypothesized that functional differences in regions of the brain may shed light on the condition. Though there exist many approaches to measure brain activity, the least invasive one is functional magnetic resonance imaging (fMRI).

While the datasets from the Auditory and Cognitive Neuroscience Laboratory were leveraged to explore biomarkers of tinnitus, data from the Human Connectome Project (HCP) was utilized to investigate what insights can be gained from the application of the cyclicity algorithms described in Section 3.1. The following sections will largely follow the trajectory of [19, 20, 21].

4.1 Zimmerman et. al

In this study[1] data was collected from neuro-typical controls as well as tinnitus afflicted participants with and without hearing loss using a 3T Siemens Magnetom Prisma MRI scanner. The data collected was in the so called *resting state*, i.e. participants are simply asked to lie awake in the scanner with their eyes open and fixated on a cross, as opposed to being asked to perform a mental or cognitive task. Such resting-state data was collected at two sessions, 1 week apart and in each session two 10 minute scans were performed. Resting state fMRI (rs-fMRI) studies have shown reorganization in resting state functional connectivity in the brain due to the presence of tinnitus [19, 20]. Rs-fMRI is an interesting candidate to study because the absence of task eliminates as cause for concern, a poorly designed/executed experiment. On the other hand, tinnitus participants are at a pseudo task state because they are presumably aware of the internal noise.

[1]Material in this subsection was previously published in [19].

4.1.1 Data collection

All imaging data were collected using a 3T Siemens Magnetom Prisma MRI scanner. A high-resolution, T1-weighted sagittal MPRAGE image (TR = 2300 ms, TE = 2.32 ms, flip angle = 89°, 192 slices, voxel size = $0.9 \times 0.9 \times 0.9$ mm^3) and a lower-resolution, T2-weighted, image (TR = 3400 ms, TE = 65 ms, flip angle = 120°, 38 slices, voxel size = $1.2 \times 1.2 \times 3.0$ mm^3) were both collected for use during preprocessing. Resting state data was collected at two sessions, one week apart. Two ten minute runs of resting state data were acquired at both sessions. Resting state BOLD acquisition used a gradient echo-planar EPI sequence with transversal orientation (TR = 2000 ms, TE = 25 ms, flip angle = 90°, 38 slices, voxel size = $2.5 \times 2.5 \times 3.0$ mm^3). During the resting state scans, participants were instructed to lie still with eyes open fixated on a cross presented to them, and to not think about anything in particular. The first 4 volumes of each run were discarded prior to preprocessing to allow for magnet stabilization. Thus, of the 304 volumes collected in each run (four runs were collected per subject), 300 were used for subsequent analysis. Cohort specific and demographic data are available in the appendix.

4.1.2 Data processing

Preprocessing was performed using SPM12 (Welcome Trust Centre for Neuroimaging[2]). Slice-time correction was first applied to the interleaved, ascending data. Functional images were realigned according to a 6-parameter rigid body transformation to correct for head motion. Seven subjects were removed from subsequent analysis due to motion exceeding a 2 mm translation or 2° rotation in one of the resting state runs. Following this, two co-registration steps were performed. First, the T2-weighted image was registered to the mean functional image generated during realignment. Second, the MPRAGE image was registered to the resulting T2-weighted image. Next, the MPRAGE image was normalized to MNI space via a nonlinear warp transformation. The resulting image was used to normalize the realigned functional data. Lastly, the functional images were smoothed using a Gaussian kernel of $8 \times 8 \times 8$ mm^3 full width at half-maximum. Thirty

[2] `http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`

three regions of interest (ROIs) were chosen based on their implication in previous fMRI based tinnitus studies. A list of these can be found in Table 4.1. The ROI time series were constrcuted by averaging the time courses acros all voxels within an ROI.

Table 4.1: Regions-of-Interest used in Zimmerman et. al[19] along with the functional networks which contain them.

Name	**Center Coordinates**	**Network**
l amygdala	-17,-2,-24	Limbic
l anterior insula	-36,3,7	Attention control
l cuneus	-4,-88,16	Visual
l frontal eye field	-25,-11,54	Dorsal attention
l inferior frontal lobe	-41,6,10	Attention control
l inferior parietal lobe	-31,68,32	Default mode
l mid frontal gyrus	-39,11,38	Attention control
l parahippocampus	-24,-22,-24	Limbic
l posterior intraparietal sulcus	26,-62,53	Dorsal attention
l primary auditory cortex	55,-27,9	Auditory
l primary visual cortex	-11,-84,1	Visual
l superior occipital lobe	-12,-80,23	Visual
l superior temporal junction	-49,-53,28	Ventral attention
l superior temporal sulcus	-56,-52,9	Ventral attention
l ventral intraparietal sulcus	-30,-83,13	Dorsal attention
medial prefrontal cortex	8,59,19	Default mode
posterior cingulate cortex	-2,-50,25	Default mode
precuneus	0,-56,50	Default mode
r amygdala	18,-7,-17	Limbic
r anterior insula	36,3,7	Attention control
r cuneus	4,-88,16	Visual
r frontal eye field	27,11,54	Dorsal attention
r inferior frontal lobe	45,-4,13	Attention control
r inferior parietal lobe	40,-67,32	Default mode
r mid frontal gyrus	39,11,38	Attention control
r parahippocampus	23,-21,-20	Limbic
r posterior intraparietal sulcus	-23,-70,46	Dorsal attention
r primary auditory cortex	-41,-27,6	Auditory
r primary visual cortex	11,-84,1	Visual
r superior occipital lobe	15,-79,23	Visual
r superior temporal junction	49,-53,28	Ventral attention
r superior temporal sulcus	56,-52,9	Ventral attention
r ventral intraparietal sulcus	30,-83,13	Dorsal attention

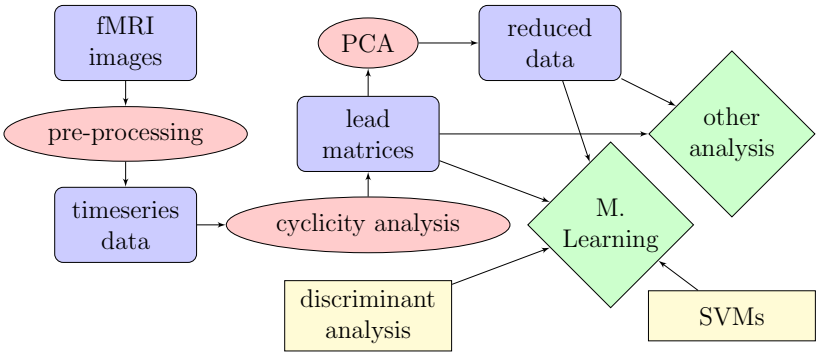


Figure 4.1: The analysis pipeline used for [19].

4.1.3 Results

Cyclicity analysis was used along with standard classifiers like SVMs and linear & quadratic discriminants to test whether it was possible to disassociate tinnitus data from control data. Since each element in the feature matrices corresponds to ROI pairs, lead matrices were used to hone in on ROI pairs that were found to be most helpful in discriminating between the groups. Moreover, since data was collected from the same participant twice one week apart, the stability, i.e. how much they differ from week to week, of the lead matrices were also tested using a 1-nearest neighbor classifier. Figure 4.1 shows the overall pipeline used in this first paper. Specifically from [19],

> Using the output from the cyclicity analysis, we were able to differentiate between these two groups with $58 - 67\%$ accuracy by using a partial least squares discriminant analysis. Stability testing yielded a 70% classification accuracy for identifying individual subjects' data across sessions 1 week apart. Additional analysis revealed that the pairs of brain regions that contributed most to the dissociation between tinnitus and controls were those connected to the amygdala. In the controls, there were consistent temporal patterns across frontal, parietal, and limbic regions and amygdalar activity, whereas in tinnitus subjects, this pattern was much more variable.

For each of the four scans corresponding to one participant a lead matrix was created and examined. Since each subject was scanned twice per session, using one session as a training set and the second set as a test set was a natural split.

Recall that in cyclicity analysis, eigenvector components with higher magnitude correspond to larger magnitude signals. To determine which ROIs were consistently 'loud', phase magnitudes were computed across the tinnitus and control population and the ROIs that consistently appeared in the top 10 magnitudes determined. This result is shown in Figure 4.2[3]. It was found that the bilateral cuneus and bilateral primary visual cortices were the strongest signal in both sets of population. These are ROIs involved in visual processing and dominated signals from other regions - perhaps an artifact of the fact that participants had their eyes open during scans. Further, PCA revealed that the greatest contributor to variance in the data were the bilateral cuneus in both sets of population. Therefore, it was adjudged that the bilateral cuneus does not provide any discriminatory information and was removed from further analysis. After removal of the bilateral cuneus, PCA was performed again and studying the loadings of the principal component revealed that [19]:

> The first and second components appear to switch between the two groups. In the normal hearing controls, the first component is weighted toward the cyclic relationships with the primary visual cortices and occipital lobe, while the second component is more weighted toward the precuneus and amygdala. In contrast, in the tinnitus group, the first component seems to be strongly weighted toward the cyclic connectivity of the precuneus while the amygdala is less constrained, and the second component seems to be weighted toward the visual areas.

We interpreted signals with large phase magnitudes having a more robust or constrained cyclic ordering compared with signals with low phase magnitudes [19]. Further, as mentioned earlier, the 'stability' of the lead matrix was investigated using a 1-nearest-neighbor classifier using the cosine metric. The classifier was trained and tested twice - trained on the first weeks data and tested on the second week and vice versa. Therefore, a subject could be correctly identified upto four times. This result is shown in Figure 4.3 with a 70 % accuracy rate for the classifier.

Further, classifcation using Linear SVMs (LSVM) and Quadratic SVMs (SVM) as well as Linear (LDA), Quadratic (QDA) and Partial-Least Squares

[3]Figures and captions in this section are reproduced from [19, 20].

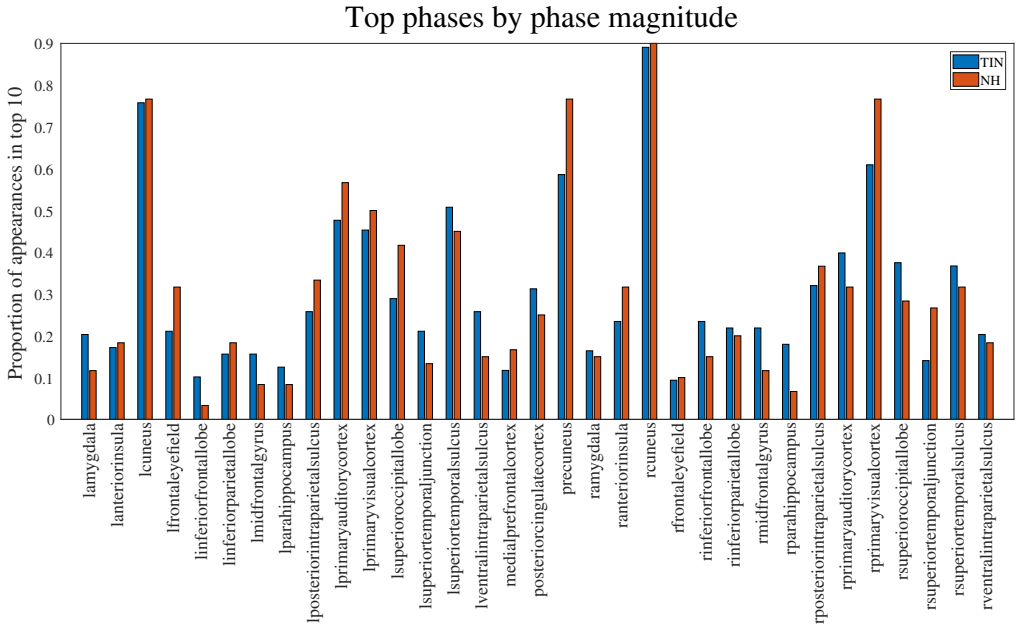


Figure 4.2: Regions with the highest magnitudes in the cyclicity analysis. The leading eigenvalue and the corresponding eigenvector of the lead matrix determine the magnitude; in particular, the elements of the eigenvector correspond to ROIs, and the larger an element's modulus is, the greater the magnitude corresponding to the signal from that ROI. The above chart shows the proportion of times each region occurred in the top 10 magnitudes for each subject. Bars are displayed for the tinnitus group and the normal hearing controls. This graph reveals that certain regions have consistently high magnitudes in the cyclicity analysis, especially in visual regions such as the right and left cuneus. This is true for both tinnitus (TIN) and controls (NH). In other regions, such as the precuneus, the phase magnitudes are more variable between groups.

Discriminant (PLS-DA) Analysis were performed. When controlled for class sizes the results were no better than a coin toss in being able to discriminate tinnitus data from control data. To see if classification results could be improved, the Wilks Lambda criterion was used to select ROI pairs that were most discriminating in their ability to seperate tinnitus data from control data. Wilks' lambda is a direct measure of the proportion of variance in the combination of dependent variables that is unaccounted for by the independent variable (the grouping variable or factor). If a large proportion of the variance is accounted for by the independent variable, then it suggests that there is an effect from the grouping variable and that the groups have different mean values [44].

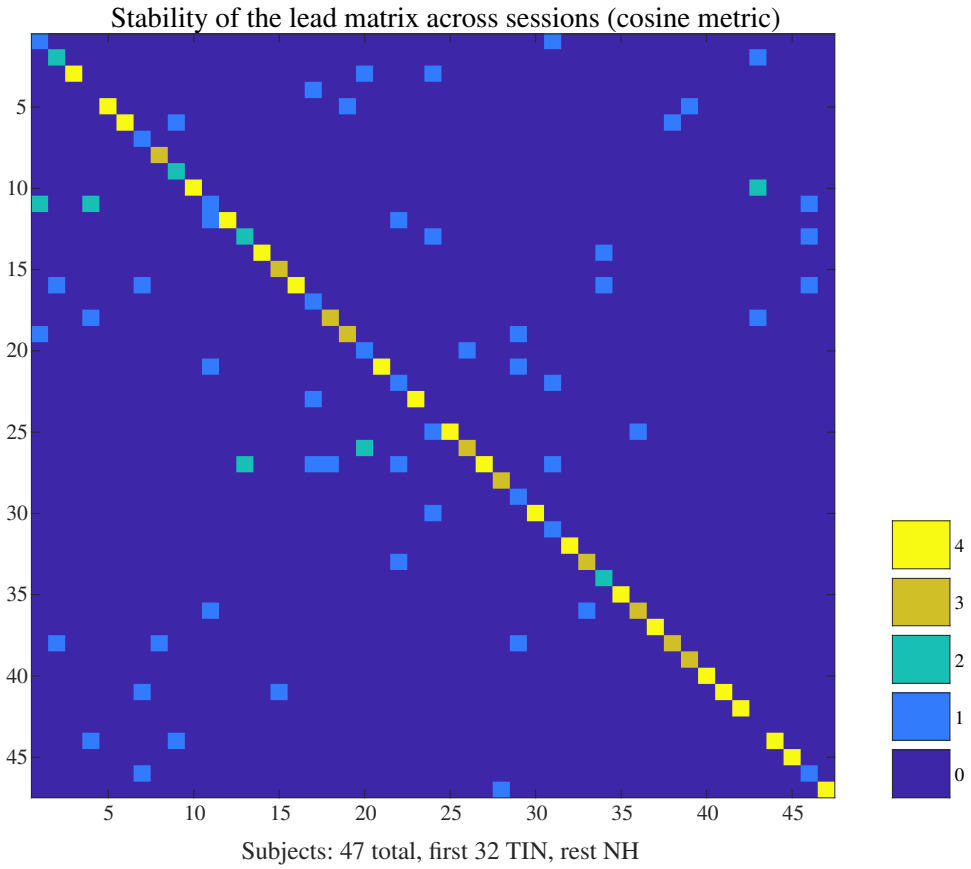


Figure 4.3: The stability of the lead matrix across sessions that were one week apart. Since the lead matrix is a feature constructed from the fMRI time series data; its consistency over time is investigated here. The figure above is a visualization of the confusion matrix arising from a classifier. Each row and column correspond to an individual subject in the analysis. The colors correspond to how many runs (out of 4 total runs) were correctly classified after training a 1-nearest neighbor classifier with the cosine metric on the other week's data. The cosine metric serves as a measure of how closely aligned vectors are in high dimensional spaces. This graph shows good stability in the cyclic patterns of data, i.e consistent leader-follower relationships between ROIs, within subjects across one week.

We used Wilks' lambda to aid analysis in two ways. First, rather than use PCA to reduce dimensions in the data set, we used Wilks' to reduce dimensions from 528 features to just 20 features by keeping the top 20 discriminatory features separating tinnitus data from control data. Classification performed this way improved accuracy to 65%. This result is shown in Table

4.2. Since the PLS-DA method utilizes class labels in finding its principal components, classification performed this way involves double dipping to an extent. To reduce this effect, 1000 randomly selected subsets of the data were analyzed and a subset of the top 20 most occurring ROI pairs were selected. Figure 4.4 visualizes this operation. The top 10 most occurring ROI pairs were then used to re-perform classification. Results obtained in this fashion are also presented in Table 4.2.

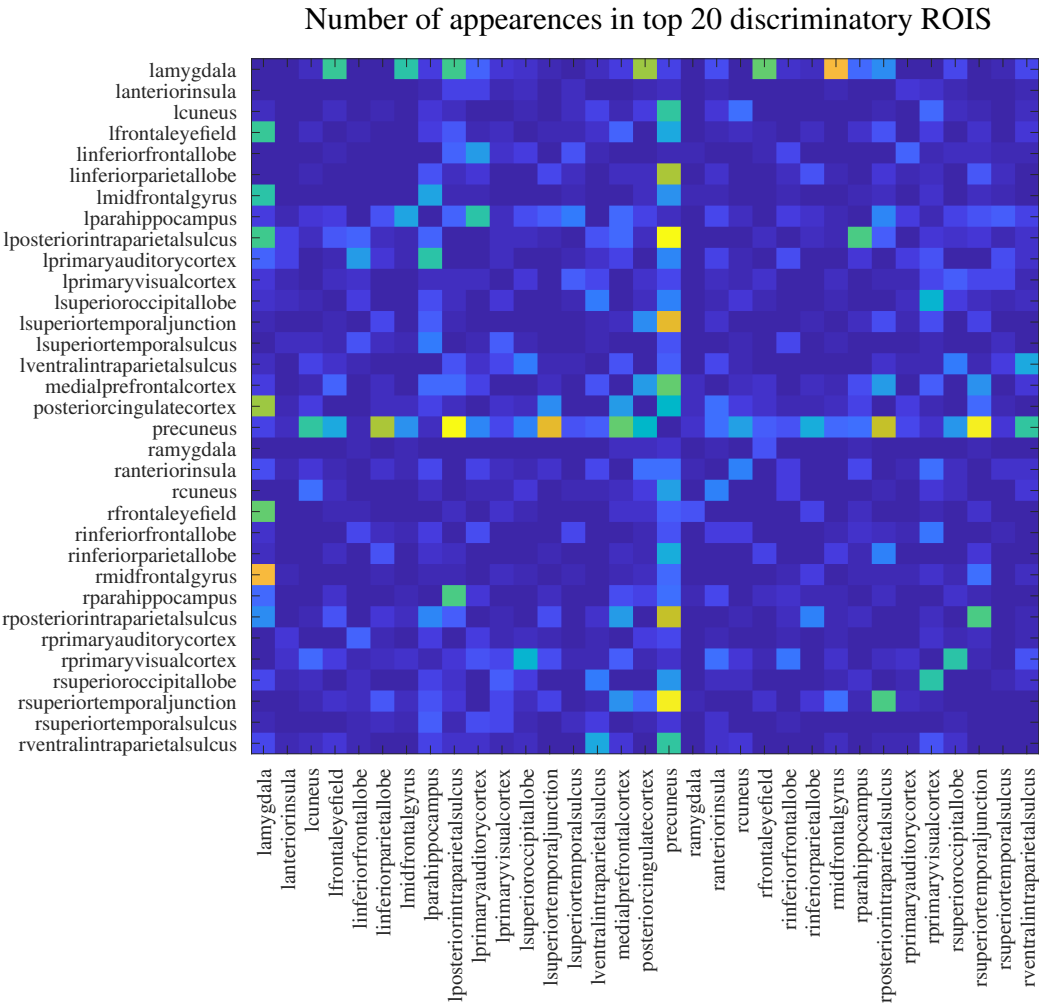


Figure 4.4: The most stable ROI pairs with respect to discriminatory ability across the data set. The Wilks lambda criterion can be used to determine which features in data have more discriminatory ability for classification. Thousand random and equally sized subsets of the data were examined and the top 20 ranked ROI pairs (with respect to discriminatory ability) were recorded. The figure shows how many times an ROI pair appeared in the ranking.

Table 4.2: Classification results using top 20 (left) and 10 (right) most discriminating ROI pairs

Linear SVM		
	Controls	**Tinnitus**
Controls	23.53%	76.47%
Tinnitus	17.25%	82.75%
		Accuracy: 63.85%

Linear SVM		
	Controls	**Tinnitus**
Controls	20.47%	79.53%
Tinnitus	11.25%	88.75%
		Accuracy: 66.96%

Linear Discriminant		
	Controls	**Tinnitus**
Controls	34.07%	65.93%
Tinnitus	29.00%	71.00%
		Accuracy: 59.21%

Linear Discriminant		
	Controls	**Tinnitus**
Controls	30.73%	69.27%
Tinnitus	19.16%	80.84%
		Accuracy: 64.85%

Quadratic SVM		
	Controls	**Tinnitus**
Controls	32.93%	67.07%
Tinnitus	24.44%	75.56%
		Accuracy: 61.96%

Quadratic SVM		
	Controls	**Tinnitus**
Controls	31.73%	68.27%
Tinnitus	24.72%	75.28%
		Accuracy: 61.38%

Quadratic Discriminant		
	Controls	**Tinnitus**
Controls	7.13%	92.87%
Tinnitus	7.81%	92.19%
		Accuracy: 65.04%

Quadratic Discriminant		
	Controls	**Tinnitus**
Controls	23.60%	76.40%
Tinnitus	27.91%	72.09%
		Accuracy: 56.62%

PLS-DA		
	Controls	**Tinnitus**
Controls	52.47%	47.53%
Tinnitus	31.31%	68.69%
		Accuracy: 63.51%

PLS-DA		
	Controls	**Tinnitus**
Controls	50.47%	49.53%
Tinnitus	25.20%	74.80%
		Accuracy: 67.03%

Analyzing the ROI pairs, the 20 ROI pairs that were found to be most discriminating allowed us to replicate some findings in extant literature. To do so, a graph was generated that visualized ROIs as nodes, and edges whose weight or size depended on the proportion of the population exhibiting that connectivity. In more detail, since each element of the lead matrix is a measure of the average leader-follower relationship between two ROIs, the lead matrix elements were thresholded to ± 1 depending on whether the entry was positive or negative. Averaging the lead matrices for each group separately then provides a matrix from which the top 20 most discriminating ROI pairs can be examined. Each pair corresponds to some entry in the matrix, and a value for the entry that is close to zero implies "no pattern." That is, the groups' population as a whole exhibited no preference for a leader-follower relation between that pair of ROIs. Conversely a value close to $+1$ or -1 implied nearly all members of the population exhibited a preference for a particular leader-follower pattern between that pair of ROIs. This result is

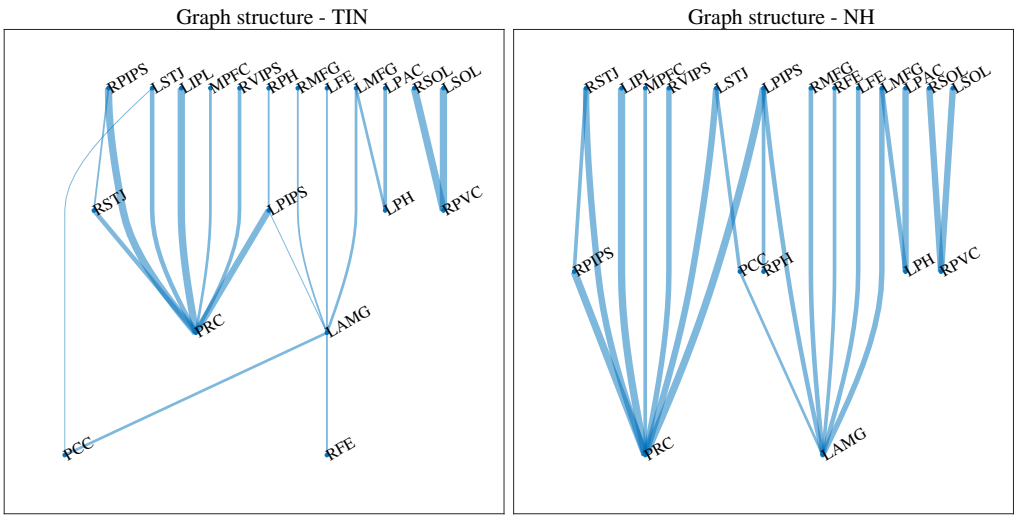


Figure 4.5: Graph structure of leader-follower connections. This figure shows the direction of leader-follower relationships between 20 ROI pairs that most help discriminate the normal hearing controls from the tinnitus subjects (activity follows downwards). A node is assigned to each layer as soon as possible provided its predecessors have already appeared. The thickness of the edges corresponds to the proportion of subjects with that direction, and thus reveals the consistency of the leader-follower connections. In normal hearing subjects, there is more consistent cyclic connectivity with the amygdala than in the tinnitus subjects.

shown in Figure 4.5.

Table 4.3: Key for reading nodes in Figure 4.5 labeltbl:acronyms

Acronym	**Expanded form**	**Acronym**	**Expanded form**
LAMG	Left amygdala	PCC	Posterior cingulate cortex
LFE	Left frontal eyefield	PRC	Precuneus
LIPL	Left inferior parietal lobe	RFE	Right frontal eyefield
LMFG	Left midfrontal gyrus	RMFG	Right midfrontal gyrus
LPAC	Left primary auditory cortex	RPH	Right parahippocampus
LPH	Left parahippocampus	RPIPS	Right posterior intraparietal sulcus
LPIPS	Left posterior intraparietal sulcus	RPVC	Right primary visual cortex
LSOL	Left superior occipital lobe	RSOL	Right superior occipital lobe
LSTJ	Left superior temporal junction	RSTJ	Right superior temporal junction
MPFC	Medial prefrontal cortex	RVIPS	Right ventral intraparietal sulcus

The most obvious result was the altered pattern of connectivity between the amygdala and other brain regions for the two groups. From [19],

> ... the clearest discriminating pattern of activity is that the cyclic connectivity with the amygdala is important in distinguishing

> tinnitus patients from controls. This result is strongly predicted from the extant literature. A review of the literature by [45] emphasizes the broad changes in neural activity across the brain in tinnitus depending on the paradigm used to study the condition. In addition, the data presented here suggests that the ordering between the amygdala and other ROIs is less constrained and more heterogeneous than that of the normal hearing controls ... Changes in amgydalar activity is largely consistent with both the literature on tinnitus as well as in other conditions with some contribution of emotional dysfunction, such as depression. In depression, there is heightened amygdala responses to emotional stimuli [46], but at the same time, there is reduced connectivity between the amygdala and the affective network [47]. The observed changes in cyclic connectivity may contribute to cognitive deficits in attention that have been seen in tinnitus [48] and may affect the overall patterns of network connectivity that change in tinnitus [49, 50]. A loss of connectivity may also correspond with the less constrained pattern of leader-follower relationships involving the amygdala, and future research should seek to better understand this correspondence.

In conclusion, [19] served as a proof of concept that cyclicity analysis could be used to study connectivity patterns in the brain using time series data.

4.2 Shahsavarani et. al

In [20], we continued analysis using the cyclicity method by comparing it with other features like correlation matrices and dynamic time warping matrices in its ability to represent group level and individual specific information on the same data set involving tinnitus afflicted participants and controls. However, this data set was slightly expanded in number because data collection was still ongoing when the manuscripts were prepared[4].

4.2.1 Data collection & processing

Similar to Section 4.1.1 resting state BOLD data were obtained using a 3T Siemens Magnetom Prisma MRI scanner (TR=2000 ms) in two visits, one week apart. At each visit, two 10 minute resting state data were acquired while participants were lying supine inside the scanner with eyes open looking at a white fixed point on the center of a black screen. Further, earplugs were used to reduce the scanner noise. More details about the specifications of the scanner are given in [19]. Data collection was conducted with the approval of the University of Illinois Urbana-Champaign Institutional Review Board and each participant provided informed consent prior to image acquisition in the first visit.

Data were first corrected for slice timing and head motion, and then were normalized to a stereotactic space. Further, the normalized images were spatially smoothed using a Gaussian kernel ($8 \times 8 \times 8$ mm^3). MarsBaR [51], a toolbox for SPM12, was used to extract the time series with 300 time points from ROIs. The time series of each ROI was obtained by averaging the time courses across all voxels within the ROI. After generation of said BOLD fMRI time series, GSR was presented as an optional prepossessing step to be additionally applied. It has been argued that frequencies higher than 0.1 Hz do not contribute to resting state regional coherency as they are often related to physiological noise including cardiac and respiratory cycles [52]. As a result, the time series data were band-pass filtered, using a Bessel filter with low and high cutoff frequencies of 0.008 Hz and 0.08 Hz as a baseline. Before filtering, the time series were mean centered, de-trended, and scaled. Three scaling methods were considered: scaling relative to: the quadratic

[4]This section contains material previously published in [20].

variation, the norm, and the standard deviation of the time series. In the cyclicity analysis, the signals were also end-matched before scaling.

Compared to Table A.1 now the tinnitus group included 50 patients with mean age 52.96 years ($\sigma = 10.29$) of whom 38 had hearing loss and 21 were women. As for the control group, it had 29 participants with mean age 47.75 years ($\sigma = 11.06$) of whome 12 presented with hearing loss and 15 were women.

4.2.2 Motivation

While features like LCM and DM account for the drawback that CM does not take lag-structure into account, they still fail to express information pertinent to the collective temporal ordering of the time series data like cyclicity does [20]. Therefore, we felt it necessary to perform an analysis comparing these features to each other. However, unlike features like CM, LCM and DM which are measures of similarity between pairs of time series collated into a matrix, LMs are features that express an 'average interactivity' (specifically a direction for the leader-follower relationship and a measure of strength to this relationship). Therefore, rather than call it a 'functional connectivity metric,' we lumped all the features together under the term 'functional interactivity metrics.'

A second motivation to conduct the study was to better understand the effect of particular pre-processing techniques on the outcome of the analysis. For example, global signal regression or (GSR) is a particularly controversial step performed in many fMRI studies. From [20],

> Global signal is referred to as spontaneous BOLD fluctuations that are common throughout the brain and defined as the average of BOLD signals over all voxels in the brain [53, 54]. The fluctuations in the global signal have been attributed to non-neuronal origin, especially physiologically induced fluctuations such as changes in the level of arterial carbon dioxide or changes in cardiac rhythm. In the literature, it has been argued that the variances associated with global signal should be removed from the resting state data because it may inflate the connectivity measures, specifically the results from the correlation metric

[54]. Murphy et al. [53], however, showed that regressing out the fluctuations related to global signal introduces spurious negative correlations as well as reducing correlation in some areas. In contrast, Fox et al. [54] argue that using GSR indeed enhances and improves results from correlation analysis. This lack of consensus across studies necessitates the assessment of functional interactivity techniques with and without GSR by each study [55]. In a recent study, for instance, Meszlényi et al. [56] evaluated the efficacy of DTW distance as a rsFC measure by comparing it with correlation (both zero-lagged and lagged) while considering both conditions of with and without GSR. Their results demonstrated that the DTW distance is more robust to employing GSR than the correlation analyses.

Another choice in the pre-processing steps are filter parameters to use on the data, as well as scaling/detrending options. Commonly fMRI signals are band-passed between 0.008 Hz - 0.08 Hz. However as noted in [20] many other studies have used low-pass filtered data as well. To understand the effect of filtering we considered (a) band-pass filter with critical frequences at 0.008 Hz and 0.08 Hz and (b) band pass filter with critical frequencies at 0.008 Hz and 0.2 Hz. One way to understand whether a particular feature has a propensity to represent individual level identifiers is to use the feature to try and 'fingerprint' subjects. A feature that is good at fingerprinting a subject can be thought of as capturing something invariant in the time series for each subject *in time*, i.e. across visits 1 week apart in the vein of the result of Figure 4.3. Continuing from [19], classification was also a topic investigated in [20]. In particular, we considered the performance of different kinds of feature matrices with respect to their ability to aid in classification. Additionally, in this work we also investigated methods of *augmenting* the data. Recall, that in machining learning settings, for example in image classification, rotated, flipped, cropped, and resized images of the same object serve to increase the training data set. Building on this idea we used Variational Auto Encoders (VAEs) to augment the data (see Figure A.3 for details regarding the encoder/decoder model used) and investigate whether the augmented data set could be used to improve the classifiers.

4.2.3 Results

Our results indicated that the stability of the features were in general dependent on the use of GSR and filter parameters [20]. Removing the global signal made all features except DTW distance based feature more stable. In contrast, the effect of band-pass filtering was feature specific. In particular, applying the filters reduced the stability of LM with and without the presence of GSR while for all other features the stability increased, especially in conjunction with GSR. For all matrices, increasing the upper cutoff frequency of the enhanced stability whereas for the scaling method, it did not affect the general trend of stability except for DM where scaling time series using their norms or standard deviation had adverse effects. These results are in Tables 4.4 and 4.5. They show the accuracy rate of 1-nearest neighbor classifier to identify individuals across two scan sessions held one week apart with different filter specifications with and without GSR. The time series were normalized using their quadratic variation. In Figure 4.6 we depict the confusion matrices of 1-nearest neighbor classifiers for the conditions in which each feature matrix had the highest reliability.

Table 4.4: The accuracy rate of 1-nearest neighbor classifier to identify individuals absent GSR.

	No GSR		
Feature	No filter	BPF ($f_{uc} = 0.08$ Hz)	BPF ($f_{uc} = 0.2$ Hz)
LM	0.68	0.45	0.67
CM	0.53	0.53	0.69
LCM	0.48	0.45	0.64
DM	0.32	0.47	0.56

Here BPF stands for the band-pass filter applied and f_{uc} is its upper cutoff frequency and the features are represented by their usual acronyms.

Table 4.5: The accuracy rate of 1-nearest neighbor classifier to identify individuals with GSR.

	With GSR		
Feature	No filter	BPF ($f_{uc} = 0.08$ Hz)	BPF ($f_{uc} = 0.2$ Hz)
LM	0.83	0.56	0.79
CM	0.78	0.82	0.92
LCM	0.76	0.80	0.90
DM	0.28	0.42	0.53

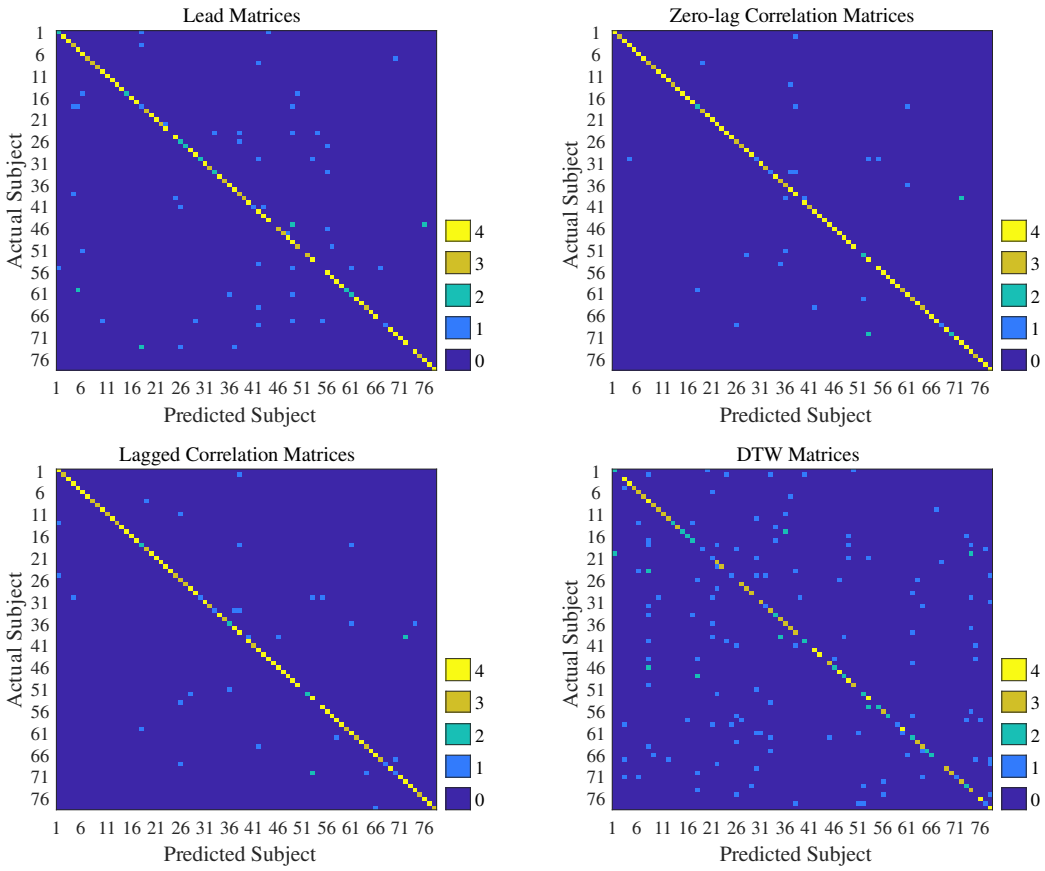


Figure 4.6: Confusion matrices of 1-nearest neighbor classifiers with the highest accuracy for each interactivity feature set used to identify individuals across two visits, held one week apart. DTW stands for dynamic time warping.

On the other hand, to assess whether a feature is better at capturing group level information, traditional classifiers were relied upon to determine the separability of the control data from tinnitus data for (a) each feature and combination of (b) filtering options, (c) with and without GSR. Specifically, we investigated the separability of tinnitus from controls by assessing the accuracy of three classifiers: (1) discriminant analysis, (2) support vector machines, and (3) convolutional neural networks (see Figure A.2 in the appendix). However, the performance of all classifiers was close to or slightly better than chance, indicating a poor dissociation between tinnitus patients and healthy controls regardless or pre-processing and filtering options - at least in the realm of classical machine learning methods. Augmenting the dataset with VAEs did not serve to improve the classification results. Nev-

ertheless, there were some differences across classifiers and input feature matrices. That is, the performance of SVM on LCM (sensitivity of 64 % and specificity of 54%) surpassed the performance of all other classifiers. Based on this we examined which ROI pairs (i.e. elements of the feature vector) were providing the slightly superior classification ability. Recall [20],

> ... the hinge-loss will choose a combination of the ROI interactions that form a maximum margin separating hyperplane between the two sets of labels because each element of our vector corresponds to the interaction between a pair of ROIs. However, in the sparse SVM formulation, the hinge loss is used in conjunction with a ℓ_1 norm on w. This causes the optimization to have a trade-off between maximizing margin while maintaining classification accuracy and sparsifying w. Thus, it selects those ROI interaction pairs that determine the cardinal orientation of the hyperplane (to maximize accuracy) while reducing the number of components (by sacrificing on the margin), and thereby gives us interaction pairs that play a dominant role in differentiating the two classes.

Based on the dominant features found, the feature vectors were further dimension reduced and classification tried again. We found that classification improved and SVM now had sensitivity and specificity of 62% for correlation matrices. We visualized the interaction between ROI pairs found by using graph methods. We first did this by first grouping together the ROIs into their 'large-scale brain networks' (see Table 4.1) which included the limbic system, visual network, dorsal attention network, default mode network and auditory network. Each ROI is represented by a node and the interaction between ROIs is depicted as an edge. The width of each edge signifies the importance of the pairwise interaction to seperability, which corresponds to the absolute values of the weight vector obtained in the SVM solution. See Figure A.4 for an example, and also to discover roughly what ROIs constitute what networks. This allowed us to study at a coarser level what interactions between *networks* (as opposed to individual ROIs) were helpful in discrimination. Such analysis was done for each of the feature matrices involved in the study and figures for those not shown here, i.e. Lagged Correlation Matrices, Lead Matrices, and Dynamic Time Warping matrices may be found

in [20].

Correlation analysis:

For both the correlation based methods, we found that seperating subject groups relied upon the interaction between the auditory network and the limbic system. From [20],

> The central auditory regions are neuroanatomically connected to the limbic system, which evokes emotional responses to auditory stimuli. In turn, the limbic system interacts with the auditory system, which regulates sound perception based on emotional processing. Previous studies have reported on tinnitus-related alternations in the interaction between the auditory network and the limbic system [50, 57, 49, 58, 59]

The fact that we could replicate observed results with correlation based features vectors may not seem surprising; however, cyclicity analysis also added to the results.

Cyclicity analysis:

Using lead matrices it was found that the interaction between precuneus and visual cortices contributed to the seperation between the groups [20]. Further the left frontal eyefield and parahippocampus were also found to be important in being able to distinguish between the two conditions. This serves as a second proof of concept that cyclicity analysis has its own merits. Though the precise figure from [20] is omitted for brevity, we note:

> Precuneus is a major hub in the default mode network and its interaction with other networks indicates that the coherency of the default mode network is disrupted in tinnitus. In fact, several studies have reported tinnitus-related decrease in the coherency of the default mode network [60, 61, 62, 49].

Dynamic time warping:

In DTW analysis the interaction between the left frontal eye field (LFE) and medial prefrontal cortex (MPFC) was found to be useful in differentiating tinnitus participants from healthy controls. In addition, similar to cyclicity analysis, the interaction between the limbic system and the dorsal attention network was found to be different between the tinnitus and healthy control groups [20]. The former replicates findings that tinnitus is correlated with increased functional connectivity between default mode network (containing MPFC) and the dorsal attention network (containing LFE) [50].

Finally, we also tried to visualize the space of feature matrices using the t-SNE method from Section 3.2.3 and obtained in the case of correlation based features and lead matrices, an embedding which seemed to span the space (from light to dark for correlation matrices) and two clusters when using DTW matrices. No such clustering was obtained at any scale using lead matrices. This embedding exercise was repeated with and without GSR, and we observed that GSR, no matter the filter setting, changed both CM and LCM into a continuum and obscured the boundary between the two clusters.

Figure 4.7 depicts CM and DM of both tinnitus and control populations after applying t-SNE, using band-pass filtering (0.008 Hz$\leq f_{\text{pass}} \leq$ 0.2 Hz) and with & without GSR. Each observation is shown with its corresponding feature matrix. Further investigation revealed that this observed clustering was not correlated with the subject group labels or any other metadata, such as hearing status, age, or gender.

In conclusion, the methods and results in [19, 20] served to establish cyclicity method as an useful avenue for analysis for time series data, which for the purposes of this project was capable of replicating extant results. On the other hand, with regards to bio-markers of tinnitus results suggest that newer methods of analysis maybe required to achieve traditional classification/clustering like results using fMRI data. We used this data (both paradigms) with cyclicity analysis in [21] to detect cortical waves of activity propagating in the brain with consistent leader-follower relationships. In particular, in the task paradigm we observed short bursts of directed temporal ordering between pairs of regions via Eq. 3.9.

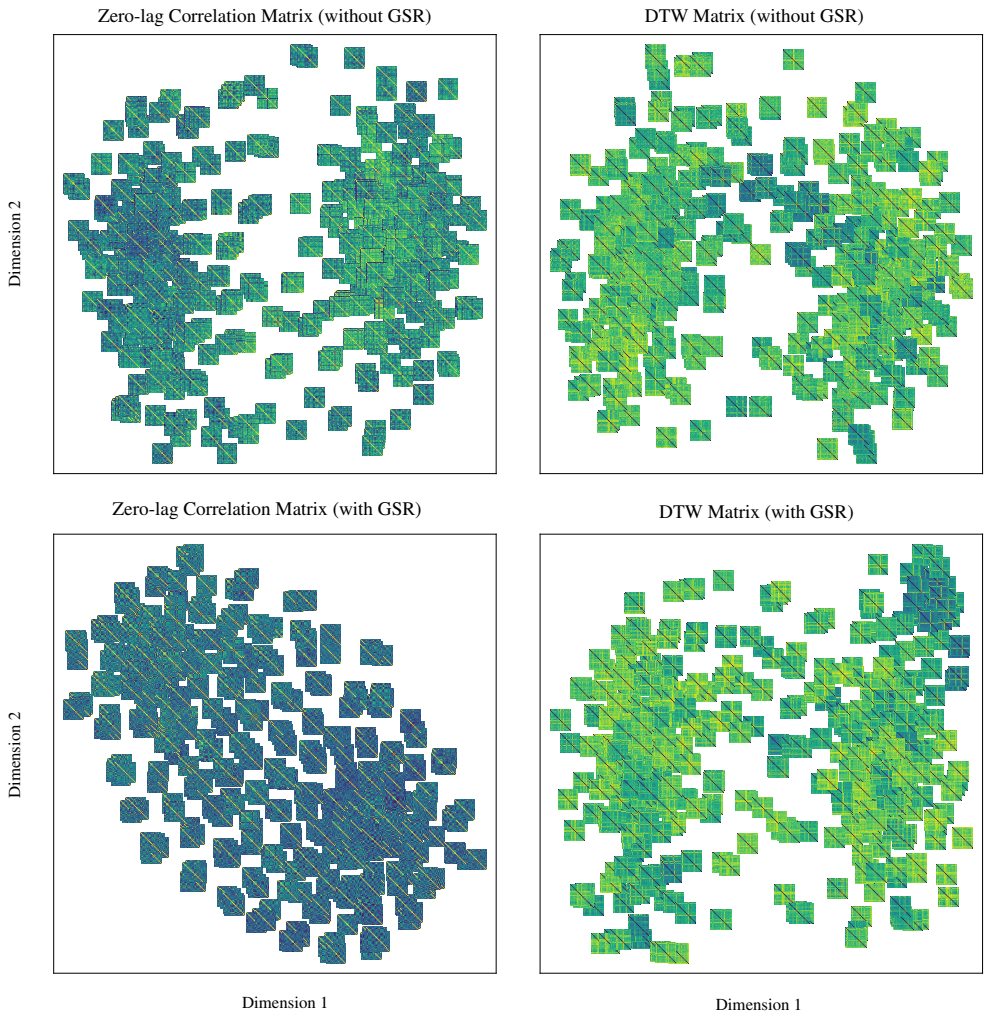


Figure 4.7: Nonlinear dimensionality reduction of feature matrices using tSNE.

4.3 Human connectome project data

While the data examined in the previous two subsections were primarily concerned with fMRI scans of tinnitus afflicted populations, the data in the following section was collected from healthy individuals aged 18 - 45 years. Moreover, the data used in [19, 20] were resting state fMRI data while the Human Connectome Project also includes data that was recorded in the task paradigm[5].

[5]The material in this section was previously published as a preprint in [21].

4.3.1 Data collection/processing

From the Human Connectome Project (HCP) website, we utilized data from $M = 862$ participants who ranged in age from 22-45 years. The data was collected at the Washington University in St. Louis, USA using a Seimens Skyra 3T MRI scanner. From the HCP S1200 release we only considered participants who had completed all structural, resting state and task paradigm scans and did not have any segmentation issues noted by the HCP Quality Control Process. The final data was denoised, minimally preprocessed and contained 464 females. Connectome Workbench software was used to extract time series corresponding to regions of interest (ROI) in the provided fMRI parcellations.

These ROIs were 34 in total and are listed in Table 4.6. Note that the dimensionality of the time series then was 68 accounting for bilateral symmetry. With a time resolution (TR) of 720 milliseconds, each resting state scan then resulted in multidimensional time series of that could be represented by arrays of size 68×1200.

ROI #	ROI full name	ROI #	ROI full name
1	Banks of superior temporal S.	18	Pars orbitalis C.
2	Caudal anterior cingulate C.	19	Pars triangularis C.
3	Caudal middle frontal C.	20	Pericalcarine C.
4	Cuneus	21	Postcentral C.
5	Entorhinal C.	22	Posterior cingulate C.
6	Fusiform C.	23	Precentral C.
7	Inferior parietal C.	24	Precuneus
8	Inferior temporal C.	25	Rostral anterior cingulate C.
9	Isthmus cingulate C.	26	Rostral middle frontal C.
10	Lateral occipital C.	27	Superior frontal C.
11	Lateral orbitofrontal C.	28	Superior parietal C.
12	Lingual C.	29	Superior temporal C.
13	Medial orbitofrontal C.	30	Supramarginal C.
14	Middle temporal C.	31	Frontal P.
15	Parahippocampal C.	32	Temporal P.
16	Paracentral C.	33	Traverse temporal C.
17	Pars opercularis C.	34	Insula

Table 4.6: Table of 68 ROIs involved in the analysis, showing numbers for left-side regions; right-side regions run their indices as 35 through 68. Here **C** - cortex, **S** - sulcus and **P** - pole.

4.3.2 Motivation & methods

While [19, 20] were intrinsically motivated to analyze fMRI scans with the objective of discovering biomarkers for tinnitus, the HCP data in [21] are all from healthy individuals. Therefore, the objective with the use of this dataset was to investigate how the cyclicity pipeline analysis fared when considerably larger datasets were fed through it (the M above is 10 times larger when compared to [19, 20]). Note that the increase in the size of the dataset is both in terms of the intrinsic as well as extrinsic dimensionality which necessitated the creation and adoption of new techniques to the pipeline.

First, lead matrices were generated from the collection of time courses D and their eigenstructure analyzed. These matrices have dimension 68×68 with each (i, j) entry denoting the average leader-follower relationship between ROI i and ROI j. See the right panel in Figure 4.8 for a representative example of a generic lead matrix. Recall that the $|\lambda_1|/|\lambda_3|$ ratio (here λ_k are eigenvalues of the lead matrix) is a measure of the dominance of the rank-two approximation described in Section 3.1.This quantity was computed for all lead matrices. A subset, $S \subset D$, of the data was identified by restricting the lead matrices to have $|\lambda_1|/|\lambda_3|$ ratios one standard deviation above the group average $\mu_D\left(|\lambda_1|/|\lambda_3|\right)$. For a visualization of the distribution of the four leading eigenvalues in D see the left image in Figure 4.9. The right image shows the distribution of $|\lambda_1|/|\lambda_3|$ ratio across D. We can see that S roughly amounts to 1/6th of all scans analyzed. The leading eigenvectors for this subset of the data was further examined. Recall that each entry v^i of the eigenvector v of a lead matrix corresponds to one of the $N = 68$ ROIs whose time courses it was created from. Since skew-symmetric matrices only admit zero or purely imaginary eigenvalues λ_k and corresponding eigenvectors v_k, the elements of an eigenvector v_k can be visualized as a *constellation* $v_k^i = \left(x_k^i + \sqrt{-1} \cdot y_k^i\right)_{i=1}^N$, a point cloud in the complex plane with each of the N points representing the complex numbers v_k^i, one for each ROI (recall we saw this earlier in the right panel of Figure 3.3).

In such a constellation, the points farthest from the origin correspond to ROI time courses that dominate the mode corresponding to the eigenvector v_k. To identify such dominant ROIs in a principled fashion, we used the following heuristics. Recall that in the case of a single harmonic, the constellations align along ellipses in the complex plane. In the general case, the

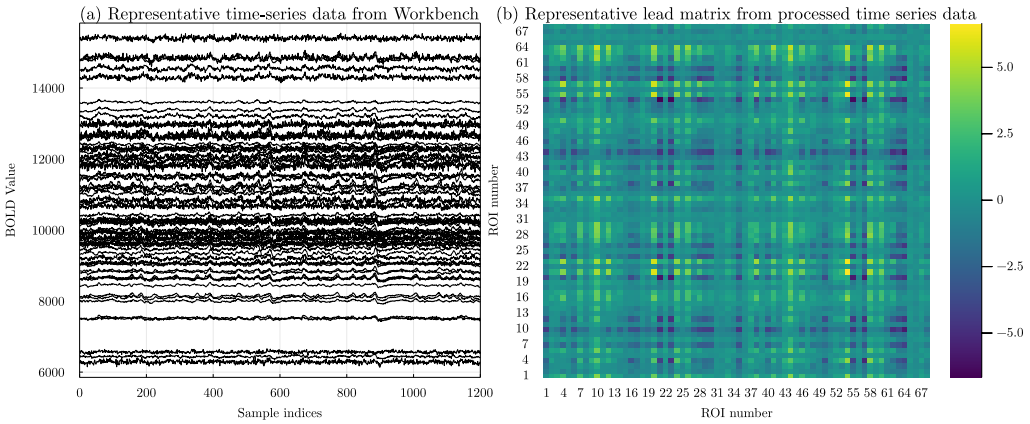


Figure 4.8: Representative time course data from the HCP dataset after processing using Connectome Workbench (left) and corresponding lead matrix (right). The lead matrix is generated after an appropriate normalization of the BOLD signal (see [19, 20]).

ellipse maybe perturbed, but one can argue that it still can be recovered, for example using least square regression. The best fitting ellipse defines a positive definite quadratic form on the complex plane, with its value evaluating to one on the ellipse. This form allows us to distinguish the significant points in the constellation (those where the quadratic form evaluates above a threshold, say the unity) vs. the insignificant ones (below unity, i.e. inside the ellipse). Therefore, one can select a collection of ROIs dominant with respect to a particular eigenvalue. We concentrated solely on the leading eigenvalues, i.e., the largest in absolute value, denoted $\lambda_{1,2}$ in our convention [21]. See the left panel of Figure 4.10 for a representative example.

4.3.3 Results

This analysis resulted in the identification of a subset R, of 14 consistently dominant ROIs for the time courses in S.

Robustness of ROIs:

Since the 14 ROI subset R was generated by considering the restriction of the data set to S it important to verify that this dominance was a feature common to the entire dataset D. To this end, random subsets $\overline{S}$ of the data

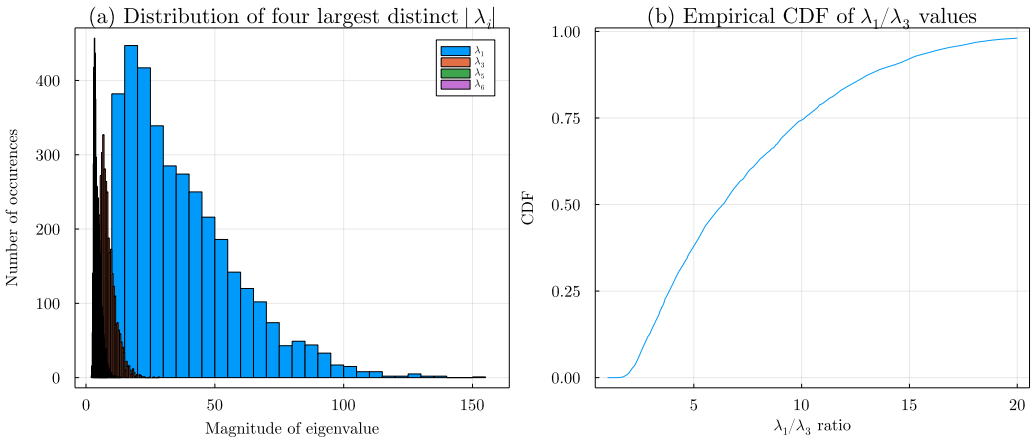


Figure 4.9: Frequency of the absolute values of four largest non-conjugate eigenvalues of the lead matrices in the data (left) along with the cumulative histogram of observed λ_1/λ_3 values (right). Higher λ_1/λ_3 values indicate more reliable outcomes from the Cyclicity Analysis pipeline.

of size $\overline{N} = 299$ were run through the pipeline instead of S and the reported dominant ROIs were tracked. Figure 4.11 below is a visualization of this as a matrix - each row corresponds to a random subset $\overline{S}$ and the values along the rows represent the dominant ROIs reported from examination of $\overline{S}$ - with the color scheme set to aide visual identification of change in values between rows. The average Levenshtein distance between each pair of rows in such a matrix (across multiple trials) was found to be less than 2 illustrating that the set R is indeed robust with respect to the choice of the subset $\overline{S}$.

Average leader-follower relationship:

As we noted in Section 3.1, the cyclicity analysis pipeline is able to extract an approximate *cyclic ordering* of the phase shifts in the Chain-of-Offsets-Model from the oriented areas rendering the leader-follower relationships between pairs of time series. It is *cyclic* because the ordering of the repeating sequence of events (such as the excitation of a node in Figure 3.4) has no intrinsic starting point; only the relative positions of the phases are of import. Therefore, to compare between obtained orderings, one needs a consistent way of presenting them. Towards this end, one can construct an axis joining the origin to the center of mass for each constellation. Choosing this as the abscissa, the cycle is then assumed to be proceeding counter clockwise

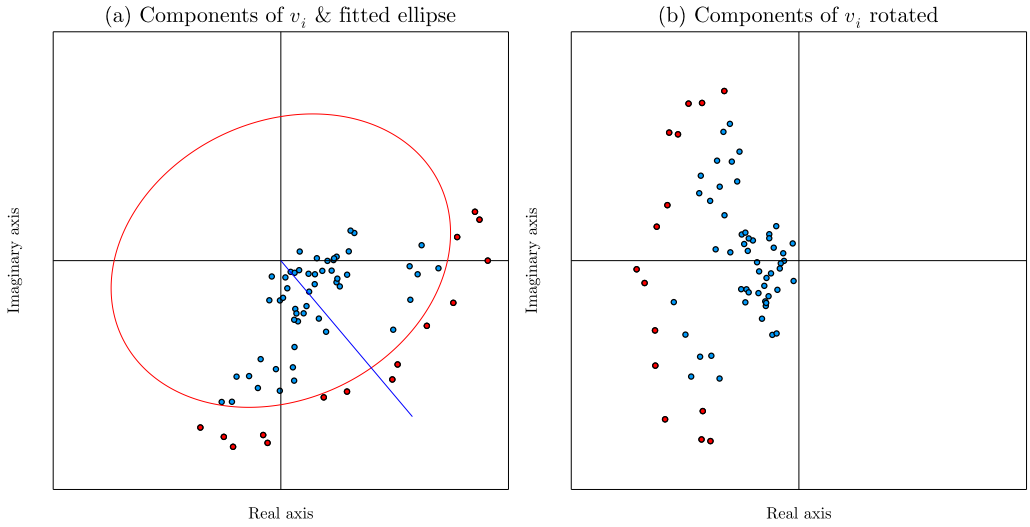


Figure 4.10: This figure shows one step in the determination of dominant ROIs in Cyclicity Analysis. The left panel shows components of the leading eigenvector visualized on the complex plane. Each point corresponds to an ROI and its BOLD signal; with greater absolute values indicating greater dominance in the multidimensional time-series. The right panel shows the constellation of points rotated about the axis identified in the left panel to provide a consistent ordering across samples (said axis being determined by the center-of-mass for the collection).

from this axis. This is shown in Figure 4.10 where the left panel shows a sample constellation of the eigenvector components along with the above-mentioned axis. The right figure shows the same constellation after the standard rotation. Applying this procedure for the regions in R to S resulted in estimated cyclic orderings for these dominant ROIs. These obtained cyclic orderings can be visualized in a square *permutation count matrix* shown on Fig. 4.12. In this matrix, the rows represent fixed ROIs and the columns represent positions in the cyclic ordering. Therefore, the (i, j) entry of this matrix represents a value showing how many times ROI i showed up in position j in all the cyclic orderings obtained. Using this matrix, it was possible to estimate the *average cyclic ordering* for the ROIS in R as shown in Table 4.7. To do so, one first obtains a list of average positions by considering the values in each row of the matrix as samplings from a distribution and estimating their mean. Then, the permutation vector that sorts this list of mean values will correspond to the average cyclic order across the subset S.The stability of signal propagation across the significant ROIs in the order

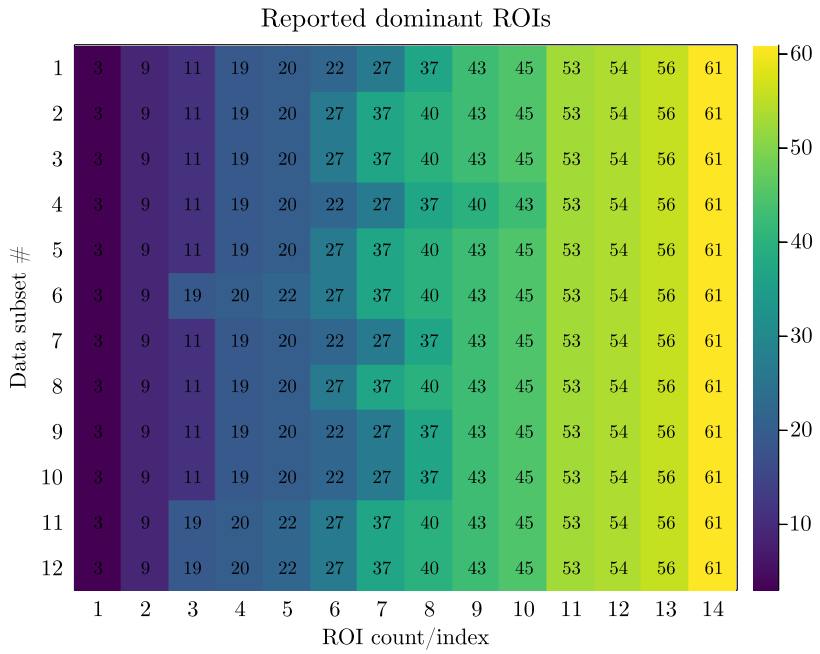


Figure 4.11: Figure showing reportedly dominant ROIs in 12 random (disjoint) partitions of the dataset. Each row of this matrix represents the most dominant ROIs reported by restricting the analysis to a random subset of the data (size $N = 299$). One can see that regardless of the subset, the same ROIs are reported to be dominant by the analysis.

shown in the Table 4.7 is one of our main findings [21].

Area integral dynamics:

As noted towards the end of Section 3.1.1, one can create yet another time-series by observing the value of Eq. 3.9 (i.e. the area involved between a pair of time series as it is built up). Figure 4.13 shows this idea in greater detail. The top left panel shows a pair of time-series corresponding to a pair of ROIs from the dominant set R.The top right panel shows the value of the above mentioned area integral computed between the pair of time-series signal. A consistent increasing or decreasing trend over the period of observation shows that there is an average leader (or follower) relationship between the pair of time series - i.e. activity in one region precedes (or lags) activity in the other region. These short periods of time where there is a significant contribution to the increase in the area integral can be termed "events" or significant periods of "directed activity" between the two pairs

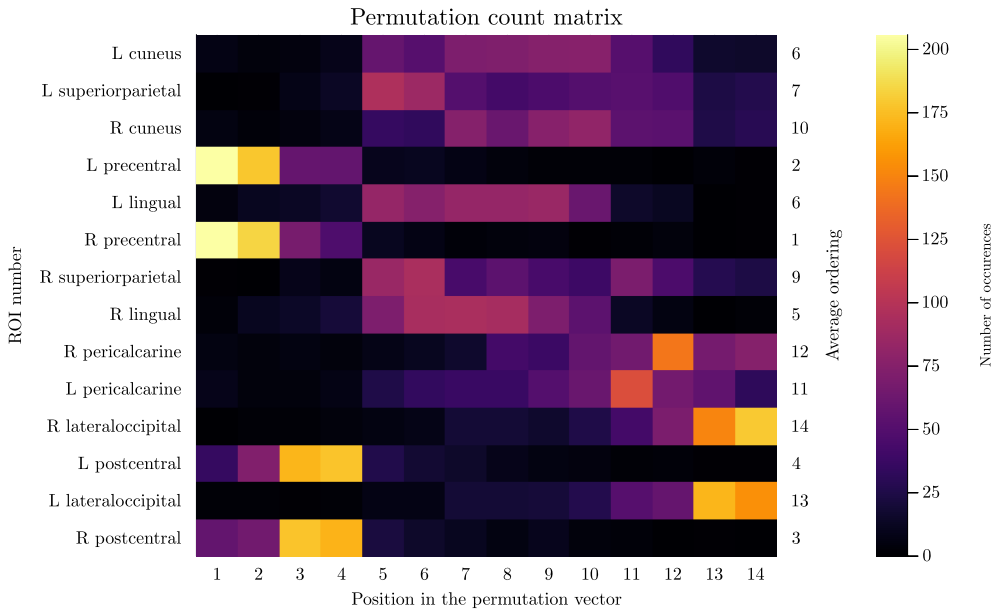


Figure 4.12: The *permutation count matrix* obtained from S by restricting ROIs considered to R. Each row of this matrix represents observed positions of the ROI in the cyclic ordering. Highlights indicate that ROI corresponding to that row *consistently* showed up in a particular position (along the columns) in the obtained cyclic orderings.

of brain signals. To methodically extract such instances in time, for each frame f_k of the time series, successive intervals of lengths 3-30 (i.e. intervals:

$$[f_k, f_{k+3}], [f_k, f_{k+4}], \ldots, [f_k, f_{k+30}]$$

were obtained and their slope computed). These slopes obtained at each frame are visualized in the bottom left image of Figure 4.13. Half of the maximum slope observed among all such intervals over all frames was taken to be a nominal threshold and *significant* jumps or contributions were deemed to be those instances when the slope of the area integral was higher than said threshold[6]. Periods of jumps identified in such manner are shown marked in red in the bottom right image of Figure 4.13. Such analysis provided for a striking observation: namely, that in the scans examined from the task paradigm (in particular the social and motor tasks), there were periods of strongly directed lead-lag relationships across multiple pairs of ROIs through-

[6]See Appendix, Section B for another alogorithm to determine significant jumps in the time series data.

ROI full name	Ordering in cycle	
bilateral order	Right	Left
Precentral cortex	1	2
Postcentral cortex	3	4
Lingual cortex	5	6
Superior parietal cortex	9	7
Cuneus	10	8
Pericalcarine	12	11
Lateral occipital cortex	14	13

Table 4.7: The fourteen dominant ROIs (including bilateral pairs) obtained and their average cyclic ordering in the dataset. The ordering shows that bilateral pairs frequently occur together in the cycle.

out the brain [21]. Recall that the tasks were as follows. For the **motor task** [63]:

> Participants are presented with visual cues that ask them to tap their left or right fingers, squeeze their left or right toes, or move their tongue to map motor areas. Each block of a movement type lasts 12 s (10 movements), and is preceded by a 3 s cue. In each of the two runs, there are 13 blocks, with 2 of tongue movements, 4 of hand movements (2 right and 2 left), 4 of foot movements (2 right and 2 left) and three 15 s fixation blocks per run.

whereas for the **social task**[63]:

> Participants are presented with short video clips (20 s) of objects (squares, circles, triangles) either interacting in some way, or moving randomly. These videos were developed by either Castelli and colleagues (Castelli et al., 2000) or Martin and colleagues (Wheatley et al., 2007). After each video clip, participants chose between 3 possibilities: whether the objects had a social interaction (an interaction that appears as if the shapes are taking into account each others feelings and thoughts),Not Sure, or No interaction (i.e., there is no obvious interaction between the shapes and the movement appears random). Each of the two task runs has 5 video blocks (2 Mental and 3 Random in one run, 3 Mental and 2 Random in the other run) and 5 fixation blocks (15 s each).

In Figure 4.14, we show the general trend observed with the resting and task paradigms. Area integral dynamics could either be plateaued or changing in the resting paradigm depending on the particular pair of ROIs chosen. Recall that the order listed in Table 4.7 is visually represented in the permutation

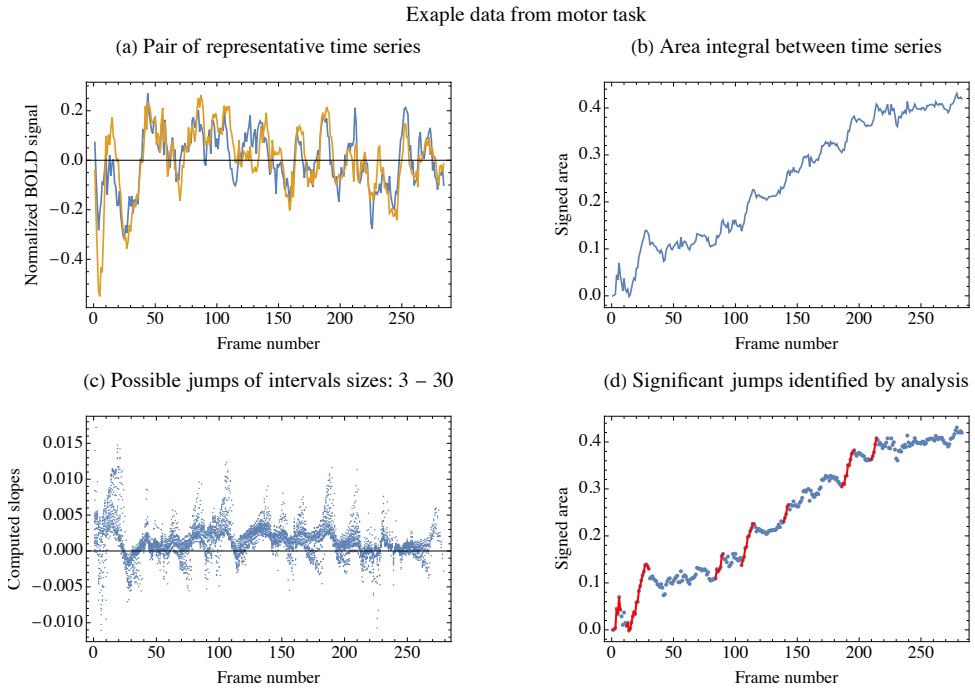


Figure 4.13: These panel shows how intense periods of directed activity between a pair of time series was identified. For each pair of timeseries as shown in the top-left panel, the area integral of Eq. 3.9 is computed as in the top-right figure. The slopes for possible jumps at each frame are identified and a threshold applied to designate significant jumps that contribute to the greatest overt increase in the value of the area integral.

count matrix of Figure 4.12. Referring to the Figure 4.12, we see that six out of fourteen ROIs are strongly constrained to be in particular positions of the ordering (which was obtained from resting state data). For resting state, integral area dynamics was highly likely to be increasing or decreasing when the pairs were chosen from the above mentioned six (c.f. panel (A) vs. (C) of Figure 4.14). With the data in the task paradigm, the dynamics depended on the task being analyzed. On one hand, for social task data (see panel (B) of Figure 4.14) periods of directed activity were interspersed with reversal in said direction in concordance with the five fixation blocks (see task description above). On the other hand, for the motor task no such general pattern was observed (see an exemplar of a generic pair in panel (A) of Figure 4.14). The observations made above are generally observable by visualization of all pairs of area integral dynamics at once as shown in Figures 4.15 and 4.16.

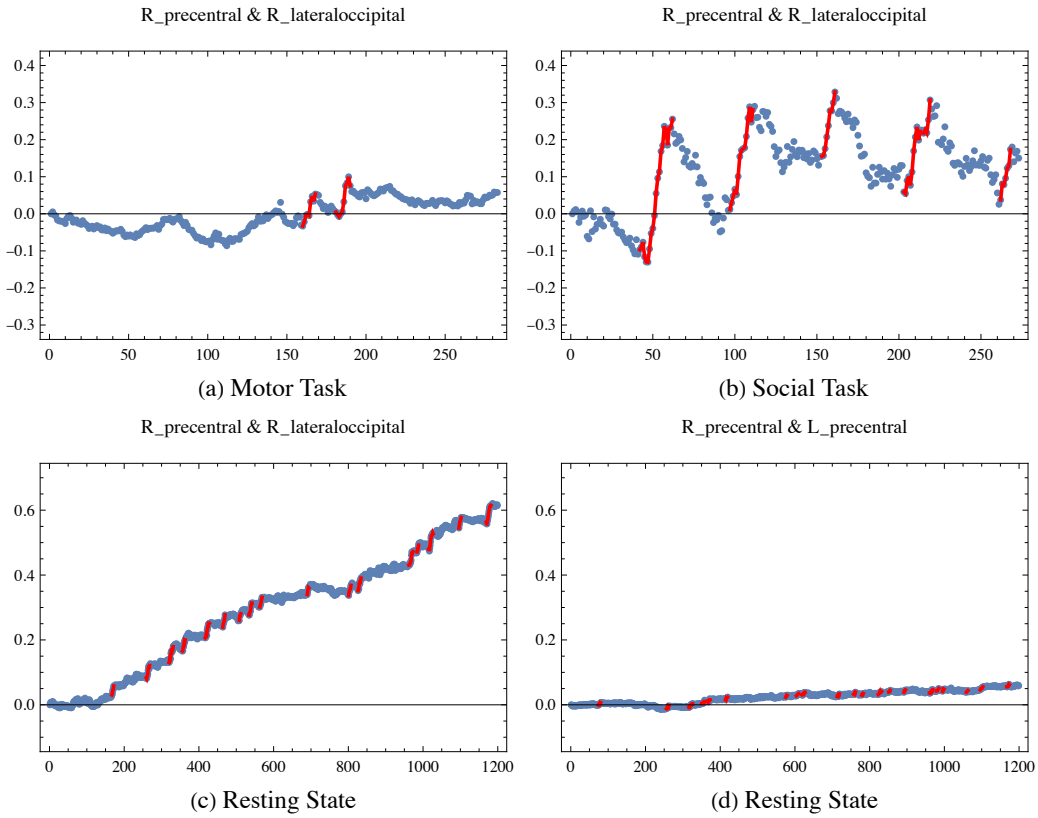


Figure 4.14: Panel of area integral dynamics. In general, the social task (B) was most likely to show periods of strongly directed activity across all pairs of task while the motor task (A) was most likely to show plateaued dynamics. Resting state data showed increasing or decreasing dynamics depending on the pairs chosen - (C) is a strongly directed pair while (D) is a weakly directed pair.

4.4 Discussion & conclusion

We conclude this chapter with a hybrid section discussing the results and concluding this current chapter dealing with fMRI data since results presented in Sections 4.1.3 and 4.2.3 were extensively discussed in [19, 20]. In [19], we analyzed the resting state data from the healthy controls and hearing disorder affected populations using cyclicity analysis. The goal there was to identify objective biomarkers of the tinnitus condition from fMRI data. We found that the tinnitus condition was characterized by more variable temporal patterns in regions connected to the amygdala. We hypothesized that this heterogeneity in the temporal patterns could be due to differences in levels of habituation to tinnitus. Some tinnitus patients may have re-

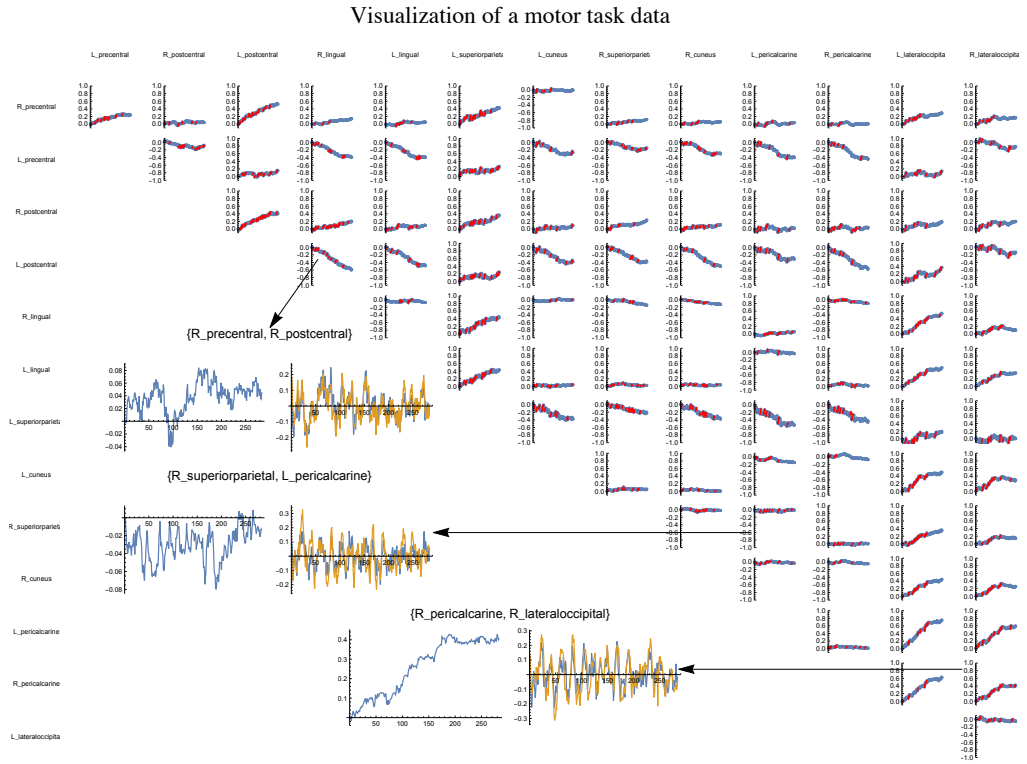


Figure 4.15: A way to visualize directed leader-follower activity between pairs of dominant ROIs obtained in the data set. A definitely increasing or decreasing trend indicates a strongly constrained directed-leader follower activity between two pairs while a variable trend indicates activity that is less constrained. In the top inset image, the blue signal leads the orange one whereas, in the bottom inset, the relationship is reversed.

duced amygdalar activation following habituation, whereas others may have increased amygdalar activation to the bothersome internal noise. Neuroscientific hypothesis for this observation and connections to extant literature were discussed in depth [19]. Traditional machine learning approaches were used to distinguish the two populations. Linear and quadratic discriminant analysis, partial least squares discriminant analysis, linear and quadratic support vector machines were the tools used. We used the lead matrix as a feature summarizing the BOLD series. We found that we could distinguish between the two with 67% accuracy using partial least squares analysis coupled with Wilks lambda criterion to choose pairs of ROIs. We also found that the lead matrix was a feature that was suitable to fingerprint the subjects in the data set. That is, we could identify the same individual across two sessions with 70% accuracy using a 1 nearest neighbor classifier with the cosine similarity

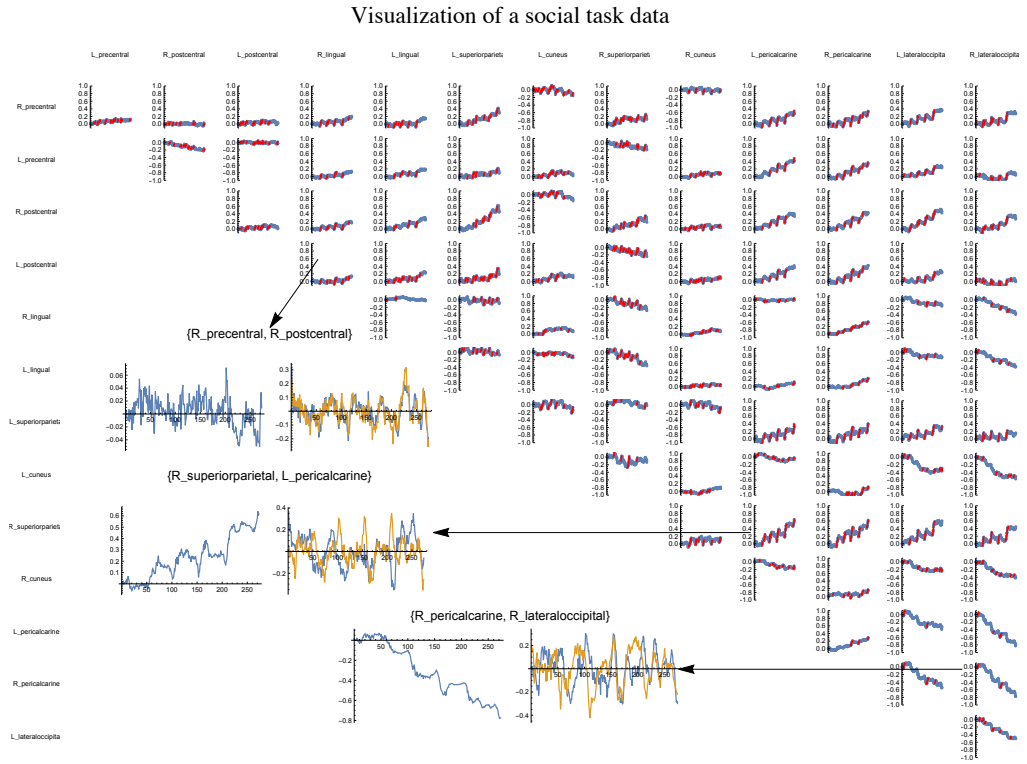


Figure 4.16: Visualization of directed activity between ROIs observed in the social cognition task. Compared to resting state or motor task scans, intermittent bursts of directed activity are a more prevalent feature in this analysis.

criterion. While group level classification was poor in [19], we noted that the study was limited by the small number of subjects in the data set. Future work here would be to combine anatomical and funtional data along with behavioral metadata collected.

While [19] was focused on the cyclicity pipeline, in [20] we investigated how different methods of generating dimension reduced features from the BOLD time series stacked up against each other on a slightly expanded dataset. In particular, we examined correlation matrices, lagged correlation matrices, dynamic time warping matrices and lead matrices and further investigated the effect of pre-processing options on the results with regards to group level classification. The preprocessing options explored were different filtering options and for correlation matrices, global signal regression. Due to the limited sample size of the data, a secondary objective of the work was to examine data augmentation methods within the context of time-series data using variational auto encoders [20]. Further, since previously in [19], we showed that

lead matrix as a feature is one suitable to be used as a *fingerprinting* method, we also explored the ability of other features to be used as fingerprints. More precisely, we utilized the fingerprinting ability of a feature generation method as a measure of its robustness/stability to variation in the preprocessing options [20]. In general, none of the four feature generation methods provided particularly good results with regard to being able to separate the two populations regardless of preprocessing and/or filtering options. With regards to fingerprinting we found that the extent to which matrices remained stable across the visits varied depending on the employment of global signal regression (GSR) and the filters as outlined in Section 4.2.3.

The major contribution of the work presented in Section 4.3 was the application of cyclicity analysis pipeline to one of the largest fMRI datasets currently available and extensions to methods via analysis of the dynamic nature of each component of the lead matrix. Possible future work includes the adoption and adaptation of network inference techniques that span the fields of graph, information, utility, and systems theory as well as geographic information systems.

Chapter 5

Electromyography data

As mentioned in the introductory chapter, on of the major problems in the field of kinesiology is one of addressing the problem of *redundancy in motor* control.

The classical approach is to adopt a multi-level theory of movement control [5, 6, 7], one level of which involves "synergies," which is the name for the mechanism by which coordinated activity amongst a set of effectors takes place. Within this model, *synergies* are patterns of joint torques or muscle activations that are either hardwired or learned with practice over time and help eliminate redundant degrees of freedom. In animal studies, it is possible to directly stimulate the nerves and spinal cord (SC) to elicit motion in the limbs & appendages. That neuronal stimulation leads to consistent and repeatable motor activity has been the central finding of such studies. These consistent and repeatable movements are taken to be *motion primitives*, various combinations of which were thought to generate all observed motion [64, 65, 66, 67, 68]. While it is possible to implant electrodes in animal muscles to record changes during the progression of some motor task, by far the most common method of detecting muscle activations in healthy humans is by the recording of electromyography (EMG) signals. Three kinds of EMG recordings are possible, (a) one in which small needle electrodes are directly implanted into a small localized region of the musculature, (b) another less localized one in which recordings are made from the surface of the skin, and (c) a more recent method utilizing a multitude of arrayed surface sensors to locally provide information about the depth at which activation occurs in the musculature [69, 70]. The latter two methods are advantageous when it is preferred to avoid penetration of the skin and musculature. It has been shown that EMG signals are directly proportional to the level of muscle activations in isometric contractions and therefore used as a standard proxy to detect muscle activity in tasks of that nature [71]. Then, one approach to the

solution of the problem of redundancy in motor control is to understand how one can relate muscle activation potentials observed as EMG signals, with forces & moments generated by extensions or contractions of the muscle so as to uncover synergies present in the system.

Various theories and models have been introduced for this purpose. In the next section we discuss some major approaches and show how their assumptions and/or reliance on linearity may hinder them from uncovering synergies in the higher dimensional space of EMG signals used to characterize muscle activity in motor tasks. In Section 5.2 we present a simple motivating example for a situation in which such *linear* methods fail. In Section 5.3 we discuss our approach to data collection and recording, and in Section 5.4 we will describe the preprocessing done on the data. Finally, in Section 5.5 we present the results of applying aforementioned nonlinear methods to our data set.

5.1 Classical models and linearity

Since the central problem being addressed is the mechanism by which the huge parameter space available to the CNS is organized and coalesced into subsets that act in concert, it is natural to assume that the data being recorded during a motor tasks is one that can be described by a low dimensional model. Various studies have borne out this assumption [72, 73, 74, 75]. Therefore, the objective then becomes to perform an appropriate dimension reduction on the recorded data, so as to recover the underlying lower dimensional structure. See [76] for a survey of dimensionality reduction techniques and [77] for models and techniques particular to neuromuscular function.

In [78] and [79], the authors introduced a paradigm to mathematically concretize the notion of synergies in motor control. Under this approach, the controller, (i.e. the combination of the brain & nervous system), rather than eliminate redundant degrees of freedom, makes use of them to extremize some functionally important performance variables. The controller assembles all the elemental variables[1] pertinent to the task at hand so as to compensate for each other's spontaneous errors and/or unpredictable environment variables.

[1] In the sense of configurable parameters or basic units, e.g. muscle lengths, joint angles, etc.

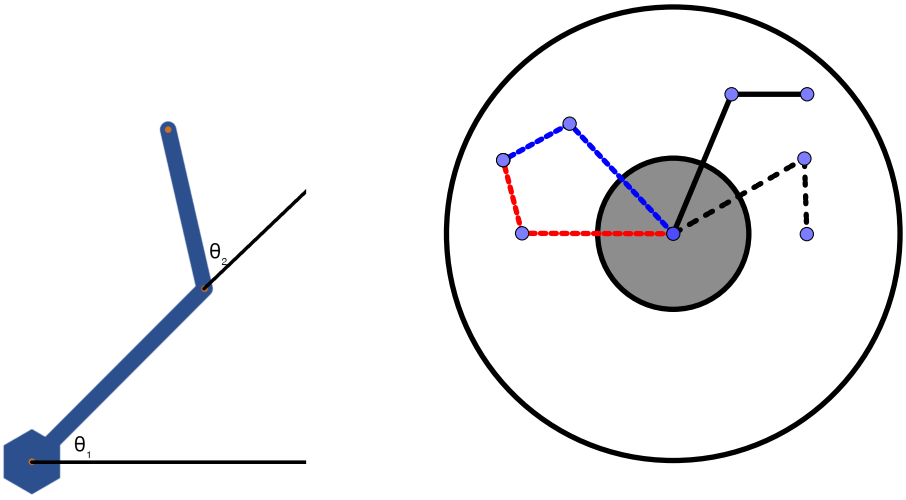
The '*uncontrolled manifold hypothesis*' (UMH) posits that the elemental variables are organized in such a way as to restrict their values to a subspace in the space of feasible ones - termed the *uncontrolled manifold*. Under this restriction, the variability of performance variables is maximized along certain *good* directions (which provide robustness) while being minimized in *bad* directions (which reduce errors) and *synergies* arise from coordination of elemental variables within this subspace. Other related theories include the minimal intervention principle (MIP) [80, 81] and optimal feedback control approaches [82, 83]. The former posits a control scheme that *approximates* a certain desired trajectory while allowing for errors in the non-task related directions. Such errors are ignored and not corrected for unless detrimental to the task performance. They argue that indeed correcting for errors in non-task related directions may incur higher costs on the controller and degrade task performance. The related latter theory casts the problem of redundancy in the framework of linear optimal feedback control.

The above and other classical approaches (e.g. principal component analysis, non-negative matrix factorizations, factor analysis etc.) used to uncover synergies in the space of EMG signals (which arise from elemental variables) are linear in nature. The UMH approach utilizes a Jacobian to map between elemental variables and the motor task space. MIP while formulated in the general form of a stochastic optimal control problem, readily admits closed form solutions only in the linear quadratic or other special cases. Even recent frameworks like that of *neural manifolds* [75], while admitting the underlying manifold maybe nonlinear, only examine planar local approximations in the neural space for each task mode. However, there is no reason to expect a-priori that synergies or coordination in biological systems are of a linear nature. The primary reason for a preference for linear frameworks has been that often, truly nonlinear phenomena are also well-approximated by linear ones. A classic example is the oscillating pendulum whose equation of motion is nonlinear but excellently approximated by a linearization. This phenomenon has also been observed in the field of neurophysiology [84, 85, 86, 87]. No doubt, another reason for our reliance on our repertoire of linear methods is one of computational simplicity and interpretability. Before the age of computers and the advent of numerical methods, this was a significant practical criterion that dictated how well the tools could be developed and/or put into practice. However, such methods can be agnostic to the source of the data

and can fail in scenarios where truly nonlinear behavior needs to be studied. For example, in the field of mechanics of materials, hydrodynamics, biology etc. essentially all phenomena are nonlinear in nature. Therefore, a reliance on purely linear methodologies may be overly restrictive in the pursuit of uncovering the workings of bio-mechanical systems. The next section describes simple situations where this may happen.

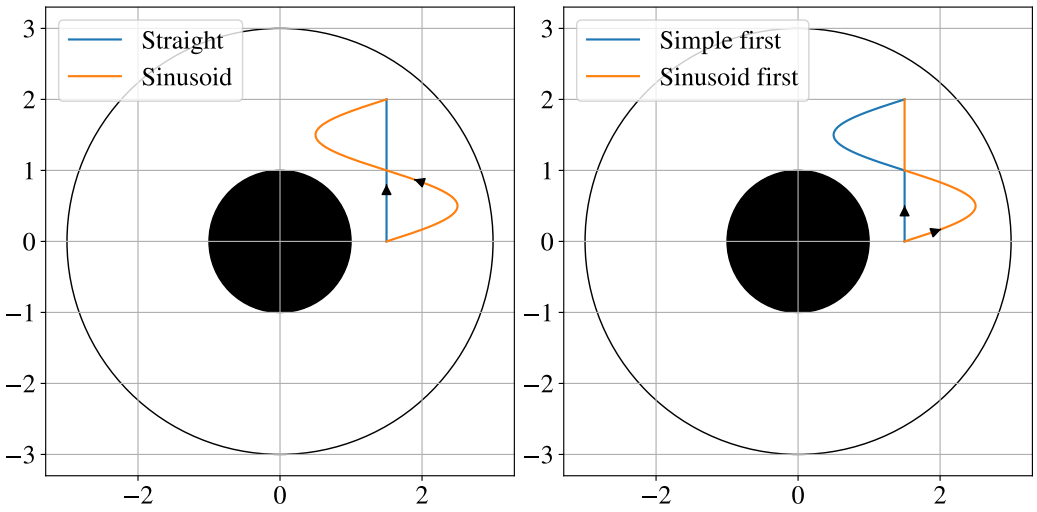
5.2 Motivation and examples

While the example of the swissroll data set may seem contrived, we now present an example from mechanics where essentially the same phenomenon is observed. Consider the two bar linkage shown in Figure 5.1(a), whose end-effector may take up position anywhere in the white unshaded region as shown in the Figure 5.1(b). Now consider a task where the end-effector needs to move from an initial point on the horizontal axis to a point that is vertically above it (see the black lines in Figure 5.1(b)).



(a) A two bar link (b) Feasible workspace

Figure 5.1: A simple mechanical example with two bars connected by a single revolute joint and the end of one bar being fixed. The left figure shows the joint angles and how they are measured while the right figure shows the feasible workspace for the end effector. Not that any point in the feasible workspace (excluding the two circles themselves), admit a pair of joint configurations as depicted by the red and blue lines.



(a) Two different trajectories for end-effector (b) Combination of trajectories on the left

Figure 5.2: Multiple different trajectories for moving the end effector of the two-bar linkage of Figure 5.1(a) from a point on the horizontal axis to one above it (see Figure 5.1(b)).

Let us consider then different ways of achieving said objective. The end effector could move smoothly in two different ways between the two starting and ending points as shown in Figure 5.2(a). On the other hand, nothing prevents us from instructing the manipulator to perform a more complicated maneuver as shown in the right panel of the same figure. Here we have combined the straight line motion with that of the sinusoidal motion, with a choice to perform the former first or second. The joint angles θ_1 and θ_2 will then be appropriately varying functions of time to facilitate this motion.

Suppose the *two modes* are the vertical straight line motion and the sinusoidal one as shown in the Figure 5.2(a), setting aside the more complicated case of Figure 5.2(b) for the rest of the section. Solving the inverse kinematics of the two-bar linkage one finds that the joint angles behave as shown in the two panels of Figure 5.3 - the top is for the simple motion and the bottom is for the sinusoidal mode. Note that since any point within the white annulus of Figure 5.1 (b) admits two pairs of joint angle configurations, there arise two solutions - (θ_1, θ_2) and (θ_1', θ_2') - for the values of the joint angles corresponding to each trajectory/mode; here we depict only one of these for simplicity.

Whereas in human activity EMG signals are in direct correlation with the level of muscle activation for isometric contractions, in this simple mechan-

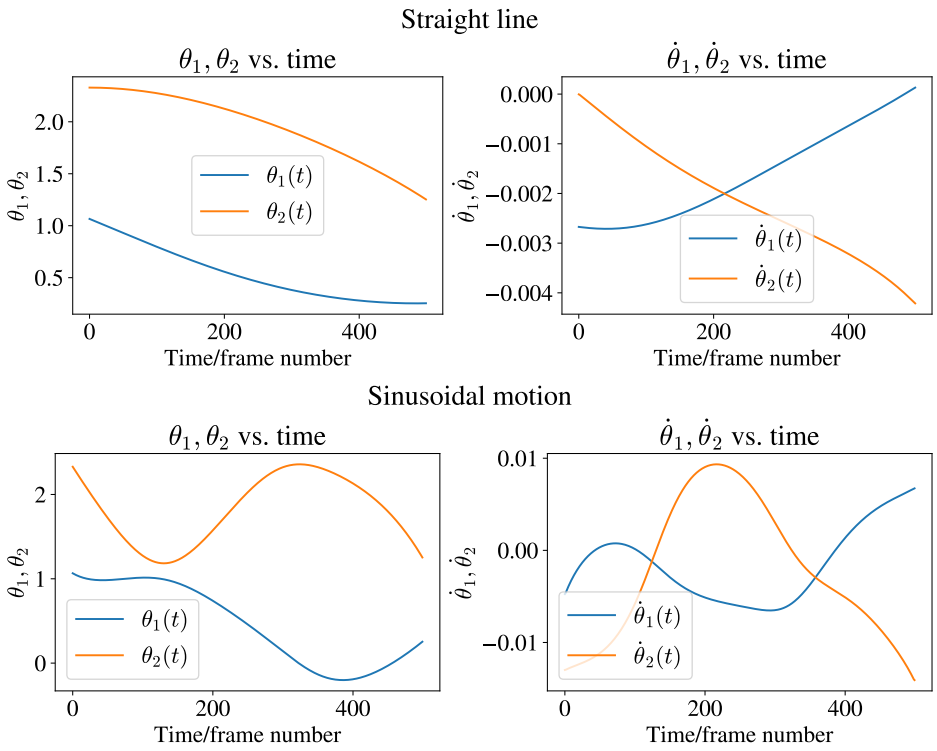


Figure 5.3: Plot of the joint angles & derivative quantities in the simple mechanical system for the trajectories shown in Figure 5.1(a) for one set of solutions.[2]

ical example, we can consider the derivatives of the joint angles as a proxy for joint activation - that is, we consider $\dot{\theta}_1, \dot{\theta}_2$ and $\ddot{\theta}_1, \ddot{\theta}_2$ as our measured signals. Application of PCA to our set of measured signals shows that a standard clustering algorithm like $k-$means would fail to separate the two *motor modes*[3] present. In Figure 5.4, a set measurements $\dot{\theta}_1, \ddot{\theta}_1, \dot{\theta}_2, \ddot{\theta}_2$ of two different trajectories in Figure 5.2(a) were fed into a standard PCA pipeline. This set corresponds to the two rows (a) and (b) in Figure 5.3 which shows one set of solutions possible to the inverse kinematics of Figure 5.1.

The figure shows that in PCA coordinates, there are two distinct *connected components* corresponding to the two different trajectories. However, they are not linearly separable in any fashion by standard techniques. In this simple example, even if we did not possess prior knowledge of the end-effector trajectories, visual inspection of the PCA data alone was sufficient to indicate the presence of two different connected components. However, in systems with more complicated dynamics when we have multi-channel data, such

[3]A crude analogue of *synergy* for the mechanical example.

may not be the case and we need algorithms to automatically detect the synergies. When linear separability is unattainable in simple examples, one cannot hope they find success in more involved ones. This necessitates that we look for nonlinear techniques for such datasets.

Towards this end, Figure 5.5 shows the result of applying standard *eigenmap* techniques followed by $k-$means clustering on our set of measurements $\dot{\theta}_1, \dot{\theta}_2, \ddot{\theta}_1, \ddot{\theta}_2$. In this case the two different colors correspond to the two different trajectories (straight vs. sinusoid) and the lower/upper panels correspond to different solutions (i.e. the left and right columns of Figure 5.3).

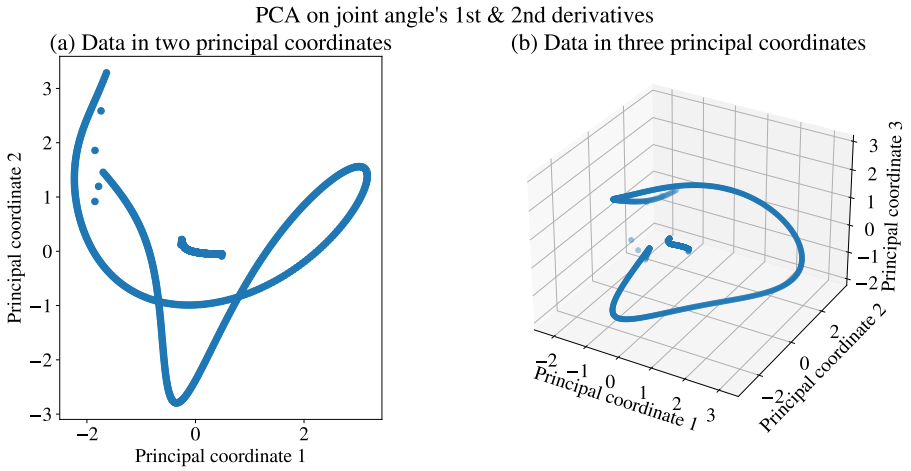


Figure 5.4: Application of PCA to derivatives of joint angles in Figure 5.3. **(L)** In two principal coordinates (PCs) and, **(R)** in three PCs. The two major connected components in the PC space correspond to the two different *motor modes* in Figure 5.2.

Here we can see that in *eigencoordinates*, k-means algorithm is able to identify clusters easily, whereas it struggles with the data in principal coordinates. Recall that PCA for dimensionality reduction, followed by k-means for clustering, is a popular choice for a clustering pipeline [88, 89]. However, *both* being linear and closely related methods [90], this combination fails on this example whereas the nonlinear method succeeds. Once again, in the above example, it is fairly obvious what is missing is an emphasis on the local structure of the data as opposed to the global one implied by a linear change of coordinates. Note that central to the idea of learning the local structure of the dataset is the idea of *distance* between two points. In the *swissroll* dataset of Figure 3.7(a), the key observation is that there an intrinsic notion of distance between two points (*on* the manifold) which is

distinct from the extrinsic notion of distance between two points (with respect to the ambient space). Successfully learning the manifold (or in effect being able to *unroll* the swissroll), therefore, will depend on techniques that rely on the notion of intrinsic distance as opposed to that of the extrinsic distance. Similarly, using a notion of some $k-$ nearest-neighbors to discover the distinct connected components, and using the neighborhood information pertinent to each point to create a new coordinate system, in which the two synergies get mapped far apart is possible. Figure 5.5 shows the result of one such approach when applied to the quantities in Figure 5.3.

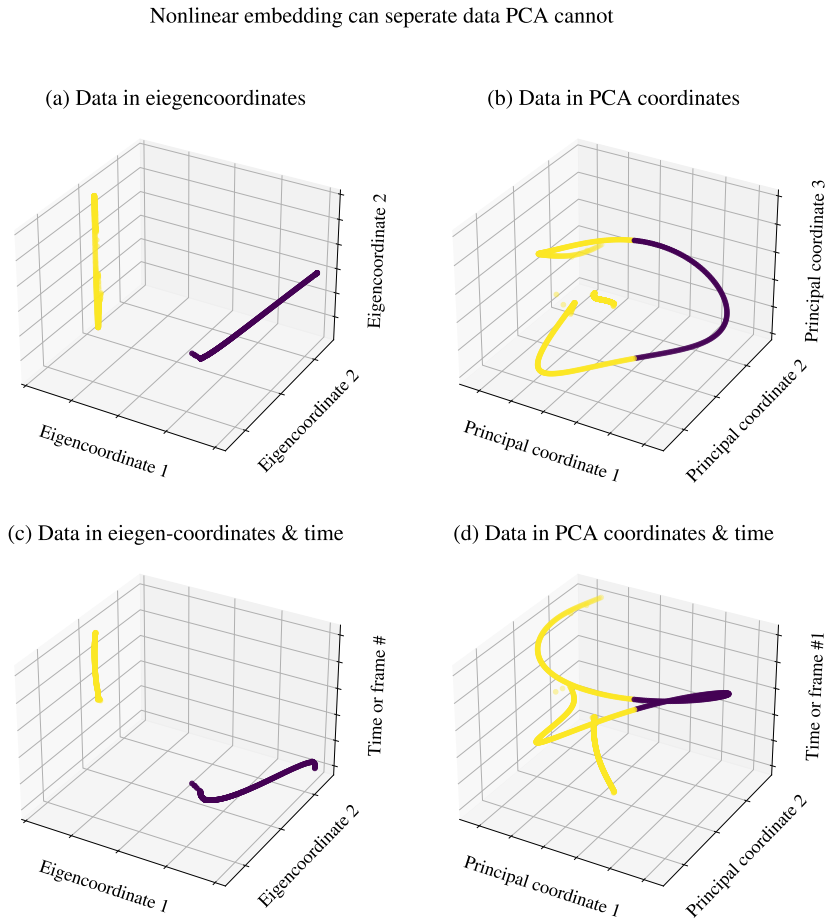


Figure 5.5: Example of nonlinear embedding technique followed by k-means algorithm being able to separate *synergies* in the observed variables for the example of Figure 5.1 while PCA fails.

5.3 Data collection

Data was collected from 11 adult participants with no history of neuromuscular disorders who gave informed written consent. All procedures were approved by the local Institutional Review Board for the protection of human subjects in research. Two participants had to be excluded because they did not complete the task as instructed. The task involved holding a predefined western classical ballet pose (second position of the feet in demi-plié and arms in first position) and exerting force using either the *forearm* or *upper arm* with a horizontal *pull* or *push* effort against a six channel force transducer (ATI Delta 660, Apex, NC, USA) held rigidly at the height of the individual's upper extremity in the described pose. The particular ballet position was chosen so as to standardize a whole body static position while voluntarily activating either the shoulder join or the elbow joint for the push and pull efforts.

Data was recorded using 32 wireless Delsys surface EMG sensors (Trigno, Delsys, Natick, MA, USA) supported by a Qualisys camera system interface (Oqus 5+ Wide, Gothenborg, Sweden). Sixteen of the EMG sensors were placed on the left side of the participants body, and sixteen were placed on the right side. The force transducer, was mounted on a wall with z-axis perpendicular to the face of the transducer, which could be adjusted vertically for the height of each participant. The force and torque recordings were made at 100 Hz while EMG was recorded at 1 kHz. The EMG channels and corresponding muscles are listed in Table 5.1.

Table 5.1: EMG Channels and corresponding muscles. All channels have bilateral symmetry, i.e left and right pairs.

Ch. #	**Muscle**	**Ch. #**	**Muscle**
01	Middle deltoid	09	Semitendinosus
02	Triceps brachii	10	Adductor magnus
03	Biceps brachii	11	Rectus femoris
04	Extensor carpi radialis	12	Vastus medialis
05	Flexor carpi radialis	13	Anterior tibialis
06	Upper trapezius	14	Medial gastrocnemius
07	Infraspinatus	15	Rectus abdominus
08	Latissmus dorsi	16	Erector spinae

We first measured the maximum voluntary contraction (MVC) for each muscle - that is the maximum force a particular muscle could generate and

recorded the corresponding EMG values [91]. Then participants were asked to complete four tasks, upper-arm push, upper-arm pull, forearm push, and forearm pull. For each of the tasks the force transducer was strapped to the body part involved in the task. The pull or push task for each upper limb segment was done ten times continuously - i.e. for example push-relax, push-relax etc. ten times. In doing so, the participants were asked to use 30% of the previously recorded maximum voluntary force magnitude for each of elbow extension, flexion, shoulder horizontal abduction, or adduction according to the task presented. They were also given real-time visual feedback on a screen as an aid.

(a) (b) (c)

Figure 5.6: (a) Frontal (b) lateral and (c) posterior views of a forearm push/pull task

5.4 Data processing

To preprocess the data, we first eliminated high frequency noise using a 6th order bi-directional low pass Butterworth filter with a cutoff frequency of 100 Hz. Note that it is commonly believed that motor control processes do not exceed 40 Hz though studies have shown that the corner frequency for the low pass filter affects the total variance accounted for by synergies [92], and therefore, a relatively high corner frequency was chosen. Further, a notch filter of 60 Hz was also applied to eliminate electrical and electronic noise. Each filtered and demeaned EMG channel's data was then normalized with respect to the maximum EMG signal value attained during the aforementioned MVC measurement or with respect to the maximum attained during the task itself. To identify regions of force generation in the recorded data,

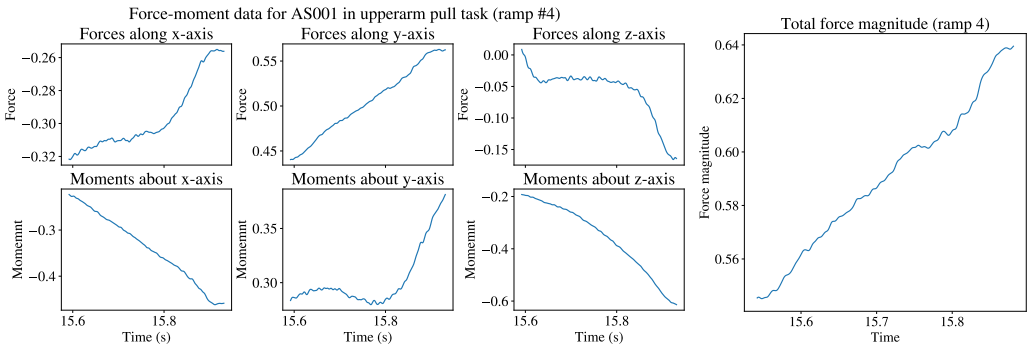


Figure 5.7: Exemplar of force-moment data that is identified to constitute a single 'ramp', i.e. a montononic increase/decrease in the force magnitude.

we began by identifying the end of force generation in total force value (i.e. magnitude of force, including all f_x, f_y and f_z components). That is, we identified peaks in the total force (signifying the beginning of a relaxation period) and worked our way backwards to the beginning of force generation. Due to noise factors present, there can be small multiple local peaks between the major ones found. We took as one *ramp-up* a clean 70% drop from the identified peak value going backwards in time. This start time was noted in the force data, and from that time, we subtracted 50 milliseconds to account for electro-mechanical delay. Each ramp was then visually inspected for monotonicity in force magnitude, with monotonicity enforced by truncation to a monotonic region. See Figure 5.7 for an exemplar ramp. Within each of the ramp intervals, the EMG data was visually checked. If there were large spikes in both the positive and negative directions for a given muscle, which would indicate a sensor malfunction, the corresponding channel was bilaterally removed from dataset. Finally, PCA and the nonlinear methods of Section 3.2.1 and 3.2.2 were applied to each ramp's EMG data.

5.5 Results

The procedure described in the previous section first results in data visualized as 'ramp-ups' in Figure 5.8 for one sample. Next, the portions corresponding to the identified ramps in the force-moment data were extracted from the EMG recordings of the 32 channels described in Section 5.3. When visualized, this data looks like Figure A.5 which shows multiple ramps identified and colored differently. Further, one can examine each of those ramps by itself

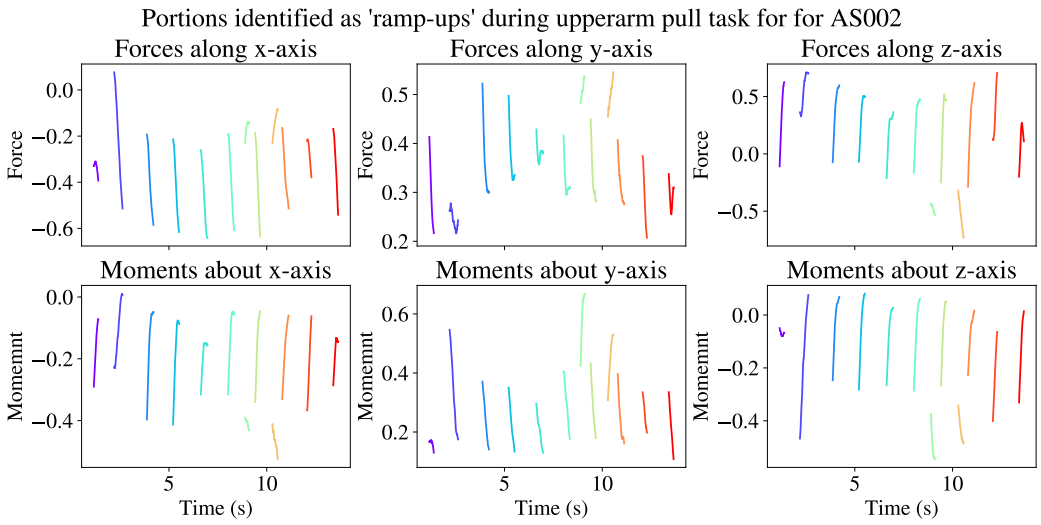


Figure 5.8: Examples of *ramp-ups* of force generation extracted from analysis of total force magnitude as described below. In the corresponding EMG signal data only data corresponding to these ramp-ups are examined.

in detail. This produces the plots shown in Figure 5.9. Each of these figures are representative images generated from the data recorded during a single task performed by a single participant.

5.5.1 Principal component analysis

Suppose that for a single task performance of a subject, the above procedure resulted in k ramp-ups being identified from the data.

Let $n_i, i = 1, \ldots, k$ denote the number of samples recorded within each ramp-up. Then it is possible to perform principal component analysis (PCA) on the extracted data in one of two ways:

(a) Stack all the ramp-ups together to form a $32 \times N$ data matrix where $N = \sum_{i=1}^{k} n_i$ and perform PCA on it

(b) Perform PCA on each data matrix D_i with sizes $32 \times n_i$ for $i = 1, \ldots, k$.

Both approaches were adopted. A prototypical example of the results obtained from the latter approach is shown in Figure 5.10. Unlike Figure 5.4 where we were at least able to see distinct connected components in the principal coordinate space, in Figure 5.10 we fully see the limitation of the

linear factor analysis methods. On the other hand, in the next section we will see markedly different results in the eigencoordinate space of the nonlinear methods.

Figure 5.9: Visualization of a single ramp EMG content after pre-processing.

It is also possible perform PCA via the former approach, i.e. stack together all ramps identified during the completion of one task (see for example, Figure A.6). However, in this case the number of components required to explain most of the variance invariably increased and visualizations were not helpful. On the other hand, since PCA is often used as an unsupervised clustering

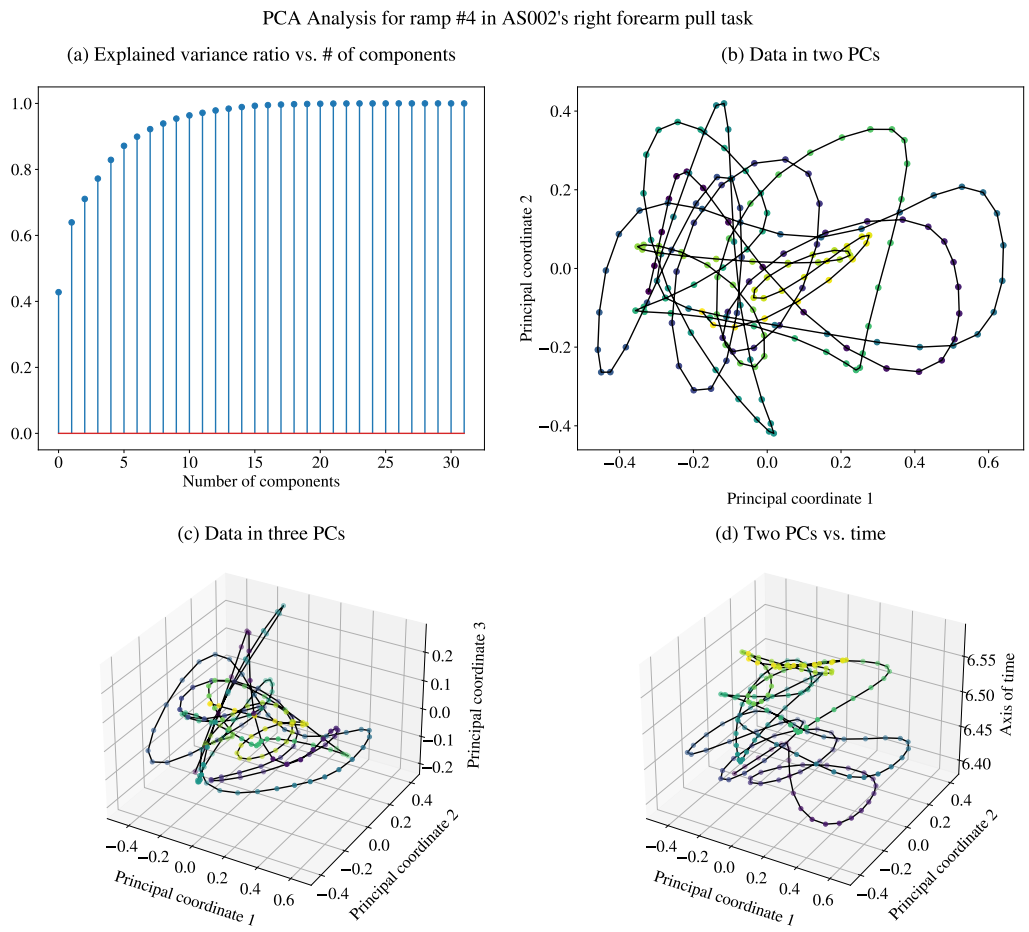
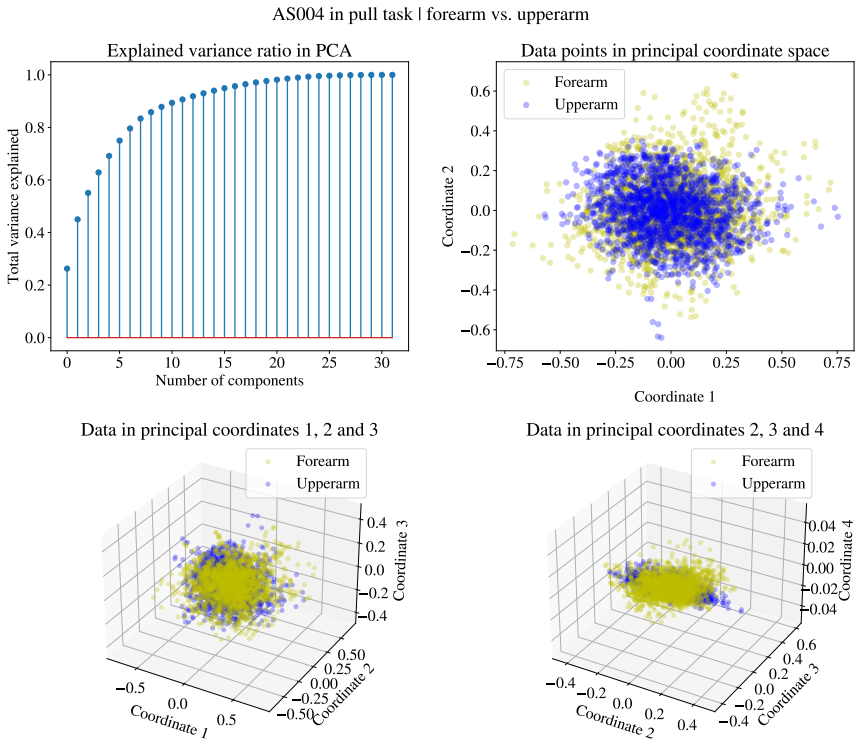
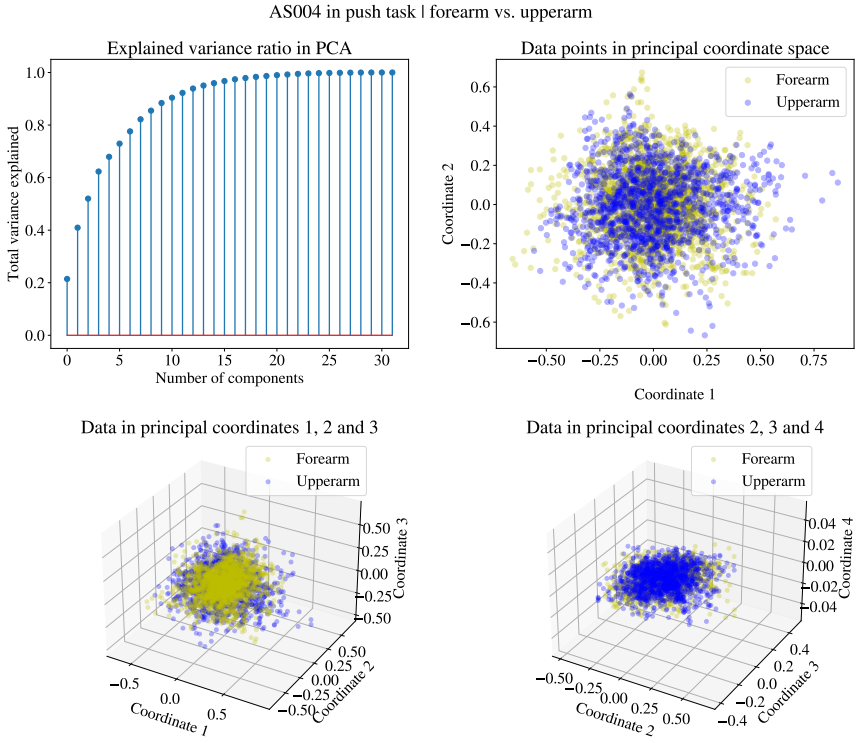


Figure 5.10: PCA analysis performed on ramp #4 identified during AS002's forearm pull task as shown in Figure 5.9. (a) Shows the total variance accounted for as a function of the number of principal components, (b) and (c) show visualization of the data in the principal coordinates and (d) shows the first two principal coordinates against time. In the latter plots, lighter colors occur later in the task.

algorithm, it was instructive to check whether the four different tasks would cluster differently. Recall that the four tasks are (a) forearm push, (b) forearm pull, (c) upper arm push, and (d) upper arm pull. Therefore, one can create four combinations of data sets for each participant, and run them through PCA to see how they would cluster. Results of this process are visualized in Figures 5.11 and 5.12. It is clear that PCA is not able perform unsupervised learning, i.e. clustering based on differences in the muscular activation when different body parts perform the same task, nor is it able to distinguish between different tasks performed by the same part of the limbs.

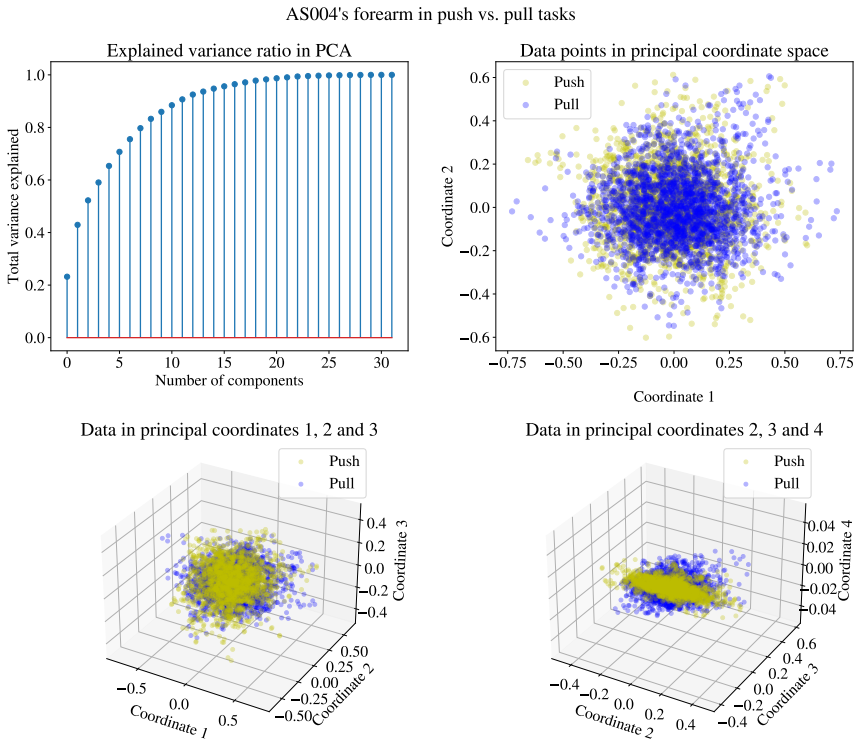


(a) Pull task - upperarm vs. forearm

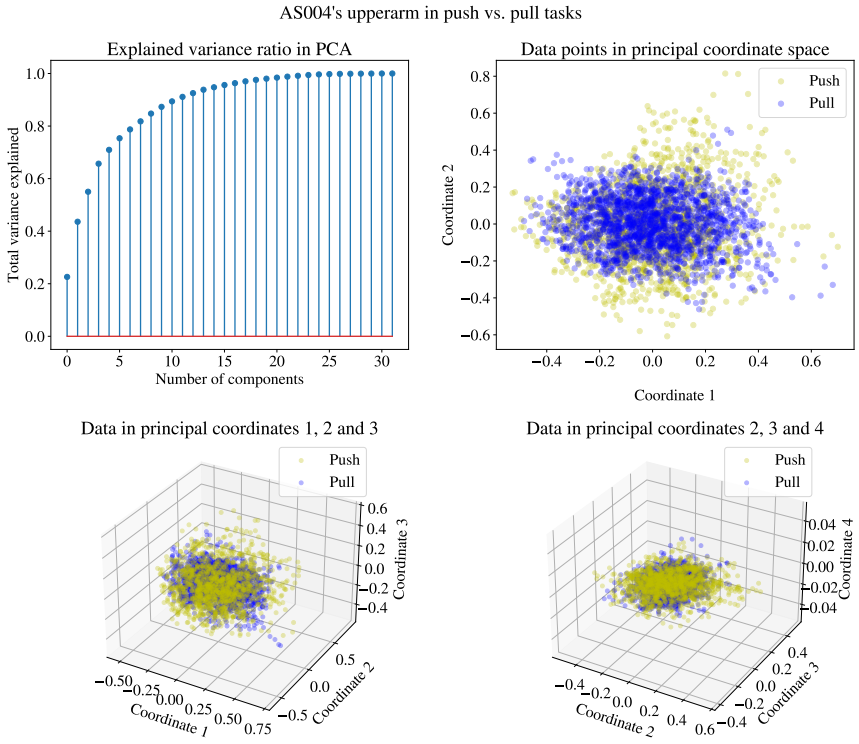


(b) Push task - upperarm vs. forearm

Figure 5.11: PCA on *all* the ramps identified during AS002's push and pull tasks.



(a) Forearm comparing push vs. pull



(b) Upperarm comparing push vs.pull

Figure 5.12: PCA on *all* the ramps identified during AS002's forearm and upperarm tasks.

5.5.2 Eigenmap & diffusion map analysis

The Laplacian eigenmap algorithm described in Section 3.2.2 was performed on the ramps ups identified in the data. Compared to PCA, the eigenmap algorithm was successful in clustering the EMG data. This is shown in Figure 5.13. Here some salient features of the eigenmap algorithm are observable.

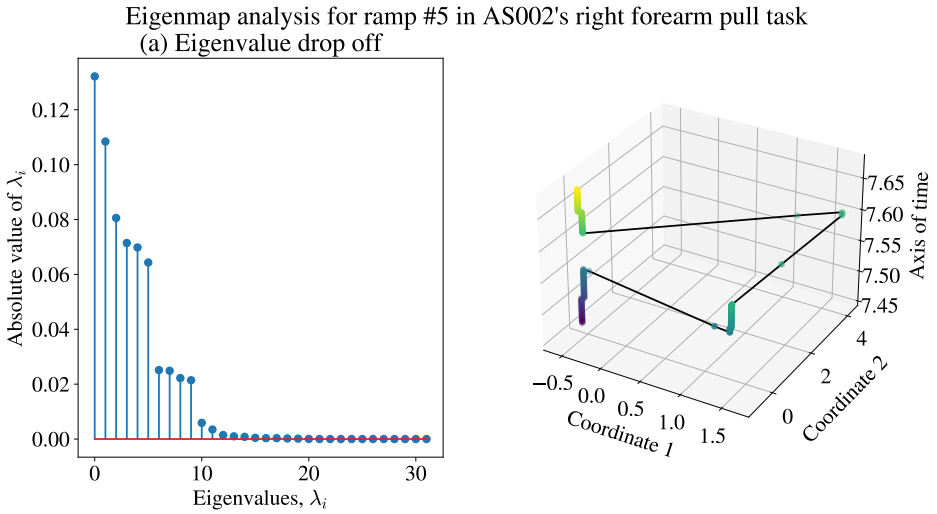


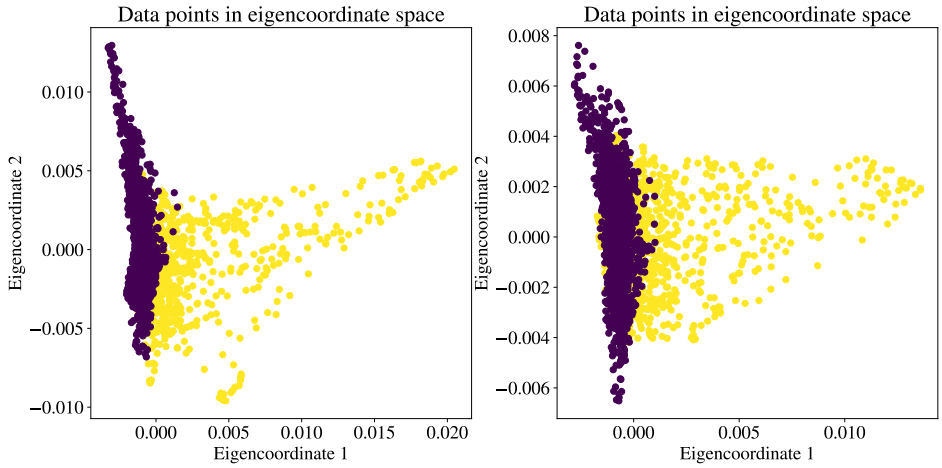
Figure 5.13: Eigenmap analysis performed on the *ramp #5* identified during AS002's forearm pull task as visualized in Figure A.5

The spectrum of the affinity matrix in Eq. 3.10 shown in the left panel of Figure 5.13 shows six dominant eigenvalues. The right panel shows that in the new *eigencoordinate* we obtain three (or four) clusters[4]. The behavior of the data in the eigencoordinates is distinctly different from that of the data in the principal coordinates of the previous section. Here the obtained groupings are from a *single* ramp showing identifiable *modes/clusters* as the task progresses. Figure 5.13 is a prototypical example of eigenmap analysis. Recall that σ in Eq. 3.10 is a hyper-parameter to choose which roughly determines the size of the *local neighborhood* used to obtain the structure of the underlying manifold. In a majority of our data, it was possible to obtain such modes/clusters by varying σ appropriately.

Similar to Figure 5.11 and 5.12 it is possible to put together different tasks or parts of the arm and see whether clustering is possible with the eigenmap

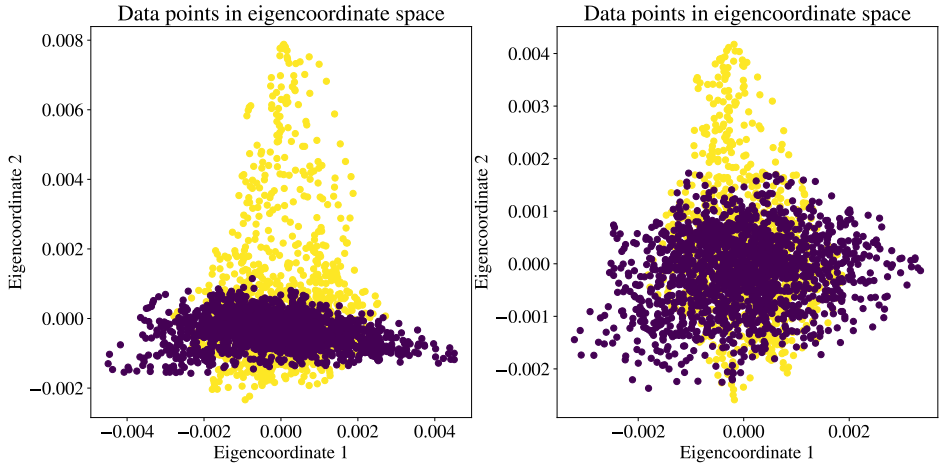
[4]The number of clusters in 3D eigencoordinates is function of spread of values in the corresponding eigenvectors, see Figure 5.17 of Section 5.6.

analysis. An exemplar result is shown in Figures 5.14 and 5.15 for different values of neighbors N in graph corresponding to the Laplacian L in Eq. 3.11.



(a) Results for $N = 50$ (b) Results for $N = 100$

Figure 5.14: Laplacian eigenmap embedding obtained for 50 and 100 nearest neighbor values on AS002's upperarm vs forearm data in a pull task.



(a) Results for $N = 250$ (b) Results for $N = 500$

Figure 5.15: Laplacian eigenmap embedding obtained for 250 and 500 nearest neighbor values on AS002's upper arm vs forearm data in a pull task.

Now we see that changing the neighborhood size allows use to qualitatively see structure of the point cloud at different scales. More importantly, we see that different tasks for the same limb or different parts of the arm for the same task tend to align along different directions in the eigencoordinate space.

5.6 Discussion

That nonlinear methods succeed where linear methods fail should come as no surprise. In effect, what they allow us to do is include higher order effects and interactions than those captured by first order approximations underlying factor analysis and other matrix methods. While various extensions to factor analysis methods exist, in this section we elucidate on the differences between the two nonlinear methods and PCA.

5.6.1 Principal components vs. nonlinear embeddings

As evidenced by their immense popularity, methods like PCA and linear regression are easy to utilize, provide us with a concrete measure of *total-variance-accounted-for* (TVAF), and are inherently highly interpretable. Compared to them, the use of nonlinear methods requires decidedly more finesse. For example, in both Laplacian eigenmaps and diffusion maps, there is a free parameter σ specifying the gaussian kernel size or N the number of nearest-neighbors which in effect controls the size of the local neighborhoods examined to recover the underlying structure. On one hand, while a standard threshold of 90% TVAF is easy to apply across the board, the determination of σ and N in the algorithms described here is an art-form dependent very much on the data at hand as well as the experience of the practitioner.

That TVAF in PCA or R^2 in regression methods do not capture the whole story is evident is markedly evident in even our own data set. Consider the principal components of the same data shown in Figure 5.9. There we see that the first sixteen PCs explain well over 90% TVAF and yet it is plain that any *synergies represented* or *encoded* by them vary in time. This is captured in the interweaving and vacillating helical structures of Figure 5.10. The time varying nature of the synergies uncovered by PCA is well known observation and various models have been proposed to address it [93, 74, 94, 68, 95]. However, note the stark contrast between the PCA output and the plot obtained via eigenmaps in Figure 5.13 where the clustering in the eigencoordinate space seems readily apparent - signifying that it is able to extract much simpler *modes* or *synergies* from the EMG signals as the task progresses.

Recall that the eigenvalues of the covariance matrix in PCA directly signify

the explanatory power of the corresponding principal components. Contrarily, the spectrum of the affinity matrices of Eq. 3.10 and Eq. 3.11 do not encode anything directly analogous. Rather, the number of nonzero components of the spectrum signify the intrinsic dimensionality of the obtained embedding.

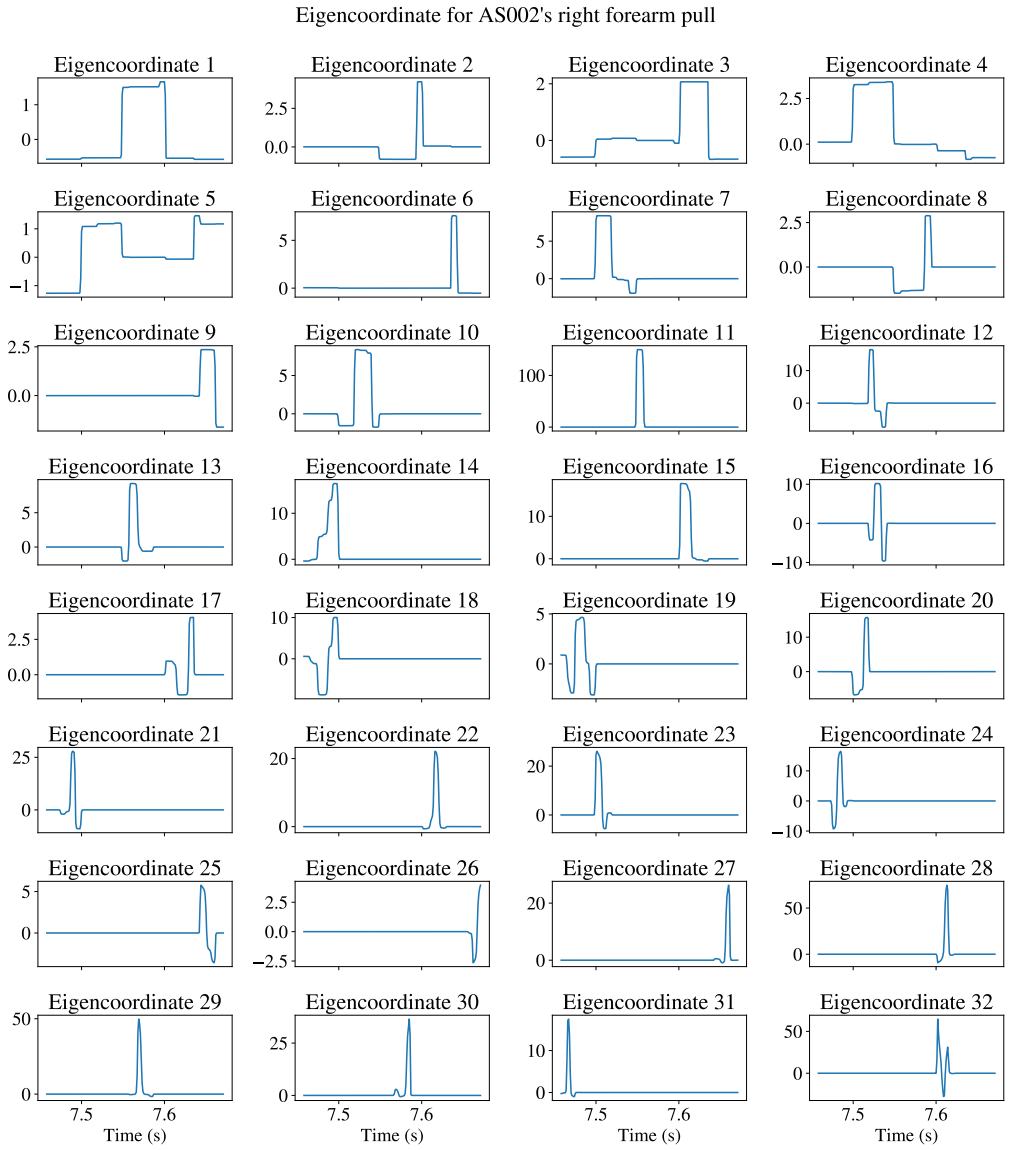


Figure 5.16: Eigencoordinates for the data which generated Figure 5.13. Note how the *three* clusters in that figure correspond to the three distinct *levels/values* exhibited in the first two eigencoordinates.

On the other hand, the number of distinct components of the eigenvectors corresponding to the largest eigenvalues indicate the number of clusters ob-

tained in the embedding. This can be seen in Figure 5.16. The three clusters of Figure 5.13 can readily seen to be a function of the three (or four) nonzero values attained by the first two eigencoordinates. The synthetic data of Figure 5.17 makes it clear why this is the case - when clustering is present in the obtained embedding the eigencoordinates necessarily adopt a *step-like* structure.

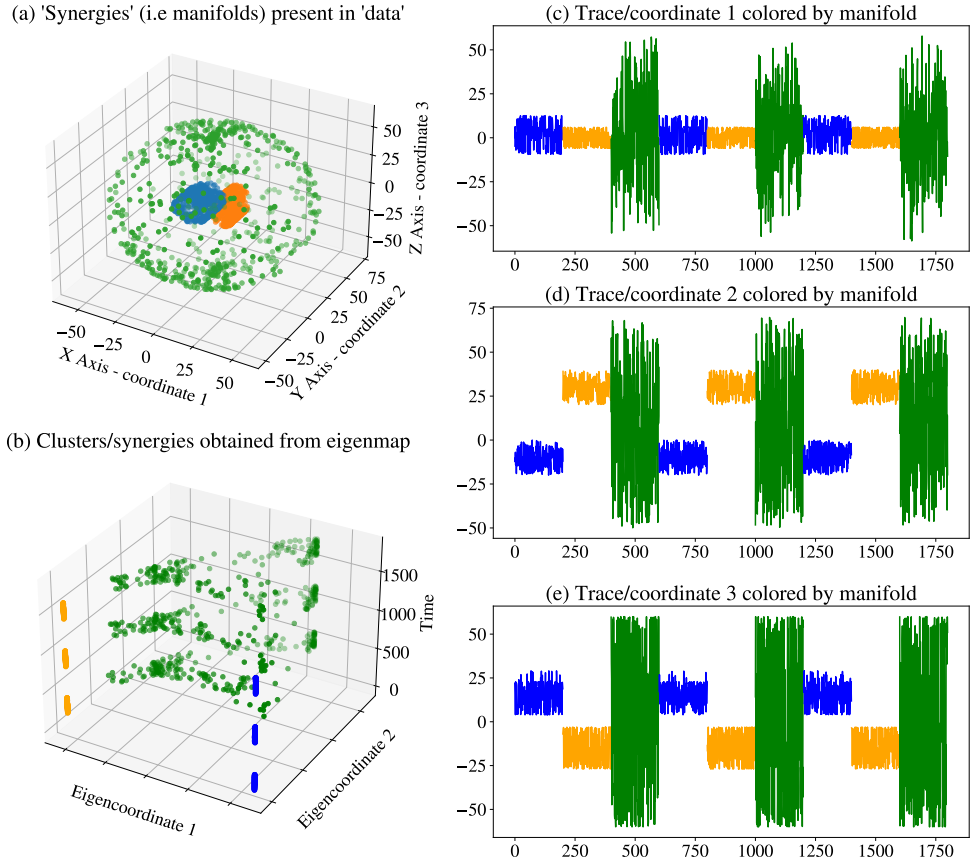


Figure 5.17: This figure shows how to interpret the clustering obtained from the eigencoordinate plots. While PCs explain components that account for most of the variation in the data, eigencoordinates provide local charts such that the obtained embedding groups together the most similar data points. Here on the top left panel we show synthetic data that simulates what happens in the EMG space. Its readouts in time coordinates are shown in the right panel. The bottom left panel shows that the manifold structure is recovered via eigenmap analysis of the traces on the right.

The left column shows two datasets, one of which exhibits clustering. Then the ordinate and abscissa when indexed against the number of points shows an approximately *piecewise-constant* structure for the dataset with clusters.

In the case of task ramps analyzed in the EMG data, the number of data points are in direct correspondence with the sampling instances and therefore the clustering observed is along the axis of time.

Eigenmap analysis of the combined forearm and upper arm data shown in Figures 5.14 and 5.15 show a marked difference from a similar analysis of PCA data presented in Figures 5.11 and 5.12. In the latter set of figures, it is clear that PCA is unable to find any separation between the two data sets in the principal coordinate space. In the former set of figures, the parameter dependent embeddings obtained in eigenmap analysis show that the two datasets occupy different (albeit overlapping) subspaces in the eigencoordinate space. The overlap is ostensibly expected to be present because the a common subset set of muscles are relevant to the completion of the task at hand.To quantify concretely the ability of the nonlinear technique to find separable embeddings for mixed datasets, classical linear and quadratic support vector machines (SVMs) were trained on subsets of the obtained embedding. On average, the nonlinear method outperformed the linear one 78% of the time.

5.6.2 Relevance to synergies

To illustrate that the clusterings obtained in the eignecoordinate space using nonlinear techniques are not an artifact of the method, it is pertinent to see whether any correlation exists between the force modes as the task progresses and the eigenmodes obtained. Figure 5.18 shows a couple of instances where it is manifestly clear that that this is indeed the case. In Figure 5.18 a right forearm pull task is being performed and the force profile changes in the x and z axes at $t \approx 5.6$s while the moment profile changes in the z and y axes at the same time. This results in the obtained nonlinear embedding changing clusters in time. While the eigencoordinates are timevarying with respect to the duration of the task, it does so in an piecewise constant fashion as opposed to the continuous fashion of Figure 5.10. Similarly Figure 5.19 shows obtained clusterings in the nonlinear embedding change at $t \approx 7.55$s and $t \approx 7.6$s. Here, remarkably, the clusters return to approximately the same eigencoordinates when the y and z axis moments return to similar values after peaking during the task. This example further illustrates how

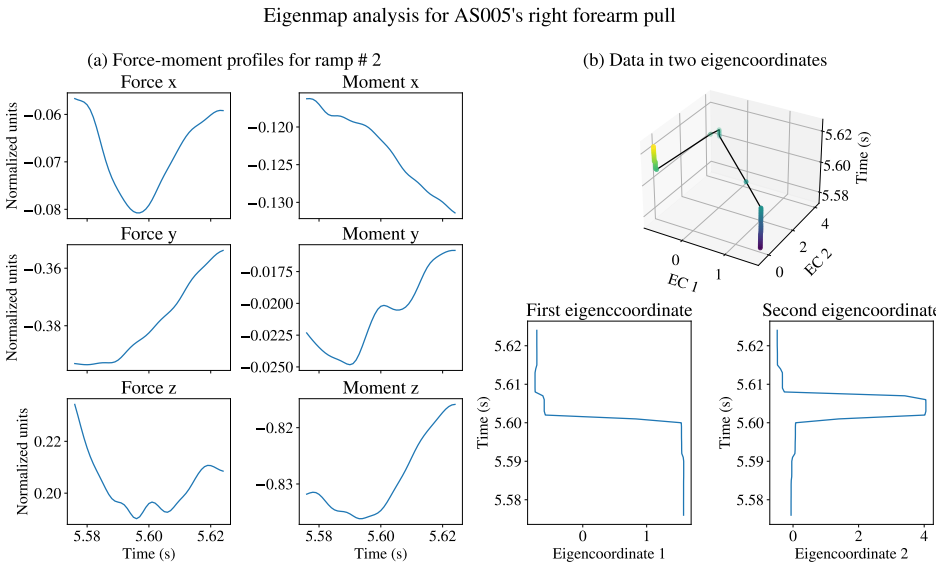


Figure 5.18: Example of correlation between force-moment profiles and the obtained nonlinear embeddings from EMG data at $t = 5.6$ secs. The *three* clusters line up in time as before, during and after the change in the x and z forces axes and the y and z moment axes.

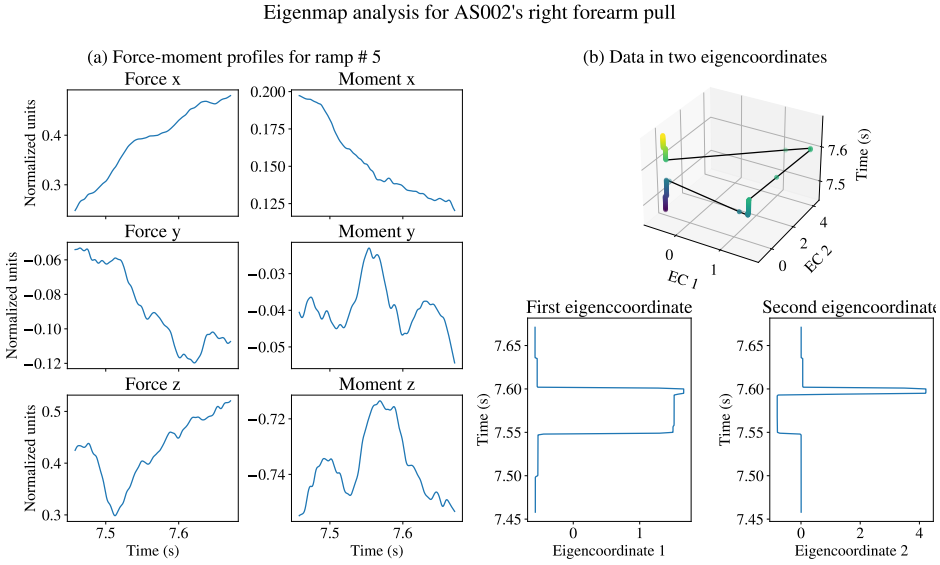


Figure 5.19: Another example of clusterings being indicative of synergies. The force-moment profiles change at $t = 7.52$s in the y and z axis. Correspondingly, the obtained clustering change at the same time. Here the clusters also change at $t = 7.59s$ corresponding to changes in the y and z moment axes at the same time instant.

changes in eigencoordinates based on EMG data alone correspond to changes in the force-moment profiles. The sharp peak in the y-axis for the moment,

is seemingly not captured in Figure 5.19. However, the addition of a third eigencoordinate as shown in Figure 5.20 resolves this discrepancy.

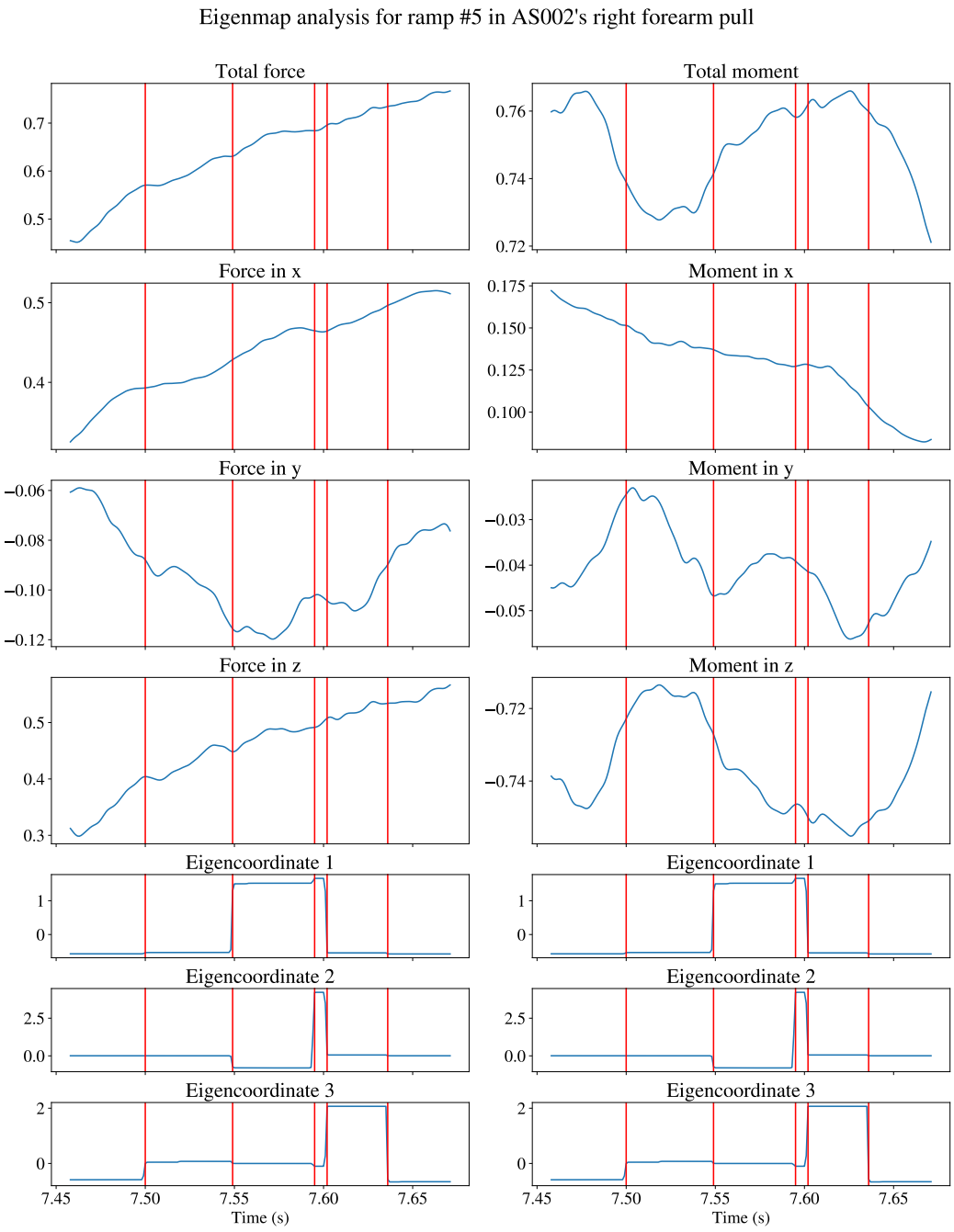


Figure 5.20: This figure visualizes the correspondance between changes in the eigencoordinate values and changes in the force-moment profile. Note that the data input into the nonlinear clustering algorithm that generates the eigencoordinates is purely EMG data and absent any force-moment data.

Note that the obtained clusters are parameter dependent in the case of both

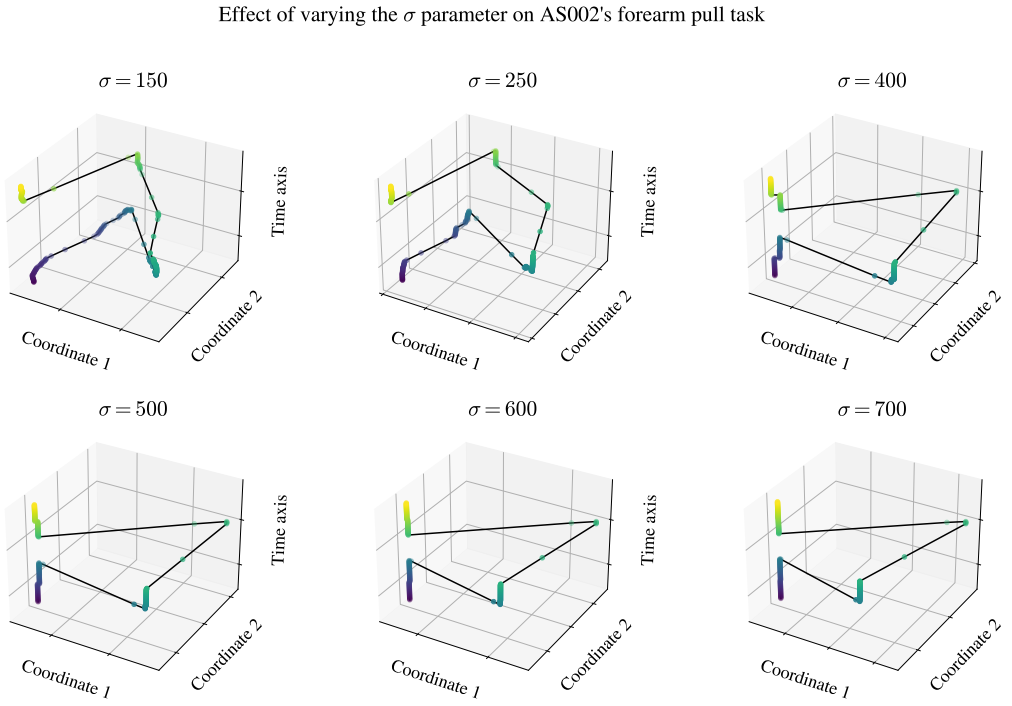


Figure 5.21: This panel shows the effect of varying σ in Eq. 3.10 on the obtained embedding. Smaller values correspond to larger neighborhoods in the original data cloud. The last figure of the panel (bottom-right) with $\sigma = 500$ results in the clustering of Figure 5.13 which correspond well with change in force-moment profile.

eigenmaps and diffusion maps; recall σ in Eq. 3.10 of diffusion maps and k in the k-nearest neighbor graph of Laplacian eigenmaps. Varying this parameter corresponds to smoothly modulating the embedding obtained in the eigencoordinate space. Therefore, this allows one to obtain and examine different clusterings admitted by the family of embeddings. The choice of a *correct* parameter being admittedly data-set dependent, one obvious strategy is to continuously vary σ or k depending on the chosen algorithm until the obtained embedding cluster points differently in the time axis - in manifest correspondence with temporal changes in the force-moment profiles (or in a large majority of them) as shown in Figure 5.18 and 5.19. Note that the eigenmap and diffusion map algorithms are agnostic to the force-moment data and take as input *only* the EMG data. The effect of changing the aforementioned parameter on the embedding is shown in the panel of Figure 5.21. As σ is increased the size of the local neighborhood around each data point in EMG space (the data cloud in the original dimensional space) is

decreased. For $\sigma = 500$ in Figure 5.21 we obtain the result of Figure 5.13 which as shown in Figure manifests changes in the force-moment profiles.

5.7 Conclusion

In the work presented in this chapter, we have shown that nonlinear analysis techniques confer distinct advantages in the analysis of EMG data compared to linear and/or factor analysis methods. We provided a small kinematic example in which PCA followed by clustering is unable to distinguish between two modes of achieving a task objective. In the context of uncovering synergies present in real EMG data, we have shown that while PCA almost invariably extracts time modulated scores that are reflective of time varying synergies, nonlinear analysis methods are able to extract piecewise constant synergies that often repeat in time. Further, in the context of analyzing combined datasets from different task paradigms, we observed that nonlinear methods lead to dimension reduced embeddings that are better amenable to successful treatment by methods like k-Means or SVMs. Finally, we noted that while traditional methods like PCA, NMF and their kin (collectively called Factor Analysis methods) are able to produce measures of quality their decompositions by constructing R^2 values from the residuals, and tallying the Total-Variance-Accounted-For (TVAF), direct analogues of the aforementioned are indeterminate for nonlinear embeddings. On the other hand, the found synergies and/or clusterings from nonlinear methods correlated extremely well with changes in externally recorded force moment profile, providing independent validation regarding presence of synergies. Possible additional/future work here includes a systematic method to vary the paramter σ in Eq. 3.10 and creation of an analogue to "goodness-of-fit" for obtained clusters.

Chapter 6

Epilogue

Humanity has always been interested in studying time-varying phenomena and making sense of the driving processes behind them. Over the last couple of centuries we have built an increasingly varied and mathematical repertoire of methods for analyzing time series data. In this work, we analyzed multi-dimensional time series data in two different settings. In the first, we examined interrelationships amongst repeating yet aperiodic signals, a type we termed *cyclic*, by relying on features that are invariant to reparametrizations of the timeline, culminating in an toolkit we called cyclicity analysis. In the second, we examined multi-dimensional time series from a geometric perspective, utilizing tools that leverage the intrinsic structure of the data cloud in higher dimensional space to establish nonlinear clustering as the natural setting for detection of covariation in the constituent signals.

Arguably, the genesis of the ideas that are utilized in cyclicity analysis began with K. T Chen's investigation of when two curves in a manifold are considered equivalent [22]. He established in a series of papers that up reparametrizations & translations (provided the pair are irreducible to begin with) their equivalence is captured in their iterated integrals [23]. Building on this idea, Lyon's seminal work established a deterministic framework for the integration of paths that traditionally arise as realizations of stochastic processes [25]. In the machinery he built to do the same, we first chance upon the *signature* of a path as an infinite dimensional object that encodes information about that path which is necessary to build a consistent integration theory. While Lyons conjured the the signature of path as an extremely insightful construction to enable *rough path theory* [25], it has generated significant interest in its own right as an object of study [26]. Various papers have explored the idea of using the signature of a time series as a basis for feature extraction [96], signal compressions [97], character recognition [98, 99], prediction [100], and analysis [26]. On the other hand, due to the infinite

dimensional nature of the signature, all such methods must rely on finite truncations of it to utilize it in their computational pipelines. In first part of this work, we have instead adopted the counterpoint to a maxim colloquially attributed to Grothendieck, namely that it is ...

> ... better to work in a nice category with nasty objects than in a nasty category with nice objects.

Our lead matrix of Eq.3.5 is a rather tame object[1] in the computationally hairy category of items presented in Eq. 3.3. We have shown how to interpret each element of Eq. 3.5 as measure of directed leader-follower relationship between the pair of time series that generate them via Eq. 3.1. Further, we have shown how spectral analysis of the lead matrix can lead to recovery of phase shifts between aperiodic and noisy signals not amenable to analysis by traditional Fourier or correlation based methods, which are adversely affected by reparametrizations of the timeline. Further, we applied the methods to real-world data, namely fMRI signals from a cohort of control and tinnitus afflicted participants, and were able to replicate observations and inferences about functional connectivity changes between brain regions that we previously reported [19]. To understand the efficacy of lead matrix as feature for automated learning methods, we compared its performance with traditional features like Dynamic Time Warping and correlation matrices [20]. Finally, we extended our analysis methods to show detection of the propagation of cortical waves in fMRI data, that were previously reported in EEG studies with much greater time resolution and on much larger datasets than previously examined using cyclicity [21].

The above studies establish cyclicity analysis as a general tool for analysis of repeating but aperiodic data, however, more avenues of analysis yet remain for the avid researcher. The field of neuroimaging is currently undergoing a sea-change in its outlook towards data collection, analysis and dissemination methods. The *replicability* crises in psychological studies that started gathering steam about a decade ago [101] has made its way into neuroscience as well [102, 103, 104]. This has spurred consensus that the field as a whole needs to invest in standardization of its methods, techniques and pipelines leading to widespread acceptance of the need to publish de-identified datasets and codebases along with publications. These and related ideas along with

[1] In the sense of being interpretable and well-understood in its matrix form.

tools to facilitate the same were already accepted practices in the realm of open source software and data science. Consequently, this is an exciting time to be working at the intersection of computer science, neuroscience & data analysis.

While cyclicity analysis has its roots in topological concerns, the methods of nonlinear dimension reduction applied here to EMG data have their roots in early 19th century physics. The heat kernel of Eq. 3.10 is a well-studied object in the field of parabolic differential equations, having first risen as the fundamental solution of the famous heat equation that describes the diffusion process for a given quantity of heat over a region as a function of time. See [105] for a comprehensive review on the genesis and history of the topic. Over time, the heat equation has become a fundamental object of study in applied mathematics and in the theory of partial differential equations. Considering its variations and extensions has produced a veritable host of results in the fields of geometric analysis and differential geometry, but not having addressed the generating partial differential equation in this work, this fact is of auxiliary concern to us, serving only to underline the deep connections the innocuous Eq. 3.10 has throughout the field of mathematics. See [106] for more information and a variety of articles illustrating its applications. More pertinently for us, results in the vein of our methods, i.e using diffusion processes over datasets to infer properties of the data appeared in a progression of papers at the turn of the current centutry: kernel PCA [107]in 1998, isometric feature mapping [108] in 2000 , Laplacian Eigenmaps [109, 110] in 2001, local tangent space alignment [111] in 2002, Hessian eigenmaps [112]in 2003 and finally diffusion maps [113] in 2004. These ideas were quickly picked up in the following years in the context of shape recognition & computer graphics [114, 115, 116] as heat kernel signature methods which studied invariants of shapes under rigid and isometric transformations.

In the latter part of this dissertation, we have shown that using the same geometric diffusion principles allows recasting the synergy detection problem associated with the motor control tasks as a nonlinear clustering problem. Both the number of synergies involved, as well as the nature of the process by which they arise, are generally unknown in the context of motor tasks. At the same time, the kinematic equations of models or mechanisms that mimic the anatomy are decidedly nonlinear. Therefore, approaches that adopt a linear framework lack a physiological basis for their choice, whereas data-driven

model free approaches like diffusion maps, which uncover the underlying structure of the datasets are more suitable. The evidence we presented for this was the following: we observed a strong correspondence between changes within eigenvectors of the normalized graph Laplacian used in the nonlinear clustering algorithm for EMG data, with that of the associated force-moment profiles, even though the algorithm never had the force-moment data as input. This external validation of synergies detected (as opposed to internal validation by examination of quality of reconstructions afforded by factor analysis methods) is a novel contribution to the field. Ostensibly, the relatively highly sophisticated mathematical pedigree of the diffusion method (compared to factor analysis models) might be the reason they are not as widely adopted in fields not adjacent to applied mathematics.

The borrowing of ideas, techniques, and tools from its different sub-realms is a common occurrence in mathematics. In what is probably the most famous example, geometry, analysis, algebra, and category theory came together in Wiles' stunning proof of Fermat's Last Theorem, a result in number theory. A similar admixing also takes place across different areas of science itself as evidenced by numerous examples throughout history. However, the pace with which integration of ideas from traditionally disparate scientific fields takes place has quickened since the turn of the century, arguably aided by availability of the internet. Here, we have demonstrated the application of concepts in applied mathematics to the setting of neuroscience and kinesiology to derive the presented insights. One expects that as our world and science becomes increasingly interconnected, such multidisciplinary collaborations will abound. This author is glad to have his humble dissertation contribute to the same.

References

[1] M. A. Belabbas and P. J. Wolfe, "On landmark selection and sampling in high-dimensional data analysis," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1906, pp. 4295–4312, November 2009.

[2] Y. Baryshnikov and E. Schlafly, "Cyclicity in multivariate time series and applications to functional mri data," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 1625–1630.

[3] T. Lyons, H. Ni, and H. Oberhauser, "A feature set for streams and an application to high-frequency financial tick data," in *Proceedings of the 2014 International Conference on Big Data Science and Computing*, ser. Big Data Science '14. New York, NY, USA: Association for Computing Machinery, 2014. [Online]. Available: https://doi.org/10.1145/2640087.2644157

[4] M. L. Latash, V. Krishnamoorthy, J. P. Scholz, and V. M. Zatsiorsky, "Postural synergies and their development," *Neural Plasticity*, vol. 12, no. 2-3, pp. 119–130, 2005.

[5] N. A. Bernstein, *The Co-ordination and Regulation of Movements*, 1st ed. Oxford: Pergamon Press, 1967. [Online]. Available: http://catalog.hathitrust.org/api/volumes/oclc/583008.html

[6] I. M. Gelfand and M. L. Tsetlin, "On the mathematical modeling of mechanisms of central nervous system," in *Models of the Structural-Functional Organization of certain Biological Systems*, I. M. Gelfand, M. L. Tsetlin, S. V. Fomin, and V. S. Gurfinkel, Eds. Cambridge, Mass: MIT Press, 1971.

[7] V. L. Profeta and M. T. Turvey, "Bernstein's levels of movement construction: A contemporary perspective," *Human Movement Science*, vol. 57, pp. 111–133, feb 2018.

[8] J. D. Hamilton, *Time Series Analysis*. Princeton, N.J: Princeton University Press, 1994.

[9] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems part i: deterministic non-linear systems," *International Journal of Control*, vol. 41, no. 2, pp. 303–328, 1985. [Online]. Available: https://doi.org/10.1080/0020718508961129

[10] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition procedure based on loess (with discussion)," *Journal of Official Statistics*, vol. 6, 1990.

[11] I. S. Dhillon, Y. Guan, and B. Kulis, "Kernel k-means: Spectral clustering and normalized cuts," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2004*. New York, NY, USA: Association for Computing Machinery, 2004. [Online]. Available: https://doi.org/10.1145/1014052.1014118 p. 551–556.

[12] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, February 1978. [Online]. Available: http://ieeexplore.ieee.org/document/1163055/

[13] M. Cuturi, "Fast global alignment kernels," in *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*. Madison, WI, USA: Omnipress, 2011, p. 929–936.

[14] M. Cuturi and M. Blondel, "Soft-DTW: A differentiable loss function for time series," in *Proceedings of the 34th International Conference on Machine Learning, ICML 2017*, vol. 2, 2017, pp. 1483–1505.

[15] S. Makridakis, "A survey of time series," *International Statistical Review / Revue Internationale de Statistique*, vol. 44, no. 1, pp. 29–70, 1976. [Online]. Available: http://www.jstor.org/stable/1402964

[16] T. Warren Liao, "Clustering of time series data—a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320305001305

[17] M. W. Woolrich, B. D. Ripley, M. Brady, and S. M. Smith, "Temporal autocorrelation in univariate linear modeling of FMRI data." *NeuroImage*, vol. 14, no. 6, pp. 1370–1386, dec 2001.

[18] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, and M. Jenkinson, "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, 2013, mapping the Connectome. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811913005053

[19] B. J. Zimmerman, I. Abraham, S. A. Schmidt, Y. Baryshnikov, and F. T. Husain, "Dissociating tinnitus patients from healthy controls using resting-state cyclicity analysis and clustering," *Network Neuroscience*, vol. 3, no. 1, pp. 67–89, January 2019. [Online]. Available: https://doi.org/10.1162/netn_a_00053https://www.mitpressjournals.org/doi/abs/10.1162/netn_a_00053

[20] S. Shahsavarani, I. T. Abraham, B. J. Zimmerman, Y. M. Baryshnikov, and F. T. Husain, "Comparing Cyclicity Analysis With Pre-established Functional Connectivity Methods to Identify Individuals and Subject Groups Using Resting State fMRI," *Frontiers in Computational Neuroscience*, vol. 13, p. 94, January 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fncom.2019.00094/full

[21] I. Abraham, S. Shahsavarani, B. Zimmerman, F. Husain, and Y. Baryshnikov, "Slow cortical waves through cyclicity analysis," *bioRxiv*, 2021. [Online]. Available: https://www.biorxiv.org/content/early/2021/07/14/2021.05.16.444387

[22] K.-T. Chen, "Integration of paths—a faithful representation of paths by non-commutative formal power series," *Transactions of the American Mathematical Society*, vol. 89, no. 2, pp. 395–407, February 1958.

[23] K.-T. Chen, "Iterated path integrals," *Bulletin of the American Mathematical Society*, vol. 83, no. 5, pp. 831–880, September 1977.

[24] T. J. Lyons, M. Caruana, and T. Levy, "Differential Equations Driven by Rough Paths," in *Ecole d'Eté de Probabilités de Saint-Flour XXXIV*, ser. Lecture Notes in Mathematics, C. J-M. Morel, G. F. Takens, and P. B. Teissier, Eds. St. Flour: Springer, 2007, vol. 1908. [Online]. Available: http://link.springer.com/10.1007/978-3-540-71285-5

[25] T. J. Lyons, "Differential equations driven by rough signals," *Revista Matemática Iberoamericana*, vol. 14, no. 2, pp. 215–310, 1998. [Online]. Available: http://eudml.org/doc/39555

[26] I. Chevyrev and A. Kormilitzin, "A Primer on the Signature Method in Machine Learning," *arXiv*, vol. [stat.ML], March 2016. [Online]. Available: http://arxiv.org/abs/1603.03788

[27] G. E. Hinton and S. Roweis, "Stochastic neighbor embedding," in *Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds., vol. 15. MIT Press, 2002. [Online]. Available: https://proceedings.neurips.cc/paper/2002/file/6150ccc6069bea6b5716254057a194ef-Paper.pdf

[28] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html

[29] V. N. Vapnik, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 774–780, 1963.

[30] A.-L. Boulesteix and K. Strimmer, "Partial least squares: a versatile tool for the analysis of high-dimensional genomic data," *Briefings in Bioinformatics*, vol. 8, no. 1, pp. 32–44, May 2006. [Online]. Available: https://doi.org/10.1093/bib/bbl016

[31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[32] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde, "Processing strategies for time-course data sets in functional mri of the human brain," *Magnetic Resonance in Medicine*, vol. 30, no. 2, pp. 161–173, 1993. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.1910300204

[33] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar mri," *Magnetic Resonance in Medicine*, vol. 34, no. 4, pp. 537–541, 1995. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.1910340409

[34] S. L. Bressler and V. Menon, "Large-scale brain networks in cognition: emerging methods and principles," *Trends in Cognitive Sciences*, vol. 14, no. 6, pp. 277 – 290, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364661310000896

[35] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," *Proceedings of the National Academy of Sciences*, vol. 102, no. 27, pp. 9673–9678, 2005. [Online]. Available: https://www.pnas.org/content/102/27/9673

[36] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, "Functional connectivity in the resting brain: A network analysis of the default mode hypothesis," *Proceedings of the National Academy of Sciences*, vol. 100, no. 1, pp. 253–258, 2003. [Online]. Available: https://www.pnas.org/content/100/1/253

[37] B. T. Thomas Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. Liu, and R. L. Buckner, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011, pMID: 21653723. [Online]. Available: https://doi.org/10.1152/jn.00338.2011

[38] A. Mitra, A. Z. Snyder, C. D. Hacker, and M. E. Raichle, "Lag structure in resting-state fMRI," *Journal of Neurophysiology*, vol. 111, no. 11, pp. 2374–2391, June 2014.

[39] A. Mitra, A. Z. Snyder, T. Blazey, and M. E. Raichle, "Lag threads organize the brain's intrinsic activity," *Proceedings of the National Academy of Sciences*, vol. 112, no. 17, pp. E2235–E2244, April 2015. [Online]. Available: http://www.pnas.org/lookup/doi/10.1073/pnas.1503960112

[40] R. T. Dean and W. T. Dunsmuir, "Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models," *Behavior Research Methods*, vol. 48, no. 2, pp. 783–802, June 2016.

[41] M. R. Arbabshirani, E. Damaraju, R. Phlypo, S. Plis, E. Allen, S. Ma, D. Mathalon, A. Preda, J. G. Vaidya, T. Adali, and V. D. Calhoun, "Impact of autocorrelation on functional connectivity," *NeuroImage*, vol. 102, pp. 294–308, November 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811914006284

[42] M. J. Jafri, G. D. Pearlson, M. Stevens, and V. D. Calhoun, "A method for functional network connectivity among spatially independent resting-state components in schizophrenia," *NeuroImage*, vol. 39, no. 4, pp. 1666–1681, February 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811907010282

[43] B. Ng, G. Varoquaux, J. B. Poline, M. Greicius, and B. Thirion, "Transport on Riemannian manifold for connectivity-based brain decoding," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 208–216, January 2016. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26259016

[44] H. P. Bartlett, V. Simonite, E. Westcott, and H. R. Taylor, "A comparison of the nursing competence of graduates and diplomates from uk nursing programmes," *Journal of Clinical Nursing*, vol. 9, no. 3, pp. 369–381, 2000. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2702.2000.00331.x

[45] P. Simonetti and J. Oiticica, "Tinnitus Neural Mechanisms and Structural Changes in the Brain: The Contribution of Neuroimaging Research," *International Archives of Otorhinolaryngology*, vol. 19, no. 03, pp. 259–265, March 2015.

[46] Y. I. Sheline, D. M. Barch, J. M. Donnelly, J. M. Ollinger, A. Z. Snyder, and M. A. Mintun, "Increased amygdala response to masked emotional faces in depressed subjects resolves with antidepressant treatment: an fMRI study," *Biological Psychiatry*, vol. 50, no. 9, pp. 651–658, November 2001.

[47] I. M. Veer, C. Beckmann, M.-J. Van Tol, L. Ferrarini, J. Milles, D. Veltman, A. Aleman, M. A. Van Buchem, N. J. A. Van Der Wee, and S. A. R. Rombouts, "Whole brain resting-state analysis reveals decreased functional connectivity in major depression," *Frontiers in Systems Neuroscience*, vol. 4, no. 41, September 2010.

[48] K. J. Trevis, N. M. McLachlan, and S. J. Wilson, "Cognitive Mechanisms in Chronic Tinnitus: Psychological Markers of a Failure to Switch Attention." *Frontiers in psychology*, vol. 7, 2016. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/27605920http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4996052

[49] S. A. Schmidt, K. Akrofi, J. R. Carpenter-Thompson, and F. T. Husain, "Default mode, dorsal attention and auditory resting state networks exhibit differential functional connectivity in tinnitus and hearing loss," *PLOS ONE*, vol. 8, October 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0076488

[50] S. A. Schmidt, J. Carpenter-Thompson, and F. T. Husain, "Connectivity of precuneus to the default mode and dorsal attention networks: A possible invariant marker of long-term tinnitus," *NeuroImage: Clinical*, vol. 16, pp. 196–204, January 2017.

[51] M. Brett, J. Anton, R. V. .... on functional mapping ..., and undefined 2002, "Region of interest analysis using an SPM toolbox," in *8th international conference on functional mapping of the human brain*, Sendai, Japan, 2002. [Online]. Available: http://matthew.dynevor.org/research/abstracts/marsbar/marsbar{_}abstract.pdf p. 497.

[52] D. Cordes, V. M. Haughton, K. Arfanakis, J. D. Carew, P. A. Turski, C. H. Moritz, M. A. Quigley, and M. E. Meyerand, "Frequencies contributing to functional connectivity in the cerebral cortex in "resting state" data." *American Journal of Neuroradiology*, vol. 22, no. 7, pp. 1326–33, August 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11498421

[53] K. Murphy, R. M. Birn, D. A. Handwerker, T. B. Jones, and P. A. Bandettini, "The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced?" *NeuroImage*, vol. 44, no. 3, pp. 893–905, February 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811908010264

[54] M. D. Fox, D. Zhang, A. Z. Snyder, and M. E. Raichle, "The Global Signal and Observed Anticorrelated Resting State Brain Networks," *Journal of Neurophysiology*, vol. 101, no. 6, pp. 3270–3283, June 2009. [Online]. Available: http://www.physiology.org/doi/10.1152/jn.90777.2008

[55] C. Caballero-Gaudes and R. C. Reynolds, "Methods for cleaning the BOLD fMRI signal," *NeuroImage*, vol. 154, pp. 128–149, July 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811916307418

[56] R. J. Meszlényi, P. Hermann, K. Buza, V. Gál, and Z. Vidnyánszky, "Resting state fMRI functional connectivity analysis using dynamic time warping," *Frontiers in Neuroscience*, vol. 11, pp. 1–17, February 2017.

[57] A. M. Leaver, A. Seydell-Greenwald, and J. P. Rauschecker, "Auditory–limbic interactions in chronic tinnitus: Challenges for neuroimaging research," *Hearing Research*, vol. 334, pp. 49–57, April 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378595515001720

[58] A. Maudoux, P. Lefebvre, J. E. Cabay, A. Demertzi, A. Vanhaudenhuyse, S. Laureys, and A. Soddu, "Connectivity graph analysis of the auditory resting state network in tinnitus," *Brain Research*, vol. 1485, pp. 10–21, November 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0006899312008438

[59] J. P. Rauschecker, A. M. Leaver, and M. Mühlau, "Tuning Out the Noise: Limbic-Auditory Interactions in Tinnitus," *Neuron*, vol. 66, no. 6, pp. 819–826, 2010.

[60] A. M. Leaver, T. K. Turesky, A. Seydell-Greenwald, S. Morgan, H. J. Kim, and J. P. Rauschecker, "Intrinsic network activity in tinnitus investigated using functional MRI," *Human Brain Mapping*, vol. 37, no. 8, pp. 2717–2735, August 2016. [Online]. Available: http://doi.wiley.com/10.1002/hbm.23204

[61] C. Lanting, A. Woźniak, P. van Dijk, and D. R. Langers, "Tinnitus- and task-related differences in resting-state networks," in *Advances in Experimental Medicine and Biology*. Springer, Cham, 2016, vol. 894, pp. 175–187. [Online]. Available: http://link.springer.com/10.1007/978-3-319-25474-6{_}19

[62] J. R. Carpenter-Thompson, S. A. Schmidt, and F. T. Husain, "Neural Plasticity of Mild Tinnitus: An fMRI Investigation Comparing Those Recently Diagnosed with Tinnitus to Those That Had Tinnitus for a Long Period of Time." *Neural plasticity*, vol. 2015, p. 161478, July 2015. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26246914http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4515536

[63] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, D. Nolan, E. Bryant, T. Hartley, O. Footer, J. M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg, A. Z. Snyder, and D. C. Van Essen, "Function in the human connectome: Task-fMRI and individual differences in behavior," *NeuroImage*, vol. 80, pp. 169–189, 2013.

[64] S. F. Giszter, F. A. Mussa-Ivaldi, and E. Bizzi, "Convergent force fields organized in the frog's spinal cord," *Journal of Neuroscience*, vol. 13, no. 2, pp. 467–491, February 1993. [Online]. Available: https://www.jneurosci.org/content/13/2/467https://www.jneurosci.org/content/13/2/467.abstract

[65] F. A. Mussa-Ivaldi, "Nonlinear force fields: a distributed system of control primitives for representing and learning movements," in *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97. 'Towards New Computational Principles for Robotics and Automation'*, July 1997, pp. 84–90.

[66] Y. P. Ivanenko, R. Grasso, M. Zago, M. Molinari, G. Scivoletto, V. Castellano, V. Macellari, and F. Lacquaniti, "Temporal Components of the Motor Patterns Expressed by the Human Spinal Cord Reflect Foot Kinematics," *Journal of Neurophysiology*, vol. 90, no. 5, pp. 3555–3565, November 2003. [Online]. Available: https://www.physiology.org/doi/10.1152/jn.00223.2003

[67] S. A. Chvatal and L. H. Ting, "Voluntary and reactive recruitment of locomotor muscle synergies during perturbed walking," *Journal of Neuroscience*, vol. 32, no. 35, pp. 12 237–12 250, August 2012. [Online]. Available: /pmc/articles/PMC3465667//pmc/articles/PMC3465667/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3465667/

[68] P. M. Hilt, I. Delis, T. Pozzo, and B. Berret, "Space-by-Time Modular Decomposition Effectively Describes Whole-Body Muscle Activity During Upright Reaching in Various Directions," *Frontiers in Computational Neuroscience*, vol. 12, April 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5891645/

[69] H. Urbanek and P. van der Smagt, "iEMG: Imaging electromyography," *Journal of Electromyography and Kinesiology*, vol. 27, pp. 1–9, April 2016.

[70] R. Merletti and S. Muceli, "Tutorial. Surface EMG detection in space and time: Best practices," *Journal of Electromyography and Kinesiology*, vol. 49, p. 102363, December 2019.

[71] B. Maton and S. Bouisset, "The distribution of activity among the muscles of a single group during isometric contraction," *European Journal of Applied Physiology and Occupational Physiology*, vol. 37, no. 2, pp. 101–109, June 1977. [Online]. Available: https://link.springer.com/article/10.1007/BF00421696

[72] J. L. McKay and L. H. Ting, "Functional muscle synergies constrain force production during postural tasks," *Journal of Biomechanics*, vol. 41, no. 2, pp. 299–306, 2008. [Online]. Available: /pmc/articles/PMC4350792//pmc/articles/PMC4350792/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4350792/

[73] N. Kuppuswamy and C. M. Harris, "Do muscle synergies reduce the dimensionality of behavior?" *Frontiers in Computational Neuroscience*, vol. 8, June 2014. [Online]. Available: /pmc/articles/PMC4066703//pmc/articles/PMC4066703/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4066703/

[74] A. D'Avella, P. Saltiel, and E. Bizzi, "Combinations of muscle synergies in the construction of a natural motor behavior," *Nature Neuroscience*, vol. 6, no. 3, pp. 300–308, March 2003. [Online]. Available: https://www.nature.com/articles/nn1010

[75] J. A. Gallego, M. G. Perich, L. E. Miller, and S. A. Solla, "Neural Manifolds for the Control of Movement," *Neuron*, vol. 94, no. 5, pp. 978–984, June 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28595054/

[76] I. K. Fodor, "A survey of dimension reduction techniques," Office of Scientific and Technical Information - U.S Department of Energy, Tech. Rep., May 2002. [Online]. Available: https://www.osti.gov/biblio/15002155

[77] F. Cuevas, H. Hoffmann, M. Kurse, J. Kutch, and E. Theodorou, "Computational Models for Neuromuscular Function," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 110–135, 2009. [Online]. Available: http://ieeexplore.ieee.org/document/5342785/

[78] I. M. Gelfand and M. L. Latash, "On the Problem of Adequate Language in Motor Control," *Motor Control*, vol. 2, no. 4, 1998.

[79] J. P. Scholz and G. Schöner, "The uncontrolled manifold concept: Identifying control variables for a functional task," *Experimental Brain Research*, vol. 126, no. 3, pp. 289–306, 1999. [Online]. Available: https://link.springer.com/article/10.1007/s002210050738

[80] E. Todorov and M. Jordan, "A minimal intervention principle for coordinated movement," in *Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds., vol. 15. MIT Press, 2003. [Online]. Available: https://proceedings.neurips.cc/paper/2002/file/8c5f6ecd29a0eb234459190ca51c16dd-Paper.pdf

[81] F. J. Valero-Cuevas, M. Venkadesan, and E. Todorov, "Structured variability of muscle activations supports the minimal intervention principle of motor control," *Journal of Neurophysiology*, vol. 102, no. 1, pp. 59–68, July 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19369362/

[82] E. Todorov and M. I. Jordan, "Optimal feedback control as a theory of motor coordination," *Nature Neuroscience*, vol. 5, no. 11, pp. 1226–1235, 2002. [Online]. Available: http://www.nature.com/natureneuroscience/nn963

[83] E. Todorov, "Optimality principles in sensorimotor control," *Nature Neuroscience*, vol. 7, no. 9, pp. 907–915, September 2004. [Online]. Available: http://www.nature.com/articles/nn1309

[84] S. Cash and R. Yuste, "Linear summation of excitatory inputs by CA1 pyramidal neurons," *Neuron*, vol. 22, no. 2, pp. 383–394, February 1999.

[85] S. J. Day and M. Hulliger, "Experimental Simulation of Cat Electromyogram: Evidence for Algebraic Summation of Motor-Unit Action-Potential Trains," *Journal of Neurophysiology*, vol. 86, no. 5, pp. 2144–2158, November 2001. [Online]. Available: https://www.physiology.org/doi/10.1152/jn.2001.86.5.2144

[86] M. A. Lemay, J. E. Galagan, N. Hogan, and E. Bizzi, "Modulation and vectorial summation of the spinalized frog's hindlimb end-point force produced by intraspinal electrical stimulation of the cord," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 9, no. 1, pp. 12–23, 2001.

[87] C. Capaday and C. van Vreeswijk, "Linear summation of outputs in a balanced network model of motor cortex," *Frontiers in Computational Neuroscience*, vol. 9, p. 63, June 2015. [Online]. Available: http://journal.frontiersin.org/Article/10.3389/fncom.2015.00063/abstract

[88] D. C. Ibes, "A multi-dimensional classification and equity analysis of an urban park system: A novel methodology and case study application," *Landscape and Urban Planning*, vol. 137, pp. 122–137, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169204614003417

[89] N. F. Jansson, R. L. Allen, G. Skogsmo, and S. Tavakoli, "Principal component analysis and k-means clustering as tools during exploration for zn skarn deposits and industrial carbonates, sala area, sweden," *Journal of Geochemical Exploration*, vol. 233, p. 106909, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0375674221001886

[90] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proceedings of the Twenty-First International Conference on Machine Learning*, ser. ICML '04. New York, NY, USA: Association for Computing Machinery, 2004. [Online]. Available: https://doi.org/10.1145/1015330.1015408 p. 29.

[91] G. Kasman and S. Wolf, *Surface EMG made easy: A beginner's guide for rehabilitation clinicians.* Scottsdale, AZ: Noraxon, 2002.

[92] B. R. Shuman, M. H. Schwartz, and K. M. Steele, "Electromyography Data Processing Impacts Muscle Synergies during Gait for Unimpaired Children and Children with Cerebral Palsy," *Frontiers in Computational Neuroscience*, vol. 0, p. 50, June 2017.

[93] A. D'Avella and M. C. Tresch, "Modularity in the Motor System: Decomposition of Muscle Patterns as Combinations of Time-Varying Synergies," in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, 2001. [Online]. Available: https://proceedings.neurips.cc/paper/2001/file/3d863b367aa379f71c7afc0c9cdca41d-Paper.pdf

[94] A. D'Avella and F. Lacquaniti, "Control of reaching movements by muscle synergy combinations," *Frontiers in Computational Neuroscience*, vol. 7, April 2013. [Online]. Available: /pmc/articles/PMC3630368//pmc/articles/PMC3630368/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3630368/

[95] J. Esmaeili and A. Maleki, "Muscle coordination analysis by time-varying muscle synergy extraction during cycling across various mechanical conditions," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 90–99, January 2020.

[96] L. G. Gyurkó, T. Lyons, M. Kontkowski, and J. Field, "Extracting information from the signature of a financial data stream," 2013. [Online]. Available: https://arxiv.org/abs/1307.7244

[97] T. J. Lyons and N. Sidorova, "Sound compression: A rough path approach," in *Proceedings of the 4th International Symposium on Information and Communication Technologies*, ser. WISICT '05. Trinity College Dublin, 2005, p. 223–228.

[98] B. Graham, "Sparse arrays of signatures for online character recognition," 2013. [Online]. Available: https://arxiv.org/abs/1308.0371

[99] W. Yang, L. Jin, and M. Liu, "Deepwriterid: An end-to-end online text-independent writer identification system," 2015.

[100] D. Levin, T. Lyons, and H. Ni, "Learning from the past, predicting the statistics for the future, learning an evolving system," 2013. [Online]. Available: https://arxiv.org/abs/1309.0260

[101] H. Pashler and E. Wagenmakers, "Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?" *Perspectives on Psychological Science*, vol. 7, no. 6, pp. 528–530, 2012, pMID: 26168108. [Online]. Available: https://doi.org/10.1177/1745691612465253

[102] K. S. Button, J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò, "Power failure: why small sample size undermines the reliability of neuroscience," *Nature Reviews Neuroscience*, vol. 14, no. 5, pp. 365–376, 2013. [Online]. Available: https://doi.org/10.1038/nrn3475

[103] B. O. Turner, E. J. Paul, M. B. Miller, and A. K. Barbey, "Small sample sizes reduce the replicability of task-based fMRI studies," *Communications Biology*, vol. 1, no. 1, p. 62, 2018. [Online]. Available: https://doi.org/10.1038/s42003-018-0073-z

[104] S. Marek, B. Tervo-Clemmens, F. J. Calabro, D. F. Montez, B. P. Kay, A. S. Hatoum, M. R. Donohue, W. Foran, R. L. Miller, T. J. Hendrickson, S. M. Malone, S. Kandala, E. Feczko, O. Miranda-Dominguez, A. M. Graham, E. A. Earl, A. J. Perrone, M. Cordova, O. Doyle, L. A. Moore, G. M. Conan, J. Uriarte, K. Snider, B. J. Lynch, J. C. Wilgenbusch, T. Pengo, A. Tam, J. Chen, D. J. Newbold, A. Zheng, N. A. Seider, A. N. Van, A. Metoki, R. J. Chauvin, T. O. Laumann, D. J. Greene, S. E. Petersen, H. Garavan, W. K. Thompson, T. E. Nichols, B. T. T. Yeo, D. M. Barch, B. Luna, D. A. Fair, and N. U. F. Dosenbach, "Reproducible brain-wide association studies require thousands of individuals," *Nature*, vol. 603, no. 7902, pp. 654–660, 2022. [Online]. Available: https://doi.org/10.1038/s41586-022-04492-9

[105] T. N. Narasimhan, "Fourier's heat conduction equation: History, influence, and connections," *Proceedings of the Indian Academy of Sciences - Earth and Planetary Sciences*, vol. 108, no. 3, pp. 117–148, 1999. [Online]. Available: https://doi.org/10.1007/BF02842327

[106] J. Jorgenson and L. Walling, Eds., *The ubiquitous heat kernel*, ser. Contemporary mathematics. Providence, RI: American Mathematical Society, Mar. 2006.

[107] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 07 1998. [Online]. Available: https://doi.org/10.1162/089976698300017467

[108] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.290.5500.2319

[109] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds., vol. 14. MIT Press, 2001. [Online]. Available: https://proceedings.neurips.cc/paper/2001/file/f106b7f99d2cb30c3db1c3cc0fde9ccb-Paper.pdf

[110] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, jun 2003. [Online]. Available: https://direct.mit.edu/neco/article/15/6/1373-1396/6730

[111] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," 2002. [Online]. Available: https://arxiv.org/abs/cs/0212008

[112] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proceedings of the National Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, 2003. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.1031596100

[113] S. S. Lafon, "Diffusion maps and geometric harmonics," Ph.D. dissertation, Yale University, 2004.

[114] J. Sun, M. Ovsjanikov, and L. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," in *Proceedings of the Symposium on Geometry Processing*, ser. SGP '09. Goslar, DEU: Eurographics Association, 2009, p. 1383–1392.

[115] Y. Aflalo, R. Kimmel, and D. Raviv, "Scale invariant geometry for nonrigid shapes," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1579–1597, 2013. [Online]. Available: https://doi.org/10.1137/120888107

[116] C. Li and A. Ben Hamza, "A multiresolution descriptor for deformable 3D shape retrieval," *The Visual Computer*, vol. 29, no. 6, pp. 513–524, 2013. [Online]. Available: https://doi.org/10.1007/s00371-013-0815-3

[117] Y. Baryshnikov, "Time series, persistent homology and chirality," 2019. [Online]. Available: https://arxiv.org/abs/1909.09846

[118] E. W. Weisstein, "Contour winding number," From *MathWorld*–A Wolfram Web Resource. [Online]. Available: https://mathworld.wolfram.com/ContourWindingNumber.html

Appendix A

Supplementary tables & figures

A.1 Chapter 3: Methods

A.1.1 Figures

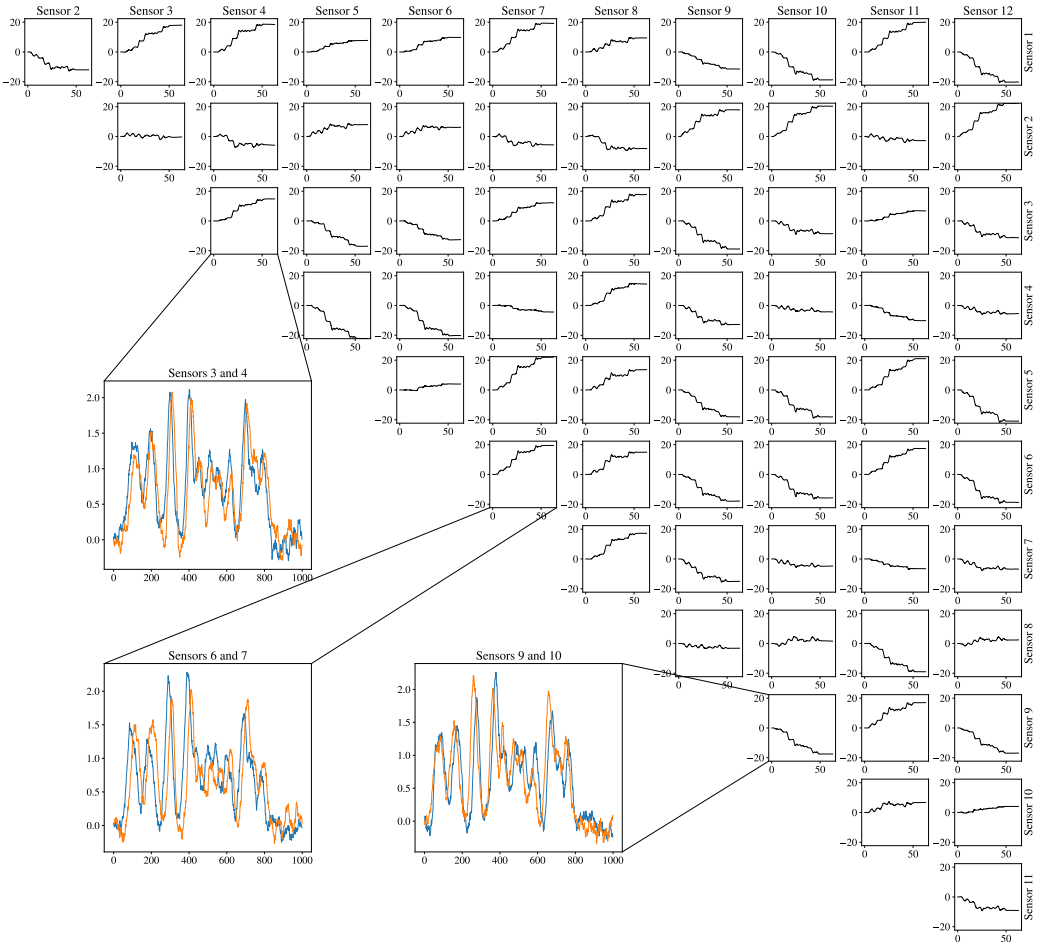


Figure A.1: This figure shows how each element of the matrix in Figure 3.6(a) is built up over time via Eq. 3.1. The insets show how sharp increases correspond to a more dominant leader-follower relationship while those that average about zero have a much more variable relationship.

A.2 Chapter 4: fMRI data

A.2.1 Tables

Table A.1: Demographics for subject groups in [19]. Note that for controls we have $N = 15$ with 10 female, and for tinnitus we have $N = 32$ with 14 female. BDI and BAI are Beck's Depression and Anxiety Index respectively

	Controls		**Tinnitus**		***p*-value**
	M	SD	M	SD	
Age	46.27	11.71	53.16	10.73	0.05
BDI	3.07	6.37	4.63	5.52	0.39
BAI	1.59	1.91	2.69	3.77	0.30
Tinnitus Functional Index					
Total	N/A	N/A	23.44	17.78	N/A
Intrusive	N/A	N/A	39.43	20.46	N/A
Sense of Control	N/A	N/A	36.77	23.01	N/A
Cognitive	N/A	N/A	22.29	19.62	N/A
Sleep	N/A	N/A	15.99	21.36	N/A
Auditory	N/A	N/A	25.89	27.13	N/A
Relaxation	N/A	N/A	26.51	24.27	N/A
Quality of Life	N/A	N/A	13.87	18.31	N/A
Emotional	N/A	N/A	9.95	10.49	N/A
Pure tone averages					
Right 250 Hz	12.00	10.14	13.28	6.04	0.59
Right 500 Hz	10.67	5.31	12.81	6.95	0.30
Right 1000 Hz	11.33	8.12	12.34	4.40	0.58
Right 2000 Hz	12.00	7.51	17.81	10.31	0.06
Right 3000 Hz	13.67	11.41	24.22	17.37	0.04*
Right 4000 Hz	16.00	16.50	27.03	18.62	0.06
Right 6000 Hz	16.67	19.24	30.78	19.06	0.02*
Right 8000 Hz	15.67	22.75	30.47	20.96	0.03*
Right 9000 Hz	21.00	19.29	38.75	21.96	0.01*
Right 10000 Hz	21.67	22.57	41.09	22.78	0.01*
Right 11200 Hz	30.00	25.64	47.50	21.02	0.02*

continued on next page ...

. . . Table A.1 continued

	Controls		**Tinnitus**		***p*-value**
	M	SD	M	SD	
Right 12500 Hz	38.33	26.70	56.88	19.58	0.01*
Right 14000 Hz	45.33	28.19	64.22	15.35	0.00*
Right 16000 Hz	44.00	16.39	49.84	13.94	0.21
Left 250 Hz	9.00	7.37	15.00	12.25	0.09
Left 500 Hz	10.00	5.35	13.75	10.78	0.21
Left 1000 Hz	10.00	6.55	12.50	9.67	0.37
Left 2000 Hz	12.00	8.41	18.59	11.59	0.06
Left 3000 Hz	15.00	12.54	27.19	16.51	0.01*
Left 4000 Hz	17.33	18.89	29.22	17.19	0.04*
Left 6000 Hz	17.67	17.51	34.38	18.65	0.01*
Left 8000 Hz	14.67	16.31	33.44	21.27	0.00*
Left 9000 Hz	17.33	15.45	43.44	22.23	0.00*
Left 10000 Hz	19.67	15.86	47.03	22.93	0.00*
Left 11200 Hz	25.33	22.08	52.81	23.21	0.00*
Left 12500 Hz	37.67	27.31	60.63	23.38	0.00*
Left 14000 Hz	46.00	29.95	63.75	18.14	0.02*
Left 16000 Hz	41.00	20.02	50.63	14.85	0.07

*Significant at the $p < .05$ level. Scores from Beck's Depression Inventory, Beck Anxiety Inventory, and the Tinnitus Functional Index were acquired at each imaging session and averaged together. Means (M) and standard deviations (SD) are presented for each group, and p-values associated with two-sample t-tests between group means are displayed. Pure tone averages at 250, 500, 1000, 2000, 3000, 4000, 6000, 8000, 9000, 10000, 11200, 12500, 14000, and 16000 Hz are presented for both right and left ears.

A.2.2 Figures

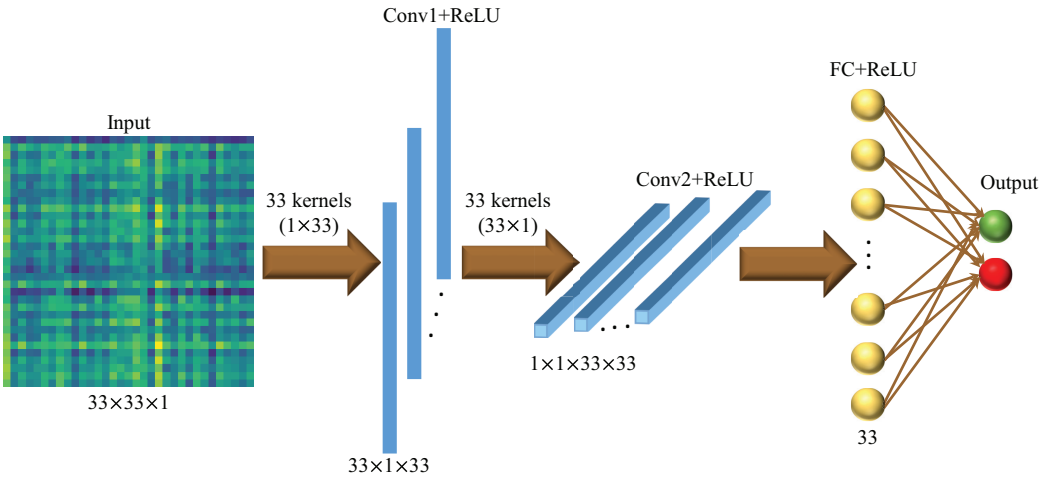


Figure A.2: This figure shows the schematics of the convolutional neural network that was used to attempt classification of feature matrices in [20].

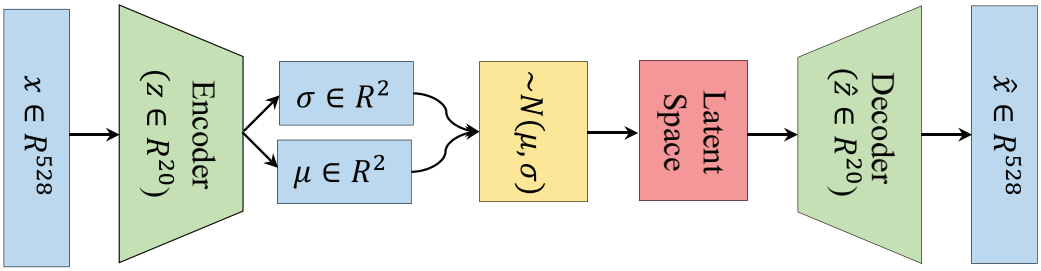


Figure A.3: This figure shows the schematics of the variational autoencoder that was used to generate synthetic data for the purposes of data augmentation in [20].

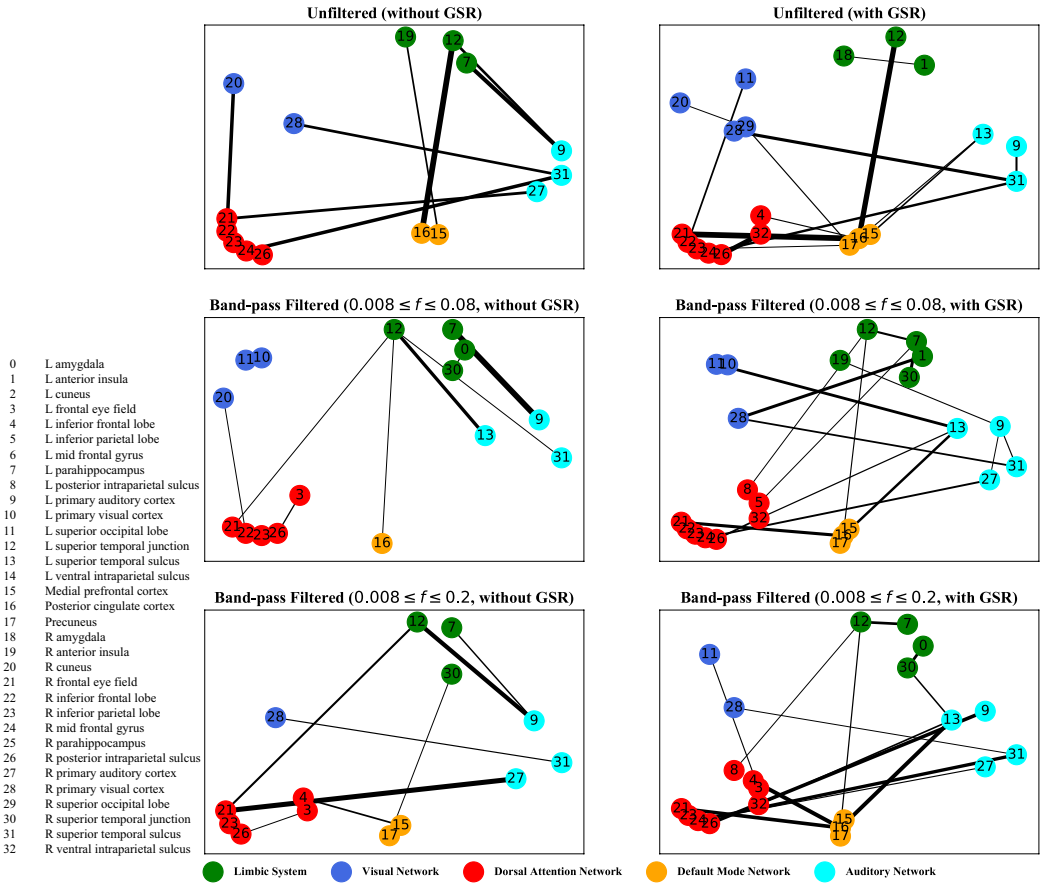
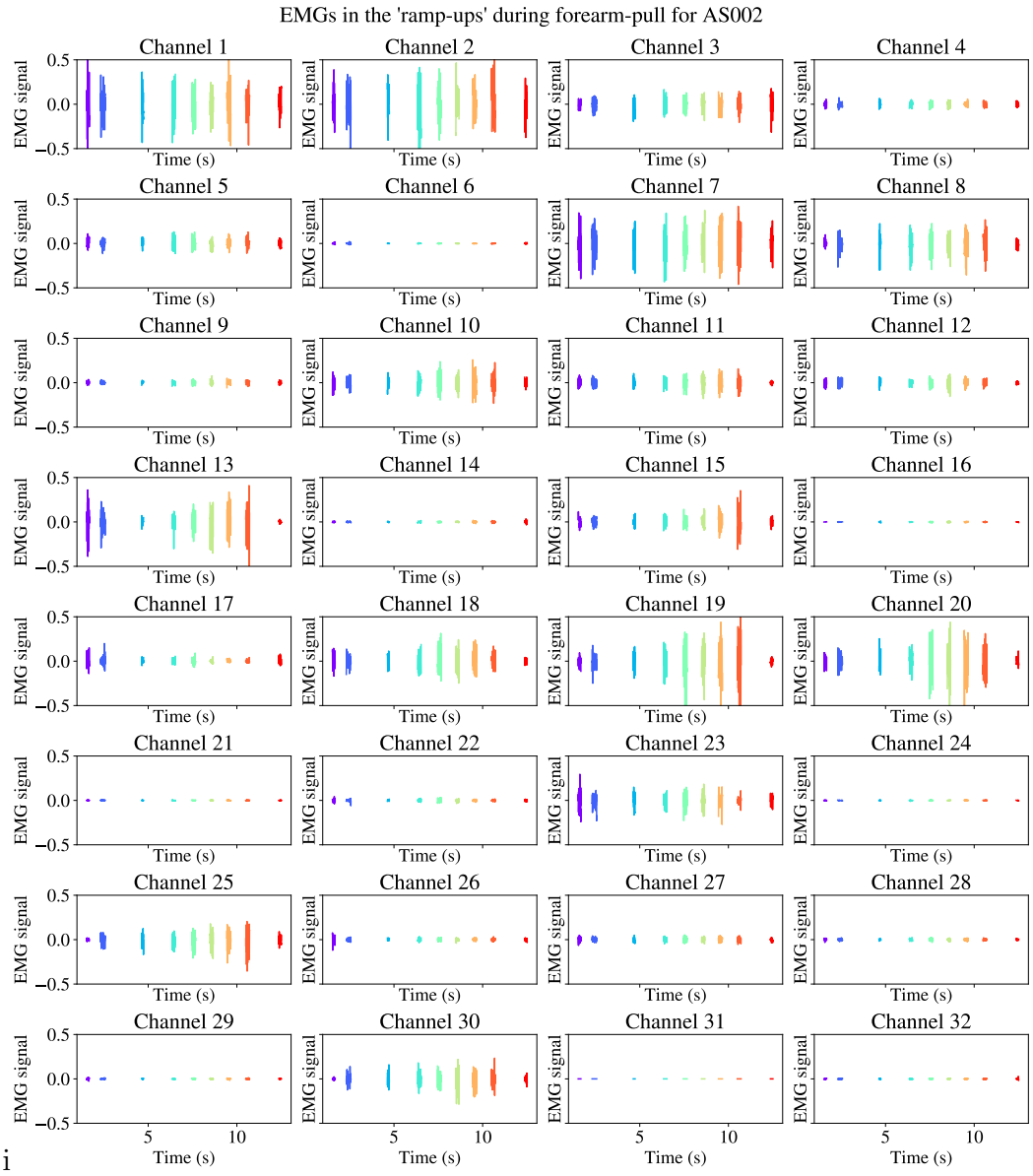


Figure A.4: Zero-lag correlation analysis—the salient interactions in separating patients and control groups, selected by the sparse SVM. The nodes correspond to the ROIs listed on the left side of the figure. The edges represents the interaction between ROIs corresponding to SVM solution of the weight vector. The width of the weights relates to the magnitude of interactions. GSR stands for global signal regression and f denotes the passed frequencies of band-pass filters.

A.3 Chapter 5: EMG data

A.3.1 Figures



i

Figure A.5: Examples of *ramp-ups* of EMG signals extracted from analysis of total force magnitude.

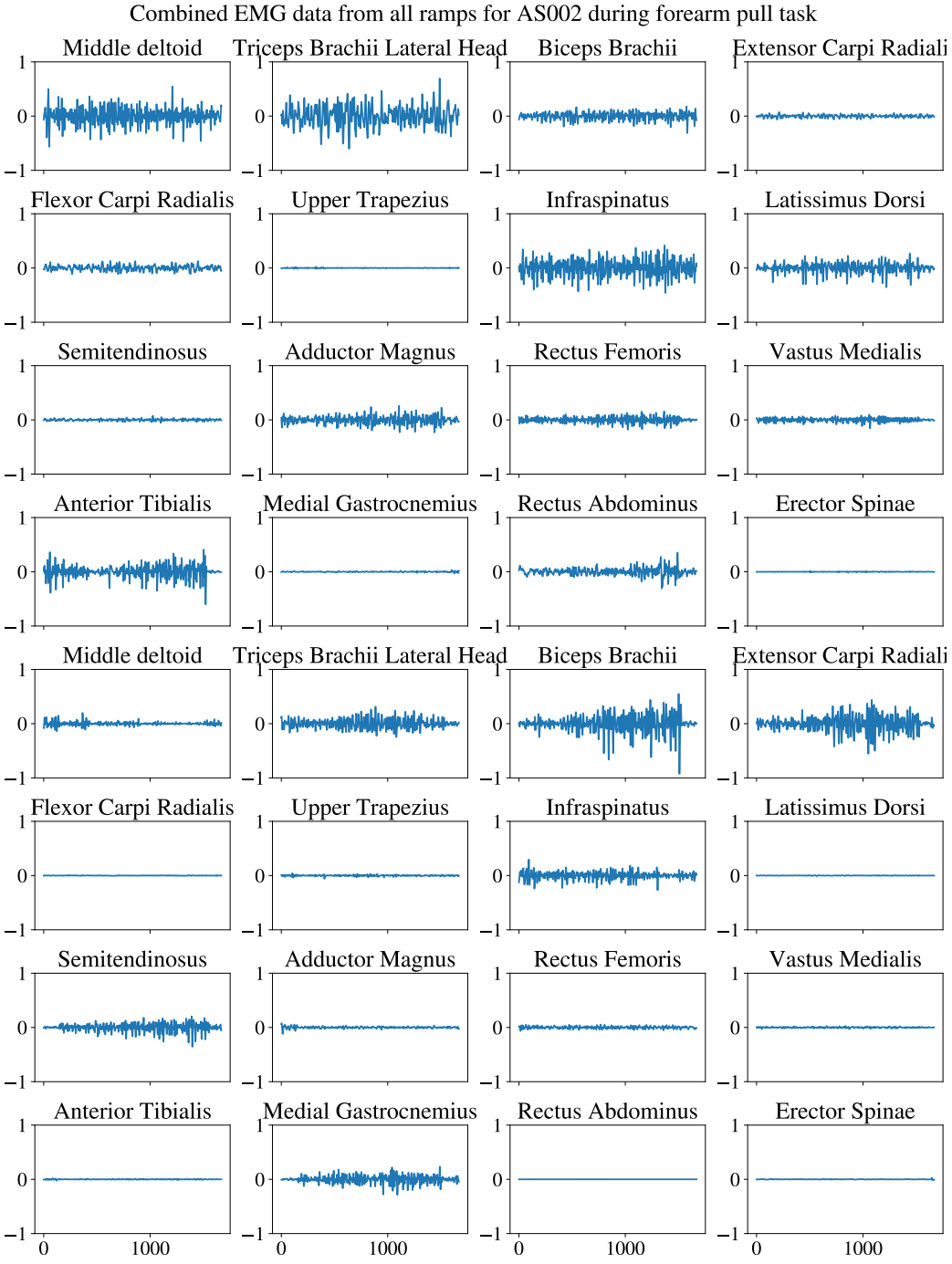


Figure A.6: Visualization of all ramps EMG content after preprocessing. Note that now rather than indicating time, the absissca corresponds to sample number.

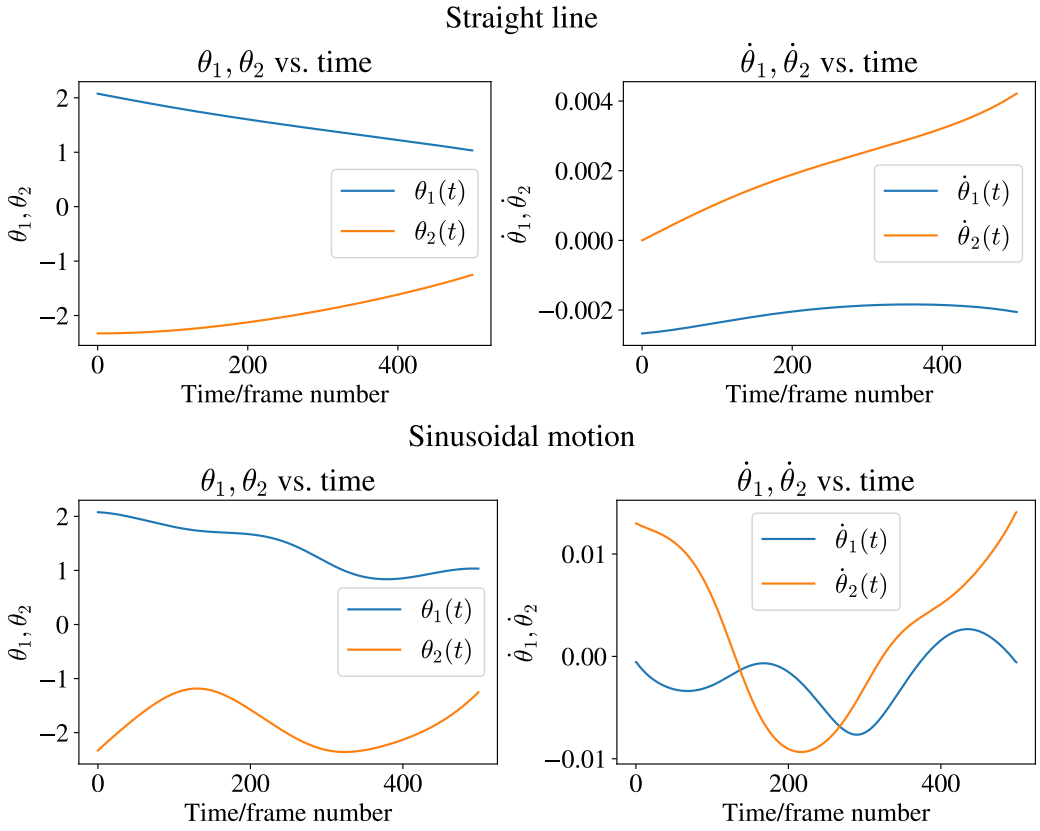


Figure A.7: Recall that the inverse kinematics of Figure 5.1 admit a pair of solutions for a trajectory. This figure shows the solution complementary to the one presented in Figure 5.3.

Appendix B

Significant jumps via persistence diagram

The algorithm we describe is due to Y. Baryshnikov and can be found in [117] and was implemented in Python with slight modifications. We first begin by finding all the minima and maxima of a function over some interval after removing any linear trends in the time series data. Then we construct the persistence diagram of the zeroth persistent homology as in [117] and keep only the corresponding maxima/minima pairs.

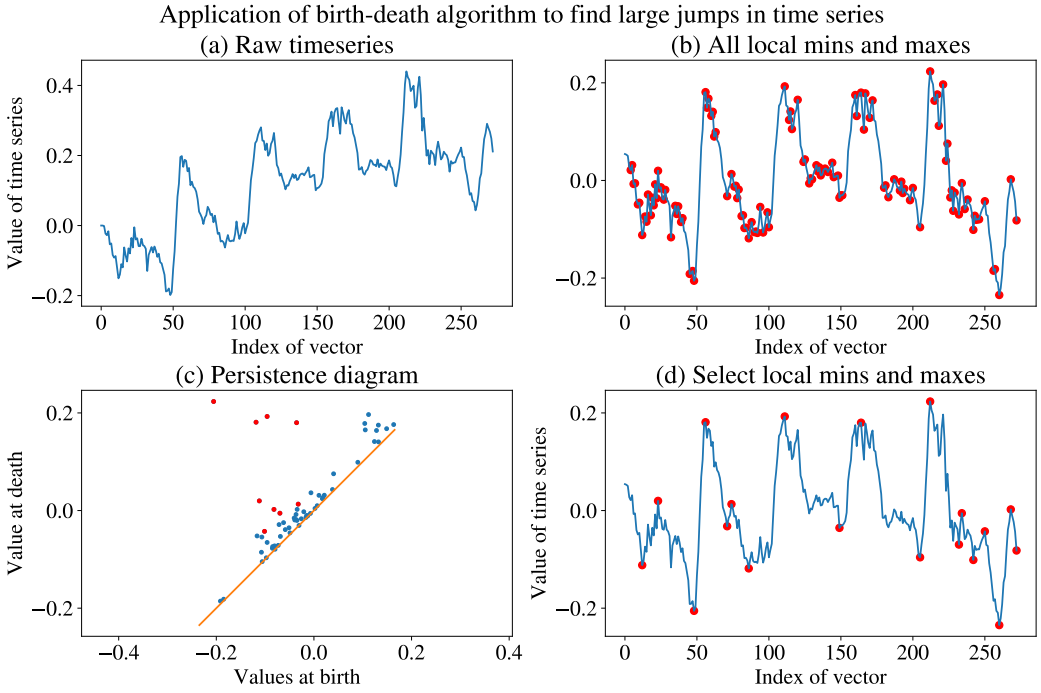


Figure B.1: In (a) we show the raw timeseries data that was (b) detrended and marked with all pairs of maxima and minima. (c) Shows the corresponding persistence diagram for the birth-death process as outlined in [117] and while (d) shows the result when restricting the birth-death pairs to be some fixed distance away from the identity.

Appendix C

A primer on winding numbers

We will elucidate in this section the connection between Eq. 3.1 and *winding numbers*. In this section let us restrict ourselves to the case where our time series is a two dimensional one. Suppose that our path $z(t)$ is a closed one as in Definition 3.4, i.e. its starting and ending points on the plane are the same. This occurs frequently when tracking coordinates on a plane; for example the flight or path of a bird or animal that leaves its home/nest to forage or hunt and then returns to it. The complete path has a natural orientation accorded to it by the direction it was traversed in time. The *winding number* $w(z) \in \mathbb{Z}$ is defined as the number of clockwise or counterclockwise turns that the path makes around the origin[1]. Since this is equivalent to rotations about the same, we can switch to polar coordinates $(r(t), \theta(t))$ to describe the trajectory of $z(t)$. Then, since $z(a) = z(b)$, the number of clockwise or anti-clockwise windings is simply

$$w(z) := \frac{\theta(b) - \theta(a)}{2\pi} \tag{C.1}$$

Considering the change in $\theta(t)$ we have from its definition that:

$$\theta\left(t\right) := \arctan\left(\frac{y\left(t\right)}{x\left(t\right)}\right) \implies \quad d\theta = \frac{x(t)dy - y(t)dx}{x^2(t) + y^2(t)}$$

Defining $r^2(t) = x^2(t) + y^2(t)$, we have that

$$\int_a^b d\theta = \int_a^b \frac{x(t)dy - y(t)dx}{r^2(t)} = \int_a^b \frac{x(t)y'(t) - y(t)x'(t)}{r^2(t)}dt \tag{C.2}$$

[1]Provided it does not pass through the origin itself

Since the integral on the left of Eq. C.2 evaluates to the numerator of the quantity in Eq. C.1 we have that in light

$$w(z) = \frac{1}{2\pi} \int_a^b \frac{x(t)y'(t) - y(t)x'(t)}{r^2(t)} dt \tag{C.3}$$

Then we see that the integrand in Eq. 3.1 and Eq. C.3 are related by the pointwise scaling transformation $r(t)$ provided $x(t)$ and $y(t)$ are not simultaneously zero. Note that there is nothing special about the origin per se, we can calculate the winding number about any point not on the curve. More generally, on the complex plane, we have from complex analysis that for any closed curve γ parametrized by $t \in [a, b]$, i.e. $\gamma : [a, b] \to \mathbb{C}$ the winding number about z_0 not in the image of $\gamma(t)$ is

$$w(\gamma, z_0) = \frac{1}{2\pi i} \int_a^b \frac{\gamma'(t)}{\gamma(t) - z_0} dt$$

The above quantity is often also called the *index* of z_0 with respect to $\gamma(t)$ and denoted $\mathrm{Ind}_\gamma(z_0)$. See Figure C.1 for some examples from [118].

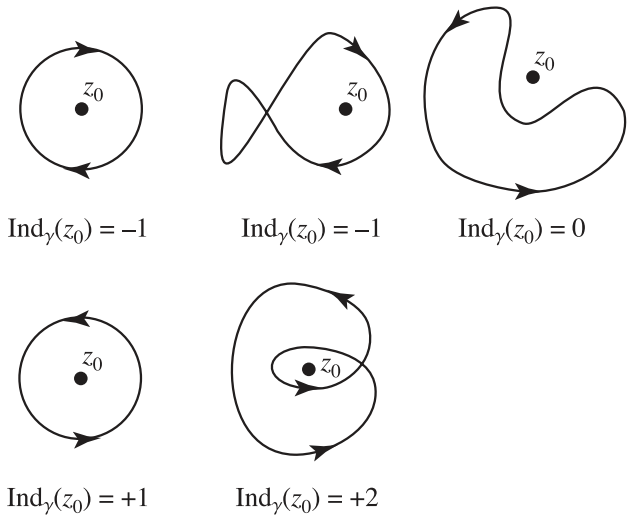


Figure C.1: The first column shows the convention in place, i.e. counterclockwise is positive. The second column shows how windings can be calculated for curves with self intersections and the final column shows that index of a point not enclosed by the curve is zero.