Engineering and Applied Science Theses & Dissertations

McKelvey School of Engineering

Winter 12-15-2016

# Understanding the Role of Dynamics in Brain Networks: Methods, Theory and Application

Mohammadmehdi Kafashan
*Washington University in St. Louis*

Follow this and additional works at: https://openscholarship.wustl.edu/eng_etds

Part of the Engineering Commons

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science

Department of Electrical and Systems Engineering

Dissertation Examination Committee:
ShiNung Ching, Chair
R. Martin Arthur
Dennis Barbour
Jr-Shin Li
Joseph A. O'Sullivan
Ben Julian A. Palanca

Understanding the Role of Dynamics in Brain Networks: Methods, Theory and Application

by

MohammadMehdi Kafashan

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2016
Saint Louis, Missouri

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| AIC | Akaike Information Criterion |
| AR | Autoregressive |
| BOLD | Blood Oxygen Level Dependent |
| CS | Compressed Sensing |
| CPEI | Composite Permutation Entropy Index |
| DAN | Dorsal Attention Network |
| dFC | dynamic Functional Connectivity |
| DLOC | Depressed Level of Consciousness |
| DMN | Default Mode Network |
| EEG | Electroencephalogram |
| EM | Expectation Maximization |
| FC | Functional Connectivity |
| fMRI | functional Magnetic Resonance Imaging |
| FPC | Frontoparietal Control Network |
| GCS | Glasgow Coma Scale |
| KF | Kalman Filter |
| K-S | Kolmogorov-Smirnov |
| LAN | Language Network |
| LMA | Lagged Microstate Analysis |
| MVAR | Multivariate Autoregressive |
| MLE | Maximum Likelihood Estimation |
| MMSE | Minimum Mean-Squared Error |
| MSE | Mean Squared Error |
| PCA | Principal Component Analysis |
| PSD | Power Spectral Density |
| RIP | Restricted Isometry Property |
| RSN | Resting State Networks |
| SEM | Standard Error of the Mean |
| SMN | Somatomotor Network |
| STRNN | Soft-Thresholding Recurrent Neural Network |
| SVM | Support Vector Machine |
| VAN | Ventral Attention Network |
| VIS | Visual Network |

# Acknowledgments

First, I would like to convey my sincere gratitude to my advisor, Dr. ShiNung Ching, for his guidance and support during my PhD studies. He has given me the opportunity to explore the research topics that interest me the most, and encouraged me in my explorations.

I wish to thank my dissertation defense committee members, Dr. R. Martin Arthur, Dr. Dennis Barbour, Dr. Jr-Shin Li, Dr. Joseph O'Sullivan, and Dr. Ben Julian A. Palanca for their constructive comments and suggestions on improving my dissertation.

I would also like to thank Dr. Ben Palanca and Dr. Terrance Kumar for their invaluable thoughts and discussions during our collaborations on exciting research projects over the past few years. Additionally, I thank Dr. Humberto Gonzalez for his helpful input and discussions.

I am thankful to current and previous members of our group who have been helpful during my studies. These include Fuqiang Huang, Dr. Gautum Kumar, Seul Ah Kim, Sensen Liu, Delsin Menolascino, Anirban Nandi, Shoko Ryu, Matthew Singh, and Akshay Thontakudi.

Words cannot express how indebted I am to my parents. I would like to thank my family on the other side of the world for their endless support and many sacrifices they endured on my behalf. Furthermore, very special thanks go to my love, Samila, who has made the journey a lot of fun, and who has inspired me striving towards achieving my goals.

MohammadMehdi Kafashan

*Washington University in Saint Louis*

*December 2016*

Dedicated to the memory of my mother.

ABSTRACT OF THE DISSERTATION

Understanding the Role of Dynamics in Brain Networks: Methods, Theory and Application

by

MohammadMehdi Kafashan

Doctor of Philosophy in Electrical Engineering

Washington University in St. Louis, December 2016

Research Advisor: Professor ShiNung Ching

The brain is inherently a dynamical system whose networks interact at multiple spatial and temporal scales. Understanding the functional role of these dynamic interactions is a fundamental question in neuroscience. In this research, we approach this question through the development of new methods for characterizing brain dynamics from real data and new theories for linking dynamics to function. We perform our study at two scales: macro (at the level of brain regions) and micro (at the level of individual neurons).

In the first part of this dissertation, we develop methods to identify the underlying dynamics at macro-scale that govern brain networks during states of health and disease in humans. First, we establish an optimization framework to actively probe connections in brain networks when the underlying network dynamics are changing over time. Then, we extend this framework to develop a data-driven approach for analyzing neurophysiological recordings without active stimulation, to describe the spatiotemporal structure of neural activity at different timescales. The overall goal is to detect how the dynamics of brain networks may change within and between particular cognitive states. We present the efficacy of this

approach in characterizing spatiotemporal motifs of correlated neural activity during the transition from wakefulness to general anesthesia in functional magnetic resonance imaging (fMRI) data. Moreover, we demonstrate how such an approach can be utilized to construct an automatic classifier for detecting different levels of coma in electroencephalogram (EEG) data.

In the second part, we study how ongoing function can constraint dynamics at micro-scale in recurrent neural networks, with particular application to sensory systems. Specifically, we develop theoretical conditions in a linear recurrent network in the presence of both disturbance and noise for exact and stable recovery of dynamic sparse stimuli applied to the network. We show how network dynamics can affect the decoding performance in such systems. Moreover, we formulate the problem of efficient encoding of an afferent input and its history in a nonlinear recurrent network. We show that a linear neural network architecture with a thresholding activation function is emergent if we assume that neurons optimize their activity based on a particular cost function. Such an architecture can enable the production of lightweight, history-sensitive encoding schemes.
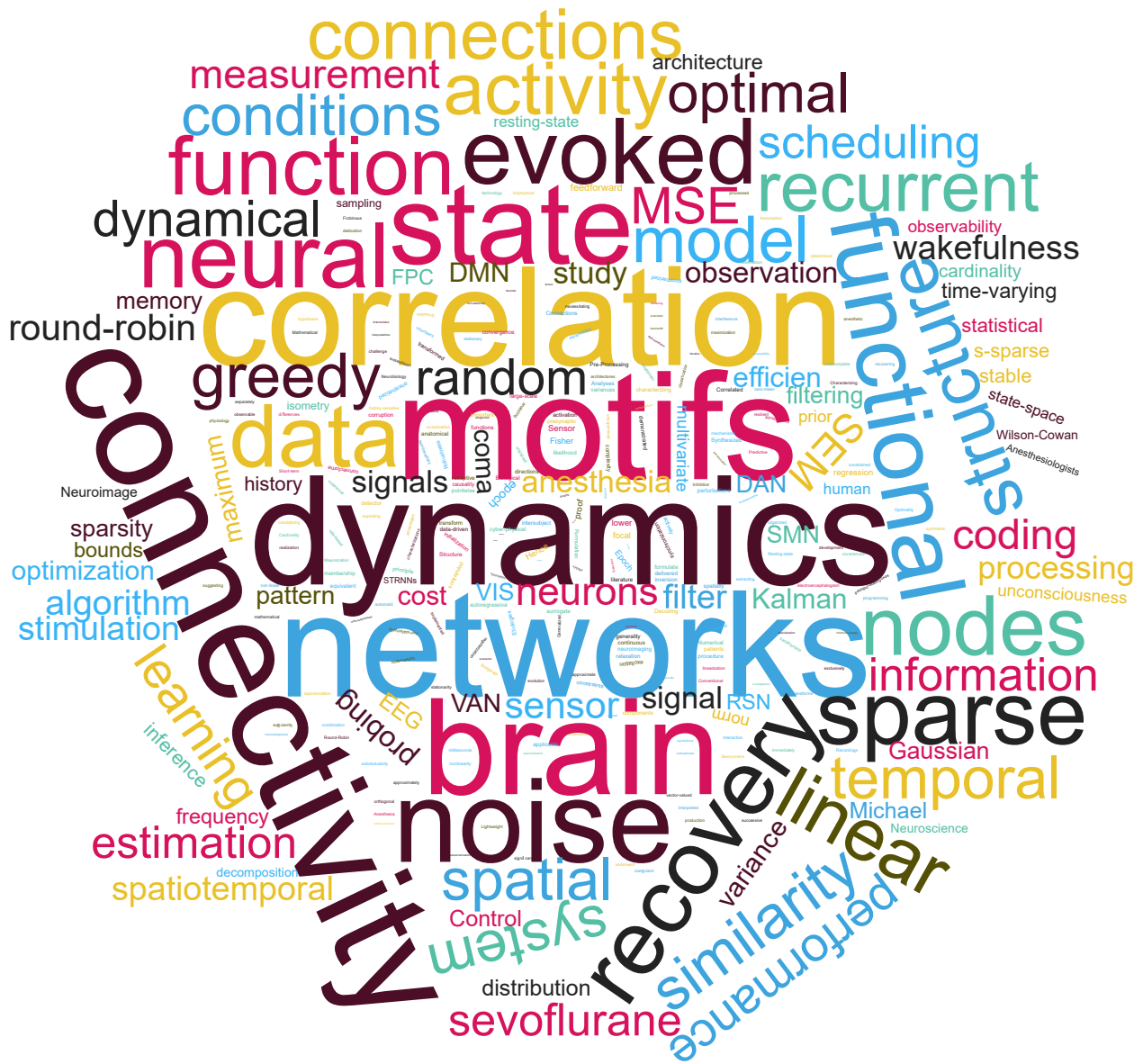
Figure 0.1: World cloud of the dissertation.

# Chapter 1

# Introduction

## 1.1  Overview

The brain is perhaps the most enigmatic organ in the human body. Within the brain are more than 100 billion neurons that communicate in trillions of connections called synapses. These neurons comprise complex networks at multiple spatial and temporal scales, whose connections are constantly changing over time. Thus, the brain is highly dynamic: its activity, and its interconnectivity, are constantly changing and adapting in ways that ultimately allow us to see, hear and think. This dissertation addresses two related questions: (i) how to best describe and uncover the dynamics within brain networks, and (ii) how to associate those dynamics with different cognitive conditions, including specific neurological disorders.

In this dissertation, we explore two principal approaches to these questions. First, in part one of the dissertation, we examine the problem of inferring time-varying connectivity between brain regions at macroscopic spatial scales. We develop, in particular, new methods
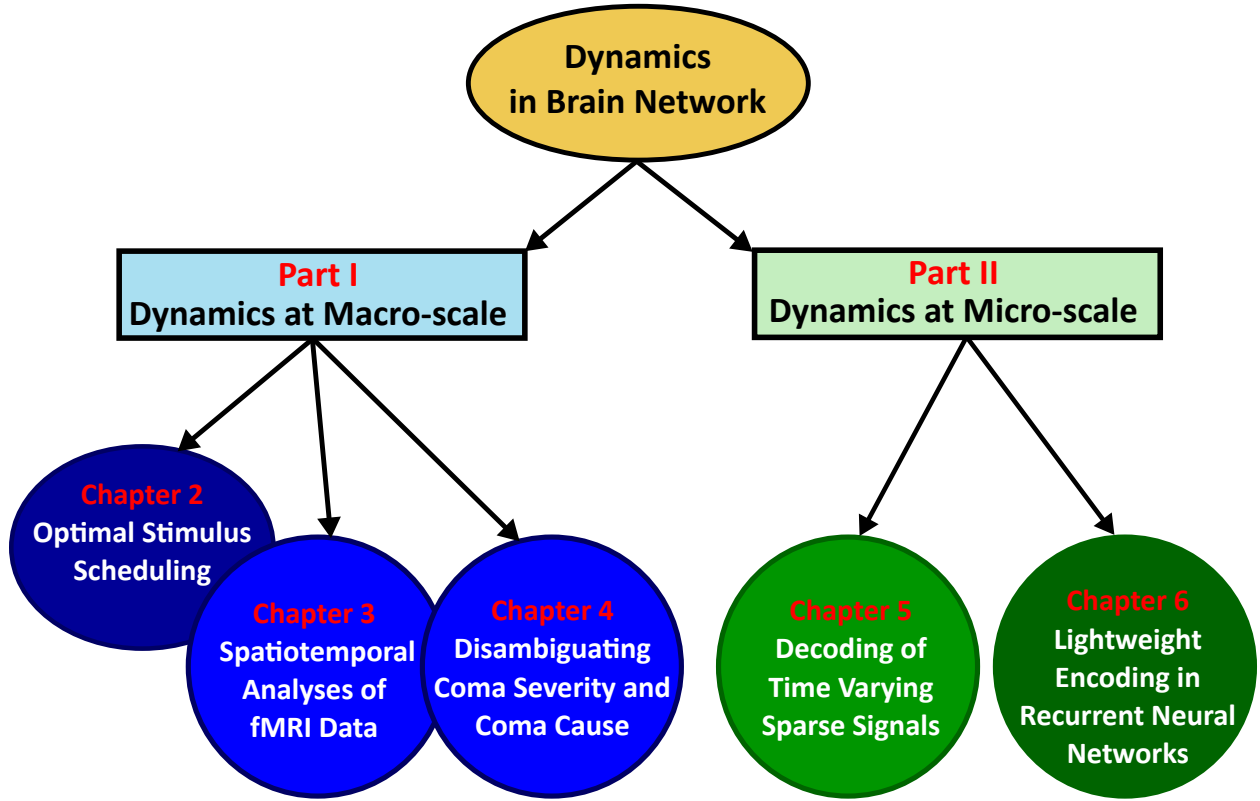
Figure 1.1: Structure of the dissertation.

for connectivity estimation, then embed these methods in a data-driven framework for characterizing the underlying network dynamics. These frameworks are deployed in the study of human brain activity from patients with altered mental status.

Following this macro-scale analytical development, we proceed to part two of the dissertation, in which we examine micro (i.e., neuronal) scale dynamics through a more theory-centric approach. Specifically, we examine the problem of neural encoding/decoding, or, how neurons in the brain enable the efficient processing of information. We consider simple dynamical models for neuronal activity and analyze, then optimize the network dynamics in order to enable information decoding.

Thus, the results of this dissertation represent a broad study that spans theoretical and data-driven development, unified under the central thematic thrust of examining links between brain dynamics and brain function.

## 1.2 Background and Motivation

Below, we provide background and motivation of the major topics represented in this dissertation. We focus on the two major parts described above, while detailed technical background on the specific contributions is left to their respective chapters.

### 1.2.1 Dynamics of Neural Connectivity

Characterizing associations between different brain regions has been a major question in neural signal analysis. The goal of this analysis is to show how different brain regions interact with each other as individuals perform different tasks, are in different cognitive states, or are in states of altered mental status due to brain injury [3–5]. The predominant metric that has been used for ascertaining association has been the Pearson correlation coefficient (the zero-lag cross-correlation of two time series). Based on this statistical metric, brain regions that are correlated in a particular context (e.g., during a resting state) are said to be *functionally connected*. Many other metrics, such as those in the frequency domain have also been used for this purpose. It is important to note that functional connectivity is distinct from actual anatomical connectivity; two regions may be functionally connected without possessing a direct physiological connection between them [1].

Despite demonstrable successes, previous characterizations of functional connectivity are limited by their largely static nature, since they are ascertained in the aggregate over long recording sessions. In other words, a single connectivity map is obtained for a subject in a cognitive state without regard for any of the temporal dynamics present within the underlying time-series of brain activity. In contrast, recent hypothesis have postulated a role for so-called dynamic functional connectivity (dFC) which relates to how association between brain regions may be changing in a second-to-second manner. Our goal in this work is to develop new methods and frameworks for ascertaining theses types of dynamics in functional connectivity, and associate such dynamical alteration in connectivity with states of neurological dysfunction.

## 1.2.2   Encoding/Decoding in Neuronal Networks

Sensory networks in the brain are organized in a layered architecture. An initial layer of neurons receives signals from the sensory periphery (e.g. sound, light, etc.), then propagate those signals to intermediate layers of neurons that are connected recurrently. These recurrent layers perform transformations of that information en-route to higher brain regions. It has been hypothesized that these sensory networks enable efficient encoding of information so as to optimally convey it and allow it to be decoded towards cognitive function [6]. However, the exact mechanisms of this information coding remain poorly understood.

Recently, parallels have been drawn between the architecture of such neuronal networks and engineered systems for efficient compression. In particular, given the temporally punctate nature of neuronal spikes, the theory of sparsity and compressed sensing (CS) has been considered as a schema for efficient encoding and decoding of information in sensory neuronal

4

networks [7, 8]. However, classical compressed sensing does not account for temporal dynamics within the forward model. Such dynamics are, of course, prevalent and pervasive in neuronal networks. Thus, we engage in a study of how recurrent neuronal dynamics in sensory networks may mediate CS-like coding. In other words, we elucidate conditions under which the dynamics of the network can enable or improve efficient information processing, particularly for time-varying signals.

## 1.3    Contributions

The original contributions in this dissertation are summarized below and in condensed form in Table 1.1:

**Optimal Stimulus Scheduling for Active Estimation of Dynamic Connectivity in Evoked Brain Networks [9, 10]:**    The first contribution relates to the estimation of time-varying weighted functional connections between brain regions. We consider the design of extrinsic stimuli that can excite the network to enable more efficient estimation of such connections. Specifically, we consider the notion of an evoked network. In such a network, nodes are defined as different brain regions. When a brain region is stimulated, it evokes a response in other nodes to which it is connected. The magnitude of the evoked response constitutes the weight of an edge between those nodes. We consider the problem of scheduling the stimulus; that is, determining the node to stimulate to best estimate these weights in a time-varying fashion. Given a state-space formulation for the connections, a solution to this problem can be obtained through a Kalman filter-like paradigm for recursive estimation of

Table 1.1: Table of contributions.

| Chapter | Technical Innovations | Major Findings | Research Output |
|---|---|---|---|
| Chapter 2 | Analytical derivation of optimal stimulus schedule using state-space formulation for network weights | Identified the equivalence of greedy and finite horizon optimal scheduling under specific conditions; Extracted network structure from evoked networks | [9, 10] |
| Chapter 3 | Established a Kalman filter-based approach for tracking correlation dynamics; Analyzed spatiotemporal structures underlying fMRI data | Identified orthogonal motifs with similar patterns of correlation dynamics that exhibit fundamental differences in awake and anesthetized conditions; Investigated altered brain microstates in resting wakefulness | [11, 12] |
| Chapter 4 | Utilized machine learning techniques toward disambiguating different coma etiologies; developed a novel lagged microstate analysis for assessing coma severity | Identified spectral and non-spectral EEG features as biomarkers for differentiating coma etiologies (focal vs. diffuse) and achieving 71% accuracy of classification; Uncovered a novel set of lagged microstates that appear specific to different levels of coma severity | [13] |
| Chapter 5 | Development of a generalized observability criterion for exact and stable input recovery in a dynamic recurrent network | Provided analytical bounds in terms of the system dynamics for recovery performance in the presence of both input disturbance and observation noise | [14] |
| Chapter 6 | Solved the problem of lightweight encoding in recurrent networks utilizing proximal gradient technique | Developed a linear network architecture with a soft-thresholding nonlinearity that is trained to efficiently encode an afferent input and its history | |

connection weights. We show the efficacy of the proposed methodology in several numerical examples, including estimation of a network of Wilson-Cowan neural mass models.

**Spatiotemporal Analysis of Functional Connectivity During Wakefulness and Sevoflurane Anesthesia [11, 12]:** Our second contribution relates to the analysis of dynamical structures that are present within functional magnetic resonance imaging (fMRI) data. Such data are typically analyzed in a static fashion as referenced above. Here, we overtly treat their temporal dynamics; that is we attempt to characterize the time-varying nature of functional connections and potential equilibria or fixed-points that might be reflected in the underlying time-series. To do this, we develop a framework that uses the filtering methodology from the first contribution together with a clustering analysis. This framework allows us to characterize the dynamics of functional networks by estimating time-varying connections between regions in a data-driven way and grouping similar patterns of time-varying connectivity between region pairs. We use this framework to show fundamental differences in the dynamics of functional connectivity of individuals in awake and anesthetized conditions. We also use this analysis to characterize different neuronal microstates present in neural data. A microstate is understood as a pattern of activation that tends to repeat in the data in a non-random fashion and, thus, may be related to the underlying equilibrium structure of the network dynamics. We show the existence of at least two microstates during resting wakefulness, and validate the methodology through a comprehensive set of positive controls.

**Characterizing Temporal Dynamics of Depressed Level of Consciousness with Applications in Detecting Coma Severity and Coma Cause [13]:** Finally, the third contribution of part one of the dissertation relates to the use of dynamical features extracted from electroencephalogram (EEG) data for the purpose of disambiguating individuals with severe brain injuries resulting in coma. We use our developed analytical frameworks as well as a range of other time-series analysis methods to disassociate the brain dynamics of individuals

7

who have coma due to focal or diffuse causes. In addition, we develop a new approach based on the aforementioned notion of microstates to characterize the temporal dynamics of EEG activity in patients in states of coma. We show fundamental differences in the dynamical structures, i.e. the microstates, of individuals who are deeply in unconscious states versus those that are maintaining certain measures of cognitive function. These analyses show the potential informativeness of temporal dynamics for clinical classification as well as their relation to different states of cognitive function.

**Decoding of Time-Varying Sparse Signals in Recurrent Linear Networks [14]:** In Part II, we move to the study of efficient coding in dynamical networks. Our first contribution here is in the study of networks with linear dynamics that receive high dimensional input. The network emits a lower dimensional set of outputs. We characterized the conditions on the network dynamics that enable a decoding of the input, from the outputs, in a time-varying manner in the presence of both disturbance and noise. Analytical bounds on performance are obtained, and demonstrated via simulation.

**Lightweight Encoding in Recurrent Networks with Soft-Thresholding Nonlinearities:** Finally, we consider how neuronal networks might enable information encoding. That is, given a raw signal (e.g., from the sensory periphery), how might a network produce a sparse, encoded version of that signal to be subsequently transformed, processed and eventually decoded. For this purpose we derive a dynamical network that, in essence, performs $\ell_1/\ell_2$ optimization so as to generate such an encoded representation from the raw input. We characterize the overall functionality of this network in terms of its ability to encode signals over a receding time horizon.

## 1.4   General Notation

We use a boldface capital letter to represent a matrix and a boldface lowercase letter to represent a column vector. For a given matrix $\mathbf{A}$, the inverse, transpose and determinant operators are given by $\mathbf{A}^{-1}$, $\mathbf{A}^T$, and $|\mathbf{A}|$, respectively. A matrix that maintain the rows of $\mathbf{A}$ with indices in set $T$ is denoted by $\mathbf{A}_T$. We use $diag(\mathbf{r})$ to return a diagonal matrix with elements of the vector $\mathbf{r}$ on its diagonal. For a given vector $\mathbf{r}$, $\|\mathbf{r}\|_0$, $\|\mathbf{r}\|_1$, and $\|\mathbf{r}\|_2$ are the $\ell_0$, $\ell_1$, and $\ell_2$ norms, respectively. We call a vector $\mathbf{r}$ as $s$-sparse if there are at most $s$ nonzero elements in $\mathbf{r}$. The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$. Finally, we denote $\|\mathbf{A}\|_2$ as the induced spectral norm of the matrix $\mathbf{A}$.

# Part I

# Dynamics at Macro-scale

# Chapter 2

# Optimal Stimulus Scheduling for Active Connectivity Inference in Evoked Brain Networks

## 2.1   Highlights

In this chapter, we consider the problem of optimal probing to learn connections in an evoked dynamic network. Such a network, in which each edge measures an input-output relationship between sites in sensor/actuator-space, is relevant to emerging applications in neural mapping and neural connectivity estimation. We show that the problem of scheduling which nodes to probe (i.e., stimulate) amounts to a problem of optimal sensor scheduling. By formulating the evoked network in state-space, we show that the solution to the greedy probing strategy has a convenient form and, under certain conditions, is optimal over a finite horizon.

We adopt an expectation maximization technique to update the state-space parameters in an online fashion and demonstrate the efficacy of the overall approach in a series of detailed numerical examples. The proposed method provides a principled means to actively probe time-varying connections in neuronal networks. The overall method can be implemented in real time and is particularly well-suited to applications in stimulation-based cortical mapping in which the underlying network dynamics are changing over time.[1]

## 2.2 Introduction

A topic of considerable and ongoing focus in neural engineering and computational neuroscience is the characterization of functional networks that describe the statistical association of the neural activity between spatially disparate brain regions [15]. In this context, the definition of a functional network requires an observation modality with multiple recording sites, and a statistical metric that can be used to quantify association between those sites. An association may, or may not, be attributed to a direct physical (anatomical) connection [11, 16, 17].

Many functional network estimation techniques are based on passive observation paradigms, wherein neuronal activity is recorded but not overtly manipulated. In contrast, the rapid emergence of neurostimulation technologies is making possible *active* network inference (circuit probing via stimulation, brain mapping), including the so-called evoked connectivity

---

[1]This chapter is a slightly amended version of the previously published articles which are listed as follows:
M. Kafashan, and S. Ching, "Optimal Stimulus Scheduling for Active Estimation of Evoked Brain Networks," *Journal of Neural Engineering*, 12(6), 066011, 2015.
M. Kafashan, and S. Ching, "Node selection for probing connections in evoked dynamic networks," *53rd IEEE Conference on Decision and Control*, 6080–6085, © 2014 IEEE.

[18–22]. In this paradigm, a third notion is added to the network definition: an input stimulus, e.g., an electrical current, which, when applied focally, leads to a detectable effect at the recording sites.

The evoked connectivity, or evoked networks, thus rely intimately on the notions of a stimulus (e.g. an input waveform) and a detector (e.g. an output response). Any evoked network is *specific* to a stimulus/detector pair, in the same way that a functional network is specific to the statistical metric used to define association. In [21], evoked networks were estimated assuming binary detectors, i.e., yielding a characterization of the probability of inducing a detected response given an input stimulus. In this study, we assume that the connections in the network follow a state-space modeling formalism. Such an approach for examining dynamic relationships between multiple brain regions is well-established in the neural connectivity modeling community. For instance, a state-space modeling approach for examining time-varying relationships between multiple brain regions was proposed in [23] and has been used in [24] to characterize temporal changes in resting-state functional connectivity. Similarly, in [25], a multivariate autoregressive model was developed using a time-varying connectivity matrix, in which each connection is assumed to follow a random walk model. This modeling paradigm has recently been adopted in several studies for modeling neuronal dynamic functional connectivity [26–28]. In this research, we generalize the theory of estimating evoked networks to consider time-varying weighted connections and, further, we address the problem of stimulus scheduling, i.e., how to determine which node to stimulate. More specifically,

1. We generalize the notion of an evoked network to consider weighted association, where the weights can vary as a function of time. For concreteness, but without loss of generality, assume that the input stimulus is a fixed input waveform applied to a single

network node, and let the detector measure the amplitude of signal deflection at all other nodes. We seek to estimate the input-to-output amplitude 'gain' for each pair of nodes. Given a state-space formulation for the connections, the solution can be obtained through a Kalman filter-like paradigm for recursive estimation of connection weights.

2. Assuming that only one site can be stimulated at a time, we determine the optimal stimulation policy (schedule) for best estimating the evoked network with the minimum number of stimulations. The assumption of single node stimulation is made for two reasons. First, it avoids the interference confound wherein, if two nodes are stimulated simultaneously, and an effect is observed in another node, then it is impossible to know which of the stimulated nodes was responsible [29]. Second, with many forms of stimulation technology, e.g., electrical microstimulation, it is difficult to stimulate and record from a node simultaneously.

   The developed theory is related to optimal design of experiments [30], but with a very specific form that allows for convenient analytical characterizations that facilitate online, adaptive evaluation of the stimulation schedule.

We show the efficacy of the proposed methodology in several numerical examples, including estimation of a network of Wilson-Cowan neural mass models. Here, we demonstrate a secondary result that if the delivered stimulus has sufficiently narrow bandwidth, the underlying causal network structure can be obtained from the evoked network through the solution of a linear inverse problem.

The remainder of this chapter is organized as follows. In Section 2.3 we formulate the scheduling problem in detail and develop its equivalence to the problem of sensor selection,
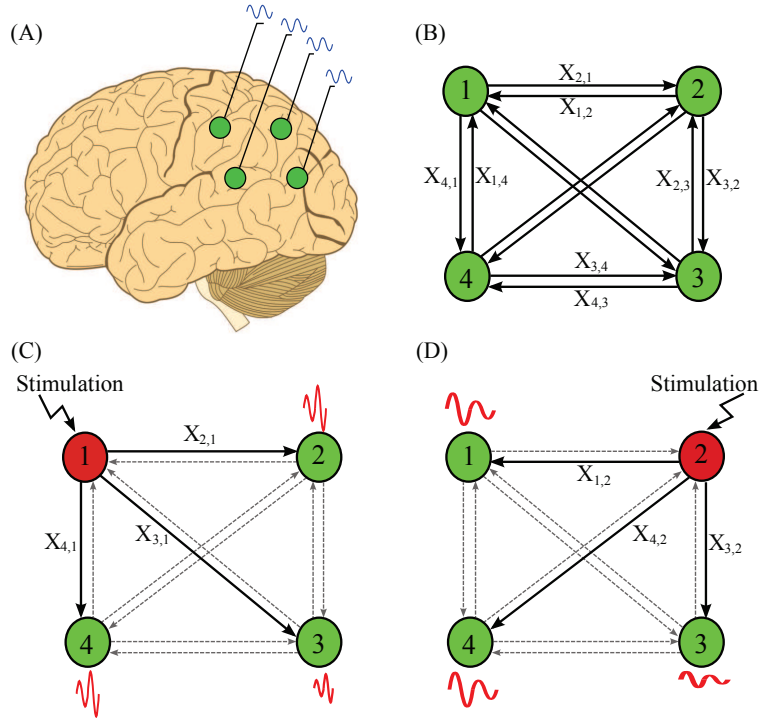
Figure 2.1: Active inference of evoked networks. (A) A functional brain network, constructed from the activity at four disparate recording/actuation sites. The resulting network consists of 12 directed edges. The weight of each edge corresponds to the magnitude of the evoked response when a 'source' node is stimulated. (B) Each probing yields observation of three edges. In this sense, the problem of determining the optimal probing for estimating weights is similar to that of sensor scheduling.

noting the relevant literature on solution methodologies. In Section 2.4 we develop methods for greedy probing and show certain results on optimality. In Section 2.5, several synthesized and biophysically inspired examples are provided. Conclusions are formulated in Section 2.6.

## 2.3 Problem Formulation

### 2.3.1 Basic Definitions

As introduced above, we consider evoked networks in the context of neural stimulation and recording. The underlying assumptions are the availability of neural data at $N$ spatially disparate sites, or nodes, and the ability to apply a prescribed exogenous input at each site. As is standard in network theory, the term *node* will be used to describe each of the $N$ sites and the term *edge* will be used to describe association between nodes.

### 2.3.2 Evoked Connectivity

Consider a network of $N$ nodes containing $N_T = N \times (N-1)$ directed edges. In a weighted evoked network, the weight of the edge $x_{j,i,k}$ corresponds to the magnitude of the detected response at node $j$ when node $i$ is stimulated at time index $k$. Let $\mathbf{x}_k$ denote the vector whose elements are $x_{j,i,k}$. Assume that $\mathbf{x}_k$ is sorted according to $\mathbf{x}_k = [\mathbf{x}_{1,k}^T \ \mathbf{x}_{2,k}^T \ ... \ \mathbf{x}_{N,k}^T]^T$ where $\mathbf{x}_{i,k}$ is vector of all edges connected to node $i$. In other words, $\mathbf{x}_{i,k}$ is $x_{j,i,k}$ for $j = 1, 2, ..., N, \ j \neq i$ at time $k$. We assume that the evoked connections in the network follow a state-space model described as

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{w}_{k-1}$$
$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k,$$

(2.1)

where $k \in \mathbb{Z}_{\geq 0}$, the state vector $\mathbf{x} \in \mathbb{R}^{N_T}$, the state noise process $\mathbf{w}_k \in \mathbb{R}^{N_T}$ is a $N_T$-dimensional random vector with multivariate Gaussian distribution having zero mean and diagonal covariance matrix $\mathbf{Q_w} \in \mathbb{R}^{N_T \times N_T}$, the measurement noise $\mathbf{v}_k \in \mathbb{R}^{N_T}$ is a zero-mean

multivariate Gaussian random vector with diagonal covariance matrix $\mathbf{R_v} \in \mathbb{R}^{N_T \times N_T}$.

$$\begin{aligned}
\mathbf{Q_w} &= diag\left(\alpha_1, ..., \alpha_{N_T}\right) & \in \mathbb{R}^{N_T \times N_T}, \\
\mathbf{R_v} &= diag\left(\beta_1, ..., \beta_{N_T}\right) & \in \mathbb{R}^{N_T \times N_T}.
\end{aligned} \qquad (2.2)$$

Two remarks are in order:

**Remark 2.1.** *The process noise $\mathbf{w}_k$ determines the extent to which evoked connectivity weights can change between time steps. A larger variance of $\mathbf{w}_k$ would correspond to larger temporal fluctuations in evoked responses, whereas a smaller variance would imply more static networks.*

**Remark 2.2.** *The model (3.4) assumes edges are independent. Clearly, this is an approximation. With prior specification of the relationship of evoked connectivity to actual physical connectivity, the independence assumption may be modified.*

In order to avoid an interference confound[2], only one node is stimulated at each time step meaning that only $N - 1$ weights are observed at a time. Thus, at each time k

$$\mathbf{C}_k \in \{\mathbf{C}_{1,k}, \mathbf{C}_{2,k}, \cdots, \mathbf{C}_{N,k}\}, \qquad (2.3)$$

where $\mathbf{C}_{i,k}$ is a diagonal matrix with $N - 1$ nonzero elements located at $jj^{th}$ element of the matrix in which $j \in \{(N-1)(i-1) + 1, ..., (N-1)i\}$. In other words, the parts of the observation matrix corresponding to states $x_{j,i,k}$, $j \neq i$, where $i$ is the node stimulated at time $k$, are non-zero.

The evoked connectivity estimation problem amounts to generating estimates of the weights $\mathbf{x}$ based on knowledge of the stimulations and the detected responses up to time step $k$. Here,

---

[2]That is, ambiguity in the source of an evoked connection when two nodes are stimulated simultaneously.

given the prior model (3.4), the best estimator in sense of mean squared error is the well-known Kalman filter (KF) [31–33]. In our specific case, the recursive KF update equations can be summarized as

$$\hat{\mathbf{x}}_{k|k-1} = \hat{\mathbf{x}}_{k-1|k-1}$$

$$\mathbf{P}_{k|k-1} = \mathbf{P}_{k-1} + \mathbf{Q_w}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{C}_k^T(\mathbf{C}_k\mathbf{P}_{k|k-1}\mathbf{C}_k^T + \mathbf{R_v})^{-1} \qquad (2.4)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}_k\hat{\mathbf{x}}_{k|k-1})$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\mathbf{P}_{k|k-1}$$

where, $\mathbf{P}_{k|k-1}$ is the estimated error covariance obtained in the propagate step using the state equation and $\mathcal{Y}_{k-1}$, observation up to time $k-1$. $\mathbf{P}_k$ is the estimated error covariance obtained in update step based on $\mathcal{Y}_k$, the observation up to time $k$.

While (3.9) enables recursive estimation of network weights, the problem of identifying the best node to stimulate at time step $k$ remains. We define a stimulation policy as follows: at each time $k$, choose a vector

$$\sigma_k = [\sigma_{1,k}, ...\sigma_{N,k}]^T \in \mathbb{R}^N$$

$$\sigma_{i,k} \in \{0, 1\}, \ i = 1, ..., N \qquad (2.5)$$

such that $\sigma_{i,k} = 1$ if node $i$ is stimulated at time $k$. Since we are limited to stimulating a single node at a time, we impose the additional constraint

$$\mathbf{1}^T\sigma_k = 1, \qquad (2.6)$$

where $\mathbf{1}$ denotes the vector of all ones, and further, noting that each of the $N$ possible choices for $\sigma_k$ corresponds to a distinct $\mathbf{C}_k$, as described above. The problem at hand is to determine a schedule for $\sigma_k$ that minimizes the estimation error covariance $\mathbf{P}_k$ over a specified time horizon, i.e., for a given schedule over horizon $M$,

$$\Sigma_M = [\sigma_1, \sigma_2, ..., \sigma_M], \tag{2.7}$$

minimize the average of the mean square error over the horizon $M$

$$J(\Sigma_M) = \frac{1}{M} \sum_{k=1}^{M} \mathbb{E}[\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|_2^2]. \tag{2.8}$$

### 2.3.3 Relationship to Sensor Selection

It is evident from the formulation that the problem of stimulation scheduling design is similar to that encountered in sensor selection, i.e., of a large number of sensors, the choice of which ones to sample at any given point in time in order to facilitate state estimation. Indeed, at each time $k$, we can conceive $\mathbf{C}_k$ as being a sample of $(N-1)$ 'sensors' that measure each evoked connection weight. More generally, such problems fall into the category of optimal design of experiments [30].

The sensor scheduling problem has received considerable attention in recent years [34–37]. The complexity of a such a problem is, generally, NP-hard [38] with optimal and near-optimal algorithms generally involving tree pruning techniques [39]. With appropriate relaxation, a dynamic programming procedure was developed in [40], however it is only applicable for an objective function that minimizes the terminal estimation error. A convex relaxation has been proposed [41], without guarantees on the performance and with complexity that

is cubic in the total number of available sensors. In [42] a selection policy according to a probability distribution is proposed to obtain the best upper bound on the expected steady state performance. In order to obviate complexity, a greedy approach is common [43, 44] and can be shown to have near optimal performance under submodularity of the cost function [45–48], which, in particular occurs if the process noise is zero [48].

In contrast to the general problem of sensor selection, our particular problem is more constrained since: (i) at any given time, we sample exactly $N - 1$ 'sensors', (ii) the set of possible samplings span the entire sensor space and (iii) the process and measurement noise covariances are diagonal. Thus, we are able to formulate relatively simple conditions under which convenient performance characterizations and guarantees can be made. We begin by analyzing the case of greedy scheduling, below.

## 2.4 Greedy and Optimal Probing in Evoked Networks

### 2.4.1 Greedy Approach

For the model (3.4), we can write $\mathbf{P}_k$ as

$$\mathbf{P}_k = diag\left(P_{1,k}, ..., P_{N_T,k}\right) \quad \in \mathbb{R}^{N_T \times N_T}. \tag{2.9}$$

From the Kalman filter (3.9), the error covariance for each state at time $k$ can be obtained recursively as

$$\begin{aligned}
P_{j,k} &= \sigma_{i,k} \frac{\beta_j(P_{j,k-1} + \alpha_j)}{P_{j,k-1} + \alpha_j + \beta_j} + (1 - \sigma_{i,k})(P_{j,k-1} + \alpha_j) \\
&= -\sigma_{i,k} \frac{(P_{j,k-1} + \alpha_j)^2}{P_{j,k-1} + \alpha_j + \beta_j} + (P_{j,k-1} + \alpha_j) \\
i &= \lceil \frac{j}{N-1} \rceil, \ j = 1, ..., N_T
\end{aligned} \tag{2.10}$$

where, the function $\lceil \ \rceil$ is the smallest integer larger than or equal to its argument. The greedy scheduling approach minimizes the 1-step mean square error

$$\begin{aligned}
&Minimize && J_k(\sigma_k) = \mathbb{E}[||\mathbf{x}_k - \hat{\mathbf{x}}_k||_2^2] \\
&Subject\ to && \mathbf{1}^T \sigma_k = 1 \\
& && \sigma_{i,k} \in \{0, 1\}, \ i = 1, ..., N.
\end{aligned} \tag{2.11}$$

The cost function in the above problem is the special case of the cost function in (2.8), with $M = 1$. The second and third lines of (2.11) represent the associated constraints that must be satisfied in the solution, i.e., stimulation of one node at a time. They mirror the definitions in (2.5) and (2.6).

**Lemma 2.1.** *The problem* (2.11) *is solved by stimulation of node* $m$, *where*

$$\begin{aligned}
m &= \underset{i}{\operatorname{argmax}} \sum_{j \in I_i} \frac{(P_{j,k-1} + \alpha_j)^2}{P_{j,k-1} + \alpha_j + \beta_j} \\
I_i &= \{(N-1)(i-1) + 1, ..., (N-1)i\}
\end{aligned} \tag{2.12}$$

*Proof.* Observe that the cost in (2.11) can be separated, resulting in

$$\textit{Minimize} \quad J_k(\sigma_k) = \hat{J}_k(\sigma_k) + \bar{J}_k$$

$$\textit{Subject to} \quad \mathbf{1}^T \sigma_k = 1$$

$$\sigma_{i,k} \in \{0,1\}, \ i = 1, ..., N.$$

$$\hat{J}_k(\sigma_k) = -\sum_{i=1}^{N} \sigma_{i,k} \sum_{j \in I_i} \frac{(P_{j,k-1} + \alpha_j)^2}{P_{j,k-1} + \alpha_j + \beta_j} \qquad (2.13)$$

$$I_i = \{(N-1)(i-1) + 1, ..., (N-1)i\}$$

$$\bar{J}_k = \sum_{j=1}^{N_T} (P_{j,k-1} + \alpha_j),$$

In (2.13), $\bar{J}_k$ is not a function of $\sigma_k$ and minimization of $J_k(\sigma_k)$ is equivalent to minimization of $\hat{J}_k(\sigma_k)$. Thus, the solution of problem (2.13) is equivalent to selection of $\sigma_k$ which results in maximum reduction of $\hat{J}_k(\sigma_k)$, i.e., the node that satisfies the criterion (2.12). □

Conceptually, the argument is understood as follows. We divide all possible edges into $N$ distinct groups. By selecting a node to stimulate, the error covariance for $N-1$ edges are updated, and the estimation error in all other groups is propagated.

We now proceed to explore conditions under which the greedy algorithm can have optimal performance over a finite horizon.

## 2.4.2 Greedy and Round-Robin Equivalence

We first present conditions under which the greedy algorithm reduces to a round-robin (i.e., periodic) probing schedule. In particular, consider the case where the process and measurement noise variances are equal, i.e.,

**Assumption 2.1** (A.1).

$$\alpha_1 = \alpha_2 = ... = \alpha_{N_T} = \alpha$$

$$\beta_1 = \beta_2 = ... = \beta_{N_T} = \beta.$$

(2.14)

Such an assumption, while seemingly strong, is reasonable when one considers applications such as probing a neuronal network using, for instance, a uniform grid of electrodes [22], with no prior assumptions on the evoked network structure or dynamics. Furthermore, consider

**Assumption 2.2** (A.2).

$$\mathbf{p}_0 = [p_{1,0}, ..., p_{N,0}]^T \in \mathbb{R}^N$$

$$\forall j \in I_i, P_{j,0} = p_{i,0}, \ i = 1, ..., N$$

$$I_i = \{(N-1)(i-1)+1, ..., (N-1)i\}.$$

(2.15)

This assumption states that the initial error covariance for all edges in the same group is equal (but edges in different groups can, in general, have different initial values).

Finally, we consider

**Assumption 2.3** (A.3).

$$p_{i,0} + \alpha > \beta, i = 1, ..., N,$$

(2.16)

which states, roughly, that the initial uncertainty is at least as large as the observation error variance.

We now state two preliminary lemmas whose proofs are strictly algebraic and thus omitted,

**Lemma 2.2.** *For $p, \alpha, \beta > 0$ and $c \in \mathbb{Z}, c > 1$ it holds that*

$$\frac{\beta(p + c\alpha)}{p + c\alpha + \beta} < \frac{\beta(p + (c - 1)\alpha)}{p + (c - 1)\alpha + \beta} + \alpha. \tag{2.17}$$

**Lemma 2.3.** *The function $f(p) = \frac{(p+\alpha)^2}{p+\alpha+\beta}$ is monotonicity increasing in $p > 0$ for $\alpha, \beta > 0$.*

Further, we consider:

**Lemma 2.4.** *Let $p_{1,0} = p_{2,0} = ... = p_{N,0}$, and let $\Sigma_N = [\sigma_1, \sigma_2, ..., \sigma_N]$ with $\sigma_{i,k} = 1$, $if$ $i = k$, then after $N$ time steps, $p_{1,N+1} > p_{2,N+1} > ... > p_{N,N+1}$, $\forall \alpha, \beta, p_{i,k} > 0$.*

*Proof.* After $N$ time steps and knowing the schedule $\Sigma_N$, the error covariance for the edges connected to node $i$, can be written as

$$p_{i,N} = \frac{\beta(p_{i,0} + i\alpha)}{p_{i,0} + i\alpha + \beta} + (N - i)\alpha \tag{2.18}$$

It can be seen based on equation (2.17) that

$$\frac{\beta(p_{1,0} + \alpha)}{p_{1,0} + \alpha + \beta} + (N - 1)\alpha > \frac{\beta(p_{2,0} + 2\alpha)}{p_{2,0} + \alpha + \beta} + (N - 2)\alpha$$
$$> ... > \frac{\beta(p_{N,0} + N\alpha)}{p_{N,0} + \alpha + \beta} \tag{2.19}$$

The above condition still holds if $p_{1,0} > p_{2,0} > ... > p_{N,0}$ because equation (2.18) is increasing with $p_{i,0}$. To make the last statement more clear, assume that $p_{1,0} > p_{2,0}$, then based on

24

equation (2.17) we have

$$\frac{\beta(p_{2,0} + \alpha)}{p_{2,0} + \alpha + \beta} + (N-1)\alpha > \frac{\beta(p_{2,0} + 2\alpha)}{p_{2,0} + \alpha + \beta} + (N-2)\alpha \tag{2.20}$$

We know that the equation (2.18) is increasing with $p_{i,0}$, therefore based on $p_{1,0} > p_{2,0}$ we can write

$$\frac{\beta(p_{1,0} + \alpha)}{p_{1,0} + \alpha + \beta} + (N-1)\alpha > \frac{\beta(p_{2,0} + \alpha)}{p_{2,0} + \alpha + \beta} + (N-1)\alpha \tag{2.21}$$

which results in

$$\frac{\beta(p_{1,0} + \alpha)}{p_{1,0} + \alpha + \beta} + (N-1)\alpha > \frac{\beta(p_{2,0} + 2\alpha)}{p_{2,0} + \alpha + \beta} + (N-2)\alpha \tag{2.22}$$

$\square$

**Theorem 2.1.** *Assume, without loss of generality, that $p_{1,0} \geq p_{2,0} \geq ... \geq p_{N,0}$. Then, under assumptions (A.1)-(A.3), the greedy optimization (2.11) reduces to a 'round-robin' strategy that stimulates node 1 to node N in sequence, periodically.*

*Proof.* From (2.12), at time $k = 1$, the new error variances can be written as

$$p_{1,1} = \frac{\beta(p_{1,0} + \alpha)}{p_{1,0} + \alpha + \beta},$$

$$p_{i,1} = p_{n,0} + \alpha, \ \ i = 2, ..., N. \tag{2.23}$$

Similarly, at time $k = 2$, because $p_{1,1} < \beta$ and $p_{i,1} = p_{n,0} + \alpha > \beta$, $i = 2, ..., N$ then (2.12) implies that node 2 should be stimulated. Similarly, by iterating forward, for the first $N$ simulations the greedy scheduling stimulates nodes 1 through $N$, in sequence.

25

After $N$ time steps, at time $k = N + 1$, it follows from Lemma 2.2 and Lemma 2.3 that the error covariances satisfy is $p_{1,N} \geq p_{2,N} \geq ... \geq p_{N,N}$ and the result follows. $\qquad\square$

### 2.4.3  Greedy and Finite Horizon Optimality Equivalence

We now explore conditions under which the finite horizon problem (2.8) reduces to the greedy problem (2.11). We begin with a preliminary lemma, stated without proof.

**Lemma 2.5.** *For $p, \alpha, \beta > 0$*

$$\begin{aligned}
\frac{(p+\alpha)^2}{p+\alpha+\beta} &> p + \alpha - \beta, \\
\frac{(p+\alpha)^2}{p+\alpha+\beta} &< p + \alpha
\end{aligned} \tag{2.24}$$

**Theorem 2.2.** *Under assumptions (A.1)-(A.3), the schedule $\Sigma_M$ that solves (2.8) contains at most one stimulation of each node where $M \leq N$.*

*Proof.* The proof is by contradiction. First, consider the case of $M = 2$. Assume, without loss of generality, that in the optimal policy $\Sigma_2$, the node $m$ is stimulated twice, i.e., $\Sigma_2 = [\sigma_1, \sigma_2]$, where $\sigma_{m,1} = \sigma_{m,2} = 1$ and all other elements of $\Sigma_2$ are zero. In this case $J_1$ and $J_2$ are

26

$$\frac{J_1(\Sigma_2)}{N-1} = -\frac{(p_{m,0} + \alpha)^2}{p_{m,0} + \alpha + \beta} + \sum_{i=1}^{N}(p_{i,0} + \alpha)$$

$$\frac{J_2(\Sigma_2)}{N-1} = -\frac{\left(-\frac{(p_{m,0}+\alpha)^2}{p_{m,0}+\alpha+\beta} + (p_{m,0} + \alpha) + \alpha\right)^2}{-\frac{(p_{m,0}+\alpha)^2}{p_{m,0}+\alpha+\beta} + (p_{m,0} + \alpha) + \alpha + \beta} \qquad (2.25)$$

$$-\frac{(p_{m,0} + \alpha)^2}{p_{m,0} + \alpha + \beta} + \sum_{i=1}^{N}((p_{i,0} + \alpha) + \alpha)$$

$$J_{Total}^2(\Sigma_2) = \frac{1}{2}\left(J_1(\Sigma_2) + J_2(\Sigma_2)\right)$$

Now, suppose that a node $m' \neq m$ was stimulated at time step $k = 2$. In this case, the schedule becomes $\Sigma_2' = [\sigma_1', \sigma_2']$, where $\sigma_{m,1}' = \sigma_{m',2}' = 1$ and all other elements of $\Sigma_2'$ are zero. $J_1$ and $J_2$ for this case can be written as

$$\frac{J_1(\Sigma_2')}{N-1} = -\frac{(p_{m,0} + \alpha)^2}{p_{m,0} + \alpha + \beta} + \sum_{i=1}^{N}(p_{i,0} + \alpha)$$

$$\frac{J_2(\Sigma_2')}{N-1} = -\frac{\left((p_{m',0} + \alpha) + \alpha\right)^2}{(p_{m',0} + \alpha) + \alpha + \beta} \qquad (2.26)$$

$$-\frac{(p_{m,0} + \alpha)^2}{p_{m,0} + \alpha + \beta} + \sum_{i=1}^{N}((p_{i,0} + \alpha) + \alpha)$$

$$J_{Total}^2(\Sigma_2') = \frac{1}{2}\left(J_1(\Sigma_2') + J_2(\Sigma_2')\right)$$

Based on Lemma 2.5, it follows that

$$-\frac{(p_{m,0} + \alpha)^2}{p_{m,0} + \alpha + \beta} + (p_{m,0} + \alpha) < \beta. \qquad (2.27)$$

We know that $p_{m',0} + \alpha > \beta$ and based on Lemma 2.3 it is straightforward to see that whenever $m' \neq m$ then $J_{Total}^2(\Sigma_2') < J_{Total}^2(\Sigma_2)$, establishing the contradiction. The remainder of the proof follows an inductive argument, noting that over a horizon $M \leq N$, probing any node

that has not previously been stimulated will result in more reduction in the total cost function (via $J_2(\cdot)$) versus one that has already been stimulated. $\qquad\square$

We now provide a sufficient condition under which the greedy algorithm is finite horizon optimal.

**Theorem 2.3.** *Let assumptions (A.1)-(A.3) hold and, additionally, let*

$$\beta > \frac{\sqrt{N-1}}{\sqrt{N}-\sqrt{N-1}}\alpha. \tag{2.28}$$

*Then, the solution for (2.8) over a horizon $M \leq N$ is the greedy solution.*

*Proof.* Based on Theorem 1, in a finite horizon with $M \leq N$, each node should be selected just once and the total cost function can be written as

$$
\begin{aligned}
\frac{J^M_{Total}(\Sigma_M)}{N-1} = &-\frac{\left(p_{m''',0}+M\alpha\right)^2}{p_{m''',0}+M\alpha+\beta} - 2\frac{\left(p_{m'',0}+(M-1)\alpha\right)^2}{p_{m'',0}+(M-1)\alpha+\beta} \\
&- \ldots - (M-1)\frac{\left(p_{m',0}+2\alpha\right)^2}{p_{m',0}+2\alpha+\beta} - M\frac{(p_{m,0}+\alpha)^2}{p_{m,0}+\alpha+\beta} \\
&+ Constant, \ m''' \neq m'' \neq m' \neq m.
\end{aligned}
\tag{2.29}
$$

It can be seen from $J^M_{Total}(\Sigma_M)$ that each term leads to a different overall reduction of the $J^M_{Total}(\Sigma_M)$. In this equation $m, m', \ldots, m'', m'''$ are the node numbers that are stimulated at time $1, 2, \ldots, M$, respectively. It can be shown via algebraic manipulations that (2.28) is sufficient to ensure,

$$(M-i)\frac{(p+(i+1)\alpha))^2}{p+(i+1)\alpha+\beta} < (M+1-i)\frac{(p+i\alpha)^2}{p+i\alpha+\beta} \tag{2.30}$$

Hence, in order to get maximal reduction for (2.29) it is sufficient to maximize the term multiplied by $M$, then the term multiplied by $M - 1$ and so on, i.e., greedy scheduling. $\square$

We are now ready to state the following result,

**Theorem 2.4.** *Under assumptions (A.1)-(A.3), along with (2.28), the solution to the finite horizon problem (2.8) where $M > N$, is a periodic, i.e., round-robin, repetition of the greedy algorithm over the first $N$ steps.*

*Proof.* The result is a direct consequence of Theorems 2-3. $\square$

## 2.4.4 Steady State Error Covariance Bound

Due to the structure of the optimal policies studied here (e.g., round-robin), it is possible to make comments on bounds for the steady state error covariance. To do so, we need to consider $\lim_{k \to \infty} p_{i,k} = \lim_{k \to \infty} p_{i,k-N} = p_i$, $i = 1, ..., N$. The steady state recursive equations for the error covariance of each edge can be written as

$$p_i = \frac{\beta(p_i + i\alpha_n)}{p_i + i\alpha + \beta} + (N - i)\alpha, \; i = 1, ...N. \tag{2.31}$$

For each $i$, (2.31) has one positive solution. Bounds are obtained by evaluating the sum over $i$. When the probing scheduling is known a priori, we can solve (2.31) to find the steady state error covariance bound. Even in the case of the pure greedy approach of Section 2.4.1, bounds can be obtained by solving recursive equations similar to (2.31) in which, for each edge, $N$ equations are solved and the maximum and minimum of $N$ possible solutions for the trace of the error covariance are obtained.

### 2.4.5 Expectation Maximization (EM) for Joint State and Noise Parameters Estimation

A final obstacle in using the scheduling policies and inference methods derived in the previous parts is knowing the noise parameters that specify the state-space model. Since such information is difficult to define *a priori*, a principled way to update them alongside the state estimates is required. Here, we face the challenge of estimating noise parameters alongside our online probing strategy. To address this issue we use an EM algorithm [49, 50] to jointly estimate the state and model parameters from measured evoked responses.

Assuming diagonal covariance matrices for model parameters, we can decouple the update procedures for each edge, resulting in a reduction of the computational burden typically associated with EM methods. Our developed algorithm is slightly different from the standard EM technique for linear models due to the irregular sampling imposed by our scheduling policy. In other words, we jointly estimate the state, $\mathbf{x}_{j,k}, k = 1, 2, \cdots, m$, $\alpha_j$ and $\beta_j$ from the (irregular) observations $\mathbf{y}_{j,k}, k \in \mathcal{K}$, $1 \leq |\mathcal{K}| \leq m$. The detailed of derivation the EM algorithm for this scenario is provided in the Appendix A.1.

## 2.5 Examples

### 2.5.1 Numerical Examples

We proceed to illustrate the results through a series of simulation studies.
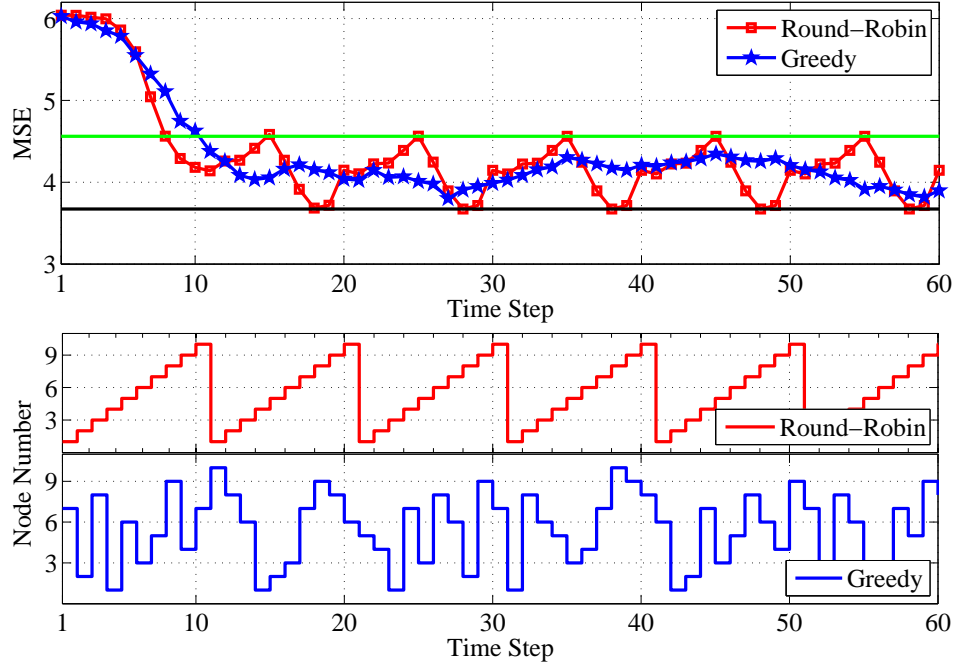
Figure 2.2: Trace of error covariance matrix with round-robin and variance greedy approach for node selection, $N = 10$ nodes. The upper and lower bound are indicated by the green and black lines, respectively.

**Greedy and Round-robin Scheduling:** Consider a network with $N = 10$ nodes with noise covariance matrices constructed by drawing $\alpha_i$ and $\beta_i$ from a uniform distribution in the intervals $[0, 0.1]$ and $[0, 0.3]$, respectively. Figure 2.2 shows the trace of the error covariance matrix (MSE) for the round-robin and greedy approach. The initial value of the error covariance for all edges are drawn from a uniform distribution in the interval of $[0.5, 1.5]$. The green and black lines in this figure constitute upper and lower steady state bounds for the round-robin policy, as obtained via (2.31).

**Optimality of the Greedy Solution:** We further consider the case where $\alpha = 0.1, \beta = 0.3$ and where initial conditions for the error covariances are assumed to be 1.5 for all edges. The results are illustrated in Figure 2.3. Clearly, the greedy and round-robin approaches are equal in this case, and the MSE curves are overlapping.

31
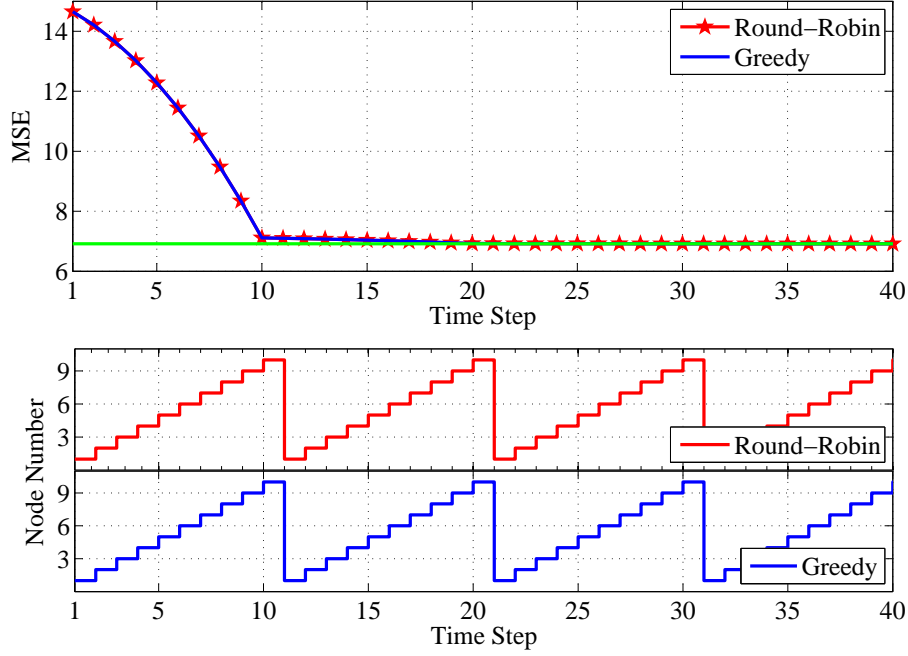
Figure 2.3: Trace of error covariance matrix with round-robin and greedy approach for node selection, $N = 10$ nodes. Here, assumption (A.1) is satisfied with $\alpha = 0.1, \beta = 0.3$.

Moreover, to illustrate the results of Section 2.4.3, we consider a network with $N = 5$ nodes, $\alpha = 0.1, \beta = 1.1$, i.e, Assumptions (A.1)-(A.3) and (2.28) are satisfied. Fig. 2.4 shows the trace of error covariance matrix for the greedy algorithm versus complete tree search over the horizon $M = 10$. We note equivalence of the two policies, and the round-robin (periodic) character of the greedy solution.

**Joint Noise and Connection Estimation via EM:** Figure 2.5 depicts the simulation results for joint model parameter estimation and network probing for a five node network. In this example, we assume $\alpha_j = 0.5$, $j = 2, 3, 5, 8, 14, 16$, while for the remaining evoked connections we assume $\alpha_j$ is drawn uniformly from the interval $(0, 0.1)$. In other words, a few edges in the network evolve with high process noise compared to other edges.
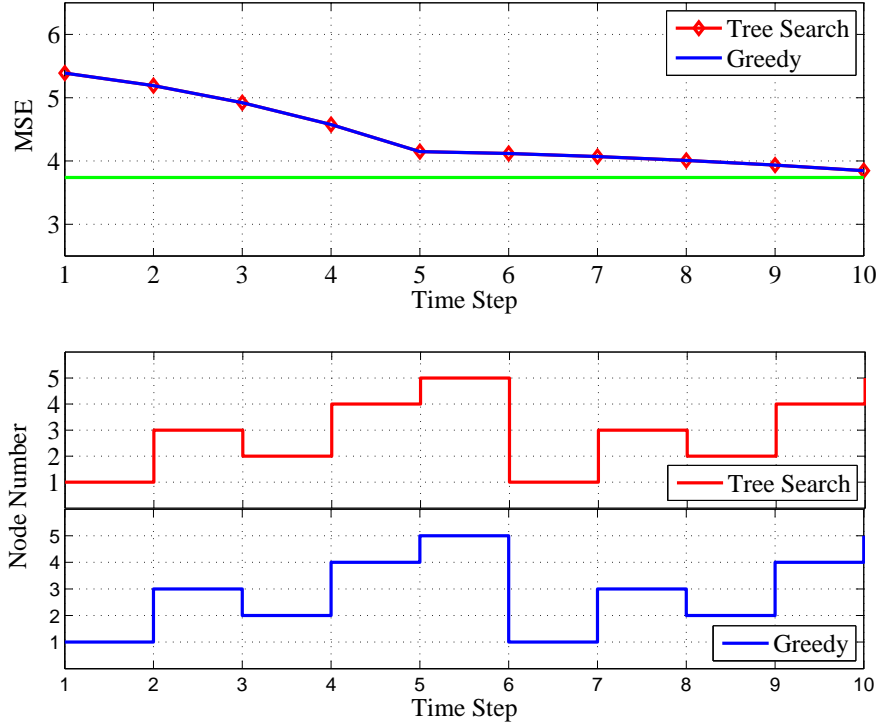
Figure 2.4: Trace of error covariance matrix with greedy algorithm and tree search for node selection assuming $\mathbf{A} = \mathbf{I}$, one for non-zero elements of $\mathbf{C_k}$ and 5 number of nodes with fixed $\alpha$ and $\beta$ satisfying assumption (A.1). $\alpha = 0.1, \beta = 1.1$ and initial error covariance drawn from uniform random distribution satisfying (A.2), (A.3) and equation (2.28).

Figure 2.6 shows the estimated process noise for the second state, $\alpha_2$, for both the round-robin and greedy approach. It can be seen from this figure that the greedy approach provides a better estimation of the noise variance compared to the round-robin. This is resulting in more samples of $\mathbf{x}_2$ compared to the uniform sampling via round-robin strategy.

Figure 2.7 illustrate true and estimated $\mathbf{x}_2$ using round-robin and greedy approach. Note, particularly, that in areas specified with dashed ellipses, the estimated state (evoked connections) with the greedy approach can track the true state closely as compared to the estimated state with the round-robin policy.

33

In the final example, we apply the developed theory to a biophysical neuronal network model, in which each node is a Wilson-Cowan mean field neuronal mass model [51–53]. In this model, the $j^{th}$ node is described by the nonlinear differential equations as

$$
\begin{aligned}
\dot{e}_j = & - e_j + (k_e - r_e e_j)\mathcal{F}(c_1 e_j - c_2 i_j + C_{e,j}(\mathbf{e}) + P_j(t)) \\
& + b_j^e u_j(t) + w(t) \\
\dot{i}_j = & - i_j + (k_i - r_i e_j)\mathcal{F}(c_3 e_j - c_4 i_j + C_{i,j}(\mathbf{e}) + Q_j(t)) \\
& + b_j^i u_j(t),
\end{aligned}
\tag{2.32}
$$
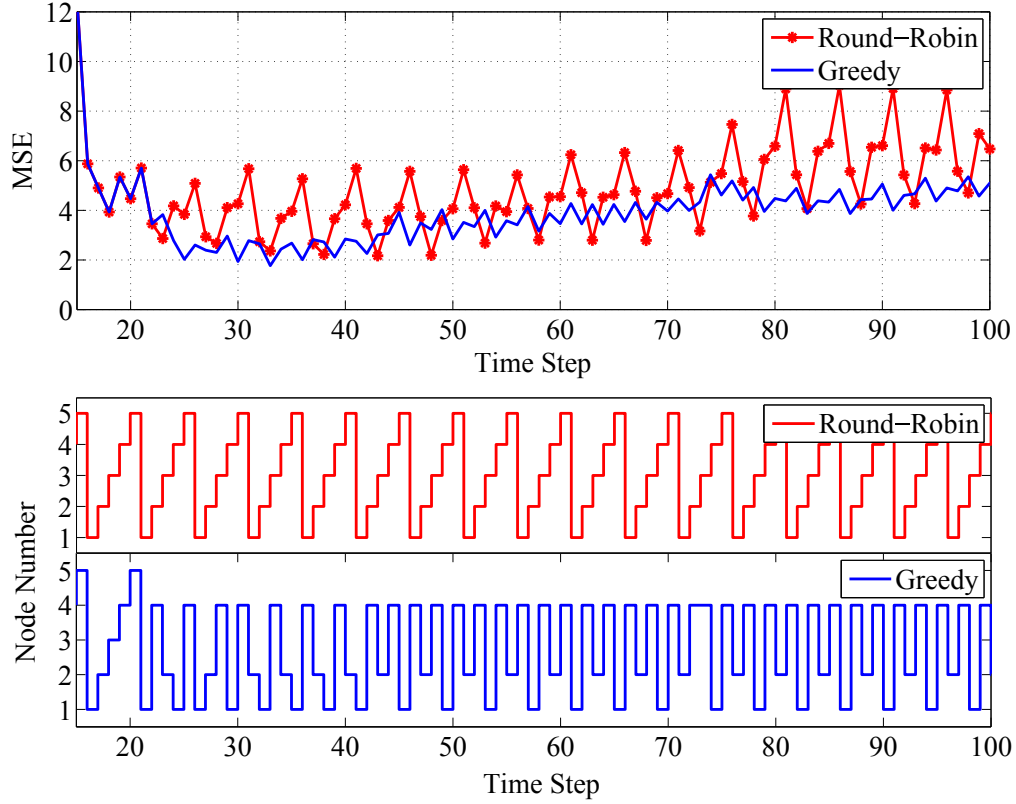


Figure 2.5: Trace of error covariance matrix with round-robin and greedy approach for node selection, $N = 5$ nodes without having information about noise variances.
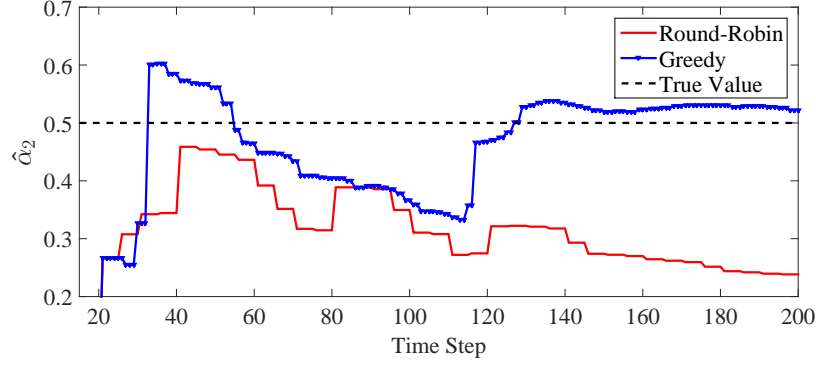
Figure 2.6: Estimated process noise for $\hat{\alpha}_2$ with round-robin and greedy approach. True value is shown with black dash line.
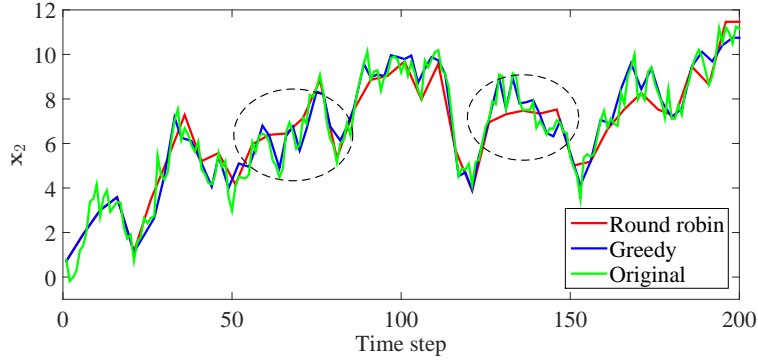


Figure 2.7: True (green) and estimated $\mathbf{x}_2$ with round-robin (red) and greedy (blue) scheduling.

where $(e_j, i_j)$ denote, respectively, the activity in excitatory and inhibitory cell populations in a cortical macrocolumn and $\mathbf{e}$ is the vector of the activity in excitatory cell populations for all macrocolumns. The connectivity between macrocolumns arises through the functions $C_{e,j}(\cdot)$ and $C_{i,j}(\cdot)$ and without loss of generality we consider the same function for both excitatory and inhibitory populations as

$$C_{x,j}(\mathbf{e}(t)) = \sum_{l=1}^{N} c_{jl} e_j(t - \tau_{jl}), x = e, i, \tag{2.33}$$

35
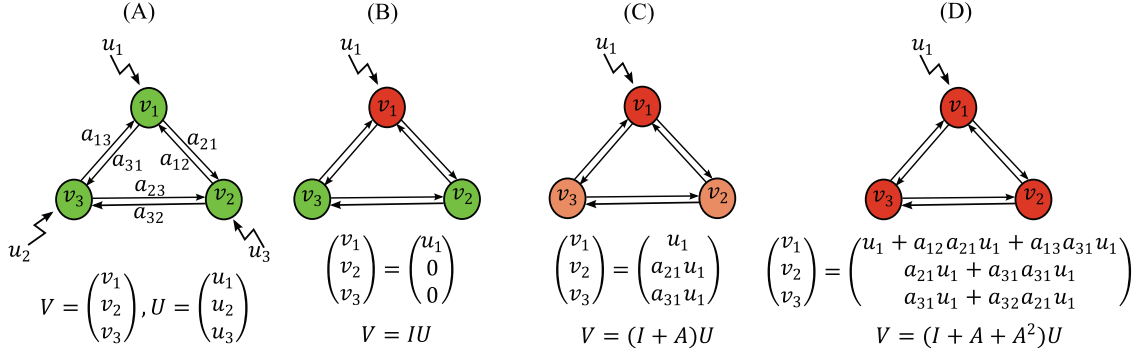
Figure 2.8: Network structure inference from evoked connectivity. (A) A three node network with structure matrix $\mathbf{A}$ where $a_{ij} = [\mathbf{A}]_{i,j}$ shows the effective gain/ attenuation and delay from node $j$ to node $i$. (B), (C) and (D) show how activity at each node, $\mathbf{V}$, evolves based on network structure when the first node is stimulated by input $u_1$.

where $c_{jl}$ and $\tau_{jl}$ are the amplitude and delay of a directed connection from node $l$ to node $j$, respectively. The term $u_j(t)$ is the exogenous input, which will facilitate active probing of the network connections.

**Evaluating Evoked Connectivity Performance in Terms of C:** In the above formulation, the matrix $\mathbf{C}$, $[\mathbf{C}]_{j,l} = c_{jl}$ represents the *actual* connectivity between nodes, which may mediate evoked connections via indirect paths (e.g., Node 1 to Node 2 to Node 3, leading to an evoked connection from Node 1 to Node 3). To the matrix $C$ we can associate a corresponding structural matrix $[\mathbf{A}]$ such that $[\mathbf{A}]_{j,l} = a_{jl}$ is the evoked connection from node $j$ to node $l$ due *only* to $c_{jl}$. In other words, $c_{jl}$ is a parameter in the dynamical system (2.32), whereas $a_{jl}$ describes the input-to-output gain (i.e., the evoked connection) that arises as a consequence of *only* $c_{jl}$ in (2.32). Note that $a_{jl}$ can be nonzero only if $c_{jl}$ is nonzero.

In order to evaluate the performance of our active inference scheme, we exploit the fact that when the delivered stimulus, i.e., $u(t)$ is sufficiently narrowband, the signal propagation via (2.32) is approximately linear; that is, a narrowband excitation travels from node to node,

36

incurring only attenuation and phase lag and not a shift in frequency. Thus, if the evoked network is defined in terms of a tightly band-limited stimulus and a similarly band-limited detector, we can exploit a formal model of the form:

$$\mathbf{V} = \left( \sum_{i=0}^{\infty} \mathbf{A}^i \right) \mathbf{U} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{U}, \tag{2.34}$$

where $i$ denotes the number of edges that mediate an evoked effect. That is, $i = 1$ describes direct connections only, $i = 2$ describes connections mediated by a cascade of two edges, etc. Figure 2.8 illustrates how evoked activity is related to the narrowband network structure $\mathbf{A}$.

Thus, the evoked connectivity matrix $\mathbf{X}$ is related to $\mathbf{A}$ via the relation

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,N} \end{pmatrix} = (\mathbf{I} - \mathbf{A})^{-1}. \tag{2.35}$$

Consequently, by inverting the evoked connectivity matrix according to (2.35), we can assess performance of our scheme against the known actual connections in $\mathbf{A}$.

**Performance in a 9 Node WC Network:** We consider a nine node network of the form (2.32) with parameters

$$c_1 = 16, c_2 = 12, c_3 = 15, c_4 = 3, a_e = 1.3, a_i = 2$$
$$\theta_e = 4, k_d = 2, \theta_i = 3.7, r_e = r_i = 1 \tag{2.36}$$
$$k_e = k_i = 1, b_j^e = b_j^i = 1, P_j(t) = 1, Q_j(t) = 0,$$

and w(t) taken as a Gaussian random process of variance 0.1. The matrix $\mathbf{C}$ was specified to be binary and symmetric, and is depicted in Figure 2.9A. With the parametrization (2.36), each node is not intrinsically oscillatory. Figure 2.10A shows the evoked response for nodes $2-9$ when the Node 1 is actively stimulated by a single harmonic input

$$u_1(t) = 10\cos(2\pi f_0 t)$$
$$u_j(t) = 0, j = 2, \cdots, 9,$$

(2.37)

where $f_0 = 1$, and the evoked response are measured over a window of 2 milliseconds. From Figure 2.10A, an induced oscillation at frequency $f_0$ is clear at the four nodes upon which node 1 impinges, i.e., the top row in Figure 2.9A. Also, it can be seen from this figure that the amplitude of the evoked response for the other nodes is low. Figure 2.10B shows the filtered version of the evoked response at $f_0$. It can be inferred from this figure that there are different delays associated with each connection.

## 2.5.2   Biophysical Network Simulation

**Inference of Static Structure:**   We proceeded to embed narrowband stimulus/detector in the overall evoked network methodology to obtain an evoked connectivity matrix, and then the structural matrix $\mathbf{A}$ via inversion. Figure 2.9B-D illustrate the performance of the inference after different time steps for a particular value of measurement noise $\beta$ (here, assumed to be known). Here, the structure itself was not allowed to change. Figure 2.11 illustrates the performance aggregated over different measurement noise intensities. As seen, the inference of network structure is quite accurate and, for low values of measurement noise, below 10% error.

38
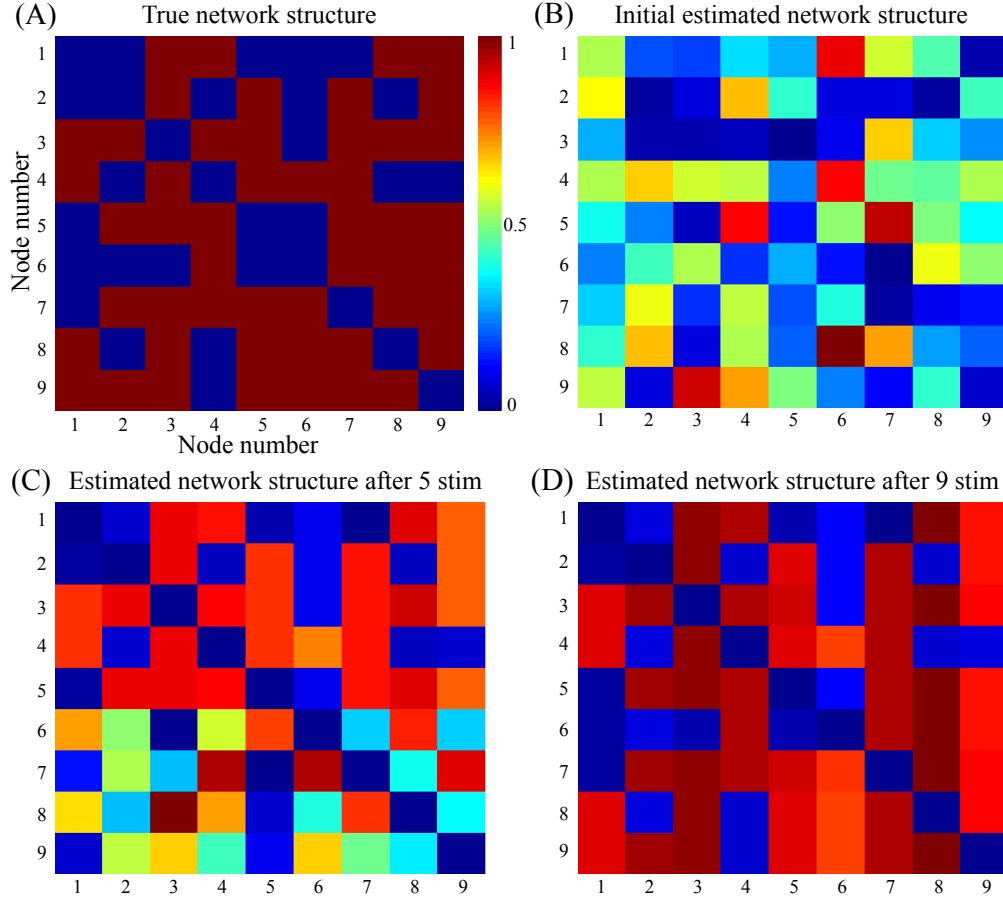
Figure 2.9: (A) Pattern of network connectivity for nine node Wilson-Cowan model. Red shows the existence of a connection while blue means no structural connection. (B) Random initialization of the estimated network structure before stimulation. (C) Estimated network structure after 5 stimulations. (D) Estimated network structure after 9 stimulations. The colorbar is the same for (A-D).

**Time-varying Connections:** Finally, we performed a simulation study of a connectivity matrix that changes over time, and there is no information about the process noise covariance matrix. Here, we assumed that delay between nodes is fixed and is not changing over time (with the same values as previous example). Therefore, the scheduling policies derived previously remain applicable. Similar to the previous example, the time window for probing and measuring the activity is specified as 2 milliseconds. At the first time step, we assume
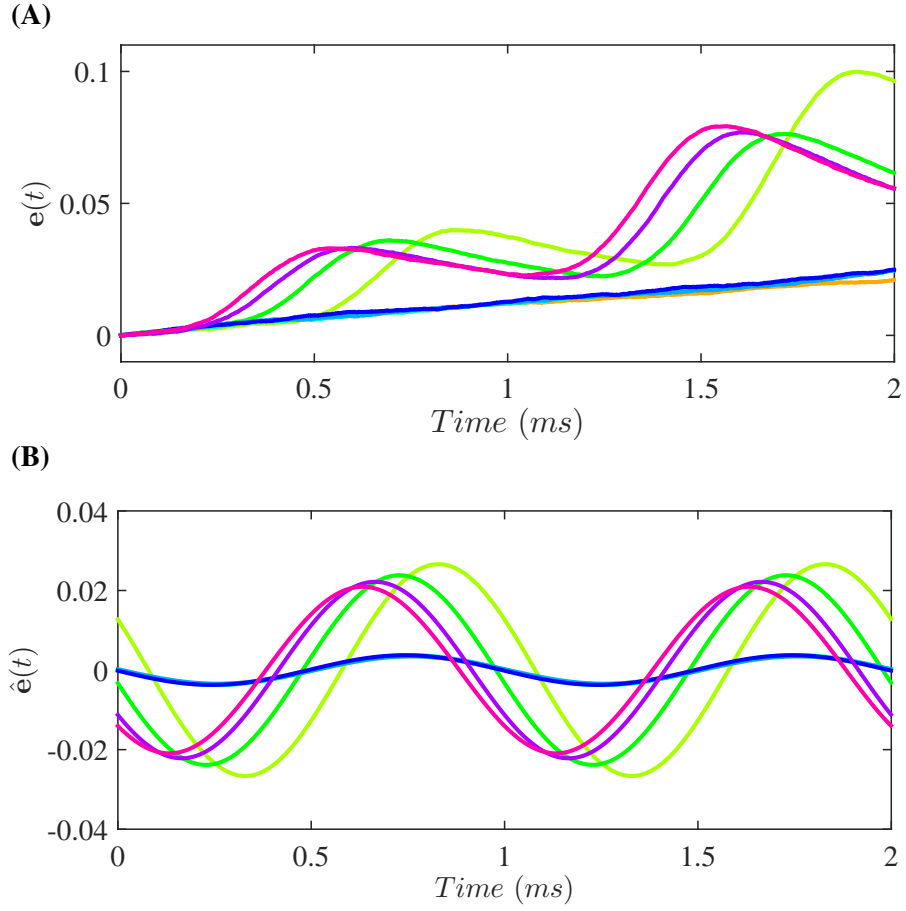
Figure 2.10: (A) Measured evoked response from nodes 2-9 when the first node is stimulated by a single tone input at frequency $f_0 = 1 \ Hz$. (B) Filtered version of the evoked response at central frequency $f_0$.

the connectivity, $\mathbf{C}(0)$, as shown in Figure 2.9A and then allow this to evolve based on the following relation:

$$\mathbf{C}_{jl}(k) = \mathbf{C}_{jl}(k-1) + w_{jl}^*(k), k = 1, \cdots, K, \tag{2.38}$$
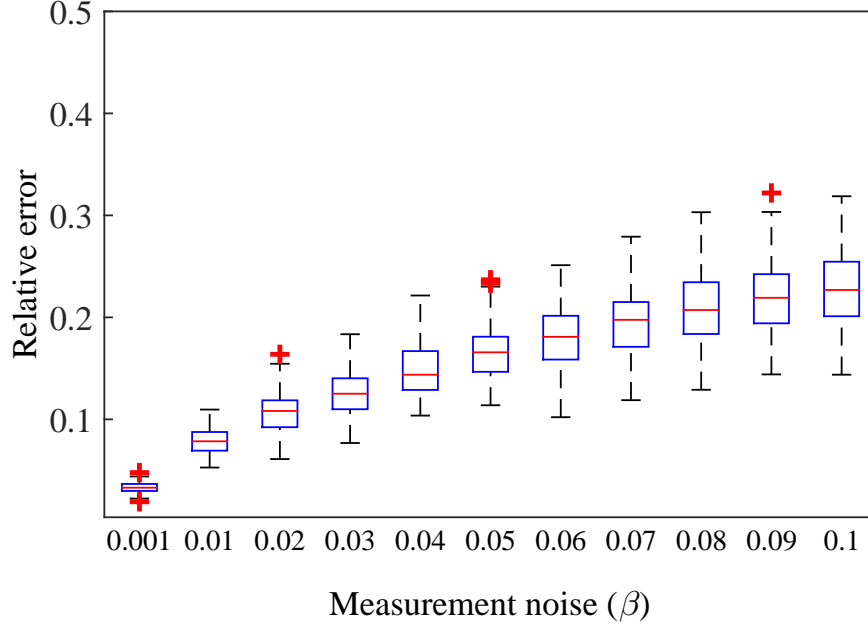
Figure 2.11: Box plot of the relative error versus measurement noise, $\beta$, for network structure inference for 100 Wilson-Cowan networks with 9 nodes and fixed network structure.

where $w_{jl}^*(k)$ is Gaussian random variable with standard deviation $\alpha_{jl}^*$. Here, $\alpha_{jl}^* = 0$ except for

$$\alpha_{1,3}^* = 0.5, \alpha_{1,4}^* = 0.5, \alpha_{3,1}^* = 0.5, \alpha_{4,1}^* = 0.5. \tag{2.39}$$

Intuitively, a good scheduling policy would be one that, after several time steps, probes exclusively the first, third and fourth nodes such that network changes can be tracked over time.

Figure 2.12 shows the results of our estimation procedure for both greedy and round-robin scheduling. It is evident from this figure that after a few time steps, the greedy policy indeed converges to an iteration between nodes 1, 3 and 4.

41
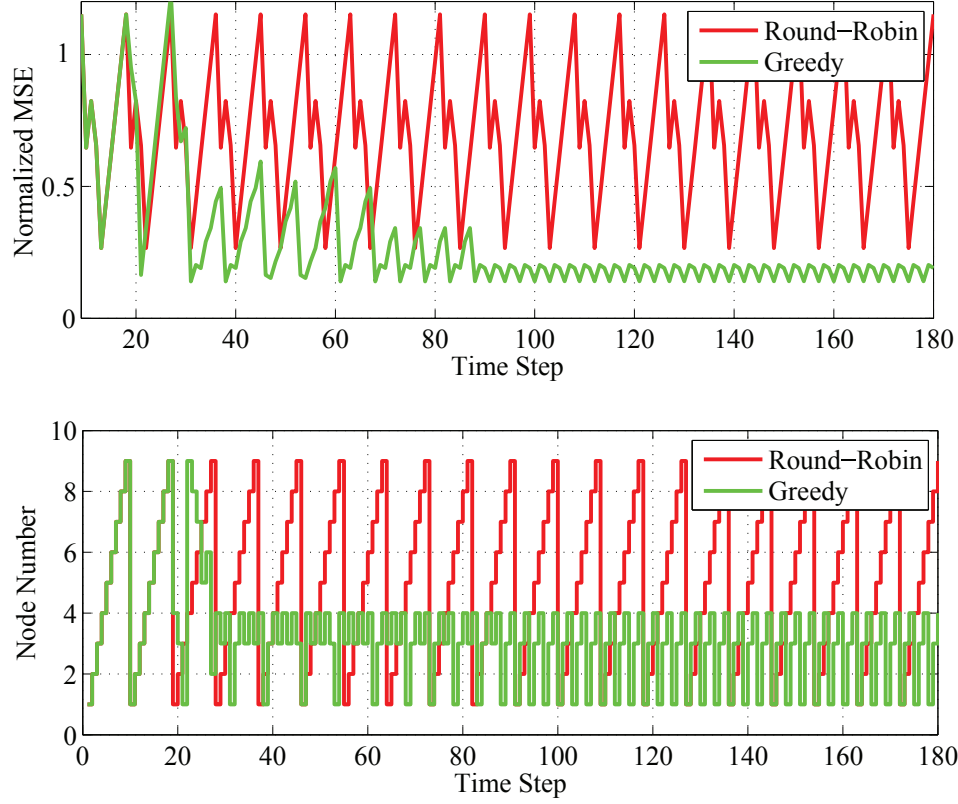
Figure 2.12: Scheduling policy and normalized MSE for round-robin versus greedy solution for Wilson-Cowan model with network structure which changes over time.

## 2.6 Chapter Summary

In this work, we have formulated the problem of node selection for probing connections in evoked networks. Here, each connection (weight) describes the node output responses when an input node is activated using a prescribed stimulus. Like 'functional' networks, such connections do not necessarily correspond to the physical network structure. However, in contrast to passive network inference methods, evoked connections are defined explicitly in both sensor *and actuator* space.

Here, we showed that the problem of deciding which node to stimulate in order to probe the evoked connections is equivalent to a problem of sensor scheduling, where each 'sensor' measures the weight of an edge relative to the delivered stimulus. Based on a Kalman filter paradigm for estimating the weights, we proceeded to derive convenient rules for greedy scheduling. Then, we showed that under certain conditions, the greedy algorithm is optimal over a finite horizon and, moreover, it amounts to a predetermined periodic 'round-robin' selection schedule. Those conditions – that measurement and process noise are equal across nodes; and that initial uncertainty of edges is grouped on a node-by-node basis – are expected to be particularly applicable in situations where the sensor/actuators are co-located, relatively homogeneous and minimal prior information is known about the weights.

We demonstrated the efficacy of the proposed methods in several examples, including a simulation study of a biophysical mean-field model. In this case, we showed that exploiting certain dynamical properties of the network, namely, approximate linearity with respect to a narrowband stimulus, allows for recovery of underlying structure from the evoked network through a linear inverse problem. Although not the focus of the present research, such a concept could certainly be generalized in future work.

The proposed methodology provides a principled way to interrogate the connections in brain networks via stimulation, with specific applications in brain mapping. More generally, our results fall into the category of perturbation analysis and system identification for complex networks, which is an emerging area of interest in the domains of biological, social and economic networks [54, 55].

# Chapter 3

# Spatiotemporal Analysis of Functional Connectivity During Wakefulness and Sevoflurane Anesthesia

## 3.1   Highlights

We provide an extensive analysis and evaluation of current dynamical functional connectivity (dFC) methods with a particular focus on the extraction of neuronal microstates, defined as repeating non-stationary patterns of network interaction. We show the existence of at least two microstates in the blood oxygen level dependent (BOLD) signal during resting wakefulness and point out key methodological considerations related to such analysis.

Further, we present a data-driven Kalman filter-based approach to tracking correlation dynamics. The filter explicitly accounts for the bounded nature of correlation measurements

through the inclusion of a Fisher transform in the measurement equation. An output linearizion facilitates a straightforward implementation of the standard recursive filter equations, including admittance of covariance identification via an autoregressive least squares method.

We demonstrate the efficacy and utility of these approaches in characterizing spatiotemporal patterns of correlated neural activity during the transition from resting wakefulness to general anesthesia. Specifically, we show how the interactions between so-called resting state networks (RSNs) become altered as individuals become sedated due to the actions of the general anesthetic drug Sevoflurane.[3]

## 3.2   Introduction

Characterization of large-scale brain networks using blood-oxygenation-level dependent (BOLD) functional magnetic resonance imaging (fMRI) is typically based on the assumption of network stationarity across the duration of a scan. As a result of the stationarity assumption, the conventional application of FC produces a single, static network of associations (between brain regions) for an entire trial or recording session. Such networks are essentially weighted graphs, where the recorded regions are the nodes and the statistical associations constitute the edge weights [15] (Fig. 3.1B). Analysis of such networks have produced groundbreaking descriptions of how different brain regions co-activate in function-specific ways.

---

Dynamic functional connectivity generalizes this methodology by allowing the associations (i.e., weights) to themselves covary in time [79]. Indeed, recent studies in humans have suggested that within-network functional connectivity fluctuates on the order of seconds to minutes [85,86]. Other studies have shown that functional networks can exhibit non-stationary, spontaneous fluctuations irrespective of cognitive processing [74,87]. These findings imply that mechanistically important network information can be missed when using average functional connectivity as a monolithic network measure. However, an opposing view is that such dynamicity is a spurious consequence of confounding factors such as head motion while subjects are recorded [94].

Thus, in this Chapter we pursue the characterization of dFC in two ways. First, we examine the basic question of whether legitimate dynamic structures can be extracted from fMRI recordings. In particular, we study the existence of neural microstates, within the conceptual framework of the brain transitioning in a series of brief movements between stationary regimes [70]. Second, we take a complementary approach to ask whether functional connections of the same *average* strength also possess the same temporal dynamics. In this regard, we introduce a new technical method for estimating time-varying correlations and use it to show fundamental changes in network dynamics during states of general anesthesia.

## 3.3   Data and Methods

### 3.3.1   Data Acquisition and Pre-Processing

We obtained fMRI recordings from a cohort of subjects who underwent recording while in an awake resting state and in a state of general anesthesia. These data were acquired and

preprocessed as previously described in [68]. Briefly, healthy human volunteers underwent BOLD fMRI (3T, repetition time, 2.2 seconds) during quiet wakefulness and while rendered unresponsive by 1.2 vol% sevoflurane. Stringent frame censoring of motion artifact and regression of whole brain global signal were applied to the BOLD signals. Cortical gray matter was segmented into 6 x 6 x 6 mm regions. A winner take all algorithm selected 1076 regions with more than 90% probability of assignment to seven RSNs: Dorsal Attention Network (DAN), Ventral Attention Network (VAN), Somatomotor Network (SMN), Visual Network (VIS), Frontoparietal Control Network (FPC), Language Network (LAN), Default Mode Network (DMN). Repeated pseudorandom resampling of these regions was performed to select 15 representatives of each RSN for each iteration. These subsampled data were used for our spatiotemporal decomposition.

## 3.3.2 Sliding Window Correlation

In order to characterize the temporal dynamics of functional connectivity, first we generate covariance matrices $\mathbf{S}_k$ over sliding windows for each subject at both 0% and 1.2% sevoflurane. Then, we estimated covariance from the regularized precision matrix using the graphical LASSO method by placing a penalty on the $\ell_1$ norm of the precision matrix ($\boldsymbol{\Theta}_k$) to promote sparsity [88, 89]. It assigns a large cost to matrices with large absolute values, thus effectively shrinking elements towards zero. For the $k^{th}$ sliding window, the following log-likelihood optimization problem is maximized:

$$\mathcal{L}(\boldsymbol{\Theta}_k) = \log\det(\boldsymbol{\Theta}_k) - \operatorname{tr}(\mathbf{S}_k\boldsymbol{\Theta}_k) + \lambda\|\boldsymbol{\Theta}_k\|_{\ell_1}, \tag{3.1}$$
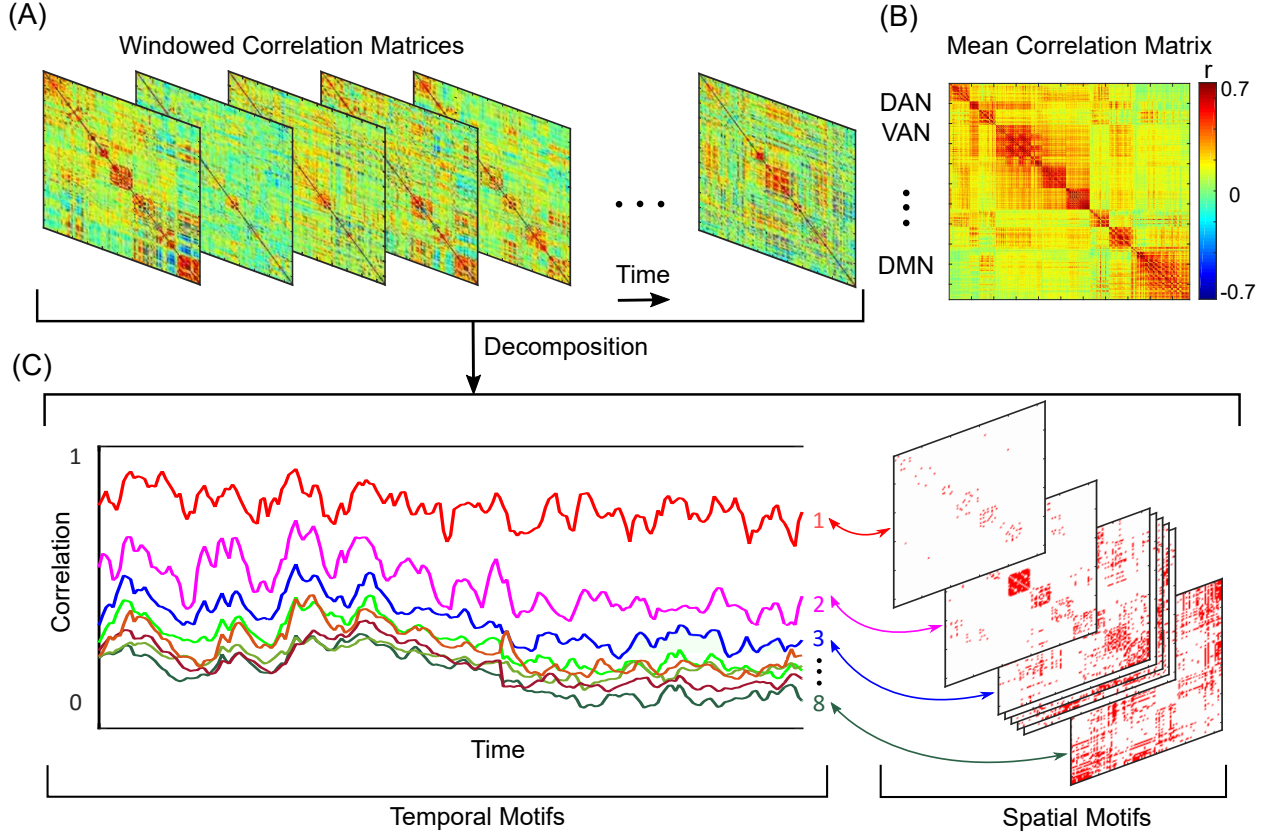
Figure 3.1: Schematic of Spatiotemporal decomposition of dynamic functional connectivity. (A) Functional connectivity over time. (B) Conventional FC over entire scan time. (C) Temporal and spatial motifs obtained from spatiotemporal analyses.

where the regularization parameter, $\lambda$, allows us to balance the trade off between log-likelihood and the number of nonzero coefficients in the inverse covariance matrix, which was set to 0.1 in our analyses. Thus, for each subject, we obtained an ordered set of (windowed) covariance matrices. The $(i, j)^{th}$ element of these matrices form the raw time-series of each weight within the functional network, which we will later treat with a novel filtering method. First, however, we examine the issue of stationarity within the set of covariance matrices.

## 3.4 Evidence of Microstate Dynamics During Resting Wakefulness

We begin with a basic investigation of whether certain spatial patterns are over-represented within the set of windowed covariance matrices. Such patterns, or microstates, would indicate bona fide dynamics within the fMRI data such as might arise if the network possessed a number of asymptotically stable equilibria. We specifically engage the negative control study performed in [94], which argues against such dynamics. Our key result here is that improperly applied dimensionality reduction can lead to substantial data misinterpretation and that, in fact, microstates exist. We show the existence of at least two microstates in the BOLD signal during resting wakefulness.

**Microstate Analysis of Resting Wakefulness:** To find microstates, we applied k-means clustering to the correlation matrices obtained from all 10 subjects (737 sliding windows in total). Correlation values between all region-pairs were included, resulting in $(105 \times 104)/2 = 5460$ features per matrix. The clustering algorithm was performed with random initialization of centroid points and was repeated 100 times to increase chances of escaping local minima.

**The Occupancy of Each Microstate and Microstates Similarity:** We define occupancy, or occurrence, of a microstate for a particular individual subject as the ratio of being at that microstate to the total number of sliding windows. We denote the occupancy of subject $i$ being at microstate $j$ as $p_i^j$ which is bounded between 0 and 1. The entropy of the occupancy for subject $i$ is defined as $E_i = -\sum_{j=1}^{2} p_i^j \log_2 p_i^j$.

**Positive Control Cases:** To investigate the existence of microstates in fMRI data of individuals during resting wakefulness, we generated different synthesized versions of the BOLD signal as follows:

1. In Case I, we generated stationary simulated BOLD fMRI time-series exactly in the way explained by Laumann et al. [94]. Briefly, for each individual subject, covariance matrix and power spectrum obtained over entire scan time of that particular subject is used to generate a surrogate multivariate time-series matched in covariance and spectral frequency of the original BOLD signal. By doing so, we can generate a stationary synthesized time -series for all the subjects in the dataset. The average functional connectivity of each subjects, calculated over entire scan time, is provided in the Appendix B.1 (Figure B.1).

2. In Case II, a non-stationary simulated BOLD fMRI time-series is generated based on the microstate analysis results on the real BOLD signal. For each subject, we considered two microstates with the assumption that the subject is jumping between these two microstates over time during resting wakefulness (see Appendix B.1 for more details).

3. In Case III, we generated stationary simulated BOLD fMRI time-series for each individual subject as in Case I, but we did not modify the power spectrum of the simulated signal.

**Statistical Analysis:** We calculated the correlation between microstates for each case separately to compare microstates obtained from the real BOLD signals to microstates obtained from different simulated fMRI time-series.

Differences in occupancy of being at a particular microstate across different cases, the real BOLD and simulated BOLD, were assessed using two sample Kolmogorov–Smirnov test (K–S test). The two sample K-S test is a nonparametric test for cases where the sample distributions are unknown which returns a test decision for the null hypothesis that two samples are from the same continuous distribution. Self-similarity of microstates across different time-series data, real and simulated, were tested using t-test on the z-values of microstates.

### 3.4.1 PCA Changes the Results of Microstate Analysis

Here, we show how applying PCA to time-series data (as performed in [94]) can fundamentally alter the results of microstate analysis. Figure 3.2A illustrates a two dimensional non-stationary data randomly generated with two distinct covariance matrices (red and blue colors). We consider the same sampling frequency as the real BOLD signals in our study. The scatter plot of the synthesized data is shown in Figure 3.2B. The covariance matrix of the two dimensional synthesized data, a $2 \times 2$ matrix, has three distinct values, because the off diagonal elements of covariance matrix are the same due to the symmetry. Figure 3.2C demonstrates three elements of the covariance matrix of the synthesized data obtained for window length of 50 samples; the same value is used to find FC in the real BOLD signal. The sliding correlation of the synthesized time-series data, which has one distinct value (diagonal elements of the correlation matrix are one), is shown in Figure 3.2D. It can clearly be seen in this plot that the correlation is going through a significant change after window number 20 (this is how we generated this data). However, applying PCA to the original synthesized time-series data conflate elements of the covariance matrix and the sliding correlation values across two different microstates as depicted in Figure 3.2E-F. In this toy example,
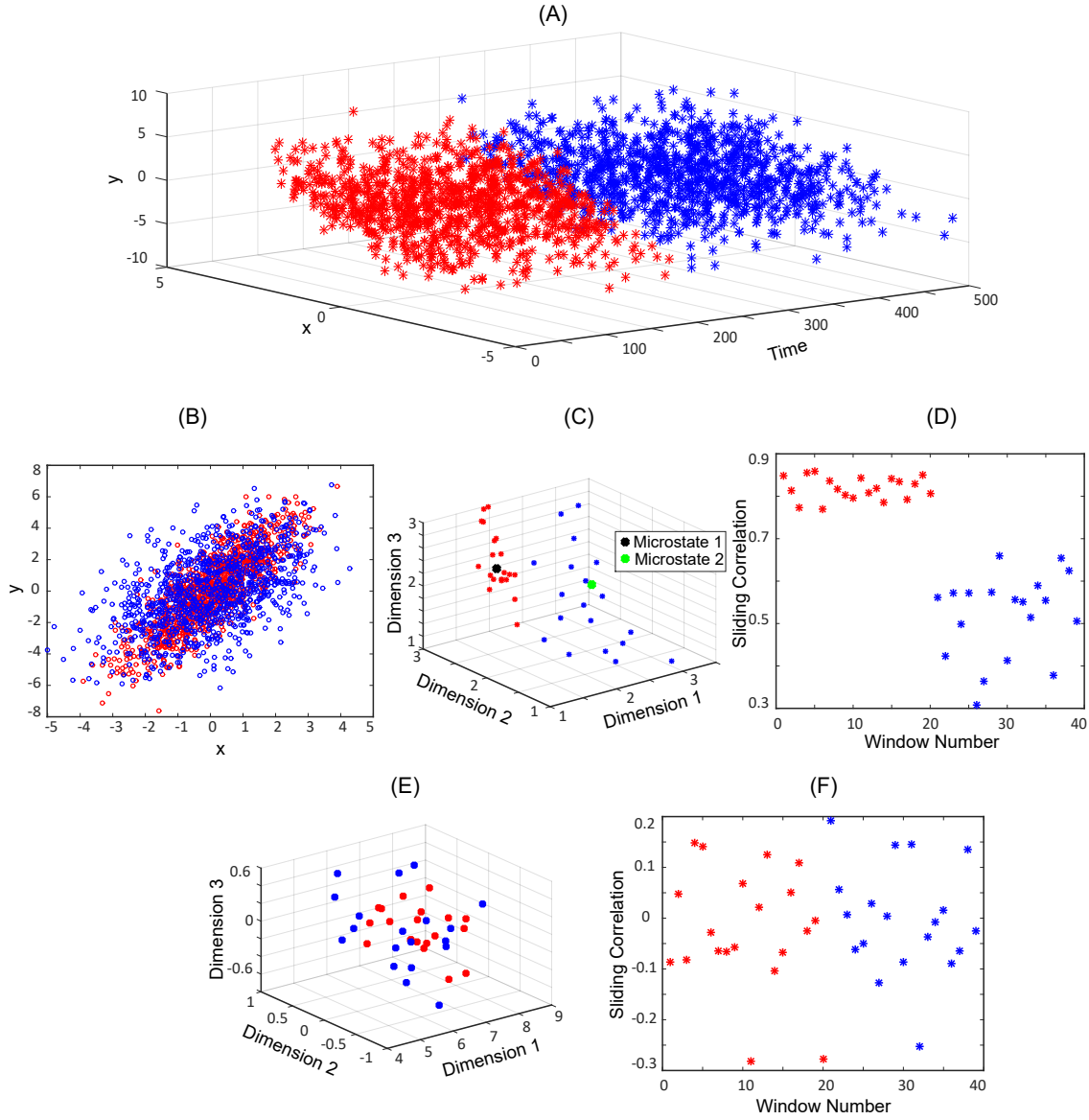
Figure 3.2: (A) A two dimensional non-stationary data randomly generated with two distinct covariance matrices (red and blue colors) over time. (B) Scatter plot of the synthesized data with a similar sampling frequency to that of real BOLD signals. (C) Three distinct elements of covariance matrix of the synthesized data obtained for window length of 50 samples, the same value is used to find FC in real BOLD signal. (D) Sliding correlation of the synthesized time-series data. (E) Three distinct elements of covariance matrix of the synthesized data obtained for window length of 50 samples after applying PCA and keeping all the dimensions (here just two dimensions). (F) Sliding correlation of the synthesized time-series data after applying PCA.
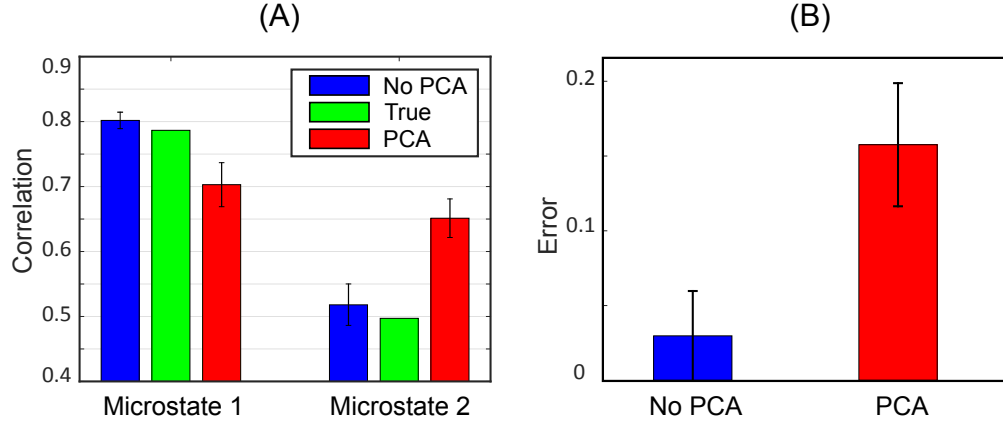
Figure 3.3: (A) Results of microstate analysis for estimating correlation for both microstates before and after applying PCA over 100 realization of random synthesized data. (B) Estimated correlation error before and after applying PCA to time-series data is highlighted in Figure 3.3B. Error bars stand for standard deviation (SD).

the dimensionality of the time-series data before and after applying the PCA were kept the same. We then applied the microstate analysis described in the section 3.4 to find the two microstates existed in the simulated data. We observed that the accuracy of finding the true microstates is reduced significantly in the case that PCA is applied to the time-series data as expected by just comparing Panel D and F in Figure 3.2. Figure 3.3A shows the results of the microstate analysis for estimating the correlation for both the microstates before and after applying PCA over 100 realization of random synthesized data. Estimated correlation error before and after applying PCA to time-series data is highlighted in Figure 3.3B. In this plot, error bars stand for standard deviation (SD). Covariance matrices of two microstates are fixed across all realizations. Exact covariance and correlation matrices, alongside the estimated ones, before and after applying PCA are reported in the Appendix B.1.
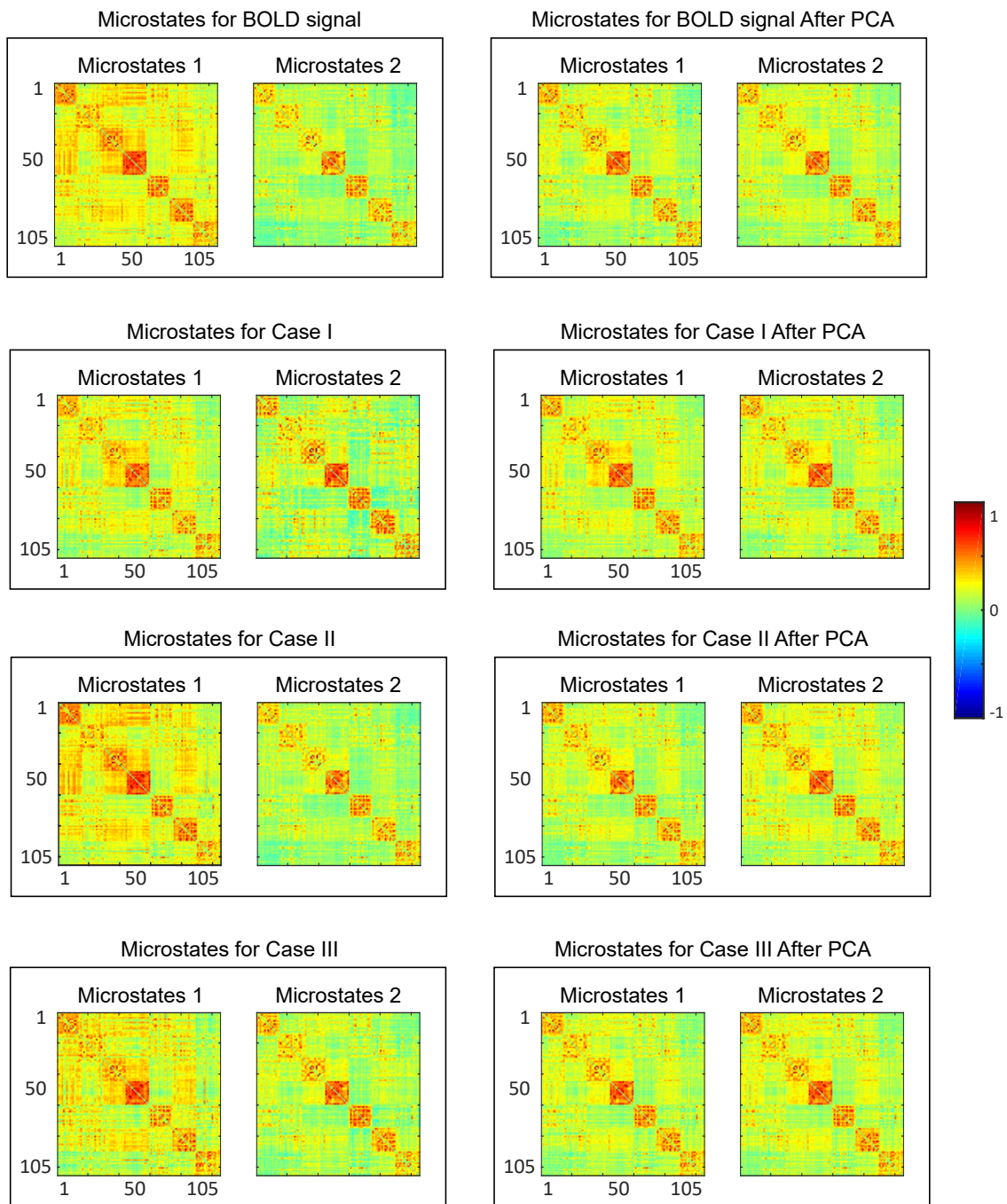
Figure 3.4: Two microstates obtained before, left panel, and after, right panel, applying PCA to the real BOLD signals, case I, case II and case III. For all the simulated cases, the covariance matrix and power spectrum of each subject is used to generate the simulated BOLD time-series for that particular subject.

## 3.4.2 Microstate Occupancy can Reveal Non-Stationarity of the BOLD Signal

Here, we present the results of microstate analysis on the real BOLD signal and different synthesized BOLD fMRI time-series, both stationary and non-stationary, to point out the importance of microstate occupancy in interpreting the salience of the obtained microstates. For all the cases introduced in the section 3.4, we present the results before and after applying PCA to the initial time-series data. Utilizing PCA analysis, the original time-series data was dimensionality reduced from 105 to 30 dimensions prior to clustering [94]. We chose $k = 2$ for our k-means analysis.

Figure 3.4 illustrates two microstates obtained before, left panel, and after, right panel, applying PCA for the real BOLD signals, and cases I-III (see section 3.4 for more detail on each cases). It can be concluded from this plot that the obtained microstates from all the cases are similar to the ones of the real BOLD signals. Also, we can see visually from this figure that by applying PCA to each case, the obtained microstates tend to be more similar to each other. For all the the simulated BOLD signals in this figure, the covariance matrix and power spectrum of each subject is used to generate the simulated BOLD time-series for that particular subject. To assess the similarity of microstate obtained from the real BOLD signals to the ones from the simulated cases, we calculated the pairwise similarity between each case. The similarity of microstates from the real BOLD signals to the ones from case I, case II and case III (with 100 realizations for the simulated BOLD signals) was 0.89±0.04, 0.97±0.01, and 0.90±0.02 (mean±SD), respectively.

Connectivity maps of two microstates for the real BOLD signal without PCA are shown in Figure 3.5. This plot demonstrates the left view of the connectivity maps. Also, the
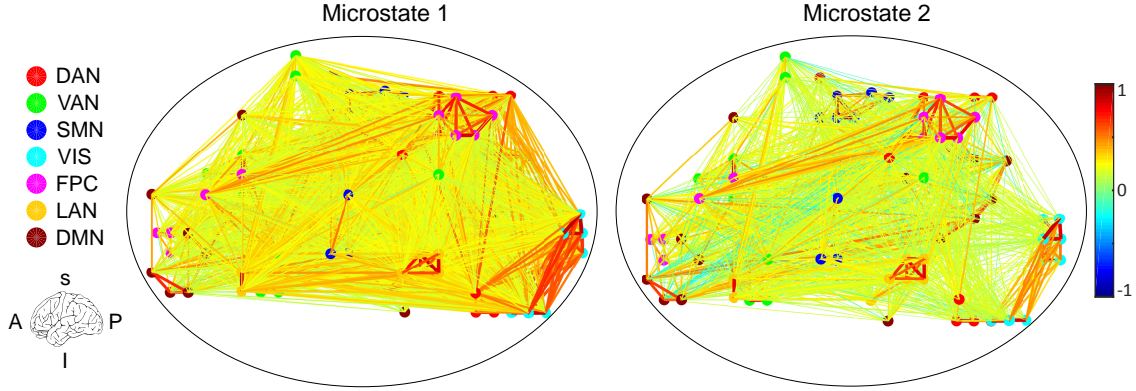
Figure 3.5: Connectivity maps of two microstates for the real BOLD signal without PCA. All connections with correlation value less than 0.1 are set to zero.

connectivity patterns of the two microstates between pairs of RSNs, for 10 pairs, are shown in Figure 3.6. In these illustrations, all connections with correlation value less than 0.1 are set to zero. We note that the FPC network and DMN are anti-correlated in the second microstate while there is no such relation between them in the first microstate. Furthermore, fairly strong correlation between VIS and DAN in the first microstate disappears in the second one.

Microstate occupancy across different subjects for the real BOLD signal, and cases I-III are demonstrated in Figure 3.7. It can be concluded from this plot that occupancy of being at either of the microstates (microstate 1 or microstate 2) in the real BOLD signal is entropic, and each subject exhibits jumping between these two microstates over the scan time. The same behavior is observed in the non-stationary simulated BOLD time-series, case II, as expected. However, for both case I and case III, the microstate occupancy before applying PCA shows mostly one microstate for each subject, which means that during the scan time that (simulated) subject does not reveal any jumping to another microstate.

As predicted from the previous section, applying PCA to the time-series data changes the microstate occupancy dramatically, especially for stationary simulated signals. Statistical
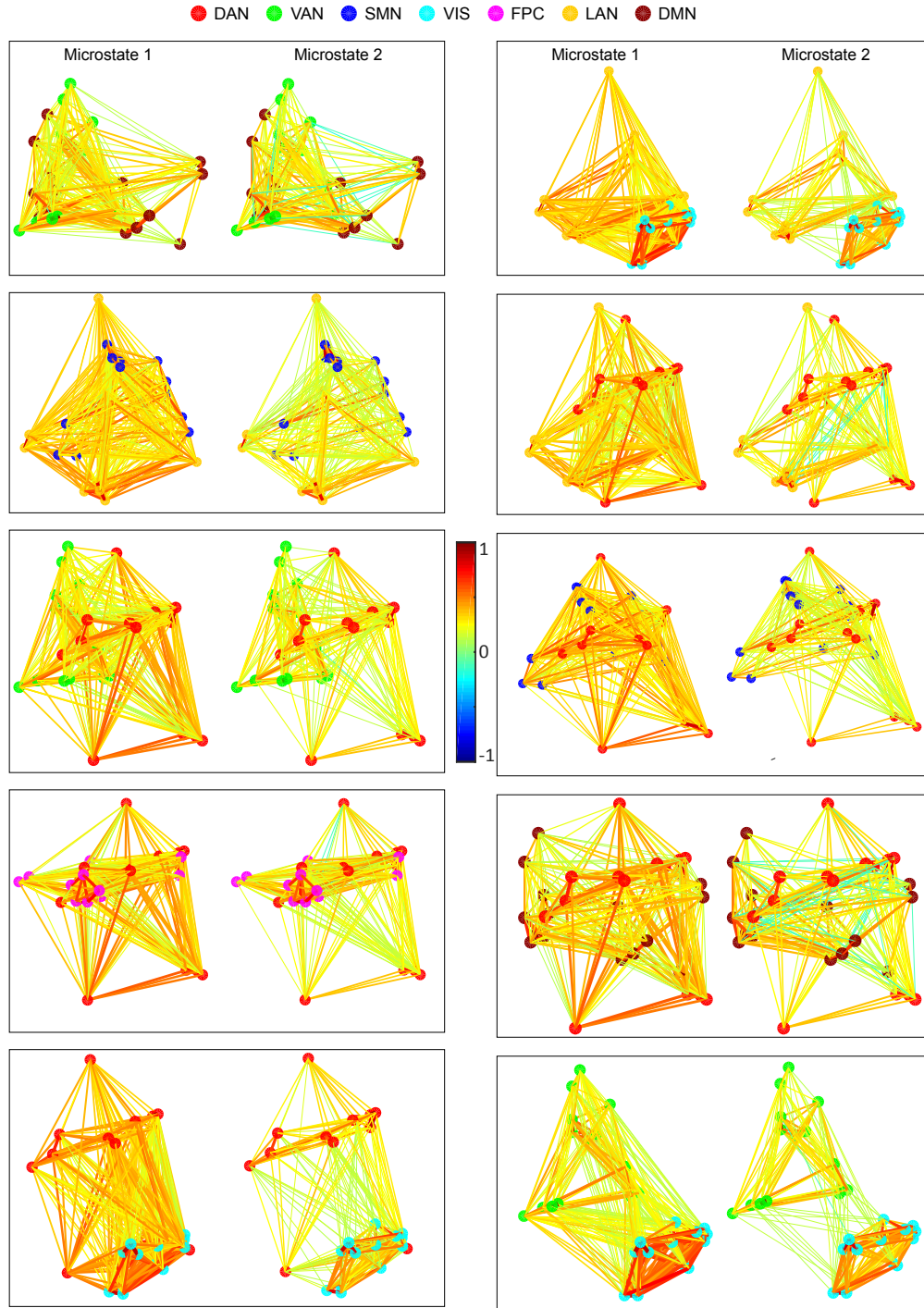
Figure 3.6: Connectivity patterns of the two microstates between pairs of RSNs, for 10 pairs. All connections with correlation value less than 0.1 are set to zero.
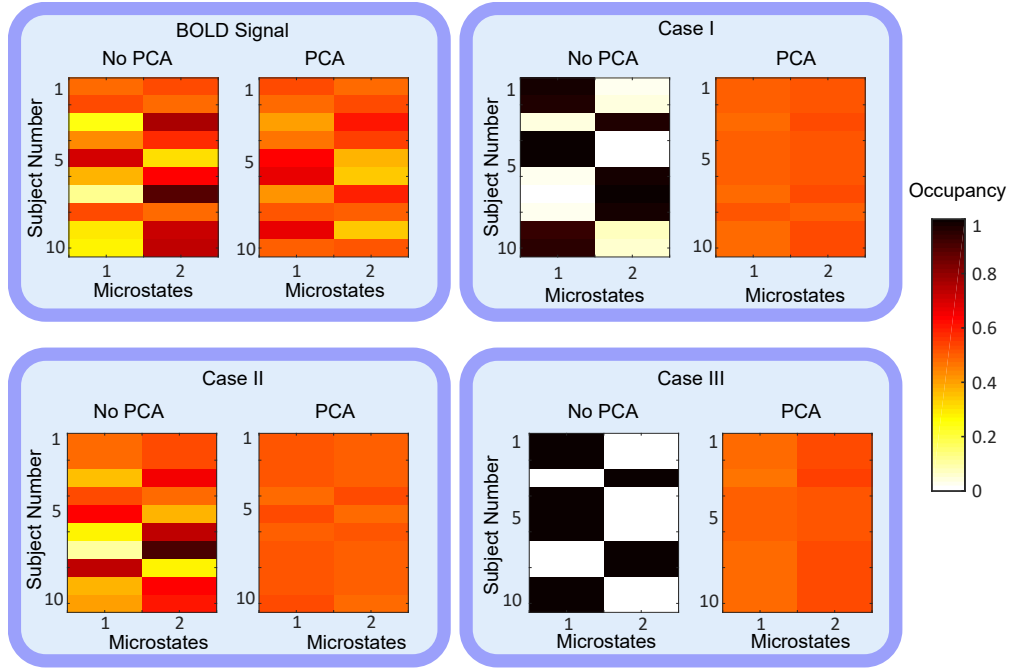
Figure 3.7: Microstate occupancy across different subject for the real BOLD signal, and cases I-III before and after applying PCA to original time-series data.

analysis revealed that the occupancy of being at the first microstate is significantly different between the real BOLD signal and case I (K-S test, p-value = 0.031, stat = 0.6). Moreover, the real BOLD signal is not significantly different from case II (K-S test, p-value = 0.975, stat = 0.2). Additionally, the real BOLD signal is significantly different from case III (K-S test, p-value = 0.007, stat = 0.7).

Sliding window correlations projected onto the first two principal components colored by microstate identity is shown in Figure 3.8. This figure clearly shows how applying PCA can change the outcome significantly. We further quantified the microstates similarity (correlation between two microstates) and microstate entropy for cases before and after applying PCA shown in Figure 3.9. We tested the difference between microstates similarity before and after PCA by utilizing t-tests on z-values of correlation (applying Fisher inverse transform to correlation values) over subjects. The statistical results show significant difference before
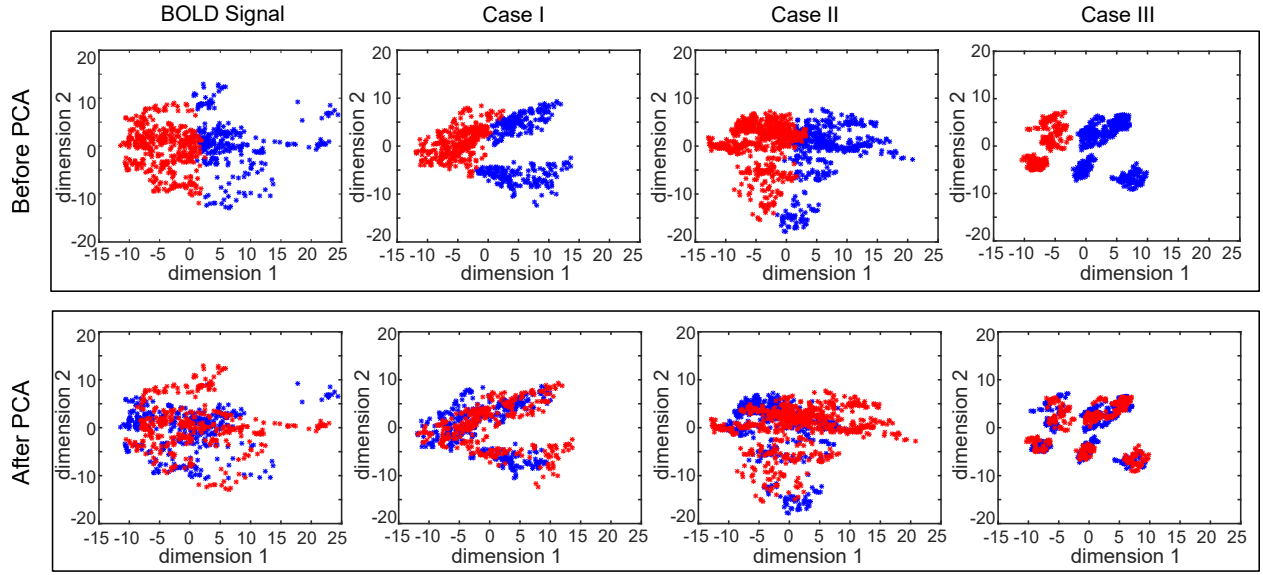
Figure 3.8: Sliding window correlations projected onto the first two principal components colored by microstate identity. Top panel shows sliding window correlations before applying PCA, while bottom panel represents microstates sliding window correlations after applying PCA to original time-series data.
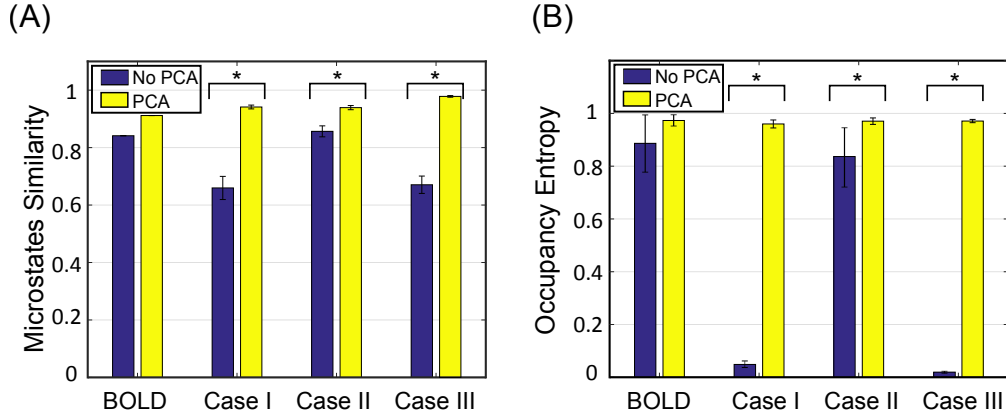


Figure 3.9: Microstates similarity (correlation between two microstates) and microstate entropy for the real BOLD, cases I-III before and after applying PCA to original time-series data. Error bars stand for standard deviation (SD).

and after applying PCA for all the cases I-III with p-value $\ll 0.001$. We further analyzed the differences in the entropy of occupancy, over subjects, before and after applying PCA to the original time-series data. The results of such statistical analyses are as follows: (K-S
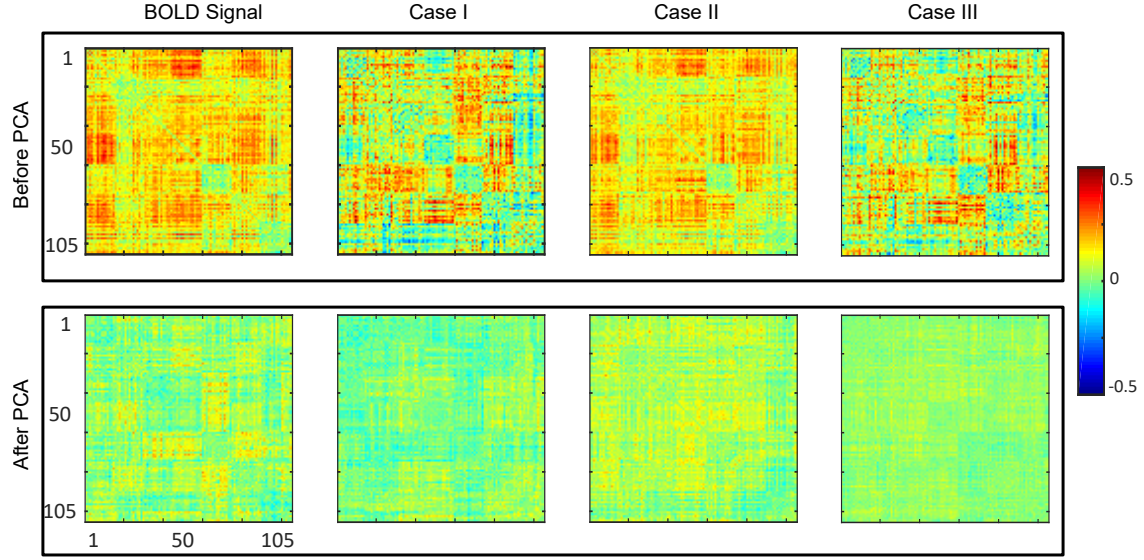
Figure 3.10: Difference between two microstates, separately, obtained from the real BOLD signal, case I-III. Top panel shows microstates difference before applying PCA, while bottom panel represents microstates difference after applying PCA to original time-series data.

test for the real BOLD, p-value $= 0.111$, stat $= 0.5$), (K-S test for the case I, p-value $\ll 0.001$, stat $= 1$), (K-S test for the case II, p-value $\ll 0.001$, stat $= 1$), (K-S test for the case III, p-value$=0.001$, stat $= 0.8$). The results of statistical tests on the occupancy, of being at a particular microstate, is reported in the Appendix B.1.

Figure 3.10 depicts the difference between two microstates, separately, obtained before and after applying PCA to time-series data from the real BOLD signal, case I-III. It is clear in this plot that before applying PCA to the original time-series, the difference between microstates for all cases is significant. Again, here we can see that applying PCA to the original time series data causes the microstate to become very similar (reduces the difference between two microstate). Figure 3.11 shows the similarity of each microstates obtained from cases I-III over 100 random realizations to the ones obtained from the real BOLD signal.
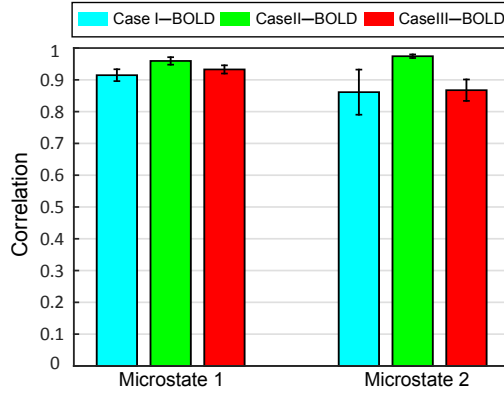
Figure 3.11: Similarity of each microstates obtained from cases I-III over 100 random realizations to the ones obtained from the real BOLD signal. Error bars stand for standard deviation (SD).

### 3.4.3 Existence of Microstates

Through the use of these positive controls, we can conclude that dynamic structure is present and can be extracted from our fMRI data, as long as dimensionality reduction is avoided on the raw time series data. While this methodological validation is itself valuable, our primary interest is in deploying the extraction of dynamics towards a better understanding of different clinical conditions, as now follows.

## 3.5 Spatiotemporal Motifs of Correlated Activity in General Anesthesia

The induction of general anesthesia incurs a dramatic change in phenotype compared to wakefulness. Substantial progress has been made at identifying the molecular targets [56]

and subcortical arousal systems [57] perturbed in this transition. Correlated electrical activity among distributed cortical regions also changes substantively, suggesting concurrent perturbation in the networks underlying cognition and behavior [58–64].

The disruption of RSNs is a candidate mechanism whereby anesthetic agents induce sedation and unconsciousness. Each RSN encompasses brain regions of presumably shared functional specialization that demonstrate marked zero time lag correlation among surrogates of neural activity. These findings evoke a complementary line of inquiring regarding which components of the functional neuroarchitecture underlying wakefulness remain preserved during the anesthetized state. In prior fMRI studies, general anesthesia appear to perturb RSNs in both animals [65] and humans [66–68].

We hypothesized that analyzing these temporal dynamics (in the sense of dynamical functional connectivity) would reveal additional anesthesia-induced disruptions within and between RSN's.

In particular, we sought to perform a clustering analysis not on the windowed covariance matrices (as in the previous subsection), but rather on the correlation time-series. That is, to group region pairs that exhibit similar temporal fluctuations in their correlation. However, given the inherent stochasticity of these time series, we first apply a filtering step to smooth the correlation trajectories (Fig. 3.1). We note that such filters have been developed for this purpose [78], however they do not constrain the observation (i.e., the cross-correlation). As a consequence, the resulting estimate may lie outside of the limits $[-1, 1]$. In the formulation presented herein, we assume that each successive correlation arises from a standard autoregressive model, transformed through a Fisher transform (a normalizing nonlinearity that constrains the random walk between $[-1, 1]$).

Following filtering, we use k-means clustering to separate the correlation trajectories into groups of region pairs that exhibit similar temporal profiles, thus yielding a number of spatiotemporal motifs that exhibit similar dynamics of correlated activity ((Fig. 3.1)C).

### 3.5.1 Dynamic Correlation Model

Here, we perform a second analysis that complements the microstate approach outlined in the previous section. Specifically, in addition to clustering the windowed matrices, we also sought to cluster the correlation time-series in order to learn whether these time-series possess systematic dynamic fluctuations. That is, if two connections have the same average weight, does those weights covary in time.

To pursue this analysis, the correlation trajectories for each region pair were calculated using a bounded Kalman filtering approach. Let $\mathbf{z}_k$, $k = 1, 2, ...$ be a vector time series of correlation values between pairs of brain regions. We assume that $\mathbf{z}_k$ arises from an underlying state $\mathbf{x}_k \in \mathbb{R}^N$ via

$$\mathbf{z}_k = \mathbf{F}^{-1}(\mathbf{x}_k), \tag{3.2}$$

where the function $\mathbf{F}^{-1}$ is vector-valued inverse Fisher transform which can be written as

$$\mathbf{F}^{-1}\left(\begin{bmatrix} x^1 \\ \vdots \\ x^N \end{bmatrix}\right) = \begin{bmatrix} \tanh\left(x^1\right) \\ \vdots \\ \tanh\left(x^N\right) \end{bmatrix}. \tag{3.3}$$

Such a function bounds the state within $[-1, 1]$ and is the standard transformation of the cross-correlation into a normally distributed variable. We assume that $\mathbf{x}_k$ obeys a standard,

linear state space model of the form

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{w}_{k-1},$$

$$\mathbf{y}_k = \mathbf{F}^{-1}(\mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k). \tag{3.4}$$

Here, the state noise process $\mathbf{w}_k \in \mathbb{R}^N$ is an $N$-dimensional random vector with multivariate Gaussian distribution having zero mean and covariance matrix $\mathbf{Q_w} \in \mathbb{R}^{N \times N}$, the noise vector $\mathbf{v}_k \in \mathbb{R}^N$ is a zero-mean multivariate Gaussian random vector with covariance matrix $\mathbf{R_v} \in \mathbb{R}^{N \times N}$. Bounded nature of correlation coefficients result in a non-Gaussian noise for the observation vector $\mathbf{y}_k$. The covariance $\mathbf{Q_w}$ determines the extent to which correlation can change in successive measurements, while $\mathbf{R_v}$ simply characterizes measurement noise. In this sense, the vector $\mathbf{y}_k$ consists of noisy observations of pairwise correlation values. We assume that state noise and measurement noise are uncorrelated.

We seek the optimal filter for obtaining the state estimate $\hat{\mathbf{z}}_k$ in the sense of minimum mean-squared error (MMSE), i.e.,

$$\min_{\hat{\mathbf{z}}_k} \mathbb{E} \left[ \|\hat{\mathbf{z}}_k - \mathbf{z}_k\|_2^2 \right]. \tag{3.5}$$

Central to this problem is the calculation of the probability density function (p.d.f.) of the state vector at any given time $k$, conditioned on $\mathcal{Y}_k = \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_k\}$, the set of all the past observations. The presence of the nonlinearity (3.3) complicates this calculation, but only slightly since $\mathbf{F}$ is smooth and invertible. Thus, it follows immediately that we can obtain a surrogate measurement

$$\mathbf{d}_k = \mathbf{F}(\mathbf{y}_k), \tag{3.6}$$

such that $\mathbf{d}_k$ is linear in the state $\mathbf{x}_k$. Note that (3.2) directly yields the p.d.f. of $\mathbf{z}_k$ from that of $\mathbf{x}_k$. Thus, the problem (3.5) reduces to the classical Kalman filter to obtain the (Gaussian) p.d.f. of $\mathbf{x}_k$ given $\mathcal{Y}_k$ [92, 93], based on the measurements $\mathbf{d}_k$.

It follows directly from (3.4)-(3.6) that the p.d.f. $p(\mathbf{y}_k|\mathbf{x}_k)$ can be written as

$$p\left(\mathbf{y}_k|\,\mathbf{x}_k\right) = \frac{1}{|\mathbf{J}_k|}p\left(\mathbf{v}_k = \mathbf{F}(\mathbf{y}_k) - \mathbf{C}_k\mathbf{x}_k\right), \tag{3.7}$$

where $p\left(\mathbf{v}_k = \mathbf{F}(\mathbf{y}_k) - \mathbf{C}_k\mathbf{x}_k\right)$ is p.d.f. of $\mathbf{v}_k$, a zero-mean multivariate Gaussian with covariance matrix $\mathbf{R}_{\mathbf{v}}$. A standard Bayesian approach thus yields the posterior density $p\left(\mathbf{x}_k|\mathcal{Y}_k\right)$, which is Gaussian with covariance matrix $(\boldsymbol{\Sigma})$ and mean $(\mu)$ as

$$\begin{aligned}
\boldsymbol{\Sigma} &= \left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{C}_k{}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{C}_k\right)^{-1}, \\
\mu &= \boldsymbol{\Sigma} \times \left(\mathbf{P}_{k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1} + \mathbf{C}_k{}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{F}(\mathbf{y}_k)\right).
\end{aligned} \tag{3.8}$$

The Kalman update equations are thus

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= \mathbf{A}_k\hat{\mathbf{x}}_{k-1|k-1}, \\
\mathbf{P}_{k|k-1} &= \mathbf{A}_k\mathbf{P}_{k-1|k-1}\mathbf{A}_k^T + \mathbf{Q}_{\mathbf{w}}, \\
\hat{\mathbf{x}}_{k|k} &= \left(\mathbf{P}_{k|k-1}{}^{-1} + \mathbf{C}_k{}^T\mathbf{R}_{\mathbf{v}}{}^{-1}\mathbf{C}_k\right)^{-1} \\
&\quad \times \left(\mathbf{P}_{k|k-1}{}^{-1}\hat{\mathbf{x}}_{k|k-1} + \mathbf{C}_k{}^T\mathbf{R}_{\mathbf{v}}{}^{-1}\mathbf{F}(\mathbf{y}_k)\right), \\
\mathbf{P}_{k|k} &= \left(\mathbf{P}_{k|k-1}{}^{-1} + \mathbf{C}_k{}^T\mathbf{R}_{\mathbf{v}}{}^{-1}\mathbf{C}_k\right)^{-1}.
\end{aligned} \tag{3.9}$$

In the subsequent results, we assume an initial multivariate Gaussian prior $\mathbf{x}_{0|-1}$ with mean $\hat{\mathbf{x}}_{0|-1}$ and covariance of $\mathbf{P}_{0|-1}$. We note that, for the problem (3.5), (3.9) returns the optimal estimate. After obtaining state estimate $\hat{\mathbf{x}}_{k|k}$, the correlation coefficients $\hat{\mathbf{z}}_k$ are approximated

using fourth order Taylor expansion of (3.2) around $\hat{\mathbf{x}}_{k|k}$ as

$$
\begin{aligned}
\hat{\mathbf{z}}_k = {} & \tanh(\hat{\mathbf{x}}_{k|k}) - \tanh(\hat{\mathbf{x}}_{k|k}) * \operatorname{sech}^2(\hat{\mathbf{x}}_{k|k}) * \operatorname{diag}(\mathbf{P}_{k|k}) \\
& + \frac{1}{24}\left(22\tanh(\hat{\mathbf{x}}_{k|k}) * \operatorname{sech}^4(\hat{\mathbf{x}}_{k|k}) - 2\sinh(3\hat{\mathbf{x}}_{k|k}) * \operatorname{sech}^5(\hat{\mathbf{x}}_{k|k})\right) * 3\operatorname{diag}(\mathbf{P})_{k|k}^2,
\end{aligned}
\tag{3.10}
$$

where all the multiplication and powers of vectors are element-wise operations. Also, diag() operator applies to a square matrix and returns a vector containing the diagonal elements of the matrix. Performance of the developed filter is evaluated on synthesized data in a numerical simulation in Appendix B.4.

A final obstacle in using the filter equations (3.9) is knowing the noise parameters that specify the state-space model. Since such information is difficult to define a priori, a principled way to update them alongside the state estimates is required. Here, we face the challenge of estimating noise parameters alongside our online filtering. To address this issue we use an EM algorithm [49, 50] to jointly estimate the state and model parameters from observed correlation data. Assuming diagonal covariance matrices for model parameters, we can decouple the update procedures for each edge, resulting in a reduction of the computational burden typically associated with EM methods. Briefly speaking, after initializing the model parameter (process and measurement noise), we perform the E-step, filtering and smoothing, and then updating the parameter in the M-step. The E and M steps alternate to convergence. In order to obviate computational issues associated with this algorithm, a maximum iteration number is set alongside the convergence criterion.

**Defining Motifs:** The filtered correlation time series were concatenated across subjects and both arousal states. Using a k-means algorithm, we defined 8 orthogonal clusters with region pair membership defined by the strength of covarying correlation trajectories. We

66

selected 8 clusters based on analysis of model fit assessed by Akaike Information Criterion (AIC) [90], Appendix B.2. More detail on this approach is provided elsewhere [91]. A schematic of spatiotemporal decomposition of dynamic functional connectivity is shown in Figure 3.1. Moreover, simulations on synthesized data demonstrate efficacy and utility of our approach for differentiating different patterns of dynamic fluctuations among spatially distributed clusters of brain regions with identical mean average correlation, Appendix B.3.

## 3.6 Spatiotemporal Functional Connectivity Motifs Persist During Wakefulness and Sevoflurane Anesthesia

### 3.6.1 Widespread Conservation of Functional Connectivity Structure under Sevoflurane Anesthesia

We first used the analysis framework to study how similar, on average, is the correlation structure during wakefulness compared to that during sevoflurane general anesthesia. This analysis complemented an earlier analysis [68] that suggested key alterations within the DMN and VAN. Using both a simple correlation (Figure 3.12) for quantifying similarity across average correlation matrices (0% and 1.2%), we observed a high degree of conservation of functional connectivity within RSNs regardless of arousal state. Matrices showing the average correlation for the two conditions are similar (Figure 3.12A). At a region level, the standard deviation across volunteers is 0.13 and mainly 0.15. Figure 3.12B shows the standard error of mean (SEM) of average correlation (0% and 1.2%). We assessed the similarity between the two conditions by calculating the correlation of the mean correlation values. The overall similarity of all correlations among brains regions at 0% and 1.2% sevoflurane was 0.76.
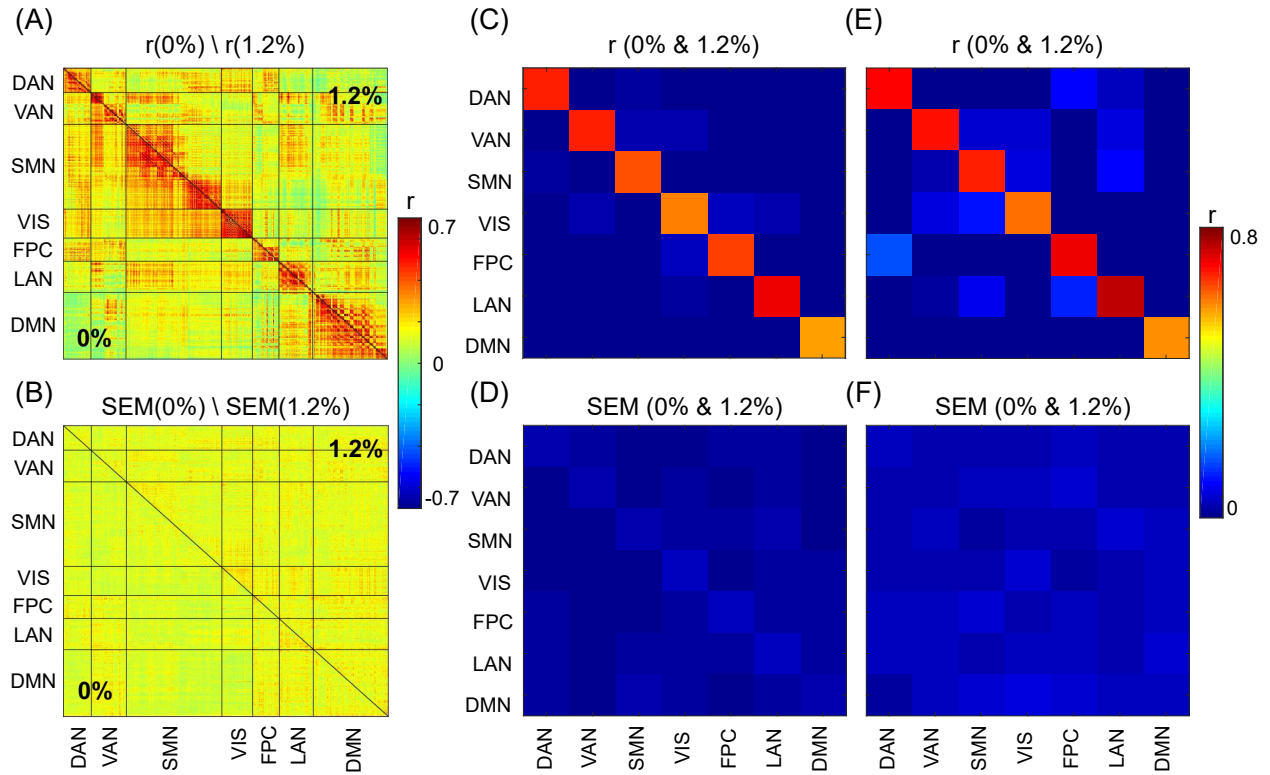
Figure 3.12: (A) Average correlation matrices (0% and 1.2%). (B) Standard error of mean (SEM) of average correlation matrices. (C) Similarity across average correlation matrices (0% and 1.2%) for entire data. (D) SEM of Similarity across average correlation matrices for entire data. (E) Similarity across average correlation matrices (0% and 1.2%) for subsampled data. (F) SEM of Similarity across average correlation matrices for subsampled data.

Thus, the correlated activity during wakefulness accounts for roughly 58% of that during sevoflurane general anesthesia. The greatest similarity was within RSNs, as summarized in Figure 3.12C, with low variance across our participants (Figure 3.12D). To allow feasibility for subsequent spatiotemporal analysis we reduced dimensionality using random subsampling of all regions, 15 per RSN. Again, similarity in the average correlation structure at 0% and 1.2% sevoflurane was conserved (Figure 3.12E), mainly for within RSN comparisons, with variance across subjects.
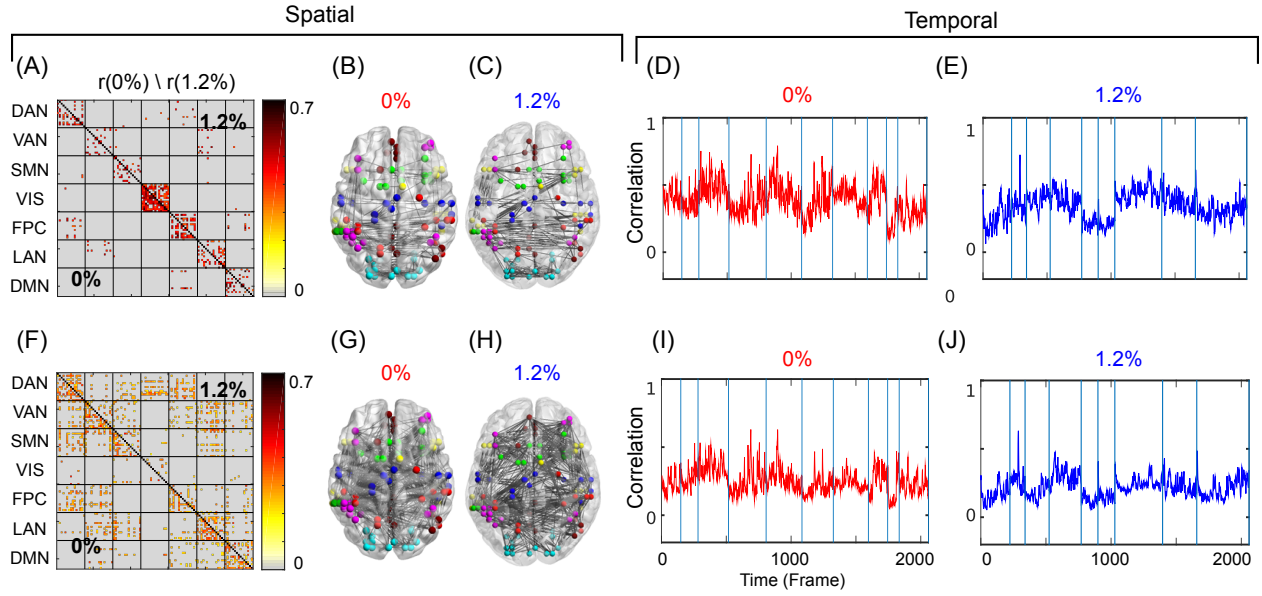
Figure 3.13: (A) Second spatial motif for 0% and 1.2% . (B) Second spatial motifs on brain surface with nodes and edges for 0%. (C) Second spatial motifs on brain surface with nodes and edges for 1.2%. Second temporal motif for (D) 0% and (E) 1.2%. (F) Third spatial motif for 0% and 1.2% . (G) Third spatial motifs on brain surface with nodes and edges for 0%. (H) Third spatial motifs on brain surface with nodes and edges for 1.2%. Second temporal motif for (I) 0% and (J) 1.2%.

## 3.6.2 Correlated Brain Activity During Sevoflurane General Anesthesia Can Be Decomposed into Dynamic Spatiotemporal Motifs

Next, we determined whether the temporal trajectory of correlation among pairs of brain regions respect RSN boundaries. We separately decomposed the correlations among fMRI BOLD signals acquired at 0% and 1.2% sevoflurane into eight clusters. Each cluster represents correlations between paired brain regions that follow a similar pattern over time. The resulting spatiotemporal motifs were ranked by average correlation among members. Figure 3.13 shows second and third spatial (Figure 3.13A,F) and temporal (Figure 3.13D,F,I,J) motifs of the decomposition.

69
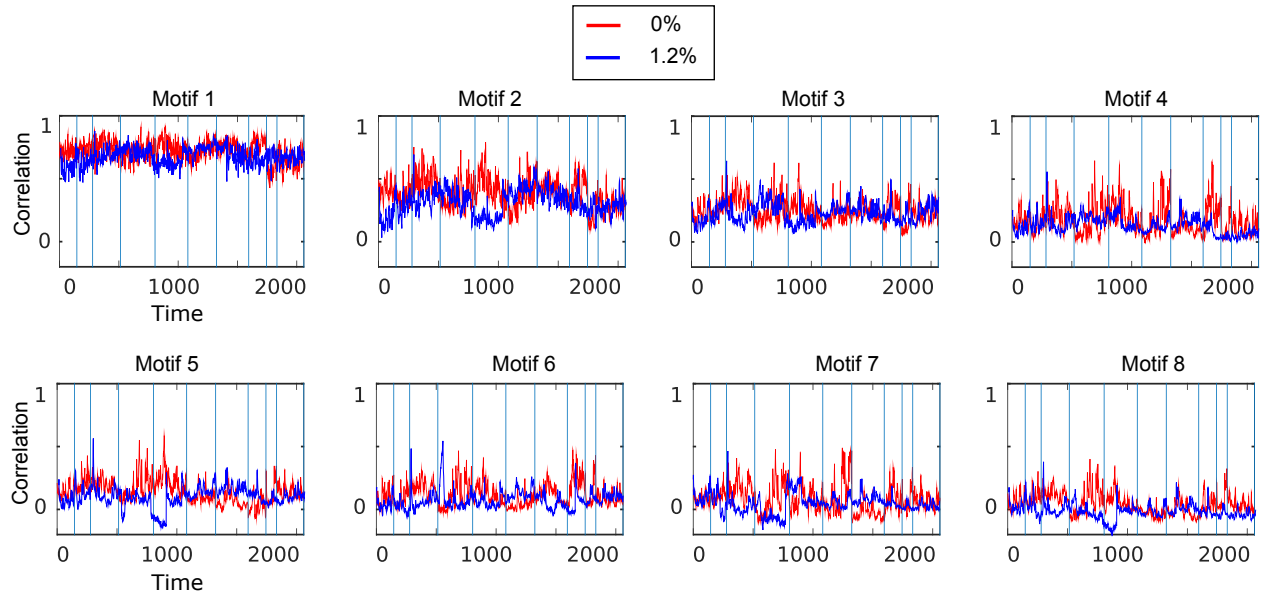
Figure 3.14: Temporal motifs for 0% and 1.2% sevoflurane
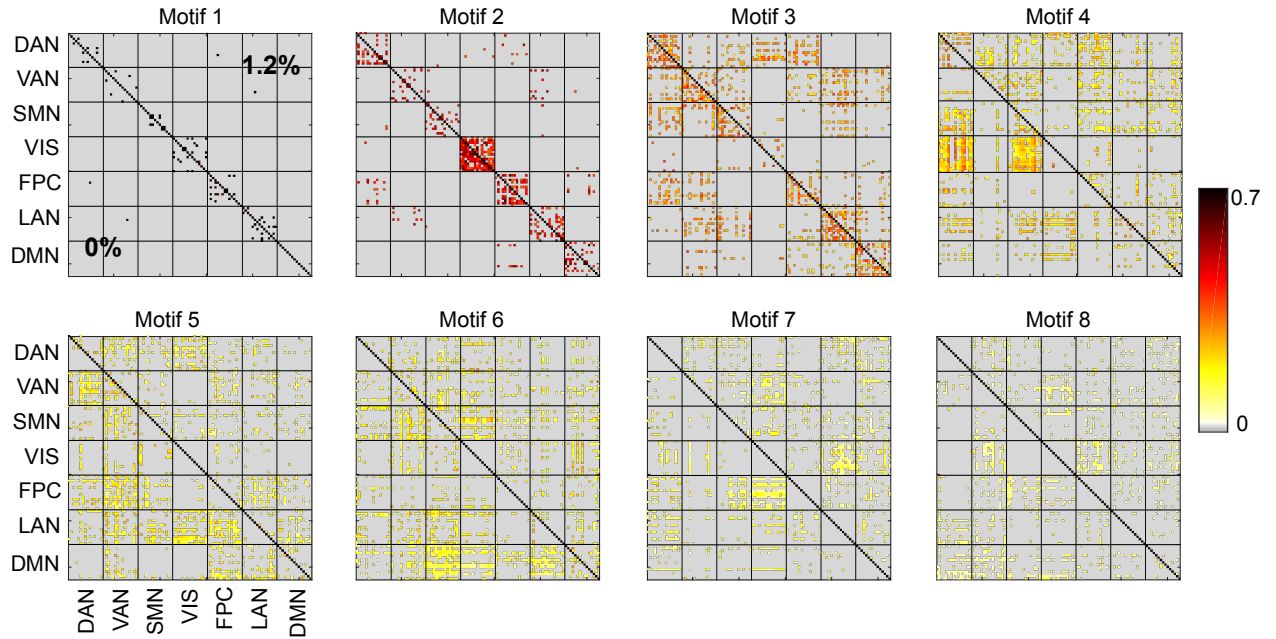


Figure 3.15: Spatial motifs for 0% and 1.2% sevoflurane

We carried out spatiotemporal decomposition of the correlated brain activity measured at 0% and 1.2% sevoflurane. These motifs were determined independently as we could not
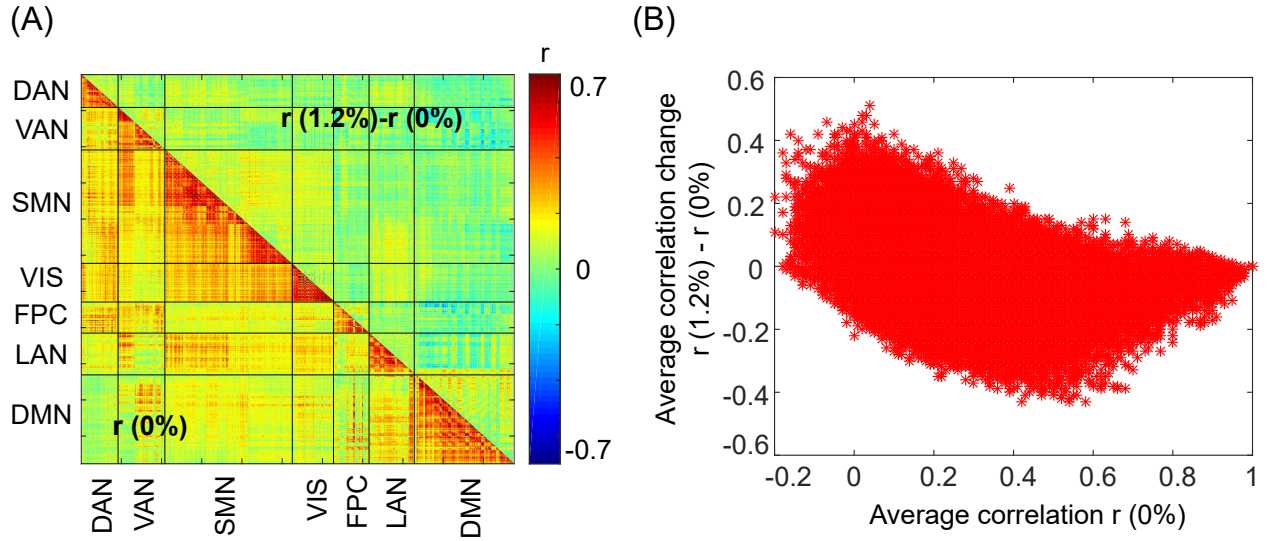
Figure 3.16: change in correlation (0%-1.2%) vs mean correlation at 0%.

assume that the motifs would be identical. Figure 3.14 shows all temporal motifs. All spatial motifs are shown in Figure 3.15. Three types of motifs were observed. Motifs 1 and 2 contained mainly within-RSN pairings with high mean correlation strength and appear relatively conserved following the transition from wakefulness to anesthetic-induced unconsciousness. Other motifs, 3-4, were composed of region pairs with intermediate correlation strength between interactions/connections within and between RSNs. Their membership and correlation strength change between 0% and 1.2% sevoflurane. Clustering also generated motifs mainly in between-RSN region pairs of low or intermediate correlation (motifs 5-7) and a motif with negligible correlation (motif 8).

We quantified the similarity between motifs during wakefulness and 1.2% sevoflurane, compared the number of edges that were in each motif among the two conditions, and found the distributions to have no statistical difference (Chi-square test, dF = 49, p = 0.23). Figure 3.16 demonstrates changes in correlation (0%-1.2%) versus mean correlation at 0%).

71

The spatiotemporal motifs obtained from this approach are explained in the following sections.

**Spatiotemporal Motifs that Persist Between Wakefulness and General Anesthesia:** As the correlations of signals within a show the greatest correlation (Figure 3.12A), one possibility is that spatiotemporal decomposition would recapitulate RSNs. Instead, our clustering generated motifs composed of brain regions spanning different RSNs but appeared to be relatively conserved between quiet wakefulness and general anesthesia, as for the motif 2 (Figure 3.13 A). This motif is composed mainly of within-RSN pairs, suggesting that changes in correlation strength covary across RSNs. The mean correlation strength does not appear to change between wakefulness and general anesthesia. The temporal profile of average correlation over time shows marginally higher correlation and variability at 0% compared to 1.2% (0%: mean 0.39, SD 0.11; 1.2%: mean 0.34, SD 0.1, Figure 3.13D-E). We use the F-test to show that the variability in variance between two conditions is statistically significant (F-test, dF1 = 2059, dF2 = 1539, p $\ll$ 0.001).

In contrast, motif 3 has varying membership of intra-RSN and inter-RSN pairs (Figure 3.13F). As with motif 2, the temporal profile of correlations across members shows greater strength and variability at 0% sevoflurane (0%: mean 0.25, SD 8.8 x 10-2; 1.2%: mean 0.21, SD 7.7 x 10-2, F-test, dF1 = 2059, dF2 = 1539, p $\ll$ 0.001; Figure 3.13I-J). As with motifs 2 and 3 (Figure 3.13D-E and 3.13I-J), the correlation strength and variance in all temporal motifs is higher at 0% (Figure 3.14).

Figure 3.17B shows the similarity between motifs, for 8 different window sizes, by computing the correlation of the spatial component of each motif. Motifs 1, 2, 3, and 8 had the greatest similarity between wakefulness and general anesthesia. In other words, the region pairs
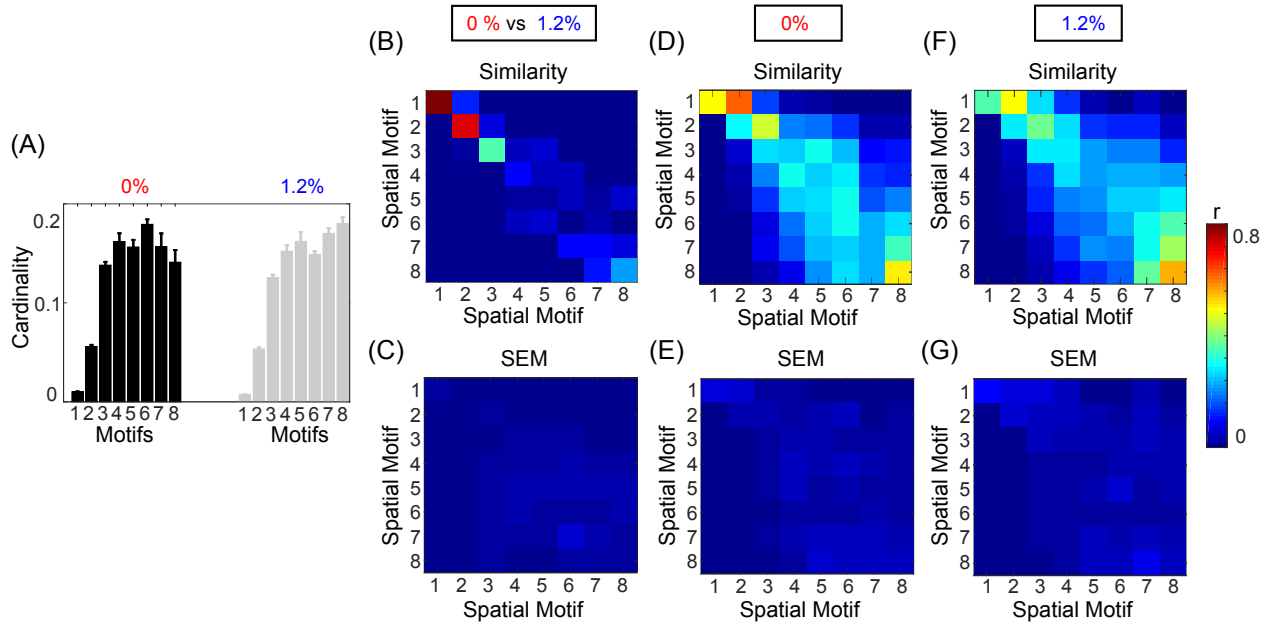
Figure 3.17: Analysis of spatial motifs. (A) Bar plot of the cardinality percentage of each motif for both condition (0% and 1.2%). Error bars stand for SEM. Average similarity (B) and SEM of average similarity (C) between motifs, for 8 different window size, by computing the correlation of spatial component of each motif. (C) Standard error of similarity between motifs. Average similarity (D) and SEM of average similarity (E) between motifs calculated from each individual at 0% sevoflurane. Average similarity (F) and SEM of average similarity (G) between motifs calculated from each individual at 1.2% sevoflurane.

with the strongest average correlation changed the least during the transition to anesthetic-induced unconsciousness.

**Spatiotemporal Motifs that Vary Between Wakefulness and General Anesthesia:**
Figure 3.17B demonstrates that there are motifs which are not similar during the transition to anesthetic-induced unconsciousness. Motifs 4, 5, 6, and 7 show very low similarity from wakefulness to anesthetic condition. An interesting result from these motifs is the observation that spatiotemporal dynamics differ between 0% and 1.2% sevoflurane mainly in interactions between RSNs.

It can be seen in motif 4 that connections/interactions between DAN and VIS networks represent similar dynamic covariation during 0% sevoflurane, which groups them in the same motif, while such a similarity is broken in 1.2% sevoflurane. This reorganization of interactions between RSNs can be seen between VIS and SMN networks in motif 4 with intermediate strength of average correlation as well as DMN and SMN in motif 6 with intermediate average correlation.

### 3.6.3   Robustness as a Function of Parameters and Intersubject Variability

We also assessed the consistency of motifs as a function across individual subjects (Figure 3.17D-G). Figure 3.17A demonstrates the average cardinality of each motif over different window length. Error bars stand for SEM of the average cardinality. Figure 3.17D represents average similarity between motifs calculated separately from each individual during wakefulness. This figure illustrates that the motifs are robust across different subjects. In other words, high value of similarity are located mostly on diagonal elements of the similarity matrix; however, because of process noise and subject variability in some motifs, medium similarity can be seen in the off diagonal elements of the similarity matrix. Standard error of the mean of similarity in this calculation is shown in Figure 3.17E. In a similar way, Figure 3.17F demonstrates average similarity between motifs calculated separately from each individual during anesthetic-induced unconsciousness. Average similarity between motifs and the deviation from the mean for a fixed window size, 11 seconds, over different individuals, without concatenation, is provided in Appendix B.5.

## 3.7 Discussion

### 3.7.1 Non-stationarity and Existence of Microstates During Resting Wakefulness

In the first part of this chapter, we evaluated dFC microstate analysis and showed that dimensionality reduction (specifically, PCA) on the original data can change the microstate result significantly. This can be understood in terms of the de-correlating effect of PCA, which may promote similarity in the post-PCA windowed correlation matrices. Understanding this methodological issue, we proceeded to show the presence of microstates within real fMRI data. In particular, we highlighted how the occupancy of the extracted microstates provides a crucial quantification that can substantiate their dynamic nature.

### 3.7.2 Spatiotemporal FC Motifs During Wakefulness and Sevoflurane Anesthesia

We quantified the extent of conservation of functional connectivity, defined by the correlation of intrinsic activity, within the cerebral cortex between wakefulness and sevoflurane general anesthesia. This was addressed in multiple ways. First, we assessed the similarity between the functional connectivity by computing correlations of the entire BOLD time-series (i.e., average correlation) for each regime. Second, we estimated how the correlation between region-pairs evolves as a function of time, thus enabling us to disambiguate region-pairs with similar average correlations into groups with different correlation dynamics. In particular, we decomposed the correlation at each moment in time into a finite set of spatiotemporal motifs

whose superposition recovers the average functional connectivity. Our results indicate that the transition from 0% to 1.2% sevoflurane is associated with altered interaction between RSNs.

**Dynamic Functional Connectivity:** Our clustering approach generates similar motifs for wakefulness compared to sevoflurane general anesthesia for region pairs within RSNs (motifs 1-3, 19% region pairs). Other motifs (4-6, 49% region pairs) with intermediate average correlation strength exhibit dissimilar membership between the baseline and anesthetic conditions. Finally, the large bulk of interactions between regions of different RSNs differ between motifs and have low mean correlation strength (motifs 7-8, 32% of region-pairs).

Our findings suggests, perhaps intuitively, that intermediate correlations between RSNs are the most sensitive to sevoflurane anesthesia. Indeed, we observe that the composition of motifs with intermediate strength of correlation is largely altered, while motifs with higher strength of average correlation are maintained at both 0% and 1.2% sevoflurane.

**Altered Correlation Dynamics in Sevoflurane-induced Sedation:** The above observations suggest a mechanism for sevoflurane-induced unconsciousness involving dis-coordination between RSNs. Specifically, whereas within-RSN interactions appear quite robust, the motifs associated with cross-RSN interactions are fundamentally altered in their composition. Thus, those regions with co-evolving correlation in awake conditions no longer exhibit the same coordination in unconsciousness. While not a causal mechanism for unconsciousness, this observation is nonetheless consistent with the integrated information theory of (loss of) consciousness [95], while also allowing for the possibility of persistent coordinated electrical activity over wide swatches of cortex during general anesthesia [96].

**Limitations:** We acknowledge limitations in our work that may inform future investigations. To identify spatiotemporal patterns, we used the k-means clustering technique. Though k-means clustering is an efficient and robust partitioning algorithm, it has several limitations, in particular difficulty separating clusters of different sizes and densities as well as a high susceptibility to outliers [97]. A possible source of variation in this approach relates to the selection of the number of clusters, k. As in other investigations, we did not have an a priori value of k and instead parametrically determined the number of groups with the best fit (Appendix B.2). An experimental limitation of this study is the number of subjects scanned under both conditions. Though we were able to scan each subject for an extended period, increasing the number of subjects will improve the robustness of the findings.

# Chapter 4

# Characterizing Temporal Dynamics of Depressed Level of Consciousness with Applications in Detecting Coma Severity and Coma Cause

## 4.1  Highlights

Determining the causes and severity of a depressed level of consciousness (DLOC) including coma is a common clinical challenge. Quantitative analysis of electroencephalogram (EEG) has the potential to improve DLOC assessment by providing readily deployable, temporally detailed characterization of the brain activity of such patients. While used commonly for seizure detection, EEG-based assessment of DLOC etiology is not well-established. We sought to determine whether salient information regarding coma severity and coma subtypes can be extracted from the EEG via time-series analyses and classification techniques.

First, we retrospectively analyzed EEG recordings from 62 patients with DLOC with consensus focal or diffuse culprit pathology. For each recording, we performed a suite of time-series and Fourier-based spectral analyses, then used a statistical framework to identify which analyses (features) could be used to distinguish between focal and diffuse cases. We identified several spectral and non-spectral EEG features that were robustly and significantly different between DLOC patients with focal vs. diffuse etiologies. A support vector machine (SVM) classifier was trained on these features to classify cases on the basis of EEG, demonstrating accuracy up to 71% correct classification. Our findings suggest that DLOC due to focal vs. diffuse injuries differ along several electrophysiological parameters. These early results may form the basis of future classification strategies for DLOC and coma.

Second, we develop a new approach termed lagged microstate analysis (LMA) to characterize temporal dynamics of brain activity in patients with DLOC. Each microstate in our definition incorporates both temporal and spatial information. Furthermore, we show how the occupancy of being at a particular microstate, extracted from this spatiotemporal analysis, can be utilized to improve automatic classification of different levels of coma severity.[4]

## 4.2   Introduction

A depressed level of consciousness (DLOC) is a near universal result of acute severe brain injury, and disorders of consciousness are among the most feared long-term sequelae of such injuries. Coma, a state of complete loss of spontaneous or stimulus-induced arousal, is the most severe form, but all forms of DLOC have substantial impacts on patient outcomes

---

[4]This chapter is a slightly amended version of the previously submitted article which is listed as follows: M. Kafashan, S. Ryu, D. Roberts, A. Thonthakudi, L. Eisenman, T. Kummer, and S. Ching, "Electroencephalographic Features for Classification of Depressed Level of Consciousness: Focal versus Diffuse Etiologies," under review in *Brain Informatics*, 2016.

[98,99]. A DLOC can result from diffuse brain injuries, or from focal insults that secondarily induce widespread alterations in cerebral function [100]. Formulating an accurate differential is crucial to the clinical management of patients with DLOC, as diagnoses drive the approach to treatment and prognosis [99–101]. Diagnostic formulation often begins with distinguishing between focal and diffuse etiologies.

In some cases a careful history, paired with a basic laboratory workup and screening neuroimaging tests are all that are required to determine the cause of coma or a related DLOC in a given patient. Often, however, these standard assessments prove inadequate to distinguish a focal from a diffuse etiology acutely. There are several common scenarios in which such ambiguity exists: A patient may have a DLOC that exceeds expectations from modest structural brain injury evident on imaging; or a patient's DLOC may result from a focal process that, due to its nature, acuity, location, or size, is not apparent on screening imaging studies. Seizures may depress consciousness and can accompany either focal or diffuse pathology. Focal and diffuse injury can also coexist, both impacting wakefulness.

More specialized testing, such as expanded laboratory assessments, specialized neuroimaging studies, and invasive procedures, are available and may help to establish a diagnosis in such cases. These tests, however, carry risks and are therefore rarely utilized in the absence of suggestive results from initial assessments. Similarly, highly specialized interventions including specific medications (e.g. specific therapy for DLOC resulting from toxin ingestion) and even surgical procedures (e.g. cerebral ventricular decompression for DLOC due to elevated intracranial pressure) are effective in some cases, but are rarely used empirically.

A non-invasive, bedside screening test that can help classify DLOC acutely, therefore, could be of significant utility in guiding both advanced diagnostic strategies and ultimately management approaches [102]. A strong suggestion of a focal etiology in the absence of initial

neuroimaging findings might prompt more advanced neuroimaging studies. Similarly, a strong suggestion of a diffuse etiology, even in the presence of distracting structural brain lesions, might prompt a more extensive toxic-metabolic work-up, or a more aggressive correction of known toxic-metabolic abnormalities.

All DLOC, and in particular coma, are characterized by pathological alterations in brain electrical activity [103]. These electrical alterations may provide valuable etiological insight. Consistent with this, in addition to its well-established role in seizure detection, the EEG has been shown to have utility in the monitoring of other large-scale alterations of neurological function. Examples include EEG monitoring of delirium [104], burst suppression [105], and cerebral ischemia [106]. Thus, EEG can non-invasively provide highly temporally-resolved data at the bedside on both structural brain injury and on non-structural injury that may not be apparent on screening neuroimaging studies.

Visual inspection of raw EEG data requires advanced training and cannot easily capture the full complexity of electrical dynamics that are potentially encoded in the EEG signal. In contrast, quantitative EEG methods use computer-assisted analysis of EEG patterns to derive quantitative metrics that are not immediately apparent upon review of raw EEG data. The use of quantitative EEG analysis in the clinical setting has seen significant recent growth, particularly in the domains of sleep [107], epilepsy [108], and general anesthesia [109]. In these scenarios, progress has been made towards translational applications including seizure detection [110], classification of sleep stages [111], and quantification of depth of anesthesia [112].

In the context of brain injury and DLOC, the deployment of EEG-based classification techniques is more limited, although its potential as a diagnostic and scientific research tool is repeatedly acknowledged. A common application of quantitative EEG to DLOC involves

classical spectral analysis, which decomposes a given signal into constituent frequencies, typically aggregated into the standard EEG 'bands' (i.e., delta, alpha, etc.) [113]. In this context, severe brain injuries are classically associated with concentration of EEG power into low frequencies ($<$1Hz). However, while approaches based on spectral analysis are commonplace in EEG studies, this form of analysis only captures sinusoidal harmonic structure in the underlying signal. Other forms of spatiotemporal time series analysis are available that may complement and augment spectral methods, and have been applied to EEG data from limited cohorts of patients with DLOC. These analyses, which include measures of signal entropy and complexity, characterize aspects of the EEG time series that are not directly captured by spectral analyses.

Significant effort has been directed at the EEG-based analysis of chronic DLOC in rehabilitative settings, such as minimally conscious and persistent vegetative states [114, 115], wherein a large number of EEG analyses have screened for their potential to disambiguate these subtypes. While these studies have yielded insights into the mechanisms of these conditions, the analyses rely on high-density research grade EEG/MEG instrumentation [114] that is generally unavailable in the acute setting. It furthermore remains unclear whether insights gleaned from these studies will prove informative in the acute setting, though a recent study did identify several acute electrophysiological correlates of outcome in coma [116], demonstrating the prognostic potential of EEG-based approaches.

## 4.3   Methods

### 4.3.1   Data Acquisition and Pre-Processing

We retrospectively collected EEG data, EEG reports, and complete medical records from 62 patients who underwent EEG for routine monitoring purposes related to a diagnosis of coma or less-severe DLOC at the time of EEG, at the Neurological and Neurosurgical Intensive Care Unit at Barnes-Jewish Hospital and Washington University School of Medicine (St. Louis, MO, USA). There were 62 patients considered and 70 total EEG studies (6 patients underwent EEG monitoring twice and one patient underwent EEG monitoring three times). In all cases, EEG was performed for the detection of non-convulsive seizures in patients with otherwise inadequately explained DLOC. Only cases in which seizures were not detected at any point in the hospitalization were analyzed. For each of the 62 patients, two neurointensivists (TTK and DER) examined all diagnostic data available including imaging to assign a focal or diffuse classification. Importantly, these assessments benefited from diagnostic data not available to the team at the time of the initial EEG. Thus in most cases we were able to determine DLOC etiology to a reasonable degree of clinical certainty despite the diagnostic ambiguity that resulted in EEG testing early on. Imaging data was given the greatest weight in etiological determinations. Evidence of herniation or direct injury to brainstem reticular activating structures resulted in assignment to a focal etiology. Less severe structural lesions were interpreted in the context of historical data and coexisting toxic-metabolic influences to determine the etiology of the DLOC.

Cases were included in the analysis of detecting coma subtypes only when the ultimate etiology (focal vs. diffuse) was apparent from clinical data. In cases of disagreement, the case was re-reviewed and discussed until a consensus was reached (most such cases classified

as indeterminate). In total 43 subjects (22 focal and 21 diffuse) were used in the analysis of detecting coma cause (24 focal and 23 diffuse, with total 47 EEG studies analyzed). Figure 4.1A summarizes the 43 subjects that were analyzed for classification of coma subtypes. To examine the utility of traditional clinical EEG metrics, the clinical EEG reports were separately scrutinized to identify reported features that could assist in the classification of cases as focal or diffuse in etiology. Specifically, any focal or lateralized abnormalities in the report were flagged as supportive of a focal as opposed to diffuse etiology.

Glasgow coma scale (GCS), which is a neurological scale that aims to give a reliable and objective way of recording the conscious state of a person, is reported for all of the cases. Hence, all the cases were included in the analysis for detecting the level of coma severity. We consider GCS = 7 to be the threshold for binarization of coma severity to low and high resulting in 42 cases with high GCS and 28 cases with low GCS. Figure 4.1B summarizes the 62 subjects that were analyzed for classification of different levels of coma. GCS level of subjects with more than one EEG study is also shown in this plot.

## 4.3.2   EEG Signal Processing and Statistical Analysis

**Dataset:** Recordings are collected at a sampling frequency of either 250 or 500 Hz using the standard 10-20 system of electrode placement. The 500Hz data were downsampled to 250Hz prior to analysis. In our analyses, we used a bipolar montage with 18 bipolar channels (FP1-F7, F7-T7, T7-P7, P7-O1, Fp1-F3, F3-C3, C3-P3, P3-O1, Fz-Cz, Cz-Pz, Fp2-F4, F4-C4, C4-P4, P4-O2, Fp2-F8, F8-T8, T8-P8, and P8-O2). Records were visually analyzed for quality control, and sections of the record containing large-amplitude artifacts were excluded from analysis. Each bipolar channel was normalized to zero-mean, unit variance. Then,
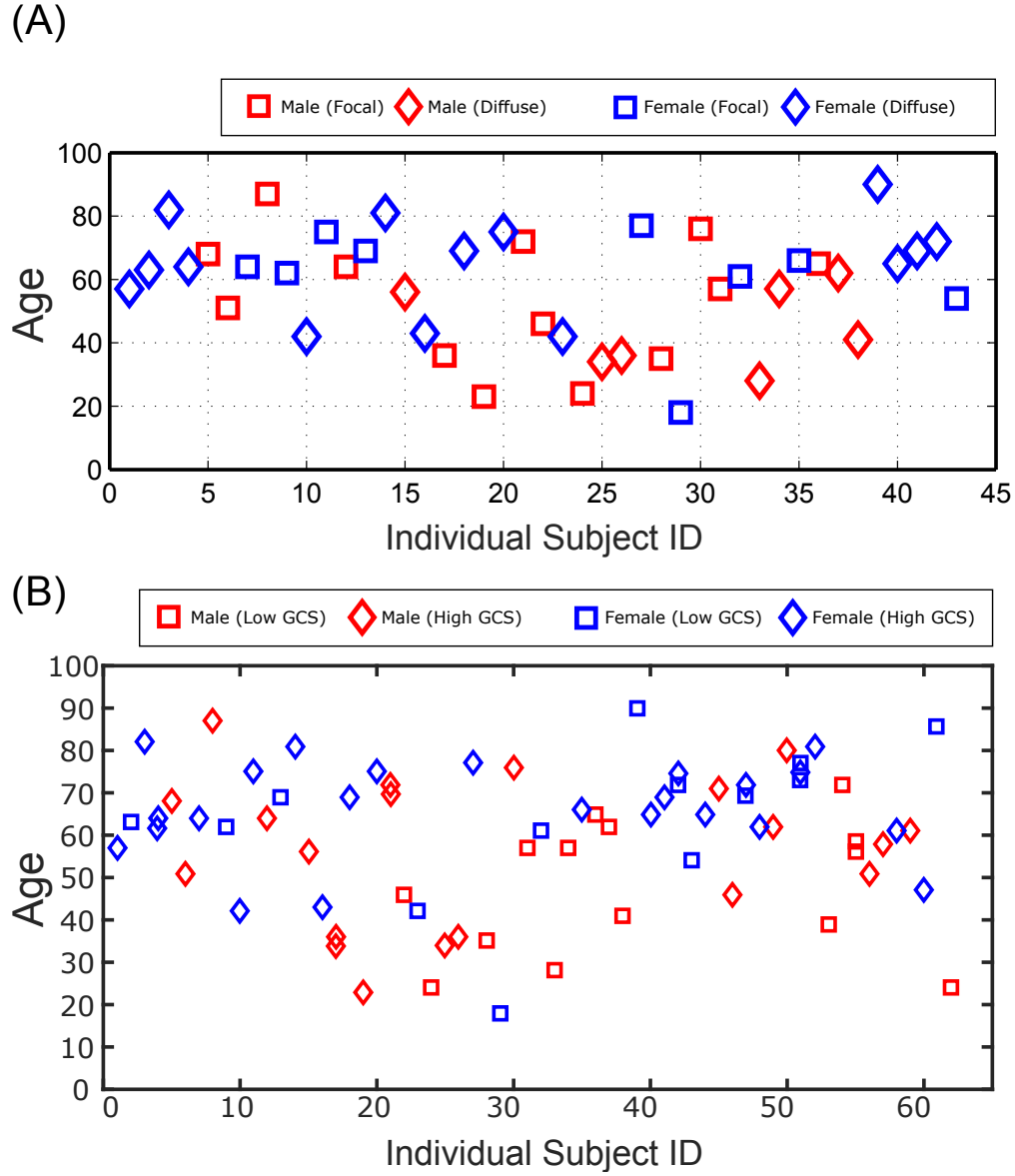
Figure 4.1: (A) Summary of Study Population for classification of different coma etiologies. (B) Summary of Study Population for spatiotemporal analysis of different levels of coma.

these channels were filtered using a 10th order Chebyshev Type I lowpass filter with cutoff frequency of 50Hz before any further analysis.

**Definition of Trials:** We divided each patient's EEG recording into separate trials for the purposes of analysis. Specifically, trials were defined on the basis of non-overlapping

sliding windows of the bipolar montaged EEG data (Figure 4.2). Dividing the EEG data into subintervals results in different predictions for each trial, thus yielding an empirical probability of focal/diffuse for each subject. We allocated the trials of a subject to either the training or testing set, ensuring that training and testing sets are independent.

**Feature Extraction:** We obtained 25 features listed in Table 4.1 for discrimination of coma severity and cause. These different features related to the dynamical (including spectral) properties of time-series data. For each subject, we extracted one 25-dimensional feature vector for each trial by averaging feature vectors over different epochs of that trial (Figure 2). Secondary statistics (e.g., higher order moments) of the features were not considered in this analysis. All signal processing and feature extractions were performed in MATLAB (Natick, MA), and feature selection and evaluation of classifiers were computed in R (version 3.1.2).

Table 4.1: List of 25 features extracted from EEG data.

| Feature ID | Description |
| --- | --- |
| 1–2 | Maximum, minimum eigenvalues of the estimated $\mathbf{A}$ matrix from MVAR fitting with $p = 1$ (4.1) |
| 3 | Number of absolute eigenvalues of matrix $\mathbf{A}$ larger than Threshold=0.95 |
| 4–6 | Statistical properties: variance, skewness, and kurtosis |
| 7–11 | Power in the delta, theta, alpha, beta, and gamma bands |
| 12 | Ratio of power in beta and gamma bands to total power |
| 13 | Ratio of power in delta and theta bands to total power |
| 14 | Hurst exponent [117] |
| 15 | Hjorth parameters [117] |
| 16–19 | Eequidistant mutual information, quiprobable mutual information, and the first minimums of both types of mutual information [117] |
| 20 | Bicorrelation |
| 21 | Median frequency [117] |
| 22–24 | Spearman autocorrelation, Pearson autocorrelation, and partial autocorrelation |
| 25 | Composite permutation entropy index (CPEI) [118] |
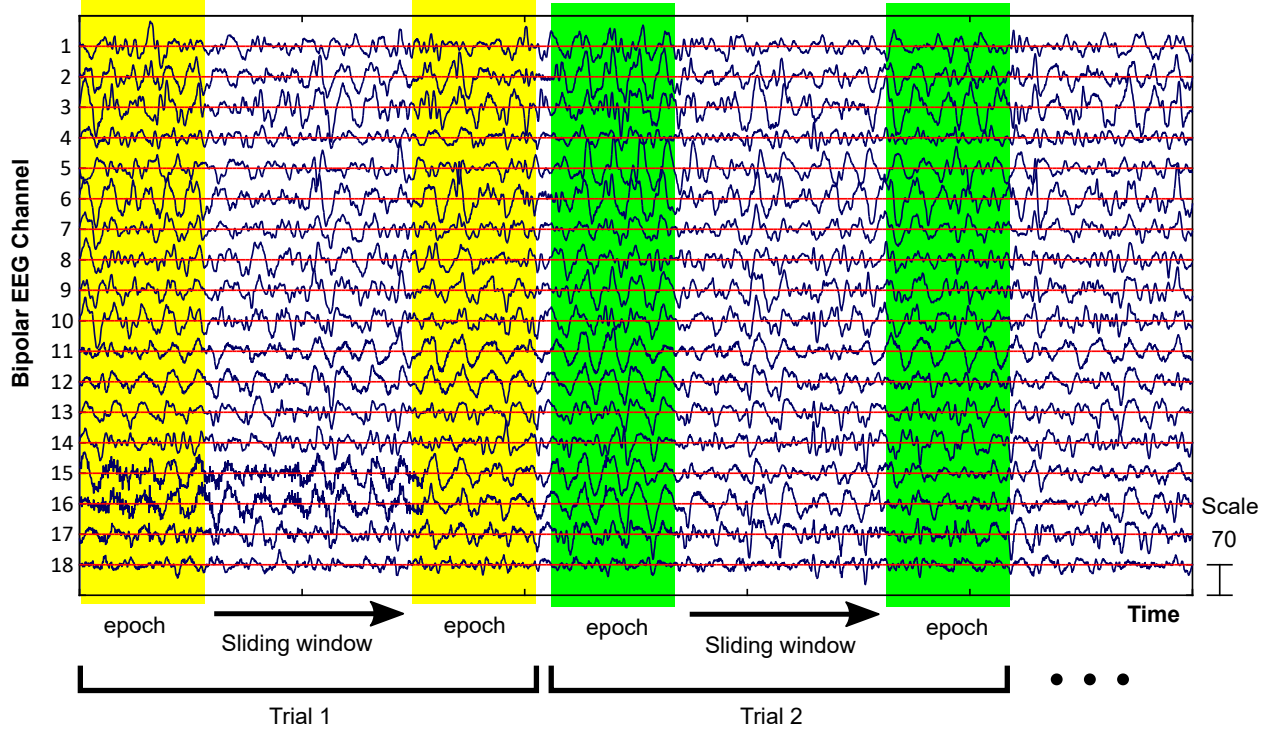
Figure 4.2: Schematic illustrating sliding window to define epoch and trial in EGG data with bipolar montage. The EEG channel number on the vertical axes are ordered as: FP1–F7, F7–T7, T7–P7, P7–O1, Fp1–F3, F3–C3, C3–P3, P3–O1, Fz–Cz, Cz–Pz, Fp2–F4, F4–C4, C4–P4, P4–O2, Fp2–F8, F8–T8, T8–P8, P8–O2.

**Training and Testing Separation:** We trained a binary classifier from our features using independent training and testing sets. Two testing paradigms were considered. In the first paradigm, we partitioned the data into two groups: two-third of total subjects were selected randomly and defined as the training set and one-third of subjects were defined as the testing set. This process was repeated within a cross-validation paradigm, in order to evaluate average classification performance. In the second paradigm, we withheld one-third of the data as a dedicated test set, and trained strictly on the remaining two-third of the patients (i.e., with no re-sampling and averaging).

**SVM Classifier with Embedded Selections of Best Predictors:** We specified a support vector machine (SVM) to function as a binary classifier to discern a patient's coma severity (i.e. low or high GCS) and DLOC etiology (i.e. focal or diffuse). The SVM approach uses a portion of data as support vectors to create a decision boundary (i.e, threshold) [119]. For the classification performance reported in this study, subjects in the testing set are different from subjects in training test.

To train the classifier, we applied principal component analysis to the primary feature vectors over trials, and only the first 20 important principal components (PCs) of feature vectors were kept. The first 20 principal component explains more than 98% of the variance in the original feature vectors. We used a data-driven approach to rank the features/PCs by importance by evaluating a family of vector quantization (LVQ) models. This procedure was completed via the train.R routine in the Caret toolbox [120], implemented in the R programming language. A 10-fold cross validation is used within this feature selection step which is resampled 50 times. We then selected a set of predictive PCs based on their importance, to train an SVM with a linear kernel according to different etiologies agreed upon by the neurologist. The use of a linear kernel aids in reducing overfitting.

**Multivariate Autoregressive Model of EEG Data:** Features 1-3 in Table 4.1 use a multivariate autoregressive model, wherein the (multivariate) EEG signal is modeled as a linear sum of previous samples. For a multivariate N-channel process $\mathbf{x}(t) = [\mathbf{x}_1(t), \mathbf{x}_2(t), \cdots, \mathbf{x}_N(t)]^T$, a MVAR model of order $p$ takes the following form:

$$\mathbf{x}(t) = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{x}(t - k) + \mathbf{w}(t), \tag{4.1}$$

where $\mathbf{w}(t) \in \mathbb{R}^N$ are additive noise vectors (innovation process) and $\mathbf{A}_k \in \mathbb{R}^{N \times N}$ are the MVAR model coefficients. Here, we used a standard Least-Squares approach to implement the model fit [121].

**Statistical Evaluation:** We used a two-sample t-test to compare feature distributions. In this pilot study, our goal was to generate hypotheses regarding which of the screened features were informative with respect to the considered coma etiologies. Since the PCs of the primary features are uncorrelated, we compared etiology distributions of each PC independently to a nominal significance level of $p = 0.05$.

### 4.3.3 Lagged Microstate Analysis (LMA) of Depressed Level of Consciousness

We utilized the lagged cross-correlation of EEG data to characterize the temporal dynamics of brain connectivity in coma. We sought to find out how coupling between different regions changes over time, and define microstates which can explain these fluctuations. To identify microstates in coma, we concatenated all data together, and then applied $k$-means clustering to extract salient microstates which are consistent over all subjects. To capture the time delays that may exist between EEG time series, we incorporated the lagged cross-correlation, as well as zero-lag Pearson correlation, to amount for connectivity structure over different time windows. Correlation values between all region-pairs over 4 different lags were included, resulting in $4 \times (18 \times 17)/2 - 18 = 666$ features per sliding window. The diagonal elements of zero-lag correlation were not included in the feature vector. To reduce the computational burden, we applied a dimensionality reduction to reduce the number of features to 30 before clustering analysis. Note that the dimensionality reduction is performed on the correlation

trajectories and not the raw EEG data. The clustering analysis was performed with random initialization of centroids and was repeated 100 times to increase chances of escaping local minimum. Finally microstates are defined by averaging the lagged cross-correlation matrices corresponding to clusters obtained from microstate analysis over sliding windows.

## 4.4 EEG Features for Classification of Depressed Level of Consciousness: Focal versus Diffuse Etiologies

### 4.4.1 Several Time-Series Metrics, including Non-Spectral Features, Discriminate Focal from Diffuse DLOC

We first screened and ranked the importance of the primary features, i.e., without applying PCA (Figure 4.3A). As illustrated in Figure 4.3B, these primary features contain substantial redundancy, particularly in the entropic features (i.e., features 15-20 in Table 4.1). Thus, we transformed the primary features into their uncorrelated principal components, then ranked these PCs (Figure 4.3C-D).

Figure 4.4 shows the distributions, portrayed as notched box plots, of the 6 most informative PCs as identified by our iterative selection method. Recall that a PC is informative if its distribution over subjects is different between focal and diffuse DLOC, so as to enable SVM classification. The means for each feature distribution in Figure 4.4 are significantly different ($p \leq 0.05$ in all cases, t-test, uncorrected, see also methods) between the focal and diffuse cases. It is important to note that the difference in these PC distributions between diffuse and focal DLOC is not contingent on the feature selection paradigm. In other words, these
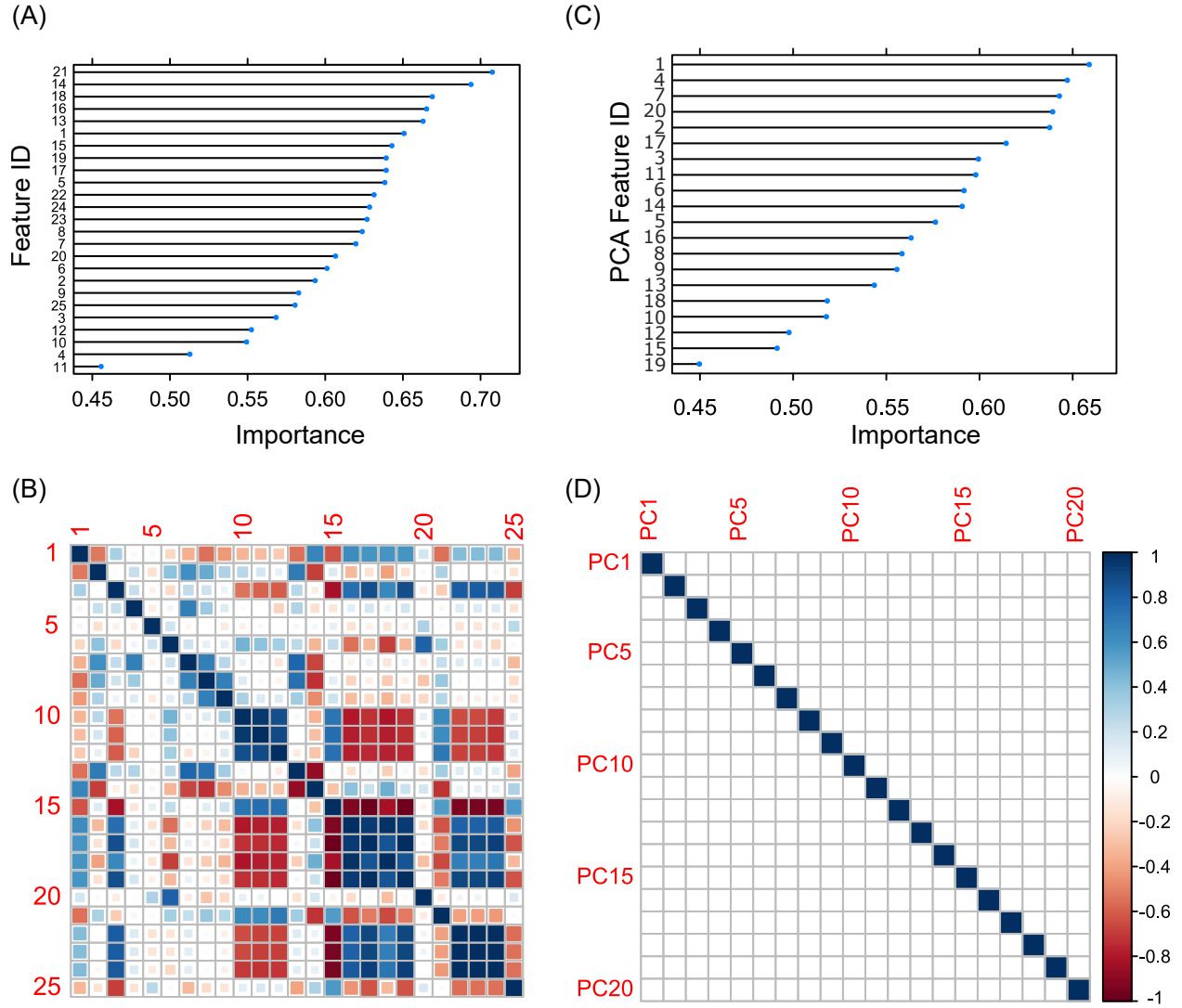
Figure 4.3: Feature ranking by importance (A) before and (C) after applying PCA to original feature vector. Correlation among feature (B) before and (D) after applying PCA.

differences are properties of the data, absent any automated selection algorithm. Here, we used epoch length of 5 seconds with 40 epochs in a trial. We repeated our analysis using different feature epoch lengths (1s, 10s, 20s, data not shown), without any qualitative change to the overall results.
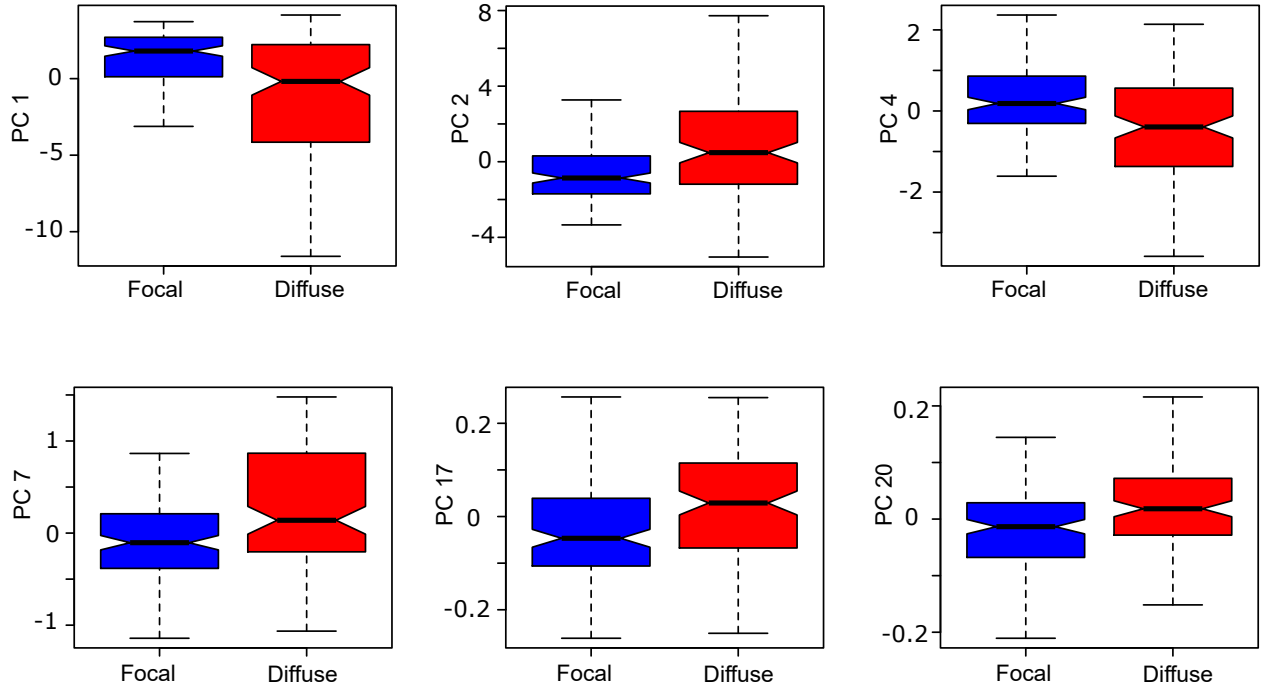
Figure 4.4: Notched box plots for the distributions of 6 most informative PCA predictors.

Figure 4.5A-C illustrates the PC loading, i.e., the linear contribution of the primary features on the six most informative PCs portrayed in a pairwise fashion. Figure 4.5D summarizes the loading across the first 20 PCs. Of note is that a combination of spectral and non-spectral time series analysis provides discriminative (focal/diffuse) power in this dataset. For instance, PC1 (most informative) is comprised largely from entropic features (15-24 in Table 4.1), whereas PC4 (second most informative) has a large contribution from spectral features (10-12 in Table 4.1). This observation indicates that these features are providing complementary information regarding DLOC etiology.
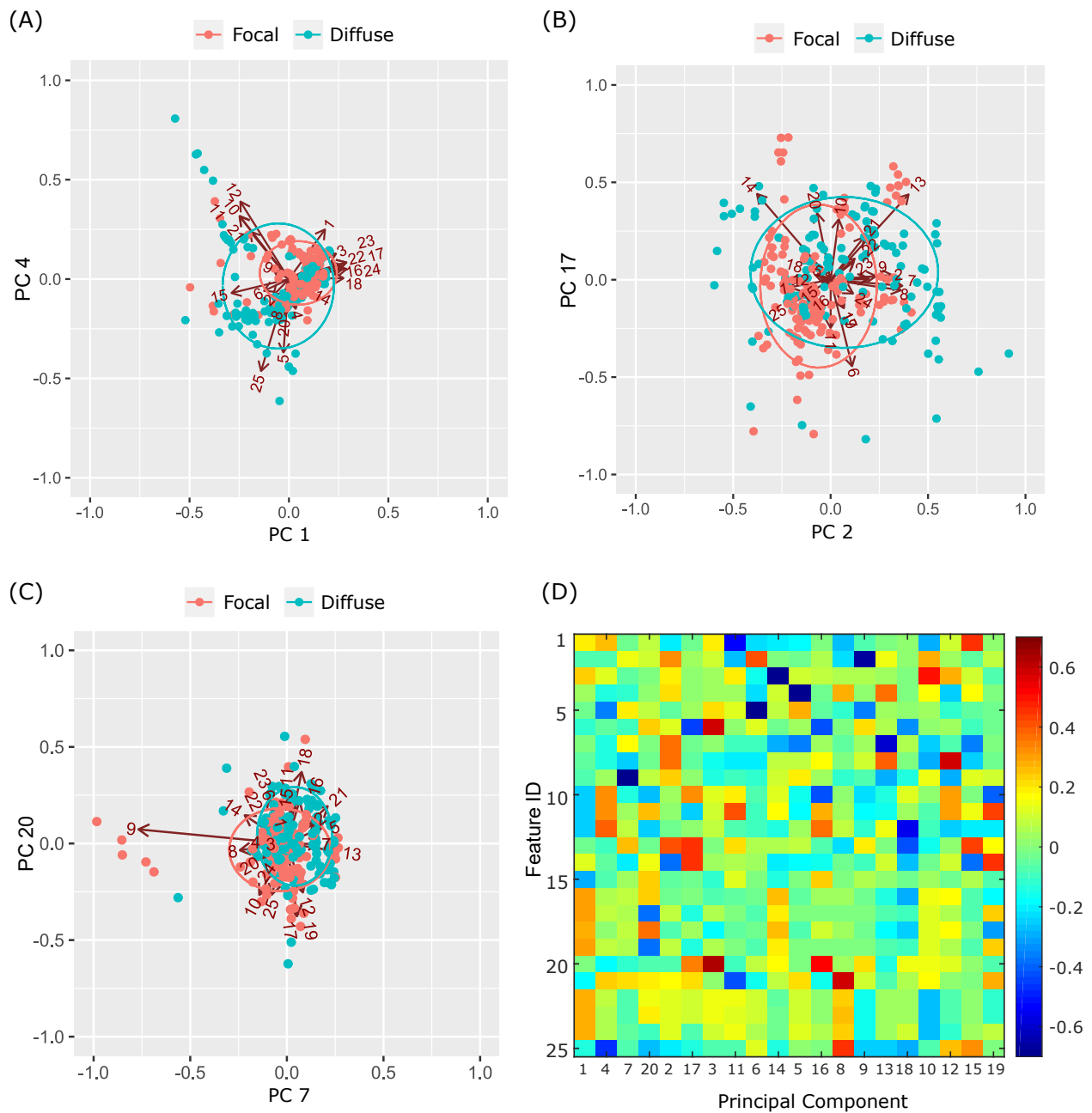
Figure 4.5: Principal components loadings for focal and diffuse cases. (A)-(C) Contribution of each initial feature on different principal component. Red and blue ellipses show the direction of variation in the component scores of focal and diffuse, respectively. (D) Loading matrix for 20 principal components. Order of principal components are based on their importance.

Table 4.2: Averaged Classification Performance before Applying PCA on Initial Features. Classification results are for different epoch length (1s, 5s, and 20s) and different number of epochs in a trial (10, 20, 40, 100, and all epochs). All the results are averaged classifier performance over 500 random training and testing sets. In each realization, the subjects in training and testing sets are different.

| epoch length | Number of epochs in a trial | Accuracy (Hard) | Accuracy (Soft) | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1s | 10 | 0.56 | 0.57 | 0.42 | 0.7 |
| | 20 | 0.57 | 0.58 | 0.45 | 0.7 |
| | 40 | 0.57 | 0.58 | 0.5 | 0.68 |
| | 100 | 0.58 | 0.59 | 0.48 | 0.69 |
| | All | 0.57 | 0.57 | 0.47 | 0.69 |
| 5s | 10 | 0.56 | 0.59 | 0.41 | 0.73 |
| | 20 | 0.56 | 0.59 | 0.42 | 0.72 |
| | 40 | 0.57 | 0.57 | 0.4 | 0.71 |
| | 100 | 0.58 | 0.56 | 0.37 | 0.71 |
| | All | 0.57 | 0.6 | 0.41 | 0.78 |
| 20s | 10 | 0.54 | 0.55 | 0.3 | 0.74 |
| | 20 | 0.58 | 0.55 | 0.3 | 0.74 |
| | 40 | 0.55 | 0.57 | 0.4 | 0.73 |
| | All | 0.58 | 0.58 | 0.39 | 0.77 |

## 4.4.2 Focal and Diffuse DLOC are Classified with Good Accuracy using a Limited Number of PCA Features

We designed a classifier for focal versus diffuse DLOC on the basis of first the primary features and then the PCs, as outlined in the Methods. Table 4.2 shows the classifier performance using the primary features only (i.e., without applying PCA) for different epoch lengths and number of epochs in a trial. In this analysis, the classifier is trained based on the 10 most informative features on two third of subjects and then tested on the remaining patients (14 subjects). We performed training and testing for 500 repetitions (each repetition, subjects in the training set are totally different from subjects in testing set). The values reported in this table are averaged over repetitions. In this Table, hard accuracy refers to deciding etiology

94

Table 4.3: Averaged Classification Performance after Applying PCA on Initial Features. Classification results are for different epoch length (1s, 5s, and 20s) and different number of epochs in a trial (10, 20, 40, 100, and all epochs). All the results are averaged classifier performance over 500 random training and testing sets. In each realization, the subjects in training and testing sets are different.

| epoch length | Number of epochs in a trial | Accuracy (Hard) | Accuracy (Soft) | Specificity | Sensitivity |
|---|---|---|---|---|---|
| | 10 | 0.65 | 0.65 | 0.58 | 0.73 |
| | 20 | 0.64 | 0.64 | 0.56 | 0.73 |
| 1s | 40 | 0.61 | 0.62 | 0.6 | 0.68 |
| | 100 | 0.65 | 0.66 | 0.62 | 0.71 |
| | All | 0.6 | 0.6 | 0.52 | 0.69 |
| | 10 | 0.62 | 0.62 | 0.53 | 0.72 |
| | 20 | 0.63 | 0.63 | 0.54 | 0.75 |
| 5s | 40 | 0.67 | 0.66 | 0.6 | 0.74 |
| | 100 | 0.69 | 0.64 | 0.54 | 0.76 |
| | All | 0.64 | 0.64 | 0.55 | 0.73 |
| | 10 | 0.61 | 0.62 | 0.51 | 0.74 |
| | 20 | 0.62 | 0.58 | 0.6 | 0.52 |
| 20s | 40 | 0.65 | 0.66 | 0.6 | 0.72 |
| | All | 0.64 | 0.64 | 0.62 | 0.68 |

on the basis of the majority of trials, while soft accuracy characterized performance over trials. The classifier exhibits higher sensitivity (true positives; here 'positive' is defined as a focal etiology) compare to specificity (true negatives, i.e., diffuse trials that were classified as diffuse). In this analysis, false positives (misclassification of diffuse as focal etiology) constitute the predominant source of inaccuracy.

Classification performance after applying PCA to initial feature is reported in Table 4.3. Only the first six PCs are used for classification. As shown in the Table, working the PCA space results in substantial improvement to the overall classifier design. Specifically, it can be seen in this table that the false positive rate is improved on average by 0.16, and average performance approaches 70% for many parameterizations.

Table 4.4: Classification Performance after Applying PCA on Initial Features with first 14 Subjects as Testing set. Classification results are for different epoch length (1s, 5s, and 20s) and different number of epochs in a trial (10, 20, 40, 100, and all epochs).

| epoch length | Number of epochs in a trial | Accuracy (Hard) | Accuracy (Soft) | Specificity | Sensitivity |
|---|---|---|---|---|---|
| | 10 | 0.64 | 0.65 | 0.61 | 0.68 |
| | 20 | 0.64 | 0.67 | 0.65 | 0.69 |
| 1s | 40 | 0.71 | 0.69 | 0.63 | 0.73 |
| | 100 | 0.71 | 0.69 | 0.56 | 0.79 |
| | All | 0.64 | 0.64 | 0.5 | 0.75 |
| | 10 | 0.64 | 0.68 | 0.59 | 0.75 |
| | 20 | 0.64 | 0.64 | 0.57 | 0.71 |
| 5s | 40 | 0.71 | 0.72 | 0.69 | 0.79 |
| | 100 | 0.71 | 0.68 | 0.55 | 0.77 |
| | All | 0.64 | 0.64 | 0.55 | 0.73 |
| | 10 | 0.57 | 0.59 | 0.51 | 0.64 |
| | 20 | 0.64 | 0.61 | 0.51 | 0.69 |
| 20s | 40 | 0.71 | 0.65 | 0.55 | 0.73 |
| | All | 0.64 | 0.64 | 0.67 | 0.62 |

Finally, we performed a final test of the classifier by withholding 14 subjects (1/3 of data) as a dedicated testing set, and trained the classifier on only the remaining cases. That is, we trained a classifier on 2/3 of the data, then evaluated it on a separate, withheld patient cohort. Table 4.4 reports the classification performance in this scenario, where again performance approaches 71% with high sensitivity, but only moderate specificity.

### 4.4.3   Clinical EEG Interpretation

Lastly, we examined the EEG reports of our study population to determine whether a similar classification performance could be achieved from a more straightforward clinical decision process. Specifically, we cataloged observations of focal or lateralized abnormalities in the EEG reports (slowing or isolated epileptiform discharges); an indication of a spatially local

Table 4.5: Clinical EEG Interpretation Offers Poor Discrimination of Focal Etiology. Of twenty one recordings from patients with consensus focal disease, only seven were interpreted as focal/lateralized in the clinical EEG report.

| Patient ID | Focal/Lateralized Slowing |
|:---:|:---:|
| 5 | Y |
| 6 | N |
| 7 | N |
| 8 | N |
| 9 | N |
| 12 | N |
| 17 | Y |
| 19 | N |
| 20 | N |
| 21 | N |
| 22 | N |
| 24 | N |
| 27 | N |
| 28 | N |
| 29 | N |
| 30 | N |
| 31 | N |
| 32 | N |
| 35 | N |
| 36 | N |
| 43 | N |

electrophysiological phenomenon correlating with focal injury. As shown in Table 4.5, of the 21 patients with focal injury in our study, a report of focal/lateralized slowing was only present in seven instances (i.e., corresponding to a sensitivity of $7/21 = 0.33$), demonstrating that the clinical EEG report was a poor indicator of focal etiology. Generalized slowing was observed in all study patients.

## 4.5 Lagged Microstate Analysis of Depressed Level of Consciousness: Signature of Coma Severity

Figure 4.6 shows the microstates (over four lags) and Figure 4.7 illustrates the connectivity pattern projected onto a scalp map. In this illustration, all connections with correlation value less than a threshold of 0.05 are set to zero (see Appendix C.2, Figure C.2 for the head map plot with 0.1 as the threshold). The occupancy of each microstate for low and high GCS is shown in Figure 4.8. The number of microstates is determined on the basis of the ratio of within and between cluster distance (see Appendix C.1). We consider a window length of 20 seconds.

The microstates are sorted based on their average correlation strength. It can be seen from this plot that State 5 is more prevalent in low GCS while State 1 is more specific to high GCS. State 5 shows high anterior-posterior anti-correlations over lags. In contrast, State 1 represents stronger scalp-wide correlation is relatively persistent over 240 milliseconds (all lags). This entire analysis was repeated for different numbers of microstates in the Appendix C.2.

To further interpret these states, we queried there specificity to particular neuroactive medications. Figure 4.9 illiterates the microstate occupancy with and without using neuroactive medication for patients with DLOC. Here, the considered medications are propofol, midazolam, and fentanyl, which are commonly used in this patient population. From the figure, we observe no specificity of the microstates of this class of drugs.

We next obtained the average lagged cross-correlation for a separate cohort of 108 healthy subjects, obtained from an open, online database [122] of resting state EEG. We selected the
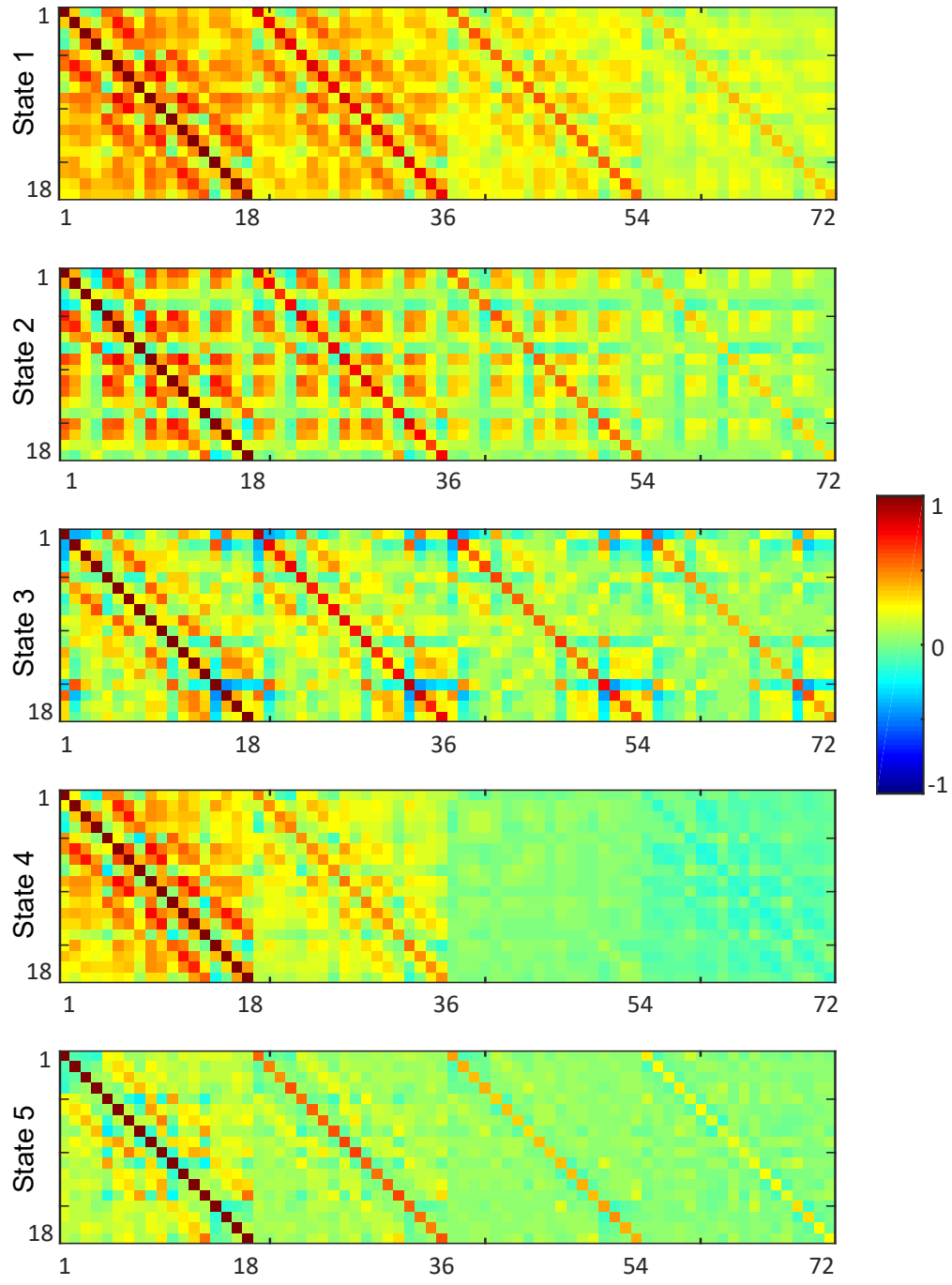
Figure 4.6: Spatiotemporal microstates extracted from subjects with different coma severity and coma subtypes based on lagged cross-correlation of EEG data.
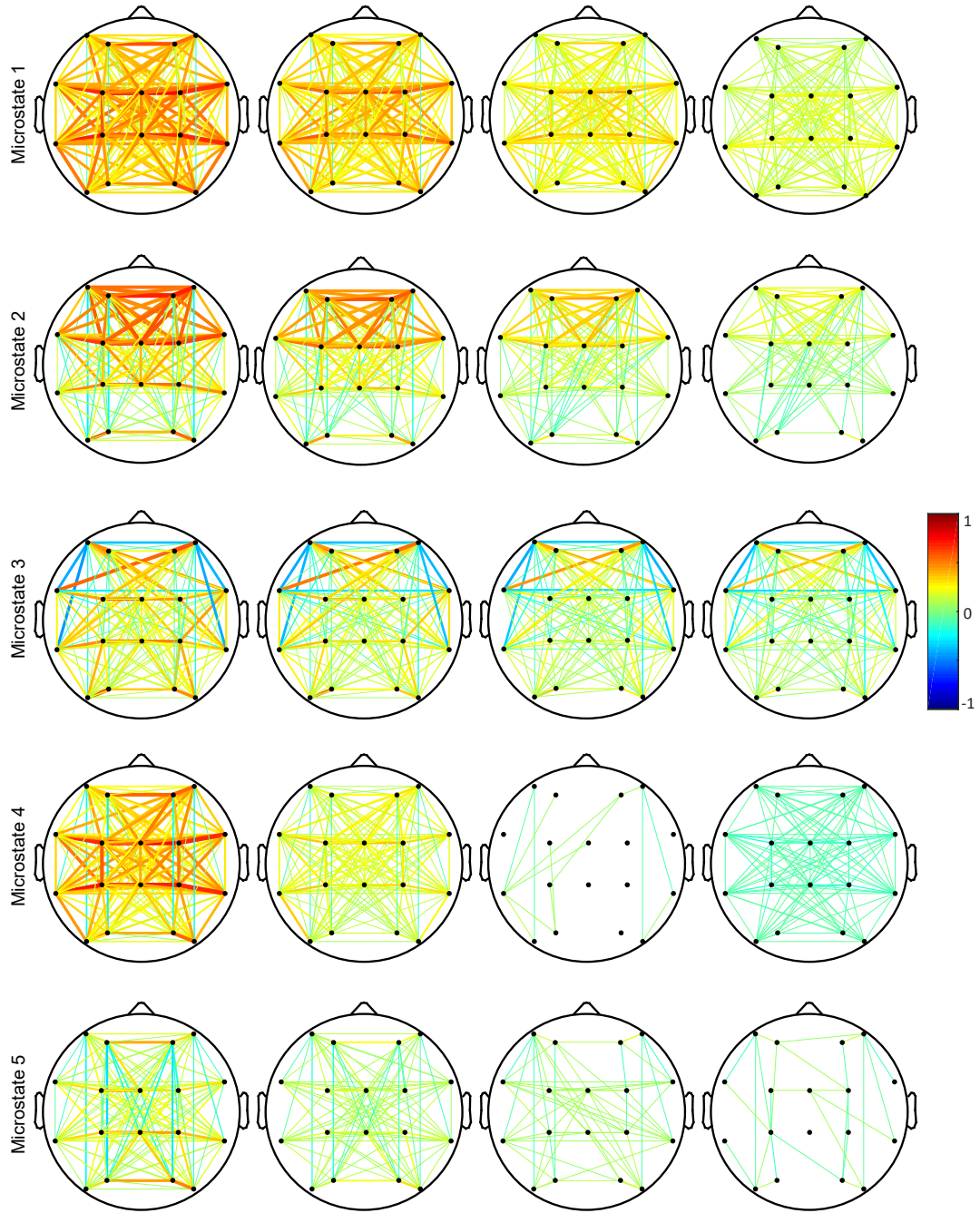
Figure 4.7: Connectivity Map of all microstates with k=5. First, second, third and fourth columns are associated to zero, first, second, and third lag connectivity. All connections with correlation value less than 0.05 are set to zero.
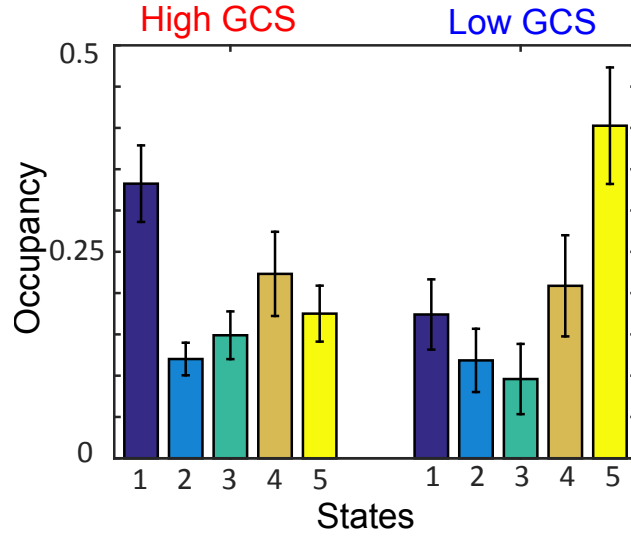
Figure 4.8: Occupancy of being at a particular microstate averaged over the subjects using lagged cross-correlation to find microstates for low and high GCS. Error bar shows the standard error of the mean.
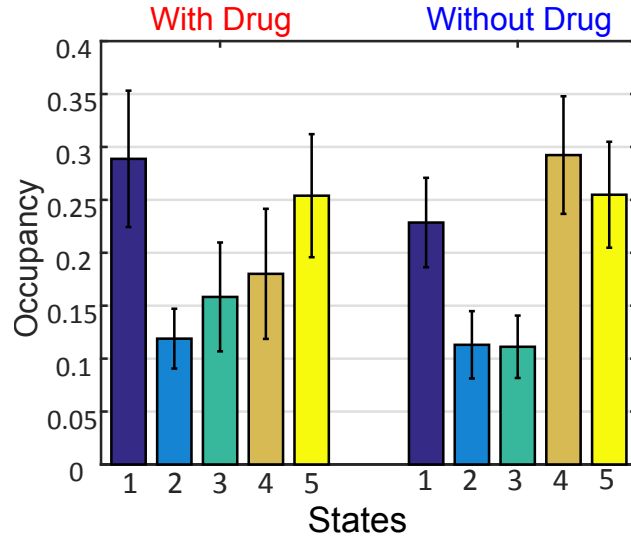


Figure 4.9: Occupancy of being at a particular microstate averaged over the subjects using lagged cross-correlation to find microstates with and without drugs (propofol, midazolam, and fentanyl). Error bar shows the standard error of the mean.

same bipolar channels as used in our clinical population. Figure 4.10 portrays the average lagged cross-correlation, where connections with correlation value less than 0.05 are set to
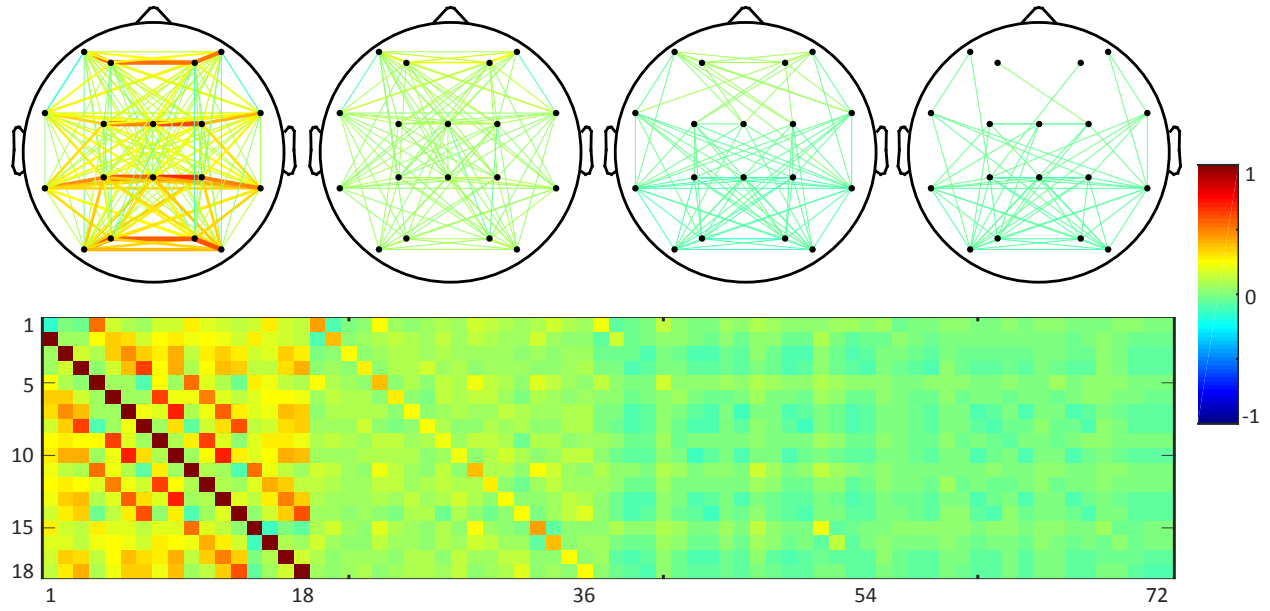
Figure 4.10: Average lagged cross-correlation for healthy subjects with four different lags. All connections with correlation value less than 0.05 are set to zero in the head map illustration.

zero in the head map illustration. This average lagged cross-correlation is mostly similar to State 4 from our patient population.

### 4.5.1   Relationship of Microstates to Channel-wise Power Spectra

The power spectral density of bipolar channels averaged over all patients with DLOC is represented in Figure 4.11. The dominant feature that unifies these power spectra is a low frequency peak in a 1-3 Hz range, which is a classical spectral characteristic of the EEG in comatose patients. Such rhythmicity may correspond to State 4, where recovery of the correlation structure is observed at $-240ms$ (i.e., the third lag), consistent with a 2Hz oscillation.
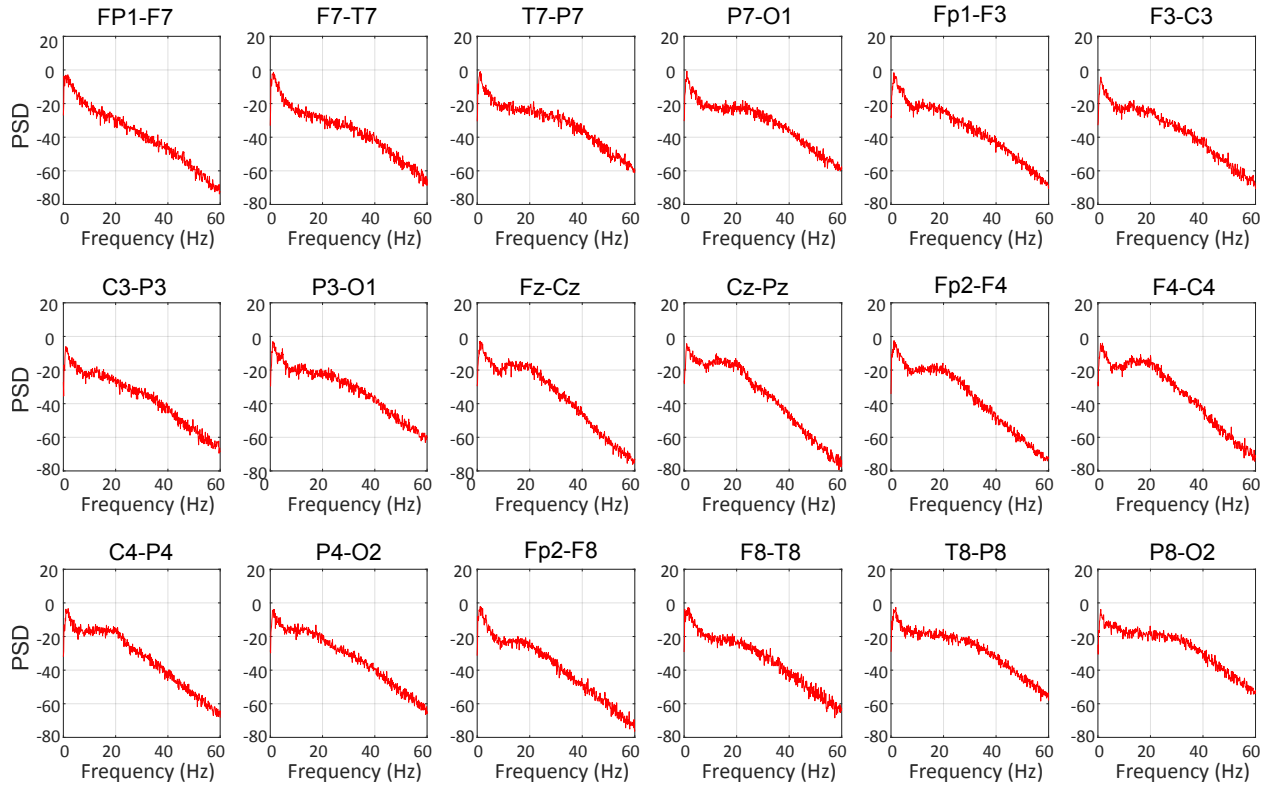
Figure 4.11: The power spectral density of 18 bipolar channels averaged over all patients with DLOC.

The power spectra for the healthy cohort is portrayed in Figure 4.12. We highlight differences in the low-frequency content of the patients and healthy subjects in Figure 4.13, specifically the lack of a 1-3Hz peak in the latter. This observation is interesting insofar as it suggests: (i) the average spatial correlation of the healthy brain (at rest) can also be found as a microstate (i.e., here, State 4) of the comatose brain; but (ii) that the temporal dynamics of this state are fundamentally different from those in the healthy condition.
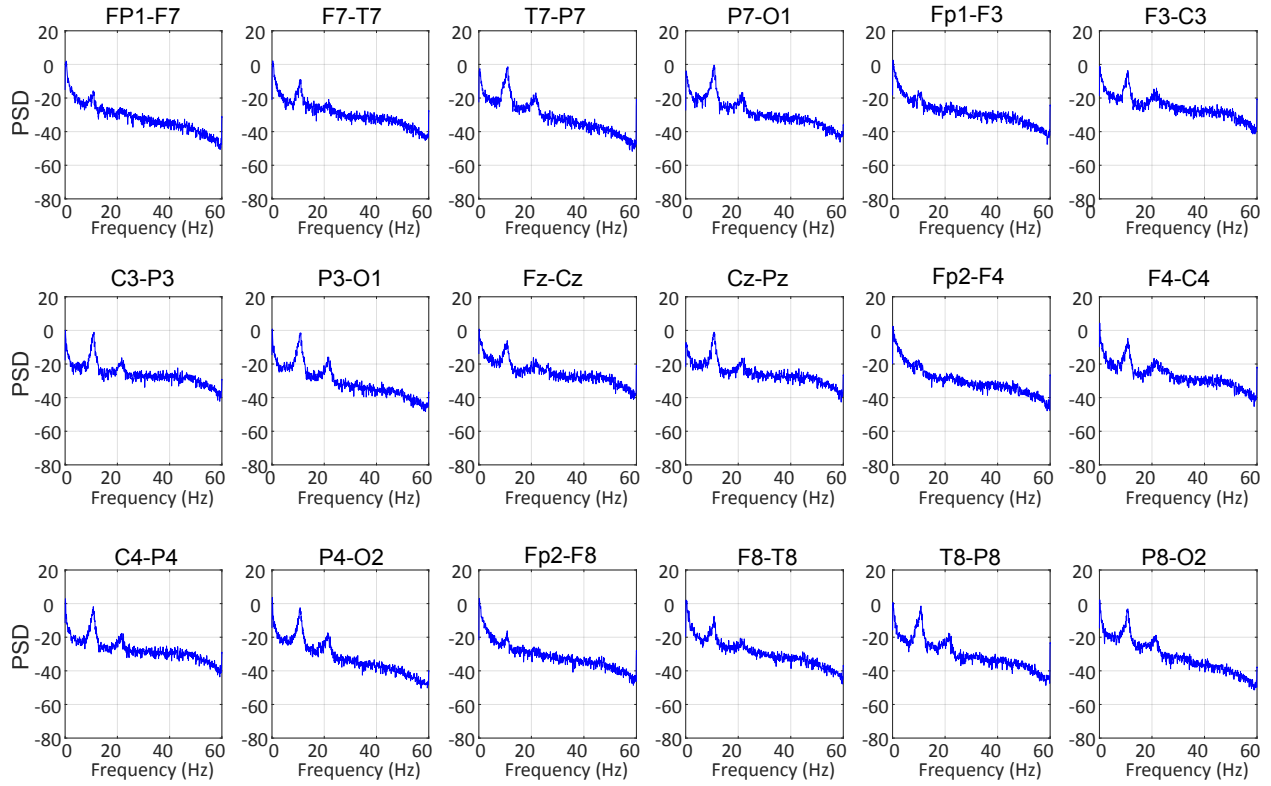
Figure 4.12: The power spectral density of 18 bipolar channels averaged over all healthy individuals in the study.
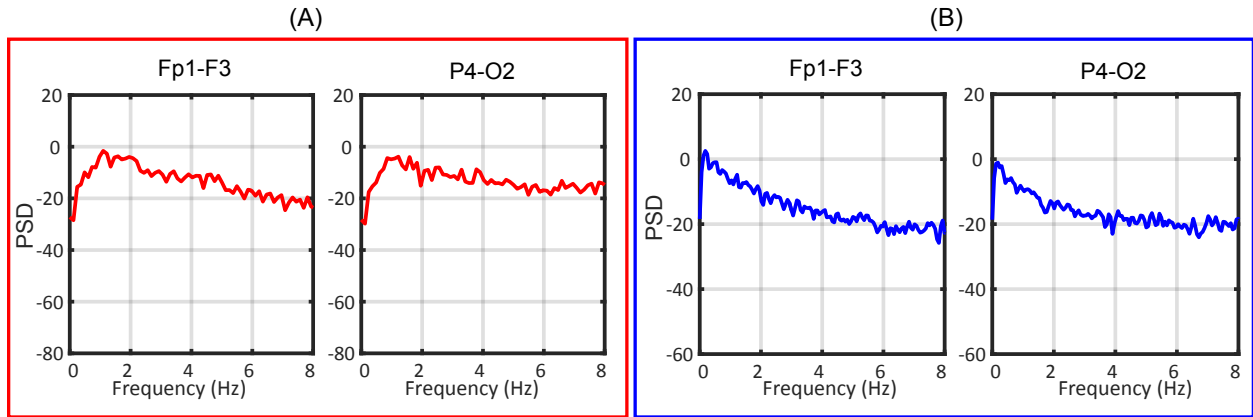


Figure 4.13: The power spectral density of 2 bipolar channels zoomed in the range of 0-8 Hz for (A) patients with coma and (B) healthy subjects. PSDs are averaged over all the subjects in the study for each case separately.

104
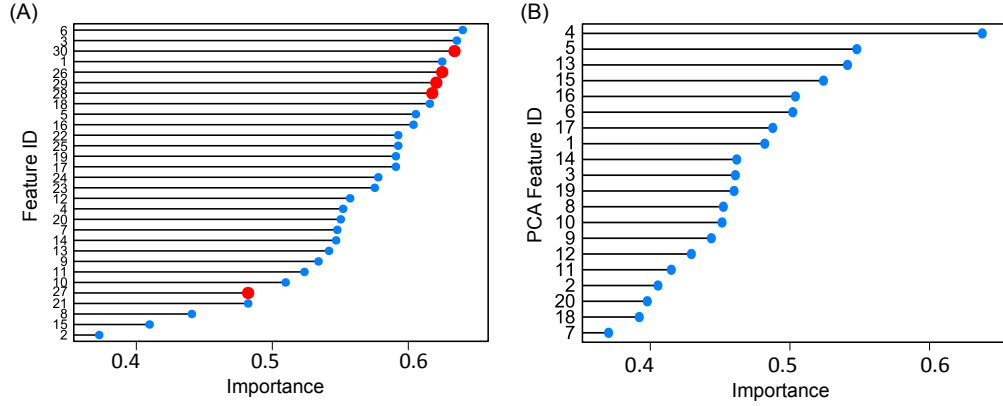
Figure 4.14: Feature ranking by importance (A) before and (B) after applying PCA to original feature vector. Occupancy features obtained from microstate analysis are shown with red filled circle in the left figure.

### 4.5.2 Classification of Coma Severity by Exploiting Microstate Occupancy

In this section, we show how the result of microstate analysis can enhance an automatic classification of different levels of coma severity. Here, the occupancy of being at a particular microstate is considered as a new feature for classification, in addition to those considered in our etiological (focal vs. diffuse) classifier. Specifically, to the features in Table 4.1 we add the occupancy of being at each of 5 different microstates (Feature IDs 26-30).

Figure 4.14A illustrates the ranking of all features for a severity classifier. It can be seen from this figure that the new defined features based on microstate analysis, shown with red filled circles, are among the most important features.

To obtain the actual classifier, we apply PCA to entire feature vector as was done previously for the etiological classifier. The rankings of principal component based on importance is shown in Figure 4.14B.

105

(A)

| Prediction | Reference | |
|---|---|---|
| | Low GCS | High GCS |
| Low | 8 | 2 |
| High | 5 | 6 |

Accuracy: 0.667
Sensitivity: 0.750
Specificity: 0.615

(B)

| Prediction | Reference | |
|---|---|---|
| | Low GCS | High GCS |
| Low | 9 | 1 |
| High | 4 | 7 |

Accuracy: 0.762
Sensitivity: 0.875
Specificity: 0.692

Figure 4.15: (A) Classification performance using 25 features extracted from 70 cases with two levels of coma severity, low and high, identified based on GCS score. (B) Classification performance using both 25 time series features in combination of five occupancy features from microstate analysis.

We compared the classification accuracy with and without utilizing the occupancy features in two ways. First, we performed a test of the classifier by withholding 21 cases as a dedicated testing set, and trained the classifier on only the remaining cases. The classifier retained only the first six PCs obtained from the with- and without-occupancy feature sets. Figure 4.15 shows the classification performance of the two scenarios.

Second, we performed training and testing for 500 repetitions (in each repetition, subjects in the training set are totally different from subjects in testing set). That is, we trained a classifier on 2/3 of the subjects, then evaluated it on a separate, withheld patient cohort. By doing so, we observed that the classifier with occupancy features outperforms the classifiers without occupancy features by an average of $5.96\% \pm 2.28$ (mean $\pm$ SD). The average

classification performance without occupancy features were as follows: accuracy=67.04, sensitivity=75.38, and specificity=57.07. After adding occupancy features, the classification performance enhanced as: accuracy=73.0, sensitivity=81.05, and specificity=62.63.

## 4.6   Discussion

### 4.6.1   Disambiguating Focal and Diffuse DLOC Etiologies using EEG Time Series Analysis

Our results demonstrate proof-of-concept for EEG-based segregation of focal versus diffuse DLOC etiology based on time-series analysis and support-vector machine classification. We evaluated the performance of this system by segmenting our data into separate training and testing sets and then using cross-validation to minimize model overfitting. The results demonstrate performance up to 71% accuracy when compared to classification based on clinical and imaging parameters. In many cases the ultimate cause of the DLOC only becomes clear in retrospect, and is ambiguous at the time the EEG is ordered. Thus even in cases where a clinical diagnosis can eventually be made, an EEG-based diagnostic method could assist with the timely delivery of care.

The design of our classifier reveals that potentially clinically-useful information regarding the level of DLOC may be embedded in both spectral and non-spectral features of the EEG signals of patients with DLOC. Our results further suggest that spectral analysis alone (e.g., band-limited power) may not capture all aspects of the underlying neuronal dynamics of coma etiology. In considering the performance of our derived classifier, it is notable
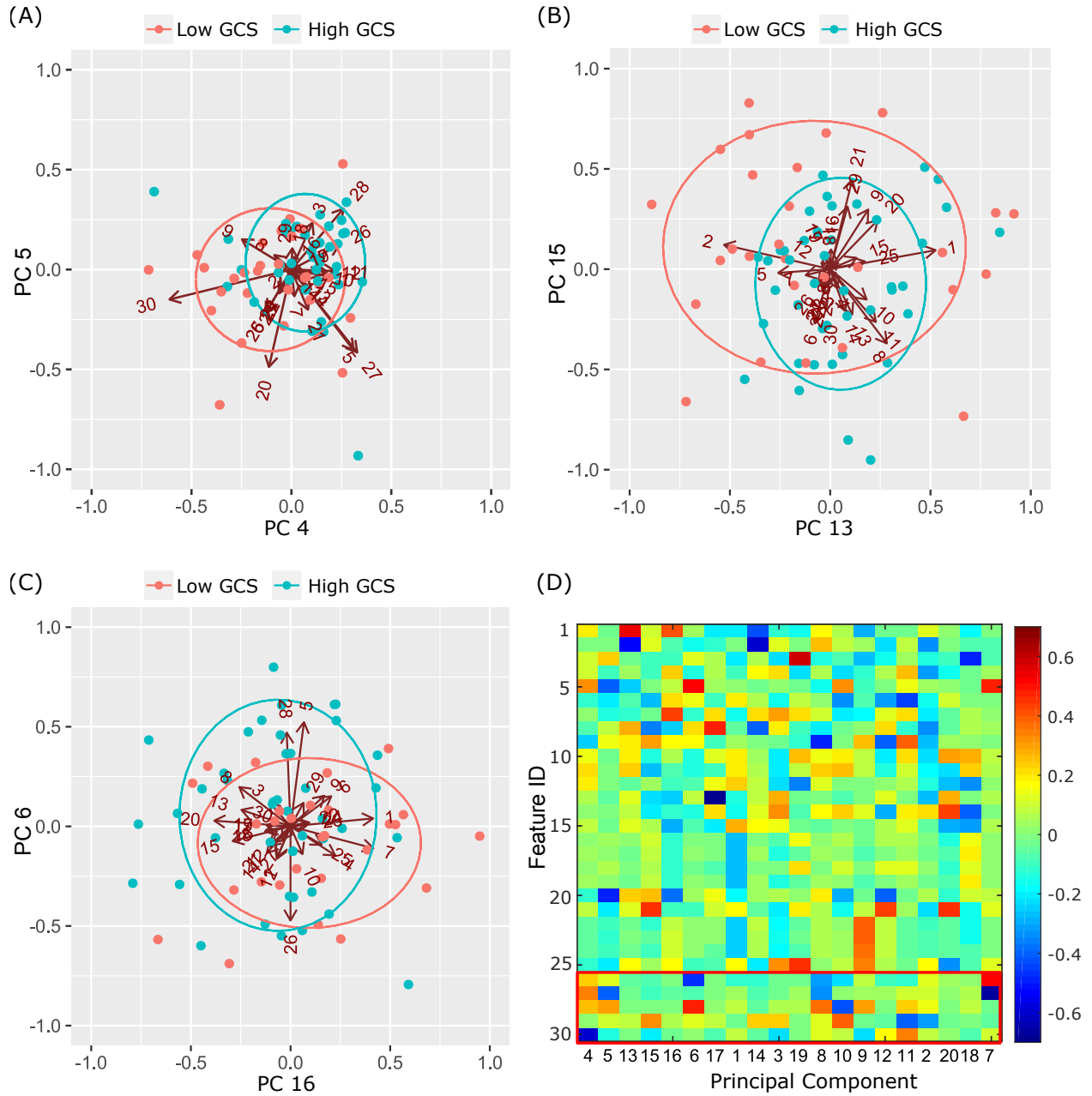
Figure 4.16: Principal components loadings for low and high GCS cases. (A)-(C) Contribution of each initial feature on different principal component. Red and blue ellipses show the direction of variation in the component scores of focal and diffuse, respectively. (D) Loading matrix for 20 principal components. Order of principal components are based on their importance.

that our primary features (Table 4.1) do not include any overt spatial analysis (e.g., cross-correlation), focusing on analysis at the individual channel level only. Thus, an interesting inference resulting from our approach is that channel-wise temporal dynamics may embed differences in coma etiology. So, inferences on diffuse/focal DLOC can potentially be made without explicitly describing between-channel spatial relationships.

In contrast, other methods for detecting focal versus diffuse pathology from the EEG include the brain symmetry index (BSI), which examines inter-hemispheric symmetries in the power-spectral density [123]. This method has been used in the detection of focal seizures and hemispheric strokes [124]. EEG-based synchronization indices, also derived from the power-spectrum, have also been suggested as a means of detecting diffuse electrophysiological phenomena [125]. To date, none of these methods has been used in the context of DLOC or coma, and none of them incorporate the time-series analysis and automatic feature selection framework deployed herein. Our goal in this study was to screen several minimally-processed dynamical features using a data-driven framework. Thus we did not evaluate indices such as the BSI, though band-limited power was included in our analysis framework. However, it is possible that these processed indices may provide discriminability regarding focal and diffuse DLOC. Moreover, the incorporation of active stimuli to assess EEG reactivity may provide additional information about DLOC [126]. A comparative study of existing indices applied to DLOC with potential inclusion of active stimulation is an important direction for future work.

**Limitations:** The main limitation in the development of our algorithm is the lack of a true gold standard upon which to train our classifier. Our cases were independently diagnosed by two neurointensivists on the basis of all available retrospective data, including neuroimaging, and only cases in which both felt a clear diagnosis was evident were included.

In this retrospective study we relied on the GCS for identification of coma severity for patients with DLOC. The GCS is an imperfect measure of DLOC: its measurement is incomplete in intubated patients and many features that are likely to be clinically significant are not assessed. Several other metrics such as the Full Outline of UnResponsiveness score [127] and the JFK Coma Recover Scale, Revised [128] are likely to provide superior differentiation of patients with DLOC. Unfortunately our data did not permit such assessments retrospectively, though a prospective trial is underway including assessment of these metrics in addition to GCS.

Another drawback of our framework is that direct mechanistic interpretation from descriptive time-series analysis is lacking. The support vector machine approach aggregates all of the data/features (because the best predictors are combination of the primary features), then generates a set of predictors that may, or may not, be overtly linked to an underlying circuit mechanism.

Lastly, we have not taken specific steps to manage confounds introduced by the administration of medications such as antiepileptic drugs, including benzodiazepines, on the brain's electrical activity. It is well-established that these medications, among other factors such as sleep [129], can lead to confounding effects on the EEG. In the absence of a much larger trial matched for specific agents or classes of agents, it will be challenging to fully understand the impact of such confounds. Indeed, our study size of 43 patients with known etiology may not be sufficient to sample the range of EEG variation associated with these drugs and we were not able to control for a bias in the types of drugs administered to focal versus diffuse DLOC. We hypothesize that the confounding effects of these drugs constitute a major source of classification error in our dataset. In a future prospective trial it may be feasible

to more specifically test the effect of classes of medications on the discriminatory power of our classification scheme in individual subjects.

**Design Considerations:** The feature selection framework reported above requires no manual specification of thresholds or other detection rules. The method can be applied to any number of EEG channels. However, it should be noted that a pervasive challenge with any clinical EEG recording is overall signal quality and presence of artifacts (e.g., due to patient motion), which is expected to be compounded with added channels.

It is important to note that with our feature selection scheme, any change to the design parameters may lead to a different set of PCA features being chosen. Nevertheless, in our study, we found that across a range of design parameterizations (e.g., changing the trial length from 5 minutes to 10 minutes) the set of best predictors was largely comprised of the same descriptive time-series statistics. Thus, we conclude that these particular PCA features are robustly informative with respect to our two DLOC variants.

## 4.6.2 Spatiotemporal Characterization of Different Levels of Coma Severity

We presented a new spatiotemporal feature termed lagged microstate analysis. Our analysis shows that these microstates ranked among the most important predictors of level of coma severity. Specifically, we quantified the extent of the appearance of each microstate for different levels of coma severity, from which we can also draw some basic conclusions regarding differences between low and high GCS scores in patients with DLOC. Our findings suggest that patients with low GCS scores (deeper levels of unconsciousness) tend to be a state of

111

frontal-posterior anti-correlation, whereas the patients with high GCS tend to exhibit more persistent correlation across the scalp.

As mentioned, above, the microstates can in some cases be related to the harmonic structure (e.g., 1-3Hz rhythmicity) of the EEG. However, we can at this point only conjecture as to the physiological basis of these observed dynamics. In particular, our microstates, while spatially cogent, depict an enigmatic mix of correlations and anti-correlations. The latter remain the topic of substantial debate in terms of their cognitive salience. For example, previous studies have demonstrated anti-correlation between ventromedial prefrontal cortex (vmPFC) and networks activated in studies of visual and spatial attention such as superior/inferior parietal and extrastriate cortices [133–135]. vmPFC is often associated with mentalizing, evaluative, and self-referential processing. Anti-correlation between the default mode network and task positive network has been reported in several studies, and it has been suggested that it represent a "division of labor" between networks with seemingly opposite functions [136]. It has already been suggested that certain neuropathologies are associated with long term anti-correlation, whereas the normal brain is more stochastic [138] in terms of its activity. Hence, persistence of anti-correlation over several lags might be an evidence of such long term anti-correlation in patients with DLOC.

# Part II

# Dynamics at Micro-scale

# Chapter 5

# Decoding of Time-Varying Sparse Signals in Recurrent Linear Networks

## 5.1 Highlights

In this chapter, we study how the dynamics of recurrent networks, formulated as general dynamical systems, mediate the recovery of sparse, time-varying signals. Our formulation resembles the well-described problem of compressed sensing, but in a dynamic setting. We specifically consider the problem of recovering a high-dimensional network input, over time, from observation of only a subset of the network states (i.e., the network output). Our goal is to ascertain how the network dynamics may enable recovery, even if classical methods fail at each time instant. We are particularly interested in understanding performance in scenarios where both the input and output are corrupted by disturbance and noise, respectively. Our main results consist of the development of analytical conditions, including a generalized

observability criterion, that ensure exact and stable input recovery in a dynamic, recurrent network setting.[5]

## 5.2   Introduction

### 5.2.1   Background

The notion of efficient coding dictates that information processing in neuronal networks is optimal according to a cost function that penalizes excessive neuronal activation [139–141]. Given the temporally punctate nature of neuronal spikes, sparsity has been frequently considered within the schema of efficient coding [6, 142]. Notable within the sparse coding framework [143–145] are analyses that leverage the theory of compressed sensing (CS) [146–148] for underdetermined linear systems (e.g., [7, 8, 149, 150]). CS outlines conditions under which a linear system admits exact and stable recovery (i.e., decoding) of a sparse input [151–154]. Classical CS, however, does not typically consider temporal dynamics within the recovery problem.

Since networks in the brain are inherently recurrent, interest has also grown around so-called *dynamic CS* and, specifically, on the recovery of signals subjected to transformation via a dynamical system (or, network). In this context, sparsity has been formulated in three ways: 1) In the network states (state sparsity) [155–157]; 2) In the structure/parameters of the network model (model sparsity) [158, 159]; and 3) In the inputs to the network (input sparsity) [160–170]. In this chapter, we focus on this latter category of recovery problems.

---

[5]This chapter is a slightly amended version of the previously published article which is listed as follows: M. Kafashan, A. Nandi, and S. Ching, "Relating Observability and Compressed Sensing of Time-Varying Signals in Recurrent Linear Networks," *Neural Networks*, 83, 11-20, 2016.
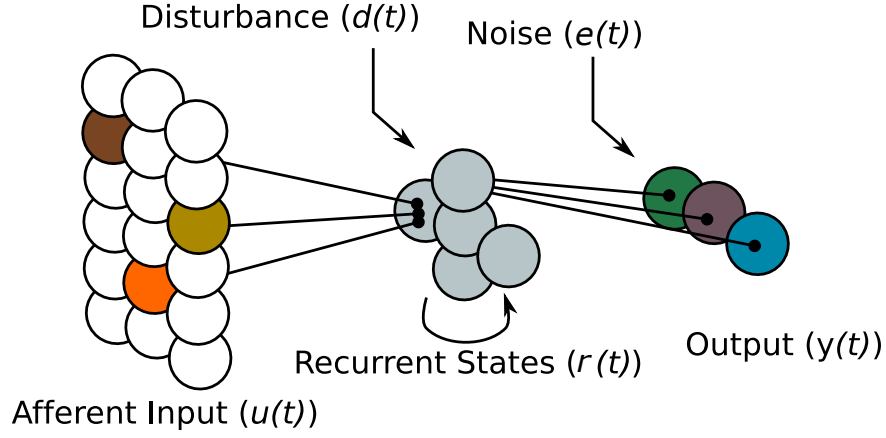
Figure 5.1: Schematic of the considered network architecture. We study how the afferent, recurrent and output stages of this architecture interplay in order to enable accurate estimation of the input $u(t)$ from $y(t)$ in the presence of both disturbance and noise.

Our motivation is to understand how three stages of a generic network architecture – an afferent stage, a recurrent stage, and an output stage (see Figure 5.1) – interplay in order to enable an observer, sampling the output, to recover the (sparse) input in the presence of both disturbance and noise. Such an architecture is pervasive in sensory networks in the brain wherein a large number of sensory neurons, receiving excitation from the periphery, impinge on an early recurrent network layer that transforms the afferent excitation en route to higher brain regions [171, 172]. Moreover, beyond neural contexts, understanding network characteristics for dynamic CS may aid in the analysis of systems for efficient processing of naturally sparse time-varying signals [162]; and in the design of resilient cyber-physical systems [160, 161, 163], wherein extrinsic perturbations are sparse and time-varying.

Specifically, we consider a network that receives an input $u_t$ and generates observations (network outputs), $y_t$ via its recurrent dynamics, i.e.,

$$r_{t+1} = f(r_t, u_t, d_t) \tag{5.1}$$

$$y_t = g(r_t, e_t) \tag{5.2}$$

where, here, $r_t$ are the network states, $d_t$ is the disturbance and $e_t$ is the noise. Our focus is on how the network dynamics, embedded in $f(\cdot), g(\cdot)$, impact the extent to which $u_t$ can be inferred from $y_t$ in the case where the dimensionality of the latter is substantially less than that of the former. We will focus exclusively on the case where these dynamics are linear.

Considered at face value, the problem of sparse recovery in a dynamic setting could be broken down into a sequential problem wherein classical CS is performed on the system frozen at each time instant. However, this approach would require the strong condition that the system admits classical CS pointwise. In a dynamic network, recovery may be possible even if the conditions for (static) CS fail at each time instant. For instance, in a linear dynamical system, if the number of observations is smaller than the dimension of the state space, then it is not possible to simply 'invert' the observations to recover the states at any particular time. However, the notion of observability provides conditions under which such recovery is possible, as long as those observations are allowed to accrue through the system dynamics over some period of time.

Here, we postulate, then derive, a set of analytical conditions for dynamic CS that involve observability, but are in general less stringent than necessitating CS at each time instant. Associated with these conditions, we also derive bounds on the $l_2$-norm of the reconstruction error over a finite time horizon, which can be defined in terms of both the disturbance and

117

the noise. Salient interpretations of the conditions, such as a basic trade-off between the ability of a network to simultaneously reject input disturbances while still admitting stable recovery, are highlighted and demonstrated via simulation.

## 5.2.2 Relationship to Prior Results in Sparse Time-Varying Input Recovery

The sparse input recovery problem for linear dynamical systems can be formulated in both the spatial and temporal dimension. Our contributions are related to the former. As mentioned above, in this context, previous work has considered recovery of spatially sparse inputs for network resilience [160,161] and encoding of inputs with sparse increments [162]. In [163], conditions for exact sparse recovery are formulated in terms of a coherence-based observability criterion for intended applications in cyber-physical systems. Our contributions herein provide a general set of analytical results, including performance bounds, pertaining to exact and stable sparse input recovery of linear systems in the presence of both noise and disturbance.

A second significant line of research in sparse input recovery problems pertains to the temporal dimension. There, the goal is to understand how a temporally sparse signal (i.e., one that takes the value of zero over a nontrivial portion of its history) can be recovered from the state of the network at a particular instant in time. This problem has been used to describe 'memory' in dynamical models of brain networks [164–169]. In particular, in [170] the problem of ascertaining memory is related to CS performed on the network states, over a receding horizon of a scalar-valued input signal. In contrast to these works, we consider spatial sparsity of vector-valued inputs with explicit regard for both disturbance and an overt

observation equation, i.e., states not to be directly sampled, but are transformed to an (in general lower-dimensional) output.

## 5.3   Detailed Problem Formulation

We consider a discrete-time linear network, formulated in the typical form of a linear dynamical system, i.e.,:

$$
\begin{aligned}
\mathbf{r}_{k+1} &= \mathbf{A}\mathbf{r}_k + \mathbf{B}\mathbf{u}_k + \mathbf{d}_k \\
\mathbf{y}_k &= \mathbf{C}\mathbf{r}_k + \mathbf{e}_k,
\end{aligned}
\tag{5.3}
$$

where $k$ is an integer time index, $\mathbf{r}_k \in \mathbb{R}^n$ is the activity of the network nodes (e.g., in the case of a rate-based neuronal network [9–11, 173–176], the firing rate of each neuron), $\mathbf{u}_k \in \mathbb{R}^m$ is the extrinsic input, $\mathbf{d}_k \in \mathbb{R}^n$ is the input disturbance, $\mathbf{e}_k \in \mathbb{R}^p$ is the measurement noise independent from $\mathbf{d}_k$, and $\mathbf{y}_k \in \mathbb{R}^p$ is the observation at time $k$. The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ describes connections between nodes in the network, $\mathbf{B} \in \mathbb{R}^{n \times m}$ contains weights between input and output and $\mathbf{C} \in \mathbb{R}^{p \times n}$ is the measurement matrix. Such a model is, of course, quite general and can be used to describe recurrent dynamics in neuronal networks [177–181], and machine learning applications such as pattern recognition and data mining [182–185].

We consider the case of bounded disturbance and noise, i.e., $\|\mathbf{e}_k\|_{\ell_2} \leq \epsilon$, $\|\mathbf{d}_k\|_{\ell_2} \leq \epsilon'$. Since $m$, the number of input nodes, is larger than $n$, the number of network nodes, $\mathbf{B}$ takes the form of a "wide" matrix. We assume that at each time at most $s$ input nodes are active ($s$-sparse input), leading to an $\ell_0$ constraint to (5.3) at each time point:

$$
\|\mathbf{u}_k\|_{\ell_0} \leq s.
\tag{5.4}
$$

In the absence of disturbance and noise, recovering the input of (5.3) with the $\ell_0$ constraint (5.4) amounts to the optimization problem:

$$(P0) \quad \underset{(\mathbf{r}_k)_{k=0}^{K},(\mathbf{u}_k)_{k=0}^{K-1}}{\text{minimize}} \quad \sum_{k=0}^{K-1} \|\mathbf{u}_k\|_{\ell_0}$$
$$\text{subject to} \quad \mathbf{r}_{k+1} = \mathbf{A}\mathbf{r}_k + \mathbf{B}\mathbf{u}_k$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{r}_k. \tag{5.5}$$

It is clear that Problem $(P0)$ is a non-convex discontinuous problem, which is not numerically feasible and is NP-Hard in general [186]. For static cases, such $\ell_0$ optimization problems fall into the category of combinatorial optimization which require exhaustive search to find the solution [187].

Thus, throughout this chapter, we follow the typical relaxation methodology used for such problems wherein the $\ell_0$ norm is relaxed to the $l_1$ norm, resulting in the problem:

$$(P1) \quad \underset{(\mathbf{r}_k)_{k=0}^{K},(\mathbf{u}_k)_{k=0}^{K-1}}{\text{minimize}} \quad \sum_{k=0}^{K-1} \|\mathbf{u}_k\|_{\ell_1}$$
$$\text{subject to} \quad \mathbf{r}_{k+1} = \mathbf{A}\mathbf{r}_k + \mathbf{B}\mathbf{u}_k$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{r}_k. \tag{5.6}$$

In the case that either input disturbance, or measurement noise, or both exist, we solve the following convex optimization Problem $(P2)$:

$$(P2) \quad \underset{(\mathbf{r}_k)_{k=0}^{K},(\mathbf{u}_k)_{k=0}^{K-1}}{\text{minimize}} \quad \sum_{k=0}^{K-1} \|\mathbf{u}_k\|_{\ell_1}$$
$$\text{subject to} \quad \mathbf{r}_{k+1} = \mathbf{A}\mathbf{r}_k + \mathbf{B}\mathbf{u}_k$$
$$\|\mathbf{y}_k - \mathbf{C}\mathbf{r}_k\|_{\ell_2} \leq \epsilon'', \tag{5.7}$$

where $\epsilon''$ is the $l_2$ norm of a surrogate parameter that aggregates the effects of disturbance and noise. In $P2$, the objective function is exactly the same as $P1$. However, due to noise and disturbance, the last constraints in $P2$ are quadratic inequalities instead of affine constraints in $P1$. In the case of noisy measurement with no disturbance $\epsilon'' = \epsilon$. In the next section, we show conditions for the network (5.3) under which Problems ($P1$) and ($P2$) result in exact and stable solutions.

## 5.4   Results

We will begin by presenting a general theory for problems ($P1$) and ($P2$) for underdetermined linear dynamic networks. We will then briefly show how this reduces exactly to the case of sequential application of classical CS when the system is fully determined. Finally, we present an analysis to characterize the ability of a network to reject input disturbances while still admitting stable recovery in the presence of disturbance and noise, simultaneously.

### 5.4.1   Preliminaries

We begin by recalling some basic matrix notation and matrix norm properties that will be used throughout this chapter. Given normed spaces $(\mathbb{R}^{n_1}, \|.\|_{\ell_2})$ and $(\mathbb{R}^{n_2}, \|.\|_{\ell_2})$, the corresponding induced norm or operator norm denoted by $\|.\|_{i,2}$ over linear maps $\mathbf{D} : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$, $\mathbf{D} \in \mathbb{R}^{n_2 \times n_1}$ is defined by

$$
\begin{aligned}
\|\mathbf{D}\|_{i,2} &= sup\{\frac{\|\mathbf{D}\mathbf{r}\|_{\ell_2}}{\|\mathbf{r}\|_{\ell_2}} \mid \mathbf{r} \in \mathbb{R}^{n_1}, \mathbf{r} \neq \mathbf{0}\} \\
&= max\{\sqrt{\lambda} \mid \lambda \in \sigma\left(\mathbf{D}^T\mathbf{D}\right)\},
\end{aligned}
\tag{5.8}
$$

where $\sigma(\mathbf{M})$ is the set of eigenvalues of $\mathbf{M}$ (or the spectrum of $\mathbf{M}$).

**Definition 5.1.** *A vector is said to be s-sparse if $\|\mathbf{c}\|_{\ell_0} \leq s$, in other words it has at most s nonzero entries.*

It is well known that in the static case (standard CS), exact and stable recovery of sparse inputs can be obtained under the restricted isometry property (RIP) [151–154, 188], defined as:

**Definition 5.2.** *The restricted isometry constant $\delta_s$ of a matrix $\Phi \in \mathbb{R}^{n \times m}$ is defined as the smallest number such that for all s-sparse vectors $\mathbf{c} \in \mathbb{R}^m$ the following equation holds*

$$(1 - \delta_s)\|\mathbf{c}\|_{\ell_2}^2 \leq \|\Phi\mathbf{c}\|_{\ell_2}^2 \leq (1 + \delta_s)\|\mathbf{c}\|_{\ell_2}^2. \tag{5.9}$$

It is known that many types of random matrices with independent and identically distributed entries or sub-Gaussian matrices satisfy the RIP condition (5.9) with overwhelming probability [189–191].

## 5.4.2 A Generalized Observability Criteria for Time-Varying Sparse Input Recovery

In this case, we consider (5.3) in the absence of input disturbance ($\mathbf{d}_k = 0$) with the linear map $\mathbf{C} : \mathbb{R}^n \to \mathbb{R}^p$. We will focus on the case where $p < n$ so that direct inversion of $\mathbf{C}$ is not possible. For any positive $K$, we define the standard linear observability matrix as

$$\mathbf{O}_K \equiv \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^K \end{pmatrix}. \tag{5.10}$$

If $rank(\mathbf{O}_K) = n$, then the system (5.3) is observable in the classical sense [6]. Here, we do not assume any knowledge of the input other than the fact that it is $s$-sparse at each time. Note that if we iterate the output equation in (5.3) for $K+1$ time steps and exploit the fact that the input vector is $s$-sparse as shown in Figure 5.2, we obtain:

$$\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{pmatrix} = \mathbf{O}_K \mathbf{r}_0 + \mathbf{J}_K^s \begin{pmatrix} \mathbf{u}_0^s \\ \mathbf{u}_1^s \\ \vdots \\ \mathbf{u}_{K-1}^s \end{pmatrix}, \tag{5.11}$$

where $\mathbf{u}_k^s$ represents $s$ nonzero elements of $\mathbf{u}_k$, and $\mathbf{J}_K^s$ is as follows:

$$\mathbf{J}_K^s = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CB}_0^s & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CAB}_0^s & \mathbf{CB}_1^s & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{K-1}\mathbf{B}_0^s & \mathbf{CA}^{K-2}\mathbf{B}_1^s & \cdots & \mathbf{CB}_{K-1}^s \end{pmatrix}, \tag{5.12}$$

where $\mathbf{B}_i^s$ is the $n \times s$ matrix corresponding the active columns of $\mathbf{B}$ (corresponding to nonzero input entries) at time step $i$ (see Figure 5.2). In general, we do not know where the

---

[6]The system is said to be observable if, for any initial state and for any known sequence of input there is a positive integer $K$ such that the initial state can be recovered from the outputs $\mathbf{y}_0, \mathbf{y}_1,..., \mathbf{y}_K$.
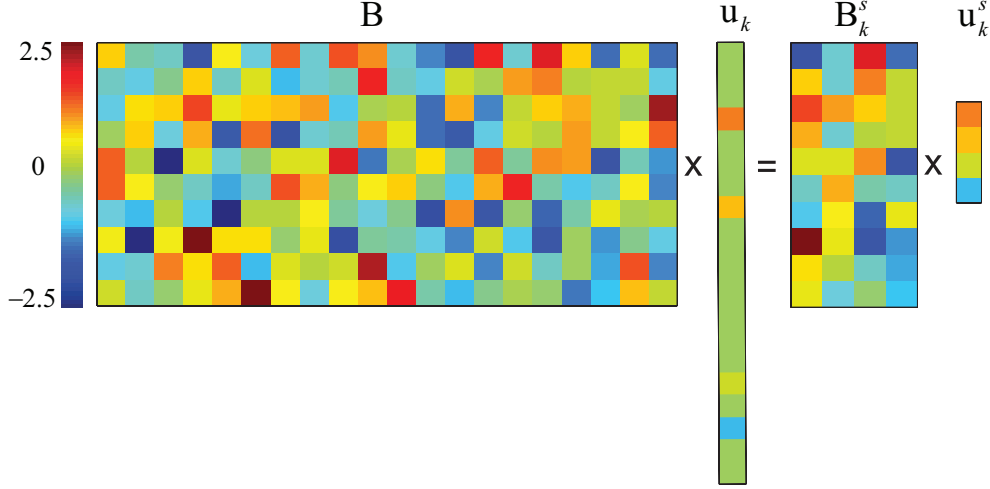
Figure 5.2: The matrix $\mathbf{B}_k^s$ is the $n \times s$ matrix corresponding the active columns of the full matrix $\mathbf{B}$ at time step $k$.

active columns of $\mathbf{B}$ are located at each time *a priori*. Thus, $\mathcal{J}_K^s$ is defined as the set of all possible matrices satisfying the structure in (5.12), where the cardinality of this set is $\binom{m}{s}^K$.

We first establish conditions under which a one to one correspondence exists between sparse input and observed output for the system (5.3).

**Lemma 5.1.** *Suppose that the sequence* $(\mathbf{y}_k)_{k=0}^K$ *from noiseless measurements is given, and* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *are known. Assume* $rank(\mathbf{O}_K) = n$ *and the matrix* $\mathbf{CB}$ *satisfies the RIP condition (5.9) with isometry constant* $\delta_{2s} < 1$. *Further, assume*

$$rank([\mathbf{O}_K \quad \mathbf{J}_K^{2s}]) = n + rank(\mathbf{J}_K^{2s}), \ \forall \mathbf{J}_K^{2s} \in \mathcal{J}_K^{2s}. \tag{5.13}$$

*Then, there is a unique s-sparse sequence of* $(\mathbf{u}_k)_{k=0}^{K-1}$ *and a unique sequence of* $(\mathbf{r}_k)_{k=0}^K$ *that generate* $(\mathbf{y}_k)_{k=0}^K$.

*Proof.* See Appendix D.1 □

124

The rank condition (5.13) implies that all columns of the observability matrix must be linearly independent of each other (i.e., the network is observable in the classical sense) and of all columns of $\mathbf{J}_K^s$. Since the exact location of the nonzero elements of the input vector are not known *a priori*, this condition is specified over all matrices in $\mathcal{J}_K^{2s}$. Thus, (5.13) is a combinatorial condition. From our simulation studies, we have observed that this condition holds for random Gaussian matrices almost always (see also Examples 1 and 2 in Section 5.5).

Having established the existence of a unique solution, we now proceed to study the convex problems $(P1)$ and $(P2)$ that recover these solutions for this case. First, we provide theoretical results for the stable recovery of the input where measurements are noisy i.e., $(P2)$.

**Theorem 5.1.** *(Noisy recovery) Assume* $rank(\mathbf{O}_K) = n$ *and the matrix* $\mathbf{CB}$ *satisfies the RIP condition (5.9) with* $\delta_{2s} < \sqrt{2} - 1$. *assume (5.13) holds. Suppose that the sequence* $(\mathbf{y}_k)_{k=0}^{K}$ *is given and generated from sequences* $(\bar{\mathbf{r}}_k)_{k=0}^{K}$ *and s-sparse* $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ *based on*

$$
\begin{aligned}
\bar{\mathbf{r}}_{k+1} &= \mathbf{A}\bar{\mathbf{r}}_k + \mathbf{B}\bar{\mathbf{u}}_k, \;\; k = 0, \cdots, K-1 \\
\mathbf{y}_k &= \mathbf{C}\bar{\mathbf{r}}_k + \mathbf{e}_k, \;\; k = 0, ..., K,
\end{aligned}
\tag{5.14}
$$

*where* $(\|\mathbf{e}_k\|_{\ell_2} \leq \epsilon)_{k=0}^{K}$ *and* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *are known. Then, the solution to Problem (P2) obeys*

$$
\sum_{k=0}^{K-1} \|\mathbf{u}_k^* - \bar{\mathbf{u}}_k\|_{\ell_2} \leq C_s \epsilon,
\tag{5.15}
$$

*where*

$$
C_s = 2\alpha C_0 K (1 - \rho)^{-1}.
\tag{5.16}
$$

125

*$C_0$, $\rho$, $\alpha$ are given explicitly below:*

$$C_0 = \left(1 + \sqrt{\frac{K\sigma_{max}\left(\mathbf{C}^T\mathbf{C}\right)\sigma_{max}\left(\mathbf{A}^T\mathbf{A}\right)\sigma_{max}\left((\mathbf{I} - \mathbf{P_{J_K^{2s}}})^T(\mathbf{I} - \mathbf{P_{J_K^{2s}}})\right)}{\sigma_{min}\left(\mathbf{O}_K{}^T\mathbf{O}_K\right)\sigma_{min}\left((\mathbf{I} - \mathbf{P_{J_K^{2s}}})^T(\mathbf{I} - \mathbf{P_{J_K^{2s}}})\right)}}\right),$$

$$\alpha = \frac{2\sqrt{1 + \delta_{2s}}}{1 - \delta_{2s}}, \tag{5.17}$$

$$\rho = \frac{\sqrt{2}\delta_{2s}}{1 - \delta_{2s}},$$

$$\mathbf{P_{J_K^{2s}}} = \mathbf{J}_K^{2s}(\mathbf{J}_K^{2s\,T}\mathbf{J}_K^{2s})^{-1}\mathbf{J}_K^{2s\,T}.$$

*Proof.* Assume that the the sequences $(\mathbf{r}_k^*)_{k=0}^K$ and sparse $(\mathbf{u}_k^*)_{k=0}^{K-1}$ are the solutions of Problem $(P2)$. First we derive the bound for the $\|\mathbf{r}_k^* - \bar{\mathbf{r}}_k\|_{\ell_2}$ in the following Lemma.

**Lemma 5.2.** *Suppose that the sequence $(\mathbf{y}_k)_{k=0}^K$ is given and generated from sequences $(\bar{\mathbf{r}}_k)_{k=0}^K$ and s-sparse $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ based on (5.14), where $(\|\mathbf{e}_k\|_{\ell_2} \leq \epsilon)_{k=0}^K$ and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are known. Then, any solution $\mathbf{r}_k^*$ to Problem (P2) obeys*

$$\|\mathbf{r}_k^* - \bar{\mathbf{r}}_k\|_{\ell_2} \leq 2\epsilon \sqrt{\frac{K\sigma_{max}\left((\mathbf{I} - \mathbf{P_{J_K^{2s}}})^T(\mathbf{I} - \mathbf{P_{J_K^{2s}}})\right)}{\sigma_{min}\left(\mathbf{O}_K{}^T\mathbf{O}_K\right)\sigma_{min}\left((\mathbf{I} - \mathbf{P_{J_K^{2s}}})^T(\mathbf{I} - \mathbf{P_{J_K^{2s}}})\right)}} \tag{5.18}$$

*Proof.* See Appendix D.2. □

From Lemma 5.2, and the equation $\mathbf{y}_k = \mathbf{CA}\bar{\mathbf{r}}_{k-1} + \mathbf{CB}\bar{\mathbf{u}}_{k-1} + \mathbf{e}_k$ we can derive a bound for $\|\mathbf{CB}\left(\mathbf{u}_k^* - \bar{\mathbf{u}}_k\right)\|_{\ell_2}$ as

$$
\begin{aligned}
\|\mathbf{CB}\left(\mathbf{u}_k^* - \bar{\mathbf{u}}_k\right)\|_{\ell_2} &= \|(\mathbf{e}_{k+1}^* + \mathbf{e}_{k+1}) + \mathbf{CA}(\bar{\mathbf{r}}_k - \mathbf{r}_k^*)\|_{\ell_2} \\
&\leq \|\mathbf{e}_{k+1} + \mathbf{e}_{k+1}^*\|_{\ell_2} + \|\mathbf{CA}(\bar{\mathbf{r}}_k - \mathbf{r}_k^*)\|_{\ell_2} \\
&\leq 2\epsilon\left(1 + \sqrt{\frac{K\sigma_{max}\left(\mathbf{C}^T\mathbf{C}\right)\sigma_{max}\left(\mathbf{A}^T\mathbf{A}\right)}{\sigma_{min}\left(\mathbf{O}_K{}^T\mathbf{O}_K\right)\sigma_{min}\left(\left(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}}\right)^T(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}}\right) \\
&= 2C_0\epsilon.
\end{aligned}
\tag{5.19}
$$

Now, denote $\mathbf{u}_k^* = \bar{\mathbf{u}}_k + \mathbf{h}_k$ where $\mathbf{h}_k$ can be decomposed into a sum of vectors $\mathbf{h}_{k,T_0(k)}$, $\mathbf{h}_{k,T_1(k)}$, $\mathbf{h}_{k,T_2(k)}$, $\cdots$ for each $k$, each of sparsity at most $s$. Here, $T_0(k)$ corresponds to the location of non-zero elements of $\bar{\mathbf{u}}_k$, $T_1(k)$ to the location of $s$ largest coefficients of $\mathbf{h}_{k,T_0^c(k)}$, $T_2(k)$ to the location of the next $s$ largest coefficients of $\mathbf{h}_{k,T_0^c(k)}$, and so on. Also, let $T_{01}(k) \equiv T_0(k) \cup T_1(k)$. Extending the technique in [154, 188], it is possible to obtain a cone constraint for linear dynamical systems.

**Lemma 5.3.** *(Cone constraint) The optimal solution for the input in Problem (P2) satisfies*

$$
\sum_{k=0}^{K-1}\|\mathbf{h}_{k,T_{01}^c(1)}\|_{\ell_2} \leq \sum_{k=0}^{K-1}\|\mathbf{h}_{k,T_0(1)}\|_{\ell_2}.
\tag{5.20}
$$

*Proof.* See Appendix D.3. □

We can further establish a bound for the right hand side of (5.20):

**Lemma 5.4.** *The optimal solution for the input in Problem (P2) satisfies the following constraint*

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(1)}\|_{\ell_2} \le K(1-\rho)^{-1}\alpha C_0 \epsilon. \tag{5.21}$$

*Proof.* See Appendix D.4. $\qquad\square$

Finally, based on Lemma 5.3 and Lemma 5.4, it is easy to see that

$$
\begin{aligned}
\sum_{k=0}^{K-1} \|\mathbf{h}_k\|_{\ell_2} &\le \sum_{k=0}^{K-1} \left( \|\mathbf{h}_{k,T_{01}(1)}\|_{\ell_2} + \|\mathbf{h}_{k,T_{01}^c(1)}\|_{\ell_2} \right) \\
&\le 2 \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(1)}\|_{\ell_2} \\
&\le 2\alpha C_0 K(1-\rho)^{-1}\epsilon = C_s \epsilon.
\end{aligned}
\tag{5.22}
$$

$\qquad\square$

**Theorem 5.2.** *(Noiseless recovery) Assume $rank(\mathbf{O}_K) = n$ and the matrix $\mathbf{CB}$ satisfies the RIP condition (5.9) with $\delta_{2s} < \sqrt{2} - 1$. Further, assume (5.13) holds. Suppose that the sequence $(\mathbf{y}_k)_{k=0}^{K}$ is given and generated from sequences $(\bar{\mathbf{r}}_k)_{k=0}^{K}$ and s-sparse inputs $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ based on dynamical equation (5.14), where $\epsilon = 0$ and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are known. Then the sequences $(\bar{\mathbf{r}}_k)_{k=0}^{K}$ and $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ are the unique minimizer to Problem (P1).*

*Proof.* It can be concluded from Lemma 5.1 and Theorem 5.1 that with the assumption stated in the theorem and $\epsilon = 0$ the sequences $(\bar{\mathbf{r}}_k)_{k=0}^{K}$ and $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ are the unique minimizer to Problem (P1). $\qquad\square$

The results in the above Theorems hinge on an RIP criteria involving the combined matrix **CB**, which can be interpreted in the context of an overcomplete dictionary in the classical

compressed sensing literature [192, 193]. In this sense, the **B** matrix (i.e., the connectivity of the afferent stage in Figure 5.1) can be interpreted as a dictionary that transforms the sparse input **u** onto the recurrent network states. The output matrix **C** enabled 'readout' of the states. Thus, the properties of **C** and **B** individually do not have direct bearing on dynamic CS. However, since these matrices dictate the signal flow to and from the recurrent stage, dynamic CS depends on whether these two matrices are matched in the sense of satisfying the RIP condition presented above.

### 5.4.3   Proper Reduction to Sequential CS

In the previous section, we focused on the case where the matrix **C** precludes direct inversion of the system states from the observations. It is worth noting that the results remain valid even without this assumption and, in fact, reduce to the case of sequential application of classical compressed sensing.

This reduction is most easily seen by assuming that **C** is square and full rank, so that the network state can be obtained directly via inversion at each instant in time. It is intuitive that in this case, the cost function in (5.7) is minimized by application of standard CS at each time instant, depending exclusively on the RIP properties of the matrix **B**. Specifically, under the assumption that **C** is full rank, Theorem 5.1 reduces exactly to:

**Theorem 5.3.**  *Assume that the matrix* **B** *satisfies the RIP condition (5.9) with* $\delta_{2s} < \sqrt{2}-1$. *Suppose that the sequence* $(\mathbf{y}_k)_{k=0}^{K}$ *is given and generated from sequences* $(\bar{\mathbf{r}}_k)_{k=0}^{K}$ *and s-sparse* $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ *based on (5.14), where* $(\|\mathbf{e}_k\|_{\ell_2} \leq \epsilon)_{k=0}^{K}$ *and* **A**, **B**, **C**, $\mathcal{N}(\mathbf{C}) = \{0\}$ *are known. Then,*

*the solution to Problem (P2) obeys*

$$\sum_{k=0}^{K-1} \|\mathbf{u}_k^* - \bar{\mathbf{u}}_k\|_{\ell_2} \le C_s \epsilon, \tag{5.23}$$

*where*

$$C_s = 2\alpha C_0 K (1 - \rho)^{-1}. \tag{5.24}$$

$C_0$, $\rho$, $\alpha$ *are given explicitly below:*

$$
\begin{aligned}
C_0 &= \frac{1}{\sqrt{\sigma}} \left( 1 + \sqrt{\frac{\sigma_{max}\left(\mathbf{C}^T\mathbf{C}\right) \sigma_{max}\left(\mathbf{A}^T\mathbf{A}\right)}{\sigma_{min}\left(\mathbf{C}^T\mathbf{C}\right)}} \right), \\
\alpha &= \frac{2\sqrt{1 + \delta_{2s}}}{1 - \delta_{2s}}, \\
\rho &= \frac{\sqrt{2}\delta_{2s}}{1 - \delta_{2s}}, \\
\sigma_{min}&\left(\mathbf{C}^T\mathbf{C}\right) < \sigma < \sigma_{max}\left(\mathbf{C}^T\mathbf{C}\right).
\end{aligned}
\tag{5.25}
$$

*Proof.* This Theorem is a special case of Theorem 5.1 and the proof can be obtained readily by following the proof of Theorem 5.1. □

An analogous reduction for exact recovery in the noiseless case ($\epsilon = 0$) can also be made.

## 5.4.4   Recovery in the Presence of Disturbance and Noise

Finally, we show how eigenstructure of the network implies a fundamental trade-off between stable recovery and rejection of disturbance (i.e., corruption).

Here, for ease of presentation, we will use of the results of Theorem 5.3, though the concepts generalize. It is easy to see from (5.23) that the upper-bound of the recovery error is reduced

by decreasing the maximum singular value of $\mathbf{A}$. Thus, we can use the upper-bound of the input recovery error as a comparative measure of performance. In the absence of both disturbance and noise, the best error performance is achieved when $\mathbf{A} = \mathbf{0}$, i.e., the network is static, which is intuitive since in this scenario any temporal effects would smear potentially salient parts of the signal.

On the other hand, having dynamics in the network should improve the error performance in the presence of the disturbance. To demonstrate this, consider (5.3) with $\mathbf{d_k}$ nonzero. When $\mathbf{A} = \mathbf{0}$, i.e., a static network, the disturbance can be exactly transformed to the measurement equation resulting in $\mathbf{Cd}_k + \mathbf{e}_k$ as a surrogate measurement noise with

$$
\|\mathbf{Cd}_k + \mathbf{e}_k\|_{\ell_2} \leq \sqrt{\sigma'}\epsilon' + \epsilon,
$$
$$
\sigma_{min}\left(\mathbf{C}^T\mathbf{C}\right) < \sigma' < \sigma_{max}\left(\mathbf{C}^T\mathbf{C}\right).
$$

(5.26)

In this case, the error upper-bound can be obtained by exploiting the result of Theorem 5.3 as

$$
\sum_{k=0}^{K-1} \|\mathbf{u}_k^* - \bar{\mathbf{u}}_k\|_{\ell_2} \leq C_s'(\sqrt{\sigma'}\epsilon' + \epsilon),
$$
$$
C_s' = \frac{2}{\sqrt{\sigma}}\alpha K(1 - \rho)^{-1}.
$$

(5.27)

When $\mathbf{A}$ is nonzero, it is not possible to exactly map the disturbance to the output as above. Nevertheless, it is straightforward to approximate the relative improvement in performance. For instance, consider a system with a symmetric $\mathbf{A}$ where the disturbance and input are in displaced frequency bands. Then it is a direct consequence of linear filtering that the power spectral density of the disturbance can be attenuated according to

$$
\mathcal{S}_{\mathbf{d}^{\text{Filt}}}(e^{j\omega}) = s_{\mathbf{d}}(e^{j\omega})(e^{j\omega}\mathbf{I}_n - \mathbf{A})^{-1}(e^{-jw}\mathbf{I}_n - \mathbf{A})^{-1},
$$

(5.28)

where $\omega$ is the frequency of the disturbance. So, for instance, if $\omega = \pi$,

$$\begin{aligned}
\text{Tr}\{\mathcal{S}_{\mathbf{d}\text{Filt}}(e^{j\pi})\} &= \text{Tr}\{s_{\mathbf{d}}(e^{j\pi})(\mathbf{I}_n + \mathbf{A})^{-2}\} \\
&= s_{\mathbf{d}}(e^{j\pi})\sum_{i=1}^{n}(1 + \lambda_i(\mathbf{A}))^{-2},
\end{aligned} \tag{5.29}$$

where $\lambda_i(\mathbf{A})$ is the $i^{th}$ eigenvalue of matrix $\mathbf{A}$. Assuming the input is sufficiently displaced in frequency from the disturbance, the error upper-bound can be then readily approximated using the results of Theorem 5.3 as follows

$$\sum_{k=0}^{K-1}\|\mathbf{u}_k^* - \bar{\mathbf{u}}_k\|_{\ell_2} \leq C_s\left(\frac{\sqrt{\sigma''}}{n}\sum_{i=1}^{n}(1 + \lambda_i(\mathbf{A}))^{-2}\epsilon' + \epsilon\right), \tag{5.30}$$

$$\sigma_{min}\left(\mathbf{C}^T\mathbf{C}\right) < \sigma'' < \sigma_{max}\left(\mathbf{C}^T\mathbf{C}\right).$$

By comparing (5.27) and (5.30), it is evident that the presence of dynamics in $\mathbf{A}$ can reduce the error upper-bound by at least a factor of two, assuming $\sigma'$ and $\sigma''$ are close to each other. In the examples below, we will show that, in fact, performance in many cases can exceed this bound considerably.

## 5.5   Examples

In this section, we present two brief examples that demonstrate the key aspects of the results. For solving our convex optimization problems, we used CVX with a MATLAB interface [194, 195]. To speed up the computations, alternative $\ell_1$ minimization algorithms could potentially be utilized [196, 197], though we do not explore these herein.
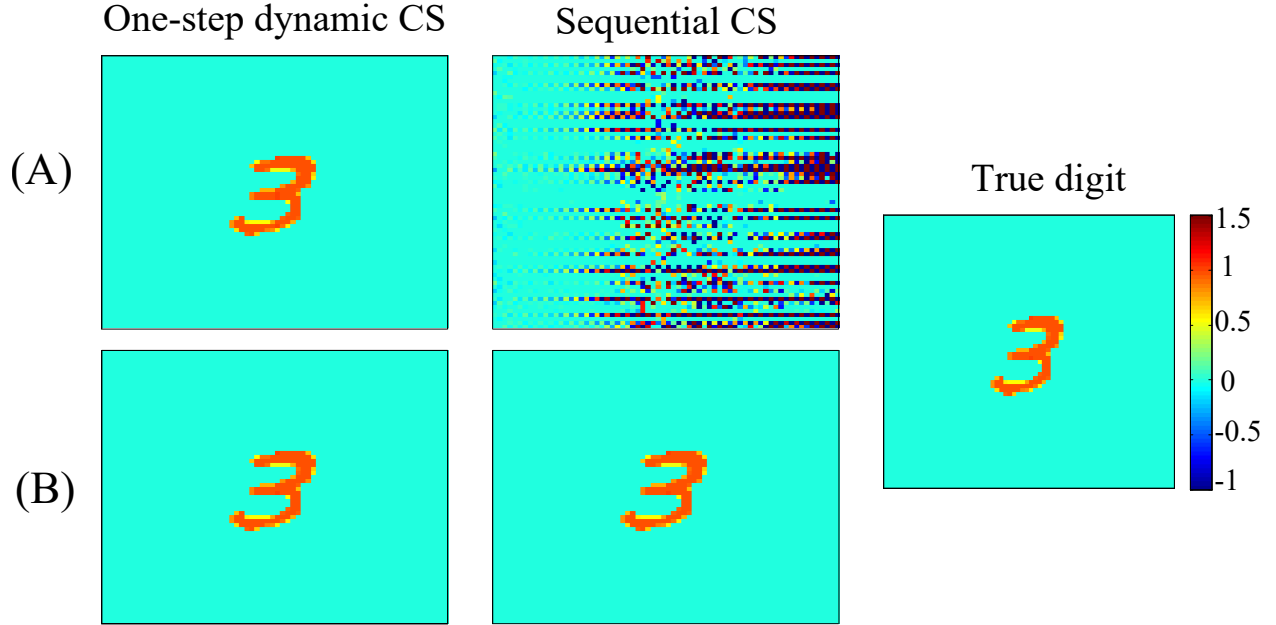
Figure 5.3: The recovered input for (A) $p = 35$ and (B) $p = 45$ for both static (middle images) and dynamic (left images) CS where $n = 45$ and $m = 68$. Original input is in the right hand side denoted as true digit.

## 5.5.1  One-step and Sequential Recovery in Recurrent Networks

In this example, we construct a recurrent network of the form in Figure 5.1, which satisfies conditions in Theorem 5.1. Specifically, we consider a random network so that the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are constructed with entries that are drawn from an independent, identically distributed normal distribution. The dimensionality considered is $n = 45$, $m = 68$ , $p = 35$.

The sparse input is defined as the image of a digit, shown in Figure 5.3 where each pixel takes the value of 0 or 1. Here, the horizontal axis is treated as time, i.e., column $k$ of the image is the input to the system at time $k$ with $K = 68$.

Since $p < n$, it is clear that sequential application of static CS is not possible in this case (i.e., since $\mathbf{C}$ cannot be inverted). However, for this network setup, we verify numerically

that the conditions of Theorem 5.2 hold generically, and thus recovery of the sparse input is expected after $K$ time steps by solving ($P1$) in one step over the entire time horizon, i.e., one-step dynamic CS. Figure 5.3A confirms this result, wherein exact recovery of the input is shown. Moreover, we confirm that sequential application of CS does not lead to correct recovery.

As a point of comparison, we also considered the case wherein $\mathbf{C}$ is full rank (i.e., $p = 45$) (Figure 5.3B), in which case sequential application of CS does indeed lead to recovery.

## 5.5.2 Recovery in the Presence of Disturbance and Noise

In this example we highlight the competing influence of disturbance and noise in a recurrent setting, and specifically our derived error bounds. We construct networks of the same dimension as in Example 1, but for simplicity, now assume that $\mathbf{C}$ is full rank.

We begin by considering the effects of noise only, wherein $\mathbf{e}_k$ is assumed to follow an uniform distribution $\mathcal{U}(-0.5, 0.5)$ while $\mathbf{d_k} = \mathbf{0}$. Figure 5.4A shows the mean square error ($MSE$) versus the maximum singular value of $\mathbf{A}$, for several random realization of $\mathbf{A}$. It can be seen from this figure that by increasing $\sqrt{\sigma_{max}(\mathbf{A}^T \mathbf{A})}$, the recovery performance is degraded, as we expect based on the derived bound for the error in (5.23).

However, based on our discussion in Section 5.4.4, we also know that dynamics embedded in $\mathbf{A}$ are beneficial for the rejection of disturbances. Thus, to contrast Figure 5.4A, we consider the case when disturbance is added to the input. In Figure 5.4B, we show the $MSE$ versus $\sqrt{\sigma_{max}(\mathbf{A}^T \mathbf{A})}$ for several random $\mathbf{A}$ when $\mathbf{e}_k \sim \mathcal{U}(-0.5, 0.5)$ and $\mathbf{d}_k \sim \mathcal{N}(0, 1)$. It is seen that increasing $\sqrt{\sigma_{max}(\mathbf{A}^T \mathbf{A})}$ allows disturbance rejection and improves performance.

(A)

(B)
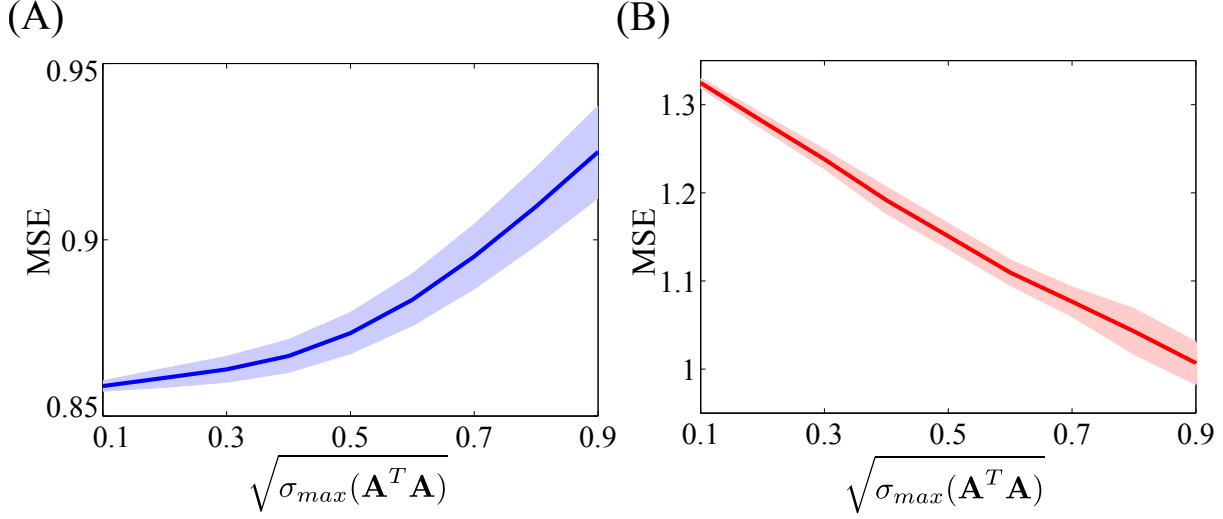
Figure 5.4: (A) *MSE* versus the maximum singular value of $\mathbf{A}$, for several random realization of $\mathbf{A}$ with noise and in the absence of disturbance. (B) *MSE* versus the maximum singular value of $\mathbf{A}$, for several random diagonal $\mathbf{A}$ with noise and disturbance.

Taken together, Figure 5.4 emphasizes the basic trade-off in recovery performance due to noise and disturbance. Thus, in the presence of both sources of corruption, an effective network must exhibit a relatively balanced spectrum.

To provide an intuitive visual demonstration of this principle, we construct an input consisting of a smoothly changing sequence of 64 frames (images). Each frame, is corrupted with disturbance at each time, and the difference between two consecutive frames is considered as the sparse input to the network. The disturbance $\mathbf{d}_k$ is assumed to be a random variable drawn from a Gaussian distribution, $\mathcal{N}(0, 0.2)$, passed through a fifth-order Chebyshev high pass filter. Each frame has $m = 400$ pixels and $K = 64$. This input is passed through a random network of dimension $n = p = 200$, i.e. random Gaussian matrices for $\mathbf{B}$ and $\mathbf{C}$.

We designed the matrix $\mathbf{A}$ to balance the performance bound (5.23) and the ability to reject the disturbance as per Section 5.4.4. Figure 5.5A shows the original frames at different times. We assumed that the first frame is known exactly. Frames recovered from the output
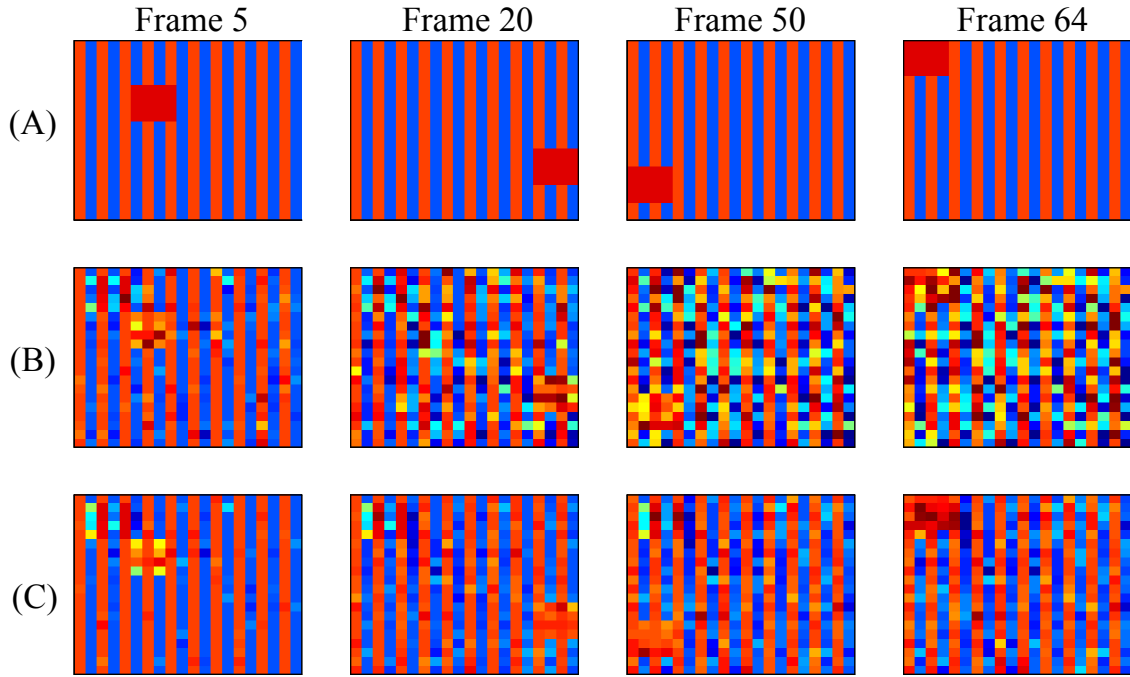
Figure 5.5: (A) Four noiseless frames of a movie. Recovery via (B) static CS and (C) dynamic CS in the presence of disturbance.
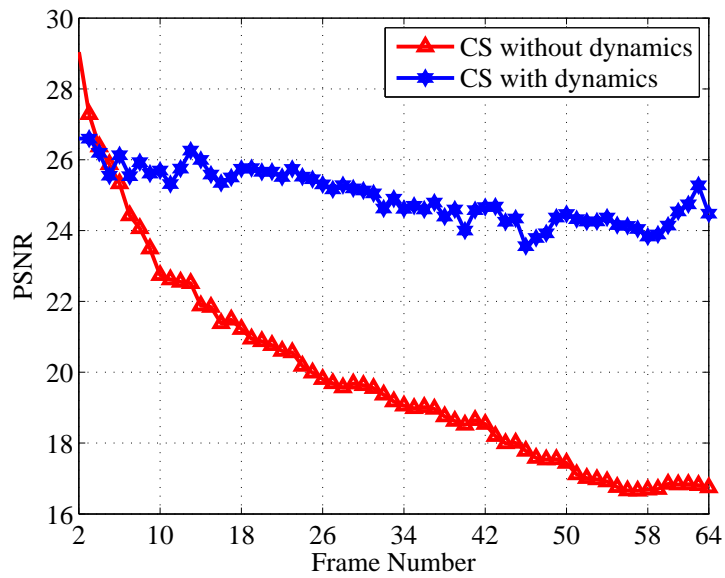


Figure 5.6: PSNR of the recovered frame versus frame number for both CS with and without dynamics.
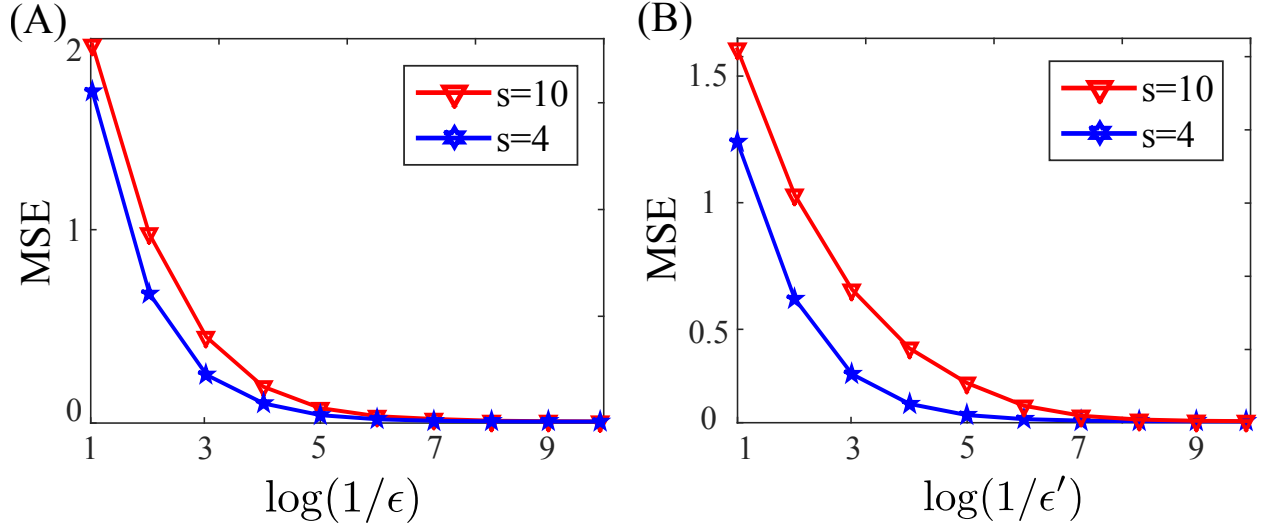
Figure 5.7: *MSE* as a function of (A) $log(1/\epsilon)$ and (B) $log(1/\epsilon')$ for the reconstructed input with $n = 50$, $m = 100$, $p = 40$, evaluated over 10 time steps for 100 random trials (different random matrices **A**, **B**, **C** in each trial).

of a static network , i.e., $\mathbf{A} = \mathbf{0}$ are depicted in Figure 5.5B. In contrast, frames recovered from the output of the designed dynamic network are shown in Figure 5.5C. It is clear from the figure that the quality of recovery is better in the latter case. Figure 5.6 illustrates the PSNR, defined as $10\log(\frac{1}{MSE})$ as a function of frame number with and without dynamics. It can be concluded from this figure that having recurrent dynamics embedded in **A** results in recovery that is more robust to disturbance, while without dynamics, error propagates over time, and the reconstruction quality is degraded.

We conducted simulation experiments to examine the effect of the noise and disturbance strength on the reconstruction error when **C** is not full. Figure 5.7 shows the average *MSE* for the reconstructed input versus $log(1/\epsilon)$ and $log(1/\epsilon')$, respectively with $n = 50$, $m = 100$, $p = 40$, evaluated over 10 time steps for 100 random trials (different random matrices **A**, **B**, **C** in each trial) with 4-sparse and 10-sparse inputs. This figure shows that the reconstruction error decreases as a function of noise energy.

### 5.5.3 Input Recovery in an Overactuated Rate-Based Neuronal Network

A fundamental question in theoretical neuroscience centers on how the architecture of brain networks enables the encoding/decoding of sensory information [6, 198, 199]. In our final example, we use the results of Theorems 5.3 and 5.2 to highlight how certain structural and dynamical features of neuronal networks may provide the substrate for sparse input decoding. Specifically, we consider a firing rate-based neuronal network [175] of the form

$$\mathbf{T_r}\frac{d\mathbf{r}}{dt} = -\mathbf{r} + \mathbf{Wu} + \mathbf{Mr}, \tag{5.31}$$

with input rates $\mathbf{u} \in \mathbb{R}^m$, output rates $\mathbf{r} \in \mathbb{R}^n$, a feed-forward synaptic weight matrix $\mathbf{W} \in \mathbb{R}^{n \times m}$, and a recurrent synaptic weight matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$. We consider $n = 50$ neurons which receive synaptic inputs from $m = 100$ afferent neurons, i.e., neurons that impinge on the network in question. Here, $\mathbf{T_r} \in \mathbb{S}_+^n$ is a diagonal matrix whose diagonal elements are the time constants of the neurons. A discrete version of (5.31), alongside a linear measurement equation can be written in the standard form (5.3) where $\mathbf{A} = \mathbf{I}_n - \Delta t \mathbf{T_r}^{-1} + \Delta t \mathbf{T_r}^{-1}\mathbf{M}$ is related to connections between nodes in the network, and $\mathbf{B} = \Delta t \mathbf{T_r}^{-1}\mathbf{W}$ contains weights between input and output nodes. For this example, we assume that the network connectivity has a Watts–Strogatz small-world topology [200] with connection probability $p_M$ and rewiring probability $q_M$.
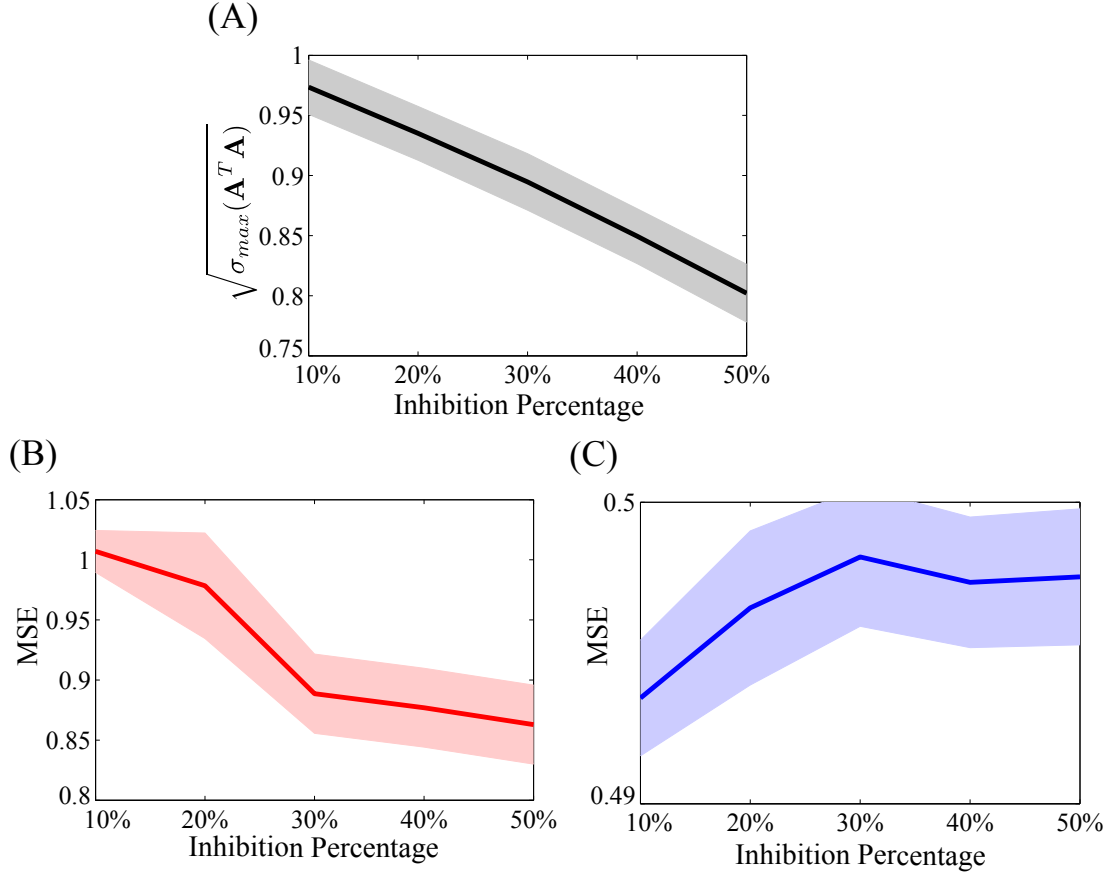
Figure 5.8: (A) The maximum singular value of **A** versus the inhibition percentage. *MSE* of input recovery in the presence of (B) Noise and (C) Disturbance as a function of the percent of inhibitory neurons.

The recurrent synaptic matrix **M** is defined as

$$
(\mathbf{M})_{ij} = \begin{cases} +m_{ij}^E & \text{if recurrent neuron } j \text{ is excitatory} \\ 0 & \text{if no connection from neuron } j \text{ to } i \\ -m_{ij}^I & \text{if recurrent neuron } j \text{ is inhibitory} \end{cases} \tag{5.32}
$$

For the purposes of illustration, we select the diagonal elements of matrix $\mathbf{T}_r$, from a uniform distribution $\mathcal{U}(0.1, 0.2)$. We study the recovery performance associated with the network over 100 time steps, assuming a timescale of milliseconds and an discretization step of 0.1 $ms$.

139

At each time step, the nonzero elements of the input vector $\mathbf{u}$, i.e., firing rate of the afferent neurons, are drawn from an uniform distribution $\mathcal{U}(0.5, 1.5)$. Moreover, we assume that elements of the observation matrix $\mathbf{C}$ are drawn from a Gaussian distribution $\mathcal{N}(0, 1)$. Finally, we assume $m_{ij}^E$ and $m_{ij}^I$ are drawn from lognormal distributions $\ln \mathcal{N}(0, 1)$ and $\ln \mathcal{N}(0, 0.1)$, respectively. The latter assumption is chosen for illustration only and is not related to known physiology.

**Recovery Performance from Error Bounds:** We proceed to conduct a Monte Carlo simulation of 100 different realizations of $\mathbf{W}$, $\mathbf{M}$ and $\mathbf{C}$. Fig. 5.8A illustrates that the maximum singular value of the matrix $\mathbf{A}$ decreases as a function of the percent of inhibitory neurons. Thus, we anticipate from our derived performance bounds that performance in terms of mean square error (MSE) should be best for networks with high inhibition in the presence of noise. This prediction bears out in Fig. 5.8B, where we indeed observe a monotone relationship between MSE and inhibition. On the one hand, low-inhibition is favorable for facilitating recovery in the presence of disturbance depicted in Fig. 5.8C. Such tradeoffs are interesting to contemplate when considering the functional advantages of network architectures observed in biology, such as the pervasive 80-20 ratio of excitatory to inhibitory neurons [175, 201]. Together, Figs. 5.8B and 5.8C illustrate how the excitatory-inhibitory ratio mediate a basic tradeoff in the capabilities of a rate-based neuronal network.

**Recoverable Sparsity based on Theorem 5.2:** Having ascertained the performance tradeoff curves, we sought to characterize in more detail the level of recoverable sparsity with specific connection to Theorem 5.2 and (5.13). We considered networks as above, but with $p = 30$ and 20/80 for the ratio of inhibitory/excitatory over 10 time steps for 100 random trials. Thus, the output of the network is of lower dimension than the network state
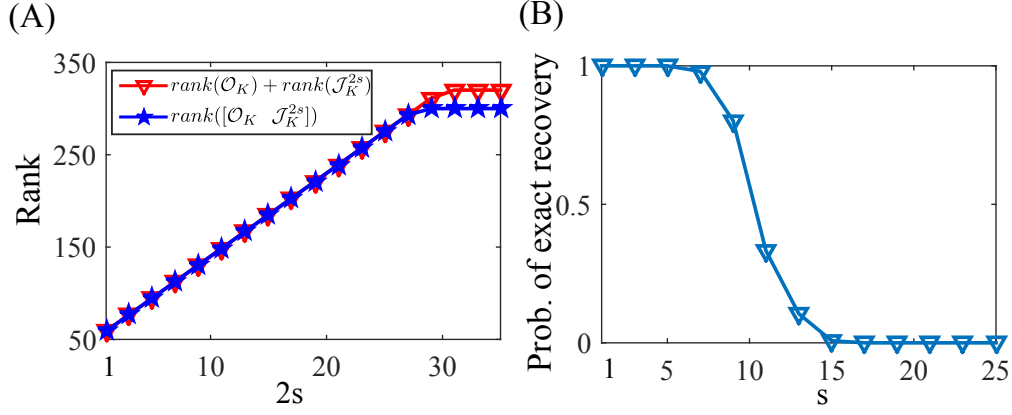
Figure 5.9: (A) Examining the rank condition (5.13) in the networks of Example 3, with $p = 30$, for different values of $s$. (B) The probability of exact recovery of dynamic sparse input to the network over $s$.

space and the observability matrix is of nontrivial construction. Fig. 5.9A shows that for this setup, the rank condition (5.13) holds up to $2s = 27$. Thus, Theorem 5.2 predicts that recover will be possible (to within the RIP condition on **CB**) for signals with 13 nonzero elements. Fig. 5.9B validates this theoretical prediction by illustrating recovery performance in the absence of disturbance and noise for different values $s$. It is observed that when the rank condition holds, reconstruction is perfect and that the probability of exact recovery is decreased by increasing $s$, as expected.

## 5.6   Discussion

In this chapter, we presented analytical conditions that are sufficient to ensure the recovery of time-varying signals processed via a recurrent linear network, wherein the network output is allowed to accrue prior to recovery. The conditions amount to a generalized observability criteria and are less stringent than the approach of performing sparse recovery pointwise, i.e., at each instant in time. Further, we derived several bounds in terms of the system dynamics

for recovery performance in the presence of both input disturbance and observation noise. We showed that dynamics can play opposing roles in mediating accurate recovery with respect to these two different sources of corruption.

**Potential Applicability:** The analysis developed herein is suited to the treatment of multivariate time-series that are modeled as the output of a network with linear dynamics. The key results are specific to contexts where extrinsic perturbations to such networks are sparse and time-varying. Both engineering and scientific impacts of these results can be postulated. In the former case, the goal may be to *design* networks to process time-varying signals that are naturally sparse, such as high-dimensional neural data. In the latter (scientific) case, the goal is to understand how the naturally occurring architectures of networks, such as those in the brain, confer advantages for processing of afferent signals. Such a paradigm may be useful in the theoretical formulation of sensory neural coding mechanisms [8]. Other application domains such as video processing [162], medical imaging [202], and resilient cyber-physical systems [160,161] may also be considered. In either the scientific or application-driven cases, a precursor to further study are a set of verifiable conditions that overtly link network characteristics/dynamics to sparse input processing, i.e., the main thrust of the present work.

**Assumptions and Limitations:** A key assumption in instantiating our analysis results is that the network dynamical matrices ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$) are known and/or specified. As well, reconstruction of inputs requires knowledge of noise bounds (i.e., upper-bounds on the magnitude of both disturbance, $\epsilon$, and noise, $\epsilon'$, as per Section 5.4). In practice, several data-driven algorithms, such as expectation maximization [49, 50], could be used in the estimation of such bounds, though we did not pursue this here. Importantly, there is no need to know the actual noise time-series in absolute terms, which would have restricted functionality.

142

Treatment of systems with nonlinear dynamics, including a more detailed examination of networks with prescribed topologies (using the developed theory), are left as subjects for future work.

# Chapter 6

# Lightweight Encoding in Recurrent Networks with Soft-Thresholding Nonlinearities

## 6.1  Highlights

A long-standing and influential hypothesis in sensory processing is that early sensory networks adapt themselves to produce efficient codes of afferent inputs. Here, we show how a nonlinear recurrent network provides an optimal solution for the efficient coding of an afferent input and its history. We specifically consider the problem of producing lightweight codes, ones that minimize both $\ell_1$ and $\ell_2$ constraints on sparsity and energy, respectively. When embedded in a linear coding paradigm, this problem results in a non-smooth convex optimization problem. We employ a proximal gradient descent technique to develop the solution, showing that the optimal code is realized through a recurrent network endowed with a nonlinear soft thresholding operator.

The training of the network connection weights is readily achieved through gradient-based local learning. If such learning is assumed to occur on a slower time-scale than the (faster) recurrent dynamics, then the network as a whole converges to an optimal set of codes and weights via what is, in effect, an alternative minimization procedure. Our results show how the addition of thresholding nonlinearities to a recurrent networks may enable the production of lightweight, history-sensitive encoding schemes.

## 6.2   Introduction

It has been hypothesized that the early stages of sensory processing have evolved to accurately encode environmental signals with minimal consumption of biological resources [6, 201, 203–206]. There are several studies on extracting salient input features, generally known as efficient coding, to explain a variety of observed properties of sensory systems [140, 207–213]. Moreover, sparse coding [214–216] is considered to be a plausible model of the visual cortex [216–218], suggesting that sensory information is encoded using a small number of active neurons at any given point in time. However, it is still far from clear how sensory neurons perform efficient encoding.

In this chapter, we develop dynamical networks that produce efficient codes by adapting (learning) their connection weights in response to properties of the input. The developed learning rules are local, ones that use only information currently available at the node or connection that is to be modified, and we incorporate efficient encoding to achieve representation patterns that are sparse, and consume minimal energy. Also, we will show how memory, the ability of the network to reconstruct previous inputs from a current representation, is related to the number of active neurons in the network. These algorithms are biologically plausible

learning rule which might be helpful in explaining the storage of memory in brain networks for sensory information processing.

The remainder of the chapter is organized as follows. In Section 6.3 we formulate the problem in detail. In Section 6.4 we introduce soft thresholding recurrent neural networks and provide learning rules to train such networks. Simulation results for several different scenarios are provided in Section 6.5. Finally, conclusions are formulated in Section 6.6.

## 6.3  Problem Statement

The goal of minimal energy sparse coding is to efficiently represent time-varying input vectors approximately as a weighted linear combination of a small number of unknown basis vectors. These basis vectors capture high-level salient structure in the input data. We consider an $m$-dimensional input signal, denoted as $\mathbf{x}$, which is represented using basis vectors $\mathbf{d}_1, \cdots, \mathbf{d}_n \in \mathbb{R}^m$ and a sparse vector of weights or firing rates $\mathbf{r} \in \mathbb{R}^n$ as

$$\mathbf{x} \simeq \hat{\mathbf{x}} = \mathbf{D}\mathbf{r}. \tag{6.1}$$

We specifically seek lightweight representations so that $\mathbf{r}$ is minimal with respect to both $\ell_1$ and $\ell_2$ criteria. First, we assume that the basis vectors (columns of matrix $\mathbf{D}$) are known. Thus, we consider the optimization of the following objective

$$J(\mathbf{r}) = \frac{1}{2} \left\| \mathbf{x} - \mathbf{D}\mathbf{r} \right\|_2^2 + \frac{\lambda_2}{2} \left\| \mathbf{r} \right\|_2^2 + \lambda_1 \left\| \mathbf{r} \right\|_1 \tag{6.2}$$

with respect to $\mathbf{r}$. Here, the first term is the data fidelity term, the error between the reconstructed input signal and the actual stimulus, the second term corresponds to the cost

146

of the signal representation, and the last term controls the sparsity level of the representation. The objective in (6.2) is a nonsmooth convex function of $\mathbf{r}$. The proximal gradient method [221–223] has recently been shown to be effective for finding the optimal solution for this class of objectives. Here, we use it to solve (6.2) and demonstrate the solution can be realized through a two-layer recurrent nonlinear network. In this method, we split the objective into two terms, one of which is differentiable. This splitting is not unique, so different splittings lead to different implementations of the proximal gradient method for the same original problem.

## 6.4 Soft-Thresholding Recurrent Neural Networks

Here, we show how a soft-thresholding recurrent neural network (STRNN) can generate a solution to the above class of optimization problems. Specifically, we note that the cost function $J(\mathbf{r})$ can be written as

$$J(\mathbf{r}) = f(\mathbf{r}) + g(\mathbf{r}), \tag{6.3}$$

where

$$
\begin{aligned}
f(\mathbf{r}) &= \frac{1}{2} \left\| \mathbf{x} - \mathbf{Dr} \right\|_2^2 + \frac{\lambda_2}{2} \left\| \mathbf{r} \right\|_2^2, \\
g(\mathbf{r}) &= \lambda_1 \left\| \mathbf{r} \right\|_1.
\end{aligned}
\tag{6.4}
$$

Further, $f$ and $g$ are closed proper convex and $f$ is differentiable. The proximal gradient method results in the following dynamic equation for $\mathbf{r}$:

$$
\begin{aligned}
\mathbf{r}^{k+1} &= \mathbf{prox}_{\mu g} \left( r^k - \mu \nabla f(\mathbf{r}^k) \right) \\
\nabla f(\mathbf{r}) &= -\mathbf{D}^\mathsf{T} \mathbf{x} + \mathbf{D}^\mathsf{T} \mathbf{Dr} + \lambda_2 \mathbf{r},
\end{aligned}
\tag{6.5}
$$

147

where the discrete step size $\mu$ is the regularization parameter of the rates.[7] At each iteration $k$, this discrete dynamical system can be interpreted as a discrete linear neural network with a "leak" $(-\lambda_2 \mathbf{r})$, feedforward connections $(\mathbf{W} = \mathbf{D}^\mathsf{T})$, and recurrent connections $(\mathbf{M} = \mathbf{D}^\mathsf{T}\mathbf{D})$, followed by a nonlinear proximal operator defined as

$$\mathbf{prox}_{\mu g}(\mathbf{r}) = \arg\min_{\mathbf{v}} \left( g(\mathbf{v}) + \frac{1}{2\mu} \|\mathbf{v} - \mathbf{r}\|_2^2 \right), \tag{6.6}$$

which is an slantwise operation on the elements of the vector $\mathbf{r} = [r_1, \cdots, r_n]^\mathsf{T}$. When $\lambda_1 = 0$, there is no $\ell_1$ constraint in the objective function, $g(\mathbf{r}) = 0$, which results in $\mathbf{prox}_{\mu g}(\mathbf{r}) = \mathbf{r}$. In this scenario, the solution for optimal firing rates matches exactly with a linear neural network. For the case that $\lambda_1 \neq 0$, the optimality condition for this minimization problem considering $g(\mathbf{r}) = \lambda_1 \|\mathbf{r}\|_1$ leads to the following solution for the proximal operator:

$$\mathbf{prox}_{\mu g}(r_i) = \begin{cases} r_i - \mu\lambda_1 & \text{if } r_i > \mu\lambda_1 \\ 0 & \text{if } |r_i| \leq \mu\lambda_1 \\ r_i + \mu\lambda_1 & \text{if } r_i < -\mu\lambda_1 \end{cases} . \tag{6.7}$$

It is easy to verify that the above proximal operator is soft thresholding. Having a soft thresholding operator applied to a linear neural network results in keeping just high values of firing rates at each time instant.

---

[7]When $\nabla \mathbf{f}$ is Lipschitz continuous with constant $L$, this method can be shown to converge with rate $\mathcal{O}(1/k)$ when $\mu \in (0, 1/L]$ is used.

## 6.4.1 Short Term Memory in STRNNs

In this section, we want to characterize the capability of SRTNNs to preserve the input history in a network. In order to train a network to be able to reconstruct the input history from the present state of the network, we need to modify the objective function in (6.2) to incorporate this ability. To do so, we first assume that along with $\mathbf{D}$, which can be thought of as a decoder to recover the current input $\mathbf{x}(t)$ from the current activity of the network $\mathbf{r}(t)$, there also exists an $\mathbf{S}$ such that $\mathbf{DS}^q$ can be considered as the decoder to reconstruct the past input $\mathbf{x}(t-q)$ from $\mathbf{r}(t)$, where

$$\mathbf{x}(t-q) \simeq \mathbf{DS}^q\mathbf{r}(t). \tag{6.8}$$

We modify (6.2) to incorporate history reconstruction ability as follows:

$$J_s(\mathbf{r}(t)) = J(\mathbf{r}(t)) + \frac{\lambda}{2} \|\mathbf{r}(t-1) - \mathbf{Sr}(t)\|_2^2. \tag{6.9}$$

The last term in (6.9) is the error between the true and reconstructed input history, one time step in the past. Similar to (6.3), $J_S$ can be split into

$$J_s(\mathbf{r}(t)) = f'(\mathbf{r}(t)) + g(\mathbf{r}(t)). \tag{6.10}$$

The optimal firing rate $\mathbf{r}(t)$ using proximal gradient method can be written as

$$
\begin{aligned}
\mathbf{r}^{k+1}(t) &= \mathbf{prox}_{\mu g}\left(\mathbf{r}^k(t) - \mu\nabla f'(\mathbf{r}^k(t))\right) \\
\nabla f'(\mathbf{r}(t)) &= -\mathbf{D}^\mathsf{T}\mathbf{x}(t) + \left(\mathbf{D}^\mathsf{T}\mathbf{D} + \lambda\mathbf{S}^\mathsf{T}\mathbf{S}\right)\mathbf{r}(t) - \lambda\mathbf{S}^\mathsf{T}\mathbf{r}(t-1) + \lambda_2\mathbf{r}(t) \\
&= -\mathbf{W}\mathbf{x}(t) + \mathbf{M}^d\mathbf{r}(t) - \mathbf{M}^s\mathbf{r}(t-1) + \lambda_2\mathbf{r}(t),
\end{aligned}
\tag{6.11}
$$
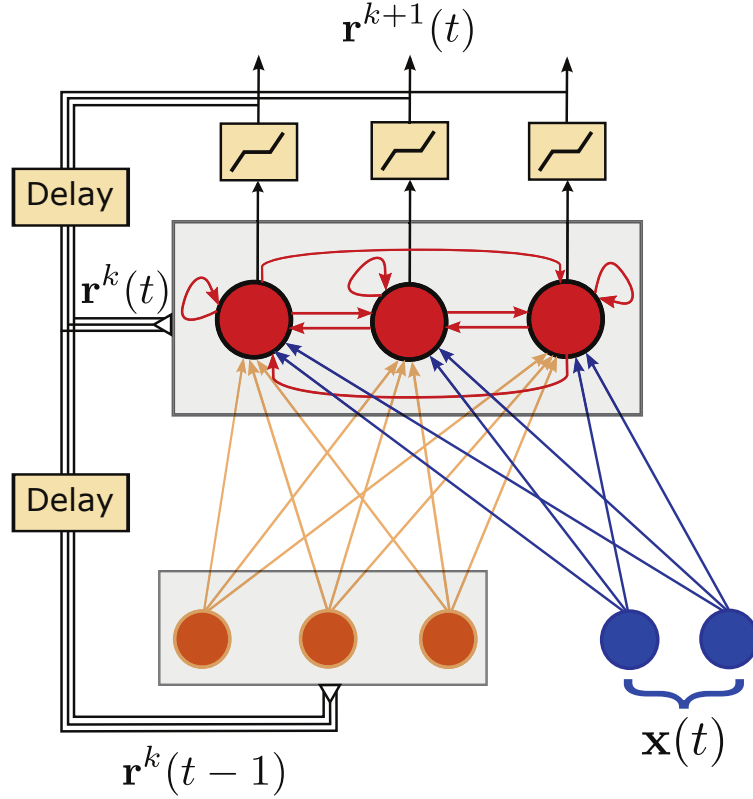
Figure 6.1: The structure of a STRNN with $m = 2$ and $n = 3$. The feedforward ($\mathbf{W}$), delayed recurrent connections ($\mathbf{M}^s$), and recurrent connections ($\mathbf{M}^d$) are shown in blue, orange and red, respectively.

where $\mathbf{M}^d$ and $\mathbf{M}^s$ contain recurrent connections in the network. The structure of a STRNN with two inputs and three network nodes is shown in Figure 6.1. The learning rules for feedforward and recurrent weight matrices are provided in the next section.

## 6.4.2 Online Learning Rules of Connections

In this section, we present the basic components of our online learning rules of connections in a STRNN. At time instant $t$, the finite training set composed of i.i.d. samples of a distribution $p(x)$ is denoted as $\mathbf{X}(t) = [\mathbf{x}(1), \cdots \mathbf{x}(t)]$. Also, the network activity over $t$

time steps is described as $\mathbf{R}(t) = [\mathbf{r}(1), \cdots .\mathbf{r}(t)]$. Learning rules for the connections in the network can be readily inferred from $\mathbf{D}$ and $\mathbf{S}$. To learn $\mathbf{D}$ and $\mathbf{S}$, we optimize the following empirical cost function at each time instant $t$:

$$
\begin{aligned}
\{\mathbf{D}(t), \mathbf{S}(t)\} &= \arg \min_{\{\mathbf{D},\mathbf{S}\}} \frac{1}{t} \sum_{i=1}^{t} J_s(\mathbf{r}(t)) \\
&= \arg \min_{\{\mathbf{D},\mathbf{S}\}} \frac{1}{t} \sum_{i=1}^{t} \frac{1}{2} \|\mathbf{x}(t) - \mathbf{D}\mathbf{r}(t)\|_2^2 + \frac{\lambda}{2} \|\mathbf{r}(t-1) - \mathbf{S}\mathbf{r}(t)\|_2^2 .
\end{aligned}
\tag{6.12}
$$

The cost function in (6.12) is separable over $\mathbf{D}$ and $\mathbf{S}$, which makes adaptation of these two variables independent of each other. It is straightforward to verify that a small change of $\mathbf{D}$ in the direction of $\Delta\mathbf{D}(t) = \tau\left(\mathbf{X}(t)\mathbf{R}(t)^\mathsf{T} - \alpha\mathbf{D}(t)\right)$ reduces the first term in (6.12). The coefficients $\alpha$ and $\tau$ control the energy of the basis vectors and the speed of the learning process, respectively. Similarly, $\Delta\mathbf{S}(t) = \tau\left(\mathbf{R}(t-1)\mathbf{R}(t)^\mathsf{T} - \alpha\mathbf{S}(t)\right)$[8] reduces the second term in (6.12). The problem with these learning rules is that storing $\mathbf{X}(t)$ and $\mathbf{R}(t)$ for large $t$ is not efficient. To remedy this issue, we define two new variables, denoted as $\mathbf{A}(t)$ and $\mathbf{B}(t)$, which are updated based on the current input and network activity. The learning rules for these variable are obtained with the assumption that the input to the network, $\mathbf{x}$, and the network activity, $\mathbf{r}$, are whitened. Even if the input is whitened the network activity is not necessarily whitened, and the choice of $\mathbf{S}$ will be dominated by second-order statistical dependencies. To resolve these dependencies, we need to take into account the covariance matrix of the network activity in the cost function. Having the update rule for $\mathbf{D}$ and $\mathbf{S}$, we can find the learning rules for the network connections. Also, different time scales can be considered for different weight matrices in the network denoted as $\tau_W$, $\tau_M^s$, and $\tau_M^d$. The leaning procedure for the network connections is summarized in Algorithm 6.1. In the

---

[8]If we want to be precise, $\mathbf{R}(t-1)\mathbf{R}(t)^\mathsf{T}$ should be replaced by $[\mathbf{0}\ \mathbf{R}(t-1)]\mathbf{R}(t)^\mathsf{T}$.

**Algorithm 6.1** Online update rules for network connections

---

**Require:** $\mathbf{x}(t), \mathbf{r}(t), \mathbf{r}(t-1), \mathbf{A}(t-1), \mathbf{B}(t-1), \mathbf{W}(t-1), \mathbf{M}^s(t-1), \mathbf{M}^d(t-1), \epsilon_a, \epsilon_b, \tau_w,$

$\tau_m^s, \tau_m^d$

1: $\mathbf{A}(t) \leftarrow \epsilon_a \mathbf{A}(t-1) + (1-\epsilon_a)\mathbf{x}(t)\mathbf{r}(t)^{\mathsf{T}}$

2: $\mathbf{B}(t) \leftarrow \epsilon_b \mathbf{B}(t-1) + (1-\epsilon_b)\mathbf{r}(t-1)\mathbf{r}(t)^{\mathsf{T}}$

$\mathbf{A}(t) = [\mathbf{a}_1(t), \cdots, \mathbf{a}_n(t)] \in \mathbb{R}^{m \times n}$

$\mathbf{B}(t) = [\mathbf{b}_1(t), \cdots, \mathbf{b}_n(t)] \in \mathbb{R}^{n \times n}$

$\mathbf{W}(t-1) = \left[\mathbf{w}_1^{\mathsf{T}}(t-1), \cdots, \mathbf{w}_n^{\mathsf{T}}(t-1)\right]^{\mathsf{T}} \in \mathbb{R}^{n \times m}$

$\mathbf{M}^s(t-1) = \left[\mathbf{m}_1^{s\mathsf{T}}(t-1), \cdots, \mathbf{m}_n^{s\mathsf{T}}(t)\right]^{\mathsf{T}} \in \mathbb{R}^{n \times n}$

$\mathbf{M}^d(t-1) = \left[\mathbf{m}_1^d(t-1), \cdots, \mathbf{m}_n^d(t-1)\right] \in \mathbb{R}^{n \times n}$

3: **for** $i = 1$ to $n$ **do**

4:      $\Delta\mathbf{w}_i(t) \leftarrow \tau_w(\mathbf{a}_i^{\mathsf{T}}(t) - \alpha\mathbf{w}_i(t-1))$

5:      $\Delta\mathbf{m}_i^s(t) \leftarrow \tau_m^s(\mathbf{b}_i^{\mathsf{T}}(t) - \alpha\mathbf{m}_i^s(t-1)\mathbf{r}(t-1)\mathbf{r}(t-1)^{\mathsf{T}})$

6:      $\Delta\mathbf{m}_i^d(t) \leftarrow \tau_m^d(\mathbf{W}\mathbf{a}_i(t) + \mathbf{M}^s\mathbf{b}_i(t) - \alpha\mathbf{m}_i^d(t-1))$

7: **end for**

8: **return** $\mathbf{W}(t), \mathbf{M}^s(t), \mathbf{M}^d(t)$

---

update rule equation for rows of $\mathbf{M}^s$, the term $\mathbf{r}(t-1)\mathbf{r}(t-1)^{\mathsf{T}}$ removes the second-order statistical dependencies.

Here, we want to show that the update rule for connection weights is local and just is related to information available to neurons on either side of the connection that is associated with that weight. The strength of the feedforward connection weight from input node $j$ to network node $i$ at time $t$ is denoted as $w_{ij}(t)$. From the update rule for feedforward weight, which is

shown in the fourth line of Algorithm 6.1, $\Delta w_{ij}(t)$ can be written as

$$\Delta w_{ij}(t) = \tau_w \left( a_{ji}(t) - \alpha w_{ij}(t) \right)$$
$$a_{ji}(t) = \epsilon_a a_{ji}(t-1) + (1 - \epsilon_a) x_j(t) r_i(t) \tag{6.13}$$

where $a_{ji}(t)$ is the $ji$-th element of $\mathbf{A}(t)$, $x_j(t)$ is the $j$-th element of $\mathbf{x}(t)$ and $r_i(t)$ is the $i$-th element of $\mathbf{r}(t)$ . For $\epsilon_a = 0$, $a_{ji} = x_j(t) r_i(t)$ while for $\epsilon_a \neq 0$ it depends on the history of input and network firing rate at input node $j$ and network node $i$, respectively as shown in Figure 6.2A for $\epsilon_a = \epsilon_b = 0$. The strength of the delayed recurrent connection weight from network node $j$ to network node $i$ at time $t$ is denoted as $m_{ij}^s(t)$. Similarly from the fifth line of Algorithm 6.1, $\Delta m_{ij}^s(t)$ can be written as

$$\Delta m_{ij}^s(t) = \tau_m^s (b_{ji}(t) - \alpha r_i^{pre}(t) r_j(t-1))$$
$$b_{ji}(t) = \epsilon_b b_{ji}(t-1) + (1 - \epsilon_b) r_j(t-1) r_i(t) \tag{6.14}$$
$$r_i^{pre}(t) = \sum_{j=1}^{n} m_{ij}^s(t-1) r_j(t-1)$$

where $b_{ji}(t)$ is the $ji$-th element of $\mathbf{B}(t)$, $r_i^{pre}(t)$ is the presynaptic firing rate of node $i$. For $\epsilon_b = 0$, $b_{ji} = r_j(t-1) r_i(t)$ while for $\epsilon_b \neq 0$ it depends on the history of network firing rate at nodes $i$ and $j$ as shown in Figure 6.2A for $\epsilon_a = \epsilon_b = 0$. Hence, the learning is just related to information available to neurons on either side of the connection. The strength of the recurrent connection weight from network node $j$ to network node $i$ at time $t$ is denoted as $m_{ij}^d(t)$. For simplicity, we show the locality of the learning rule of $m_{ij}^d(t)$ for $\epsilon_a = \epsilon_b = 0$.
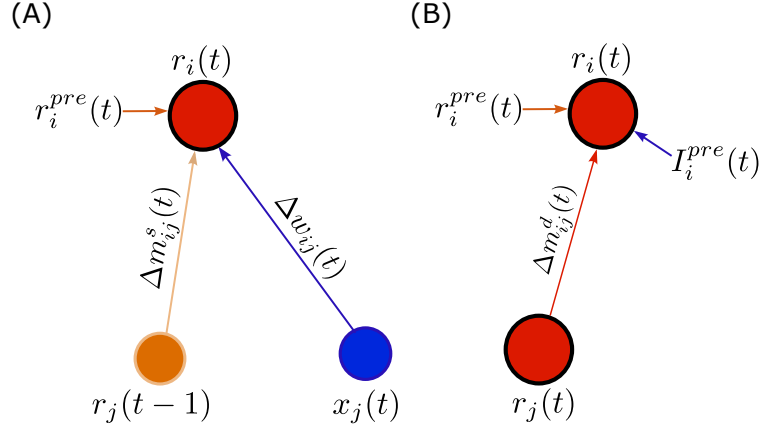
Figure 6.2: Schematics of learning rules for connection wights in a STRNN with $\epsilon_a = \epsilon_b = 0$.

From the sixth line of Algorithm 6.1, $\Delta m_{ij}^d(t)$ can be written as

$$\Delta m_{ij}^d(t) = \tau_m^d(I_i^{pre}(t)r_j(t) + r_i^{pre}(t)r_j(t) - \alpha m_{ij}^d(t-1))$$

$$I_i^{pre}(t) = \sum_{j=1}^{n} w_{ij}(t-1)x_j(t), \qquad (6.15)$$

where, $I_i^{pre}(t)$ is the presynaptic input to network node $i$. Therefore, the learning rule is local, as shown in Figure 6.2B, and related to presynaptic input, presynaptic firing rate and firing rate of a node which connection is connect to.

## 6.5    Examples

In this section, we present numerical examples to show the effectiveness of the developed soft-thresholding nonlinear network. Here, input sequences are uncorrelated Gaussian random processes with zero mean and unit variance. For the purpose of learning the STRNN, we consider $n = 15$ and a one dimensional input with 800 epochs (10000 samples per epoch, corresponding to approximately an order of magnitude time-scale separation between the
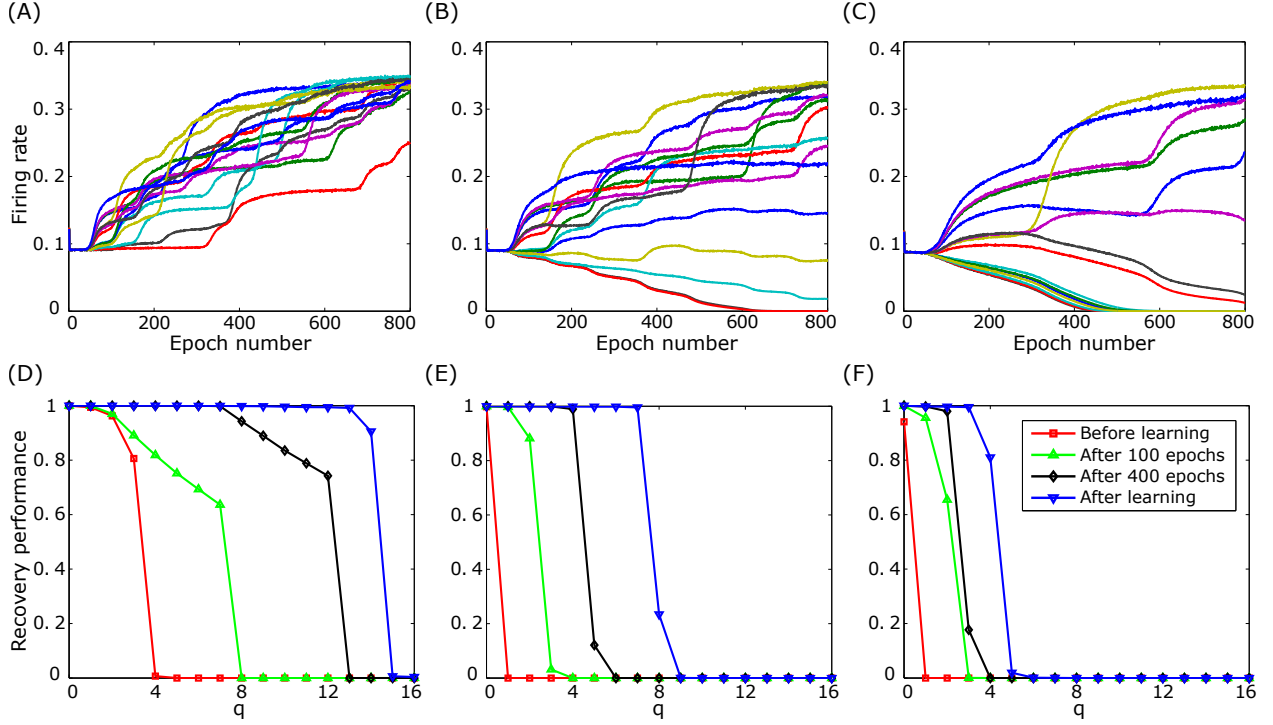
Figure 6.3: Firing rate activity of learned STRNNs over epoch number with A) $\lambda 1 = 0$, B) $\lambda 1 = 0.01$, and C) $\lambda 1 = 0.04$. Recovery performance of learned STRNNs after different number of epochs versus $q$ for D) $\lambda 1 = 0$, E) $\lambda 1 = 0.01$, and F) $\lambda 1 = 0.04$.
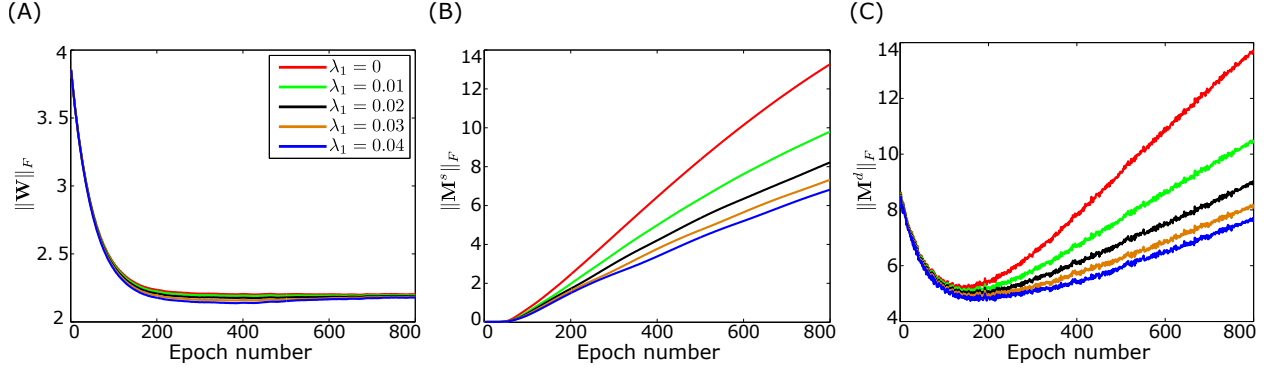


Figure 6.4: Evolution of Frobenius norm of A) feedforward, B) delayed recurrent, and C) recurrent connectivity matrices over epoch number.

learning and network dynamics). We proceed the experiment by setting $\epsilon_a = \epsilon_b = 0.1$, $\lambda = 1$, $\lambda_2 = 0.15$, $\alpha = 0.2$, $\mu = 0.1$, $\hat{\tau}_w = 1e-5$, $\hat{\tau}_m^s = 1e-5$, $\hat{\tau}_m^d = 1e-3$, zero initialization for the firing rates, and random initialization for weight matrices. Figure 6.3A-C show the
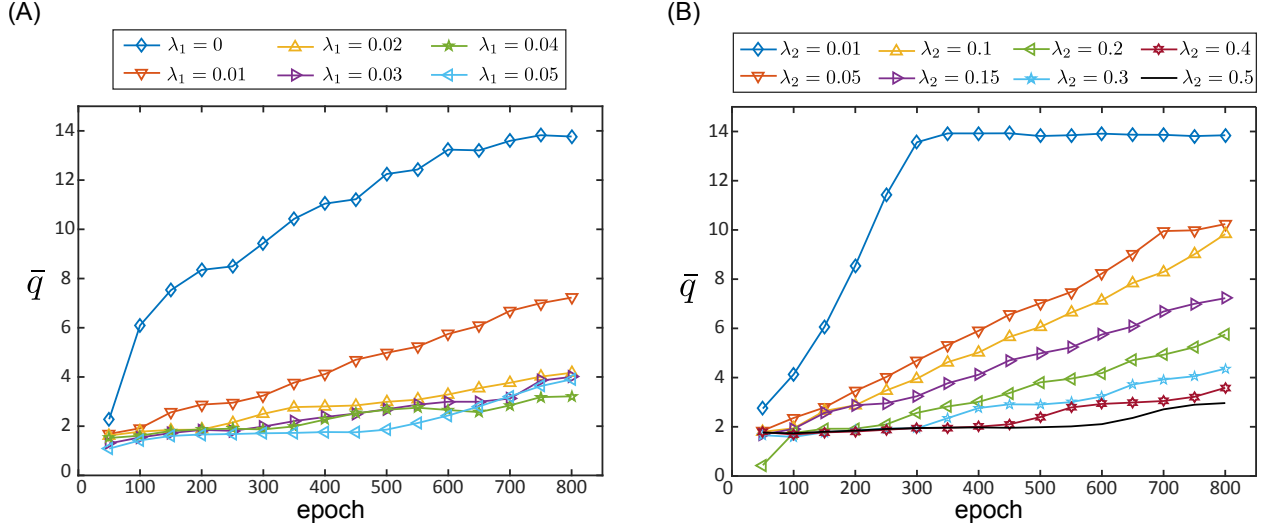
Figure 6.5: The average number of reconstructed lags as a function of epoch number for different values of (A) $\lambda_1$ and (B) $\lambda_2$. All other parameters except $\lambda_1$ and $\lambda_2$ are the same for both plots.

evolution of the firing rate activity of the network over epoch number for $\lambda 1 = 0$, $\lambda 1 = 0.01$ and $\lambda 1 = 0.04$, respectively. It is clear from these plots that different value of $\lambda_1$ causes different level of sparsity in the firing activity of the network.

For the purpose of testing the learned STRNN, the ability of the network to recover the input history from the current network activity before and after training is shown in Figure 6.3D-F. We tested the learned network using 50000 random realizations drawn from a Gaussian distribution with zero mean and unit variance. These plots also illustrate the ability of the network to reconstruct the input history over epoch number. Evolution of connections weights, Frobenius norm of connectivity matrices, over time, epoch number, is illustrated in Figure 6.4.

Figure 6.5A illustrates the average number of reconstructed lags, $\bar{q}$, as a function of epoch number for different values of $\lambda_1$. In this study $\lambda_2 = 0.15$, $\lambda = 1$, $\alpha = 0.2$, $\mu = 0.1$, $\tau_w = \hat{\tau}_w$, $\tau_m^s = \hat{\tau}_m^s$, and $\tau_m^d = \hat{\tau}_m^d$. Furthermore, $\bar{q}$ as a function of epoch number for different values

156
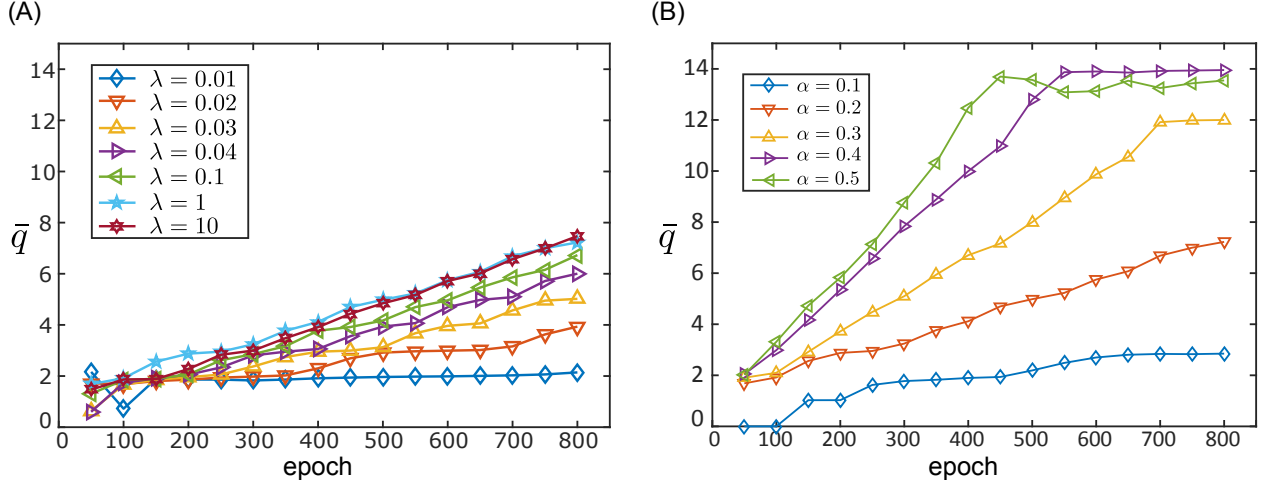
Figure 6.6: The average number of reconstructed lags as a function of epoch number for different values of (A) $\lambda$ and (B) $\alpha$. All other parameters except $\alpha$ and $\lambda$ are the same for both plots.

of $\lambda_2$ is depicted in Figure 6.5B. In this plot. $\lambda_1 = 0.01$, $\lambda = 1$, $\alpha = 0.2$, $\mu = 0.1$, $\tau_w = \hat{\tau}_w$, $\tau_m^s = \hat{\tau}_m^s$, and $\tau_m^d = \hat{\tau}_m^d$. For both of these studies, This figure represents how increasing $\lambda_2$ results in decreasing the ability to reconstruct the input history. The reason is that by fixing $\lambda$ and increasing $\lambda_2$ the optimization is forced to provide codes that are efficient mostly in terms of energy and less in terms of the history recovery.

We also studied the effect of $\alpha$ and $\lambda$ on $\bar{q}$. Figures 6.6A-B illustrate $\bar{q}$ as a function of epoch number for different values of $\lambda$ and $\alpha$, respectively. In these studies, $\lambda_1 = 0.01$, $\lambda_2 = 0.15$, $\mu = 0.1$, $\tau_w = \hat{\tau}_w$, $\tau_m^s = \hat{\tau}_m^s$, and $\tau_m^d = \hat{\tau}_m^d$.

Figure 6.7A shows $\bar{q}$ as a function of epoch number for different values of $\tau_w$ (time constant of feedforward connections) for $\lambda_1 = 0.01$. Furthermore, $\bar{q}$ as a function of epoch number for different values of $\tau_m^d$ (time constant of recurrent connections) for $\lambda_1 = 0.01$ is depicted in Figure 6.7B. In these studies, $\lambda_2 = 0.15$, $\lambda = 1$, $\alpha = 0.2$, $\mu = 0.1$, and $\tau_m^s = \hat{\tau}_m^s$.

157
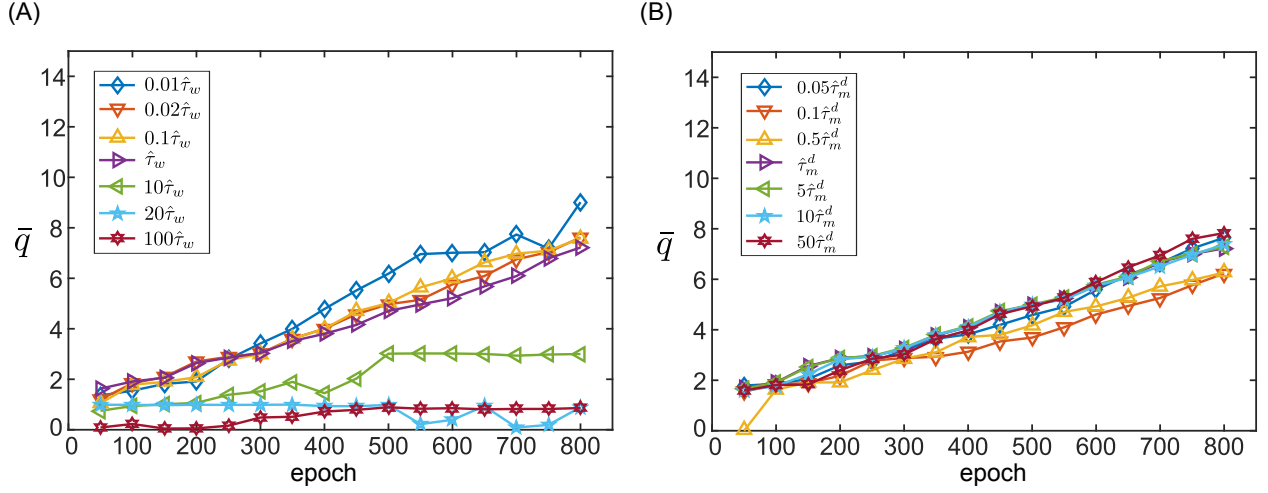
Figure 6.7: The average number of reconstructed lags as a function of epoch number for different values of (A) $\tau_w$ (time constant of feedforward connections) and (B) $\tau_m^d$ (time constant of recurrent connections). All other parameters except $\tau_w$ and $\tau_m^d$ are the same for both plots.
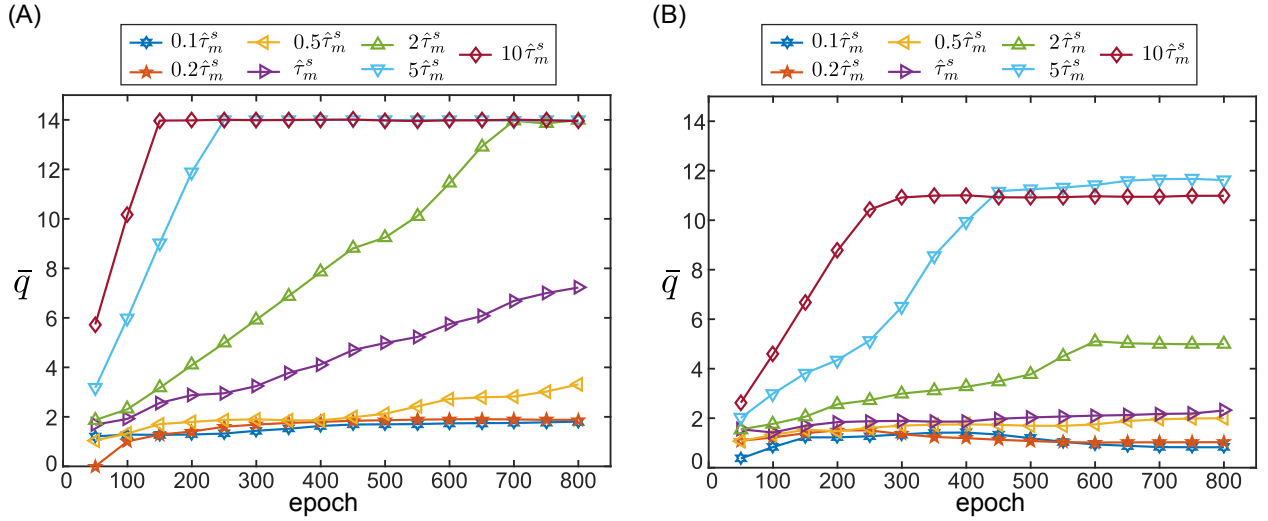


Figure 6.8: The average number of reconstructed lags as a function of epoch number for different values of $\tau_m^s$, time constant of delayed recurrent connections, for (A) $\lambda_1 = 0.01$ and (B) $\lambda_1 = 0.04$. All other parameters except $\lambda_1$ are the same for both plots.

The average number of reconstructed lags as a function of epoch number for different values

of $\tau_m^s$ (time constant of delayed recurrent connections) for $\lambda_1 = 0.01$ and $\lambda_1 = 0.04$ are shown

in Figures 6.8A-B. In these studies, $\lambda_2 = 0.15$, $\lambda = 1$, $\alpha = 0.2$, $\mu = 0.1$, , $\tau_w = \hat{\tau}_w$, and $\tau_m^d = \hat{\tau}_m^d$.

## 6.6   Chapter Summary

The processes by which efficient codes are produced remains a fundamental question in theoretical neuroscience. The role of recurrent networks in mediating these codes has been studied in the context of minimum-energy optimization problems. In this scenario, the synthesized networks exhibit linear dynamics. Here, we show how thresholding nonlinearities arise naturally through the optimization of a cost function that minimizes both energy and sparsity, i.e., what we term as a lightweight code. Such nonlinearities are, of course, intrinsic to the dynamics of neurons. Our complete network is able to produce lightweight codes that allow faithful reconstruction of both the current input and its history.

# Chapter 7

# Conclusions and Future Work

## 7.1   Summary and Conclusions

In this dissertation, we have explored several issues pertaining to the link between brain network dynamics and their associated function.

First, we considered the problem of node selection for probing connections in evoked networks. Here, each connection (weight) describes the node output responses when an input node is activated using a prescribed stimulus. Like 'functional' networks, such connections do not necessarily correspond to the physical (anatomical) network structure. However, in contrast to passive network inference methods, evoked connections are defined explicitly in both sensor and actuator space. We showed that the problem of deciding which node to stimulate in order to probe the evoked connections is equivalent to a problem of sensor scheduling, where each 'sensor' measures the weight of an edge relative to the delivered stimulus. Based on a Kalman filter paradigm for estimating the weights, we proceeded to derive convenient rules for greedy scheduling. Then, we showed that under certain conditions, the greedy algorithm is optimal over a finite horizon and, moreover, amounts to a predetermined periodic

'round-robin' selection schedule. Those conditions – that measurement and process noise are equal across nodes; and that initial uncertainty of edges is grouped on a node-by-node basis – are expected to be particularly applicable in situations where the sensor/actuators are co-located, relatively homogeneous and minimal prior information is known about the weights. We demonstrated the efficacy of the proposed methods in several examples, including a simulation study of a biophysical mean-field model. In this case, we showed that exploiting certain dynamical properties of the network, namely, approximate linearity with respect to a narrowband stimulus, allows for recovery of underlying structure from the evoked network through a linear inverse problem. The proposed methodology provides a principled way to interrogate the connections in brain networks via stimulation, with specific applications in brain mapping.

Next, we studied the notion of neuronal microstates in the context of resting state fMRI recordings of brain activity. For this purpose, we developed a bounded-observation Kalman filter for estimating correlation dynamics in neural recordings. The metric of interest–the Pearson correlation coefficient–is constrained to $[-1, 1]$, necessitating the inclusion of a Fisher transform in the filter equations. Utilizing the proposed technique, we quantified the extent of conservation of functional connectivity, defined by the correlation of intrinsic activity, within the cerebral cortex between wakefulness and sevoflurane general anesthesia. Our results indicated that the transition from 0% to 1.2% sevoflurane is associated with altered interaction between resting state networks. In addition, we observed evidence of altered brain microstate dynamics during resting wakefulness. Then, we evaluated the limitation of PCA for dFC analysis on original time-series data, and showed how such analysis can change the outcome significantly.

We then proceeded to demonstrate proof-of-concept for EEG-based segregation of focal versus diffuse DLOC etiology based on time-series analysis and support-vector machine classification. We highlighted how information regarding these etiologies may be embedded within the temporal dynamics of the measured EEG signals. Moreover, we developed a new spatiotemporal characterization of so-called lagged microstates, in order to better understand the mechanisms of impaired consciousness in DLOC patients. We showed that patients having severe impairments can be differentiated from those with relatively milder impairments on the basis of this new analysis. Thus, this method can be used as a feature for improving automatic classification of level of consciousness in patients with coma.

In parallel with our data-driven investigations, we also performed theory-based development of information processing in small neuronal network models. In this context, we derived analytical conditions that are sufficient to ensure the recovery of time-varying signals processed via a recurrent linear network, wherein the network output is allowed to accrue prior to recovery. The conditions amount to a generalized observability criteria and are less stringent than the approach of performing sparse recovery pointwise, i.e., at each instant in time. Further, we derived several bounds in terms of the system dynamics for recovery performance in the presence of both input disturbance and observation noise. We showed that dynamics can play opposing roles in mediating accurate recovery with respect to these two different sources of corruption.

Finally, we considered the processes by which efficient codes may be produced in sensory networks. The role of recurrent networks in mediating these codes has been studied in the context of minimum-energy optimization problems. In this scenario, the synthesized networks exhibit linear dynamics. In this topic, we showed how thresholding nonlinearities arise naturally through the optimization of a cost function that minimizes both energy and

162

sparsity, i.e., what we term as a lightweight code. Such nonlinearities are, of course, intrinsic to the dynamics of neurons. Our complete network is able to produce lightweight codes that allow faithful reconstruction of both the current input and its history.

## 7.2 Future Directions

This dissertation has outlined results along two major axes: (i) tracking connectivity within brain networks in a dynamic manner, and (ii) synthesizing dynamics for particular objective function. The former, in particular, may generalize to a number of applications related to understanding transient changes in the brain, for instance, associated with neurological disorders. Other potential future research directions that may emanate from this dissertation are discussed below:

**Tracking Dynamic Connectivity with Active Stimulation in Clinical Settings:** In this dissertation, we solved the optimal probing strategy for a scenario where connections are dynamic and are changing over time. However, we did not have the opportunity to deploy this theory in experimental contexts. With the rapid development of new tools and technology for modifying neural activity, such deployment may be accessible in the near future and, thus, constitutes a natural progression for the developed methods.

**Finding Spatiotemporal Motifs in Neural Data by Incorporating Structural Constraints from Physiology:** The spatiotemporal motifs obtained in Chapter 3 and Chapter 4 were obtained without any prior constraints. Hence, It is of interest to perform this

analyses by having specific constraints in terms of geometry and structure for each motif, potentially arising from the underlying anatomy.

**Designing Neural Network Architecture Based on Different Objective Function:** In this dissertation, we presented how a neural-like network architecture emerges as a result of solving a particular optimization problem. It would be interesting to explore how different objective functions result in different architectures in terms of observed dynamics and adaptation rules. Again, introducing prior biological constraints may add to the richness of the obtained rules.

**Data Processing and Dimensionality Reduction by Training a Neural Network with Applications in Time-series Data Processing:** Perhaps the greatest challenge in neural data analysis is in the extraction of low-dimensional structures from what are, in native form, very high dimensional recordings. Conventional methods, such as PCA, can be problematic since they may be biased by regions with high activity, without overt regard for temporal dynamics. Thus, there is a substantial need for neural data processing and dimensionality reduction methods that explicitly account for temporal structure and memory on time-series data. Such development is, perhaps, the most general opportunity for future extensions of the work developed herein.

# Appendix A

# Supplementary Materials for Chapter 2

## A.1  EM Algorithm with Irregular Observations

Here, we provide the details of EM algorithm used for a linear state space model with irregularly spaced observations. Again, the problem is to jointly estimate the state, $\mathbf{x}_{j,k}, k = 1, 2, \cdots, m$, $\alpha_j$ and $\beta_j$ from $\mathbf{y}_{j,k}, k \in \mathcal{K}, 1 \leq |\mathcal{K}| \leq m$. Because the noise covariance matrices are diagonal, we can decouple the estimation processes and neglect $j$. Therefore, the problem reduces to joint estimation of the state $x_k, k = 1, \cdots, m$, $\alpha$ (process noise variance) and $\beta$ (measurement noise variance) from $y_{k_n}, n = 1, \cdots, |\mathcal{K}| \leq m$, subject to the linear state space model in (3.4).

**E-Step:** We start with initial conditions $x_{0|0} = \mu_0$ and $P_0 = \sigma_0$. If there exist observation at time $k$, then the estimated mean and variance are updated based on the Kalman filter equations (3.9). Otherwise, the mean and variance estimates of the states are updated as

$$\hat{x}_{k|k-1} = \hat{x}_{k-1|k-1}$$

$$P_{k|k-1} = P_{k-1} + \alpha,$$

$$\hat{x}_{k|k} = \hat{x}_{k-1|k-1}$$

$$P_k = P_{k|k-1}.$$

(A.1)

Based on the obtained estimates using the forward filtering recursion, the smoothed estimates can be obtained using backward recursion in the usual way reported in [49, 50] as

$$J_{k-1} = P_{k-1}(P_{k|k-1})^{-1},$$

$$x_{k-1|m} = x_{k-1|k-1} + J_{k-1}(x_{k|m} - x_{k|k-1}),$$

$$P_{k-1|m} = P_{k-1} + J_{k-1}(P_{k|m} - P_{k|k-1})J_{k-1},$$

(A.2)

where $x_{k|m}$ is the smoothed state at time $k$ based on at most $m$ observations. Also, the lag-one covariance smoother can be written as

$$P_{m,m-1|m} = (1 - K_m)P_{m-1}$$

$$P_{k-1,k-2|m} = P_{k-1}J_{k-2} + J_{k-2}(P_{k,k-1|m} - P_{k-1})J_{k-2}$$

(A.3)

**M-Step:** In the maximization step, we perform maximum likelihood estimation of model parameters from the joint density of the states $X_m = \{x_0, x_1, \cdots, x_m\}$ and observations $Y_m = \{y_k, k \in \mathcal{K}\}$, where the cardinality of $Y_m$ is $|\mathcal{K}|$. For instance if $Y_4 = \{y_1, y_3\}$, then $|\mathcal{K}| = 2$, $k_1 = 1$ and $K_2 = 3$. The joint density of $X_m$ and $Y_m$ can be factorized as

$$f_\Theta(X_m, Y_m) = f_{\mu_0, \sigma_0}(x_0) \prod_{k=1}^{m} f_\alpha(x_k | x_{k-1}) \prod_{n=1}^{|\mathcal{K}|} f_\beta(y_{k_n} | x_{k_n}),$$

(A.4)

166

where $\Theta = \{\mu_0, \sigma_0, \alpha, \beta\}$. Under the Gaussian assumption, the log-likelihood function can be written as

$$
\begin{aligned}
L_{X,Y}(\Theta) = & -\frac{1}{2\sigma_0}(x_0 - \mu_0)^2 - \frac{1}{2}ln(\sigma_0) \\
& -\frac{1}{2\alpha}\sum_{k=1}^{m}(x_k - x_{k-1})^2 - \frac{m}{2}ln(\alpha) \\
& -\frac{1}{2\beta}\sum_{n=1}^{|\mathcal{K}|}(y_{k_n} - C_{k_n}x_{k_n})^2 - \frac{|\mathcal{K}|}{2}ln(\beta) \\
& + Constant.
\end{aligned}
\tag{A.5}
$$

Based on (A.5), if we have the complete data, i.e., $\{X_m, Y_m\}$, then the maximum likelihood estimation (MLE) of $\Theta$ can be easily obtained. However, we do not have the complete data and the MLE of $\Theta$ can be obtained iteratively using EM algorithm by successively improving the approximation of the complete data in E-step. In the M-step, we find the parameter $\Theta$ that maximizes following quantity at iteration $i$,

$$
Q(\Theta|\Theta^{(i-1)}) = \mathbb{E}_{X|Y,\Theta^{(i-1)}}\{L_{X,Y}(\Theta)|Y_m, \Theta^{(i-1)}\}
\tag{A.6}
$$

Given $\Theta^{(i-1)}$, the desired conditional expectations in (A.6) can be obtained from the E-step as

$$
\begin{aligned}
Q(\Theta|\Theta^{(i-1)}) = & -\frac{1}{2\sigma_0}\left[P_{0|m} + (x_{0|m} - \mu_0)^2\right] - \frac{1}{2}ln(\sigma_0) \\
& -\frac{1}{2\alpha}\left[S_{11} - 2S_{10} + S_{00}\right] - \frac{m}{2}ln(\alpha) \\
& -\frac{1}{2\beta}\sum_{n=1}^{|\mathcal{K}|}\left[(y_{k_n} - C_{k_n}x_{k_n})^2 + C_{k_n}^2 P_{k_n|m}\right] \\
& -\frac{|\mathcal{K}|}{2}ln(\alpha) + Constant,
\end{aligned}
\tag{A.7}
$$

167

where

$$S_{11} = \sum_{k=1}^{m} (x_{k|m}^2 + P_{k|m}),$$

$$S_{10} = \sum_{k=1}^{m} (x_{k|m} x_{k-1|m} + P_{k,k-1|m}), \tag{A.8}$$

$$S_{00} = \sum_{k=1}^{m} (x_{k-1|m}^2 + P_{k-1|m}).$$

Maximization of (A.7) with respect to $\Theta$ yields the updated estimates at iteration $i$ as

$$\alpha^i = \frac{1}{m} (S_{11} - S_{10}^2/S_{00}),$$

$$\beta^i = \frac{1}{\mathcal{K}} \sum_{n=1}^{|\mathcal{K}|} \left[ (y_{k_n} - C_{k_n} x_{k_n})^2 + C_{k_n}^2 P_{k_n|m} \right], \tag{A.9}$$

$$\mu_0^i = x_{0|m}, \sigma_0^i = P_{0|m}.$$

In summary, after initializing the parameter $\Theta$, we perform the E-step, filtering and smoothing, and then updating the parameter in the M-step. The E and M steps alternate to convergence. In order to obviate computational issues associated with this algorithm, a maximum iteration number is set alongside the convergence criterion.

# Appendix B

# Supplementary Materials for Chapter 3

## B.1 Additional Results for Section 3.4

**Toy Example:** The original covariance and correlation matrices for microstate 1 and 2 were used to generate synthesized data in toy example shown in Figures 3.2 and 3.3 of the main manuscript are as follows:

$$Cov_{State1} = \begin{bmatrix} 1.54 & 2.40 \\ 2.40 & 5.67 \end{bmatrix}, \ Corr_{State1} = \begin{bmatrix} 1 & 0.81 \\ 0.81 & 1 \end{bmatrix},$$

$$Cov_{State2} = \begin{bmatrix} 2.61 & 2.02 \\ 2.02 & 5.36 \end{bmatrix}, \ Corr_{State2} = \begin{bmatrix} 1 & 0.54 \\ 0.54 & 1 \end{bmatrix}.$$

The averaged estimation of covariance and correlation matrices after performing microstate analysis on the original time-series data (before applying PCA) on 100 realization of synthesized data are as follows:

$$\overline{Cov}_{State1} = \begin{bmatrix} 1.65 & 2.44 \\ 2.44 & 5.70 \end{bmatrix}, \ \overline{Corr}_{State1} = \begin{bmatrix} 1 & 0.80 \\ 0.80 & 1 \end{bmatrix},$$

$$\overline{Cov}_{State2} = \begin{bmatrix} 2.55 & 1.91 \\ 1.91 & 5.26 \end{bmatrix}, \ \overline{Corr}_{State2} = \begin{bmatrix} 1 & 0.52 \\ 0.52 & 1 \end{bmatrix}.$$

The averaged estimation of covariance and correlation matrices after performing microstate analysis on the original time-series data (after applying PCA) on 100 realization of synthesized data are as follows:

$$\widetilde{Cov}_{State1} = \begin{bmatrix} 1.66 & 2.16 \\ 2.16 & 5.79 \end{bmatrix}, \ \widetilde{Corr}_{State1} = \begin{bmatrix} 1 & 0.70 \\ 0.70 & 1 \end{bmatrix},$$

$$\widetilde{Cov}_{State2} = \begin{bmatrix} 2.43 & 2.27 \\ 2.27 & 5.23 \end{bmatrix}, \ \widetilde{Corr}_{State2} = \begin{bmatrix} 1 & 0.65 \\ 0.65 & 1 \end{bmatrix}.$$

**Statistical Analysis of Occupancy:** The results of statistical test on occupancy of being at a particular microstate (without loss of generality microstate 1) before and after applying PCA on the original time-series data for different cases are as follows: (K-S test for the real BOLD case, p-value $= 0.111$, stat $= 0.5$), (K-S test for the case I, p-value$=0.031$, stat$=0.6$), (K-S test for the case II, p-value$=0.007$, stat$=0.7$), (K-S test for the case III, p-value$=0.031$, stat$=0.6$).
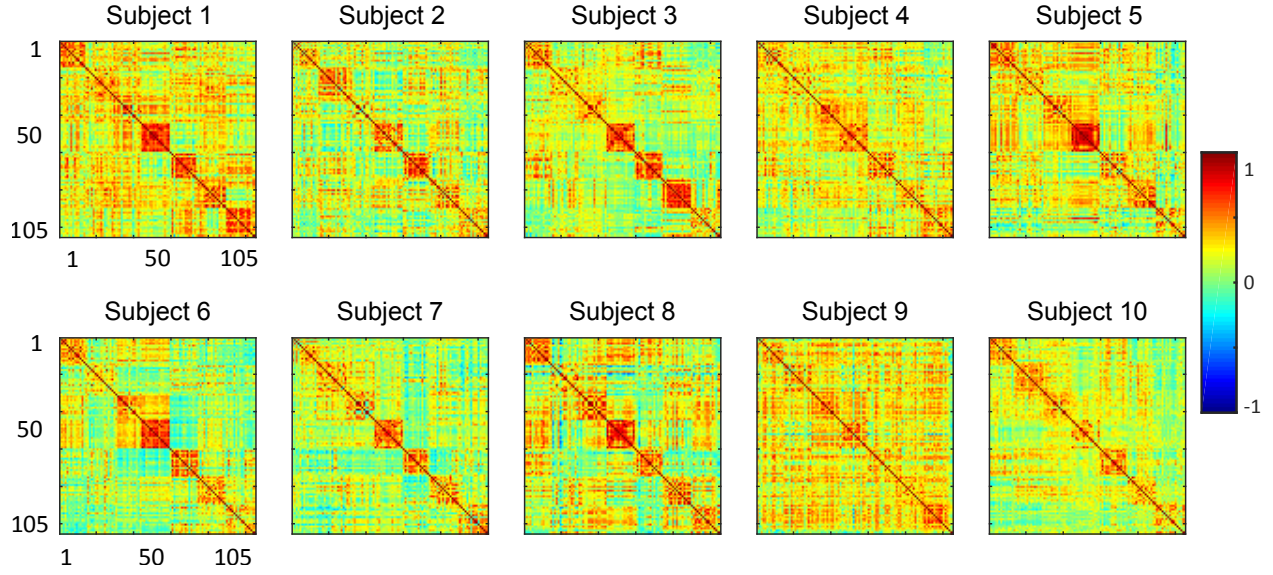
Figure B.1: Functional connectivity of each subjects, calculated over entire scan time, for all individuals in our dataset.

**Static Functional Connectivity of Each Subjects:** Figure B.1 demonstrates the functional connectivity of each subjects, calculated over entire scan time.

**Non-stationary Simulation of the BOLD Time-Series:** To simulate a non-stationary BOLD time-series data, we use the results of microstate analysis on the real BOLD signal. After applying microstate analysis to the real BOLD signal, we exactly know the microstate index for each sliding window for all subjects. Then for all sliding window which are assigned to the firs microstate, we calculate the average covariance matrices for all subjects individually. This analysis results in Figure B.2A. In a similar way, we can find covariance matrices associated to second microstate for each individual subjects (Figure B.2B). We then use these covariance matrices to simulate BOLD signal with similar non-stationarity as the real BOLD signal for each subject.
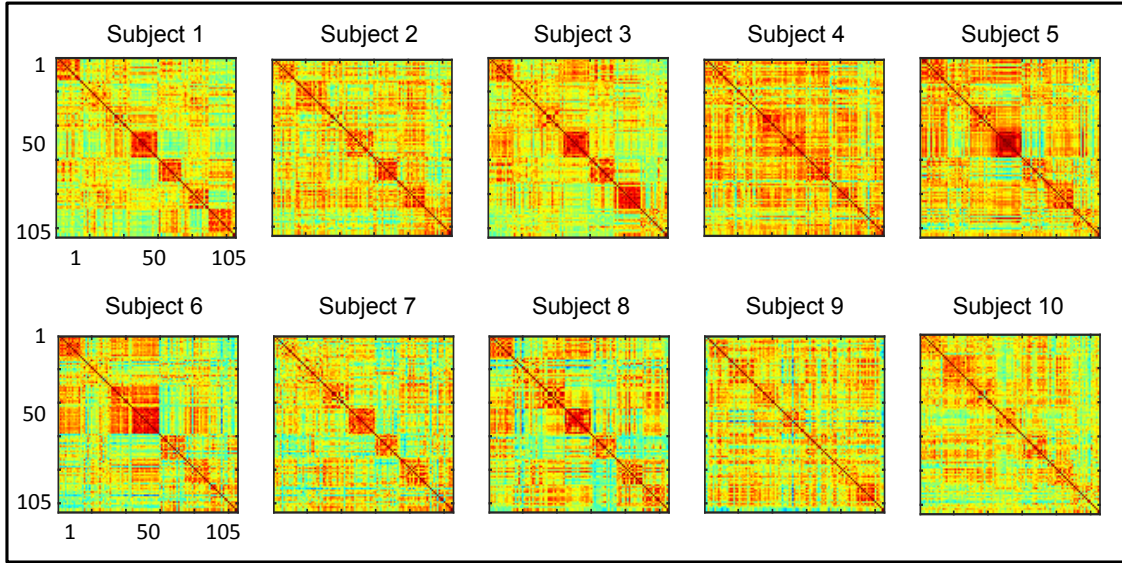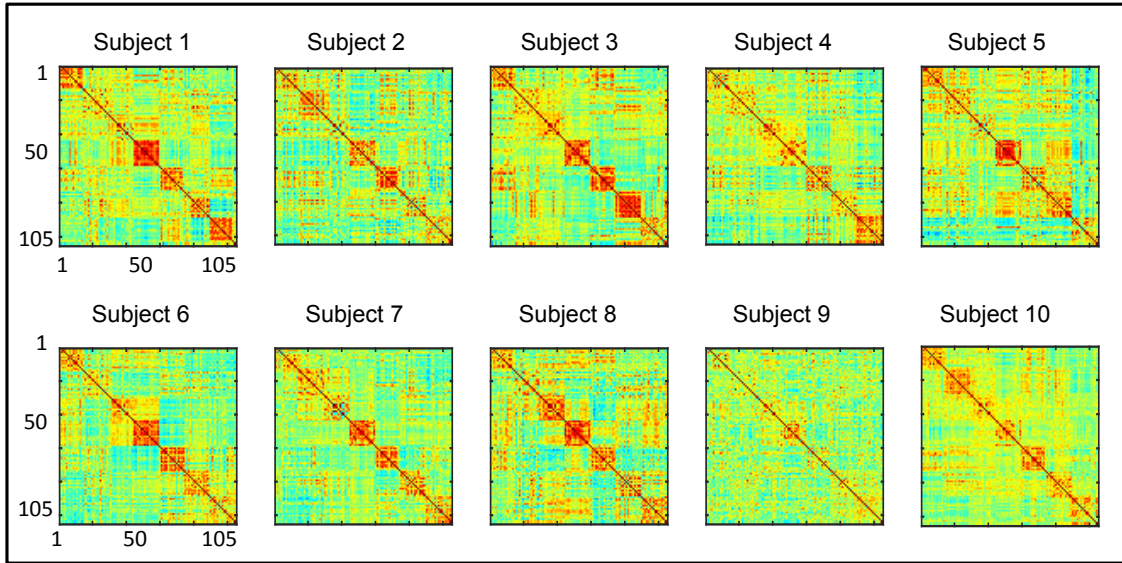
Figure B.2: Microstates used for simulation of time-series data in case II. (A) Correlation matrices associated to the first microstate for each individual obtained from microstate analysis of the BOLD signal. (B) Correlation matrices associated to the second microstate for each individual obtained from microstate analysis of the BOLD signal.
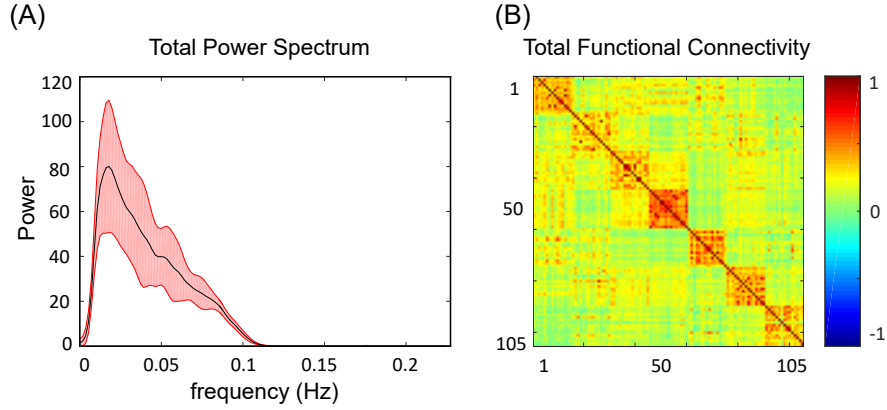
Figure B.3: (A) Total power spectrum and (B) total functional connectivity of all subjects in our dataset.

**Total Power Spectrum and Total Correlation Matrix Across All the Subjects:**
Figure B.3 illustrates total correlation matrix and Total power spectrum of all subjects in our dataset.

# B.2 Determining the Number of Motifs Using AIC

Selecting the number of clusters, k, in unsupervised learning is one of the key problem. The correct choice of K is often ambiguous. There exists several heuristic methods to facilitate the selection of k such as "The Elbow Method" by [224] and a technique based on the use of hierarchical clustering by [225]. Another set of approaches for determining $k$ are information criteria methods. In this chapter, we adopt the Akaike information criterion (AIC) [90] approach. AIC rewards goodness of fit, assessed by the likelihood function, but it also includes a penalty that is an increasing function of K. Figure B.4 shows AIC versus K for both 0% and 1.2% vol sevoflurane. In this plot, AIC values are normalized to have maximum value of unity for each condition. It can be seen from this figure that K = 8 is the optimal choice of K for both condition.
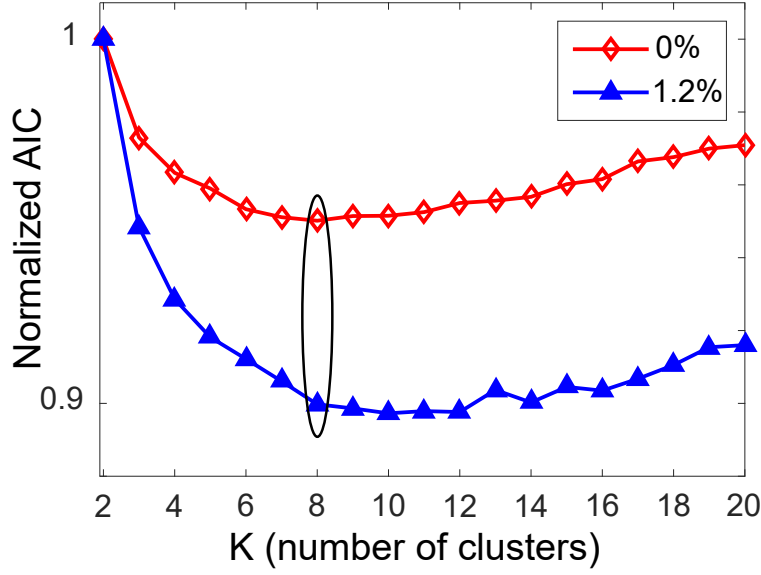
Figure B.4: Normalized AIC versus number of cluster K. AIC values are normalized to have maximum value of unity for each condition.

## B.3 Spatiotemporal Analyses of Synthesized Data

Here, we demonstrate the efficacy and utility of the approach in a simple example of a data set with 9 locations and 160 time windows. We assume there exists 5 spatial motifs shown in Figure B.5A and each motif has the same temporal correlation trajectory with the same mean over time (Green lines shown in Figure B.5B). Before decomposing the correlation trajectories, we added noise drawn from uniform distribution inside interval $(-0.1, 0.1)$ to all region pair correlations over time. Figure B.5 shows spatiotemporal decomposition of correlation time series. Spatial and temporal motifs are shown in Figure B.5A and B.5B, respectively. It can be conclude from these plots that temporal information can play an important role for extracting meaningful motifs even though if all the region pairs have the same static correlation value.
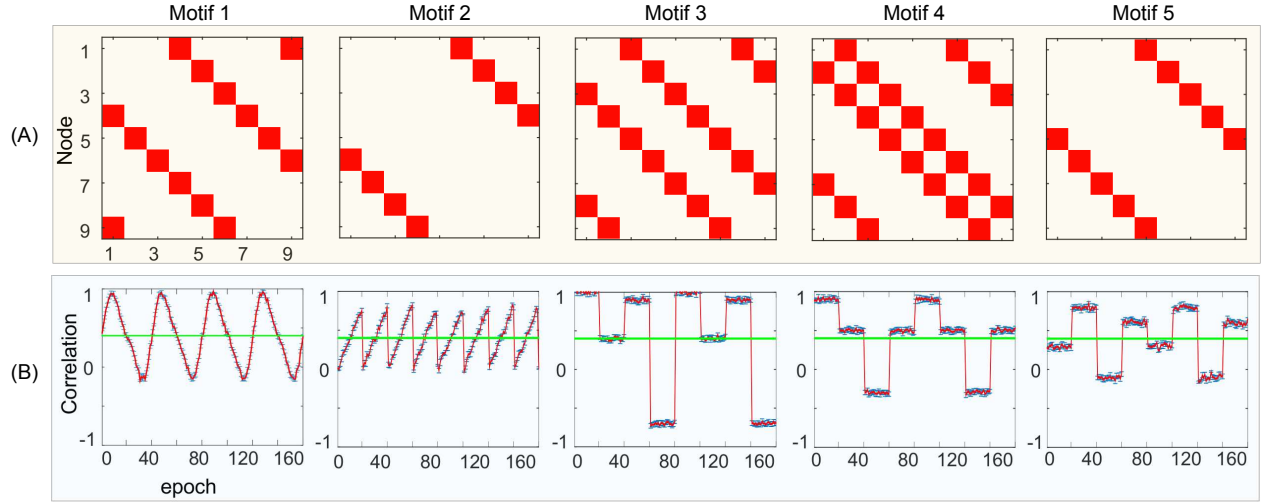
Figure B.5: (A). Spatial motifs and (B) temporal correlation trajectories associated with each cluster for synthesized data. Green line shows the average correlation over time and space for each cluster.

## B.4 Numerical Simulation of the Developed Filter on Synthesized Data

To illustrate the performance of the proposed filter, we simulate a univariate example of the system (3.4) with $A = C = 1$, $Q_w = 0.1$ and $R_v = 0.05$. Fig. B.6 shows the estimate of the correlation, i.e., $z_k$, using the proposed filter, as well as the results of naive application of the Extended Kalman Filter (EKF) based on linearization of (3.3) [226]. Clearly, state estimation using proposed optimal filter is closest to the true state of the system. This result is expected because EKF is simply a linearization of $\mathbf{F}^{-1}()$, which will produce poor results when the argument of the function is large (i.e., large correlation values).

Fig. B.7 illustrates mean-square error (MSE) for the proposed optimal filtering compared to EKF for different noise parameter values. In this simulation, a 4-dimensional system with $\mathbf{A} = \mathbf{C} = \mathbf{I_4}$ is considered. The process and measurement noise are considered in the form of
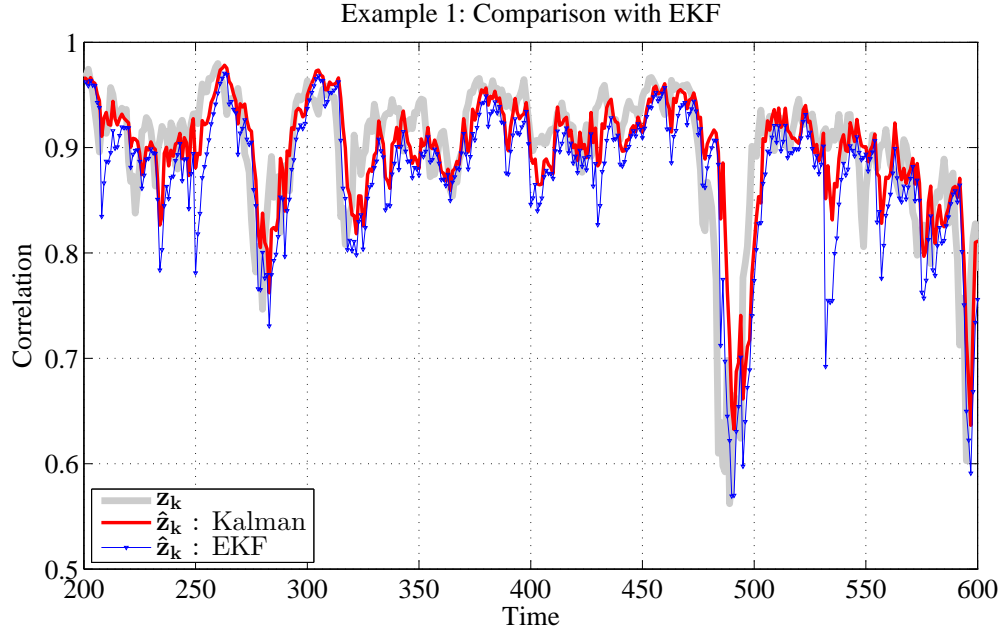
Figure B.6: Estimation of the state in system (3.4) with $A = C = 1$, $Q_w = 0.1$ and $R_v = 0.05$ using optimal non-linear filter for bounded observation (red dashed line) and EKF (green dash-dotted line).
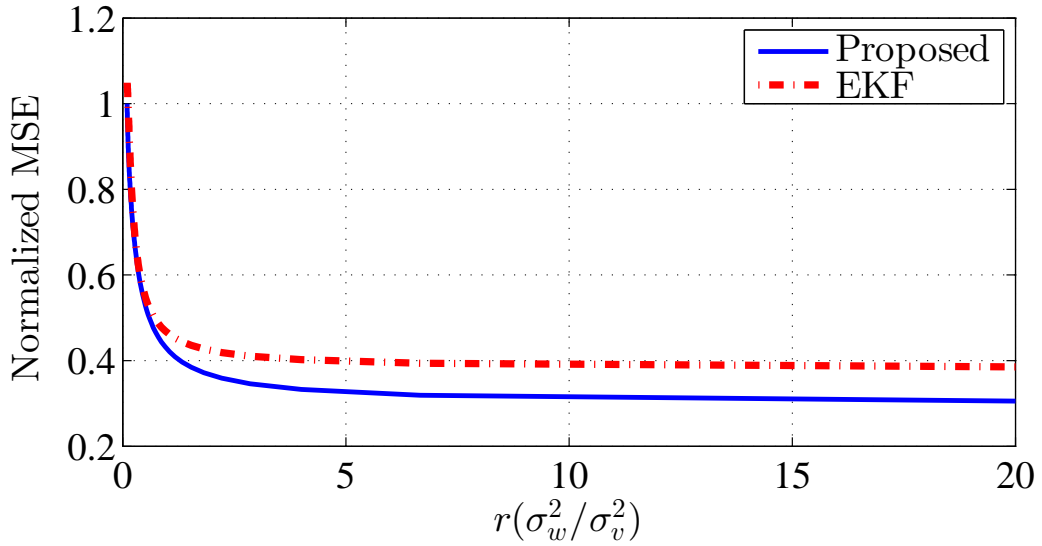


Figure B.7: Normalized MSE versus the ratio of process and measurement noise $r = \sigma_\mathbf{w}^2/\sigma_\mathbf{v}^2$ with $\mathbf{A} = \mathbf{C} = \mathbf{I_4}$, $\mathbf{Q_w} = \sigma_\mathbf{w}^2 \mathbf{I_4}$ and $\mathbf{R_v} = \sigma_\mathbf{v}^2 \mathbf{I_4}$.
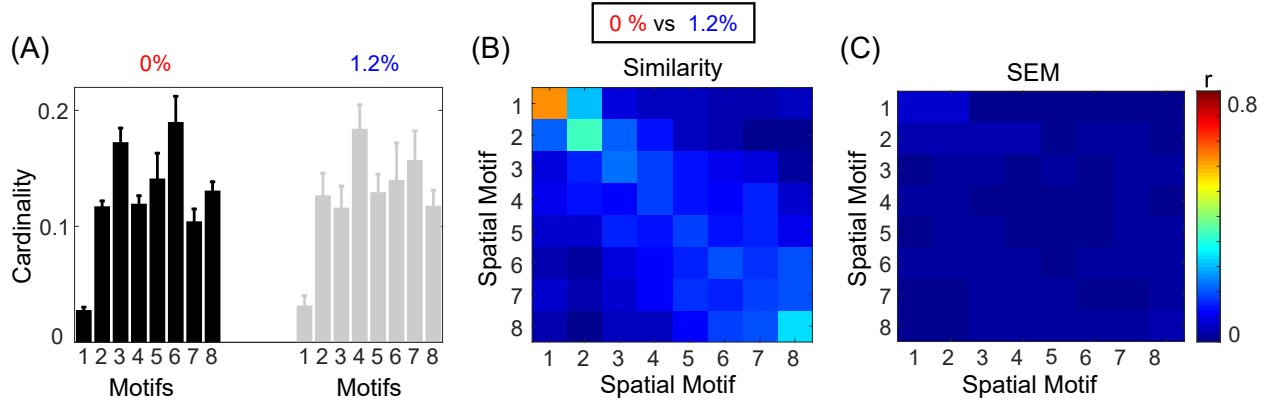
Figure B.8: (A) Bar plot of the cardinality percentage of each motif for both condition (0% and 1.2%) for a fixed window size, 11 seconds, over different individuals. Error bars stand for SEM. Average similarity (B) and SEM of average similarity (C) between motifs over different individuals, by computing the correlation of spatial component of each motif.

$\mathbf{Q_w} = \sigma_\mathbf{w}^\mathbf{2}\mathbf{I_4}$, $\mathbf{R_v} = \sigma_\mathbf{v}^\mathbf{2}\mathbf{I_4}$ and each time MSE is calculated for a specific ratio of $r = \sigma_\mathbf{v}^2/\sigma_\mathbf{v}^2$. It can be seen in this figure that by increasing $r$, MSE is decreased for both filters. Also, It can be seen that optimal filter always has better performance in term of MSE compare to EKF.

## B.5 Robustness as a Function of Intersubject Variability

To see how consistent the spatiotemporal motifs are for different conditions (0% and 1.2%), we look at how spatial motifs at 0% and 1.2% obtained from each individual, without concatenating data from all subjects for each condition, are similar. Figure B.8A demonstrates the average cardinality of each motifs over different individual. Error bars stand for SEM of the average cardinality. Figure B.8B represents average similarity between motifs at 0% and 1.2% obtained from each individual for 11 seconds window size. Standard error of the

mean of the average similarity is shown in Figure B.8C. We can see from this figures that the similarity between first, second and eighth motifs are robust over different individual.

# Appendix C

# Supplementary Materials for Chapter 4

## C.1 Determining the Number of Microstates

Figure C.1 shows the ratio of within-cluster distance to between-cluster distance versus the number of microstates/clusters based on lagged cross-correlation analysis for patients with different levels of coma severity and subtype. It can be seen from this figure that $k = 5$ is a good choice of K for different cases.

## C.2 Robustness of Results with Respect to the Choice of Microstates Number

Figure C.2 demonstrates the connectivity pattern of microstates over head map with k=5. In this illustration, all connections with correlation value less than 0.1, threshold, are set to zero.
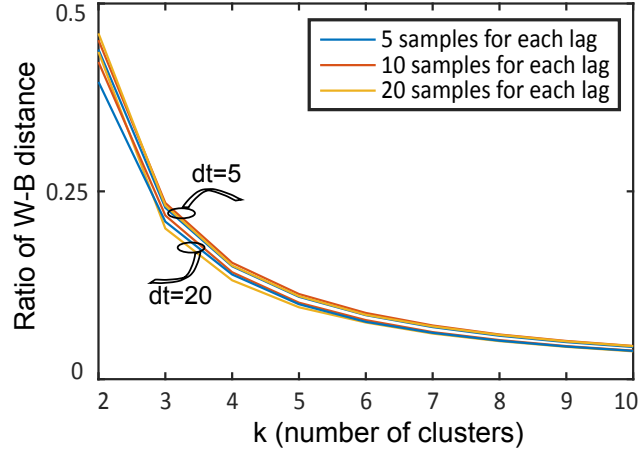
Figure C.1: Ratio of within–cluster distance to between-cluster distance versus number of clusters for microstate analysis based on lagged cross-correlation with different epoch length and number of samples for each lag. Four lags is considered for all of the cases.

We performed the microstate analysis using lagged cross-correlation of EEG data for different number of clusters. Figures C.3, C.4, and C.6 illustrate the microstates for $k = 4$, $k = 6$, and $k = 7$, respectively. All of the parameter, except the number of microstates, are exactly the same as parameter considered in section 4.5. Figure C.5 demonstrates the connectivity pattern of microstates over head map with k=6. In this illustration, all connections with correlation value less than 0.05, threshold, are set to zero. Microstate occupancy for all the cases are shown in Figure C.7. For all cases, the states are ranked based on their average strength.
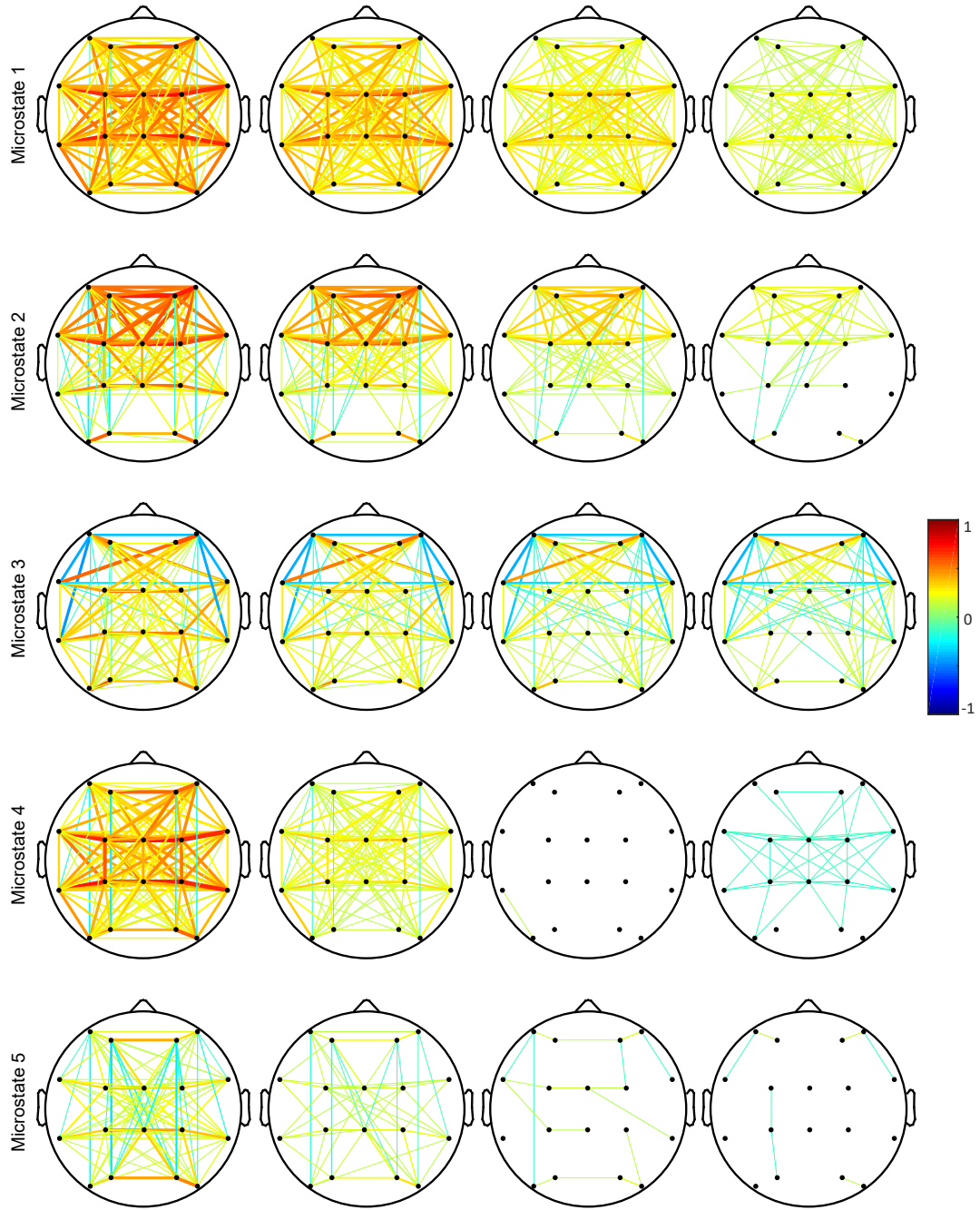
Figure C.2: Connectivity Map of all microstates with k=5. First, second, third and fourth columns are associated to zero, first, second, and third lag connectivity. All connections with correlation value less than 0.1 are set to zero.
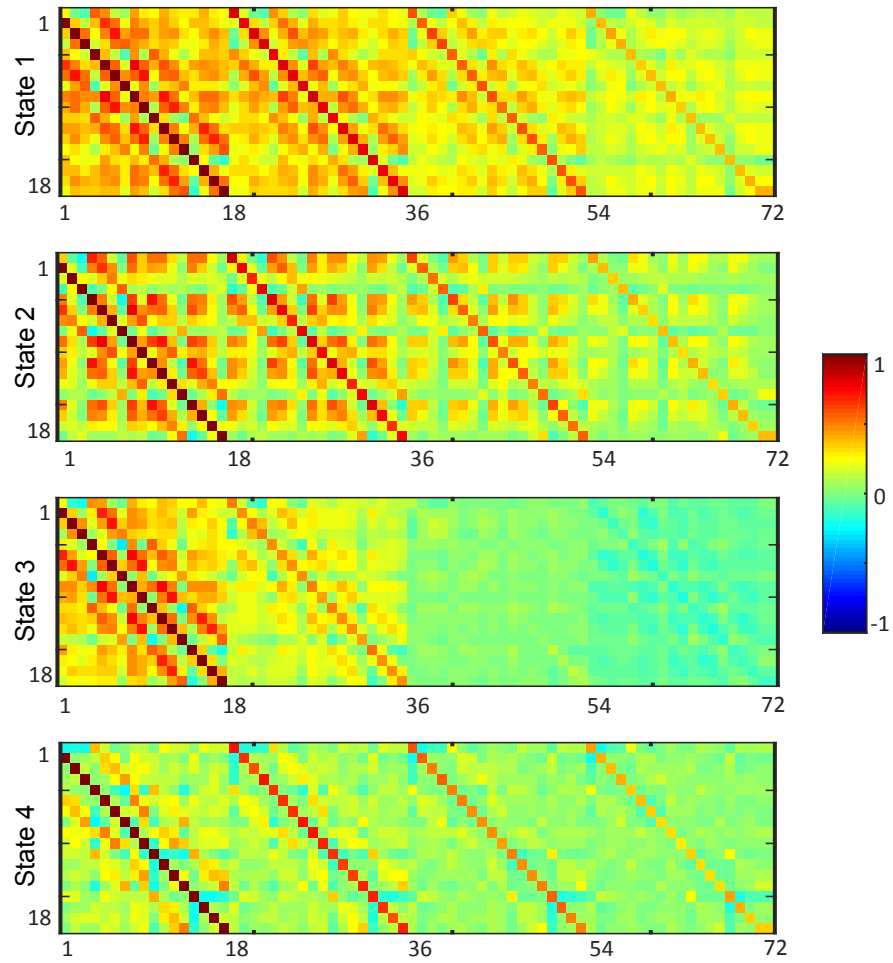
Figure C.3: Spatiotemporal microstates extracted from subjects with different coma severity and coma subtypes based on lagged cross-correlation of EEG data with k=4 microstates.
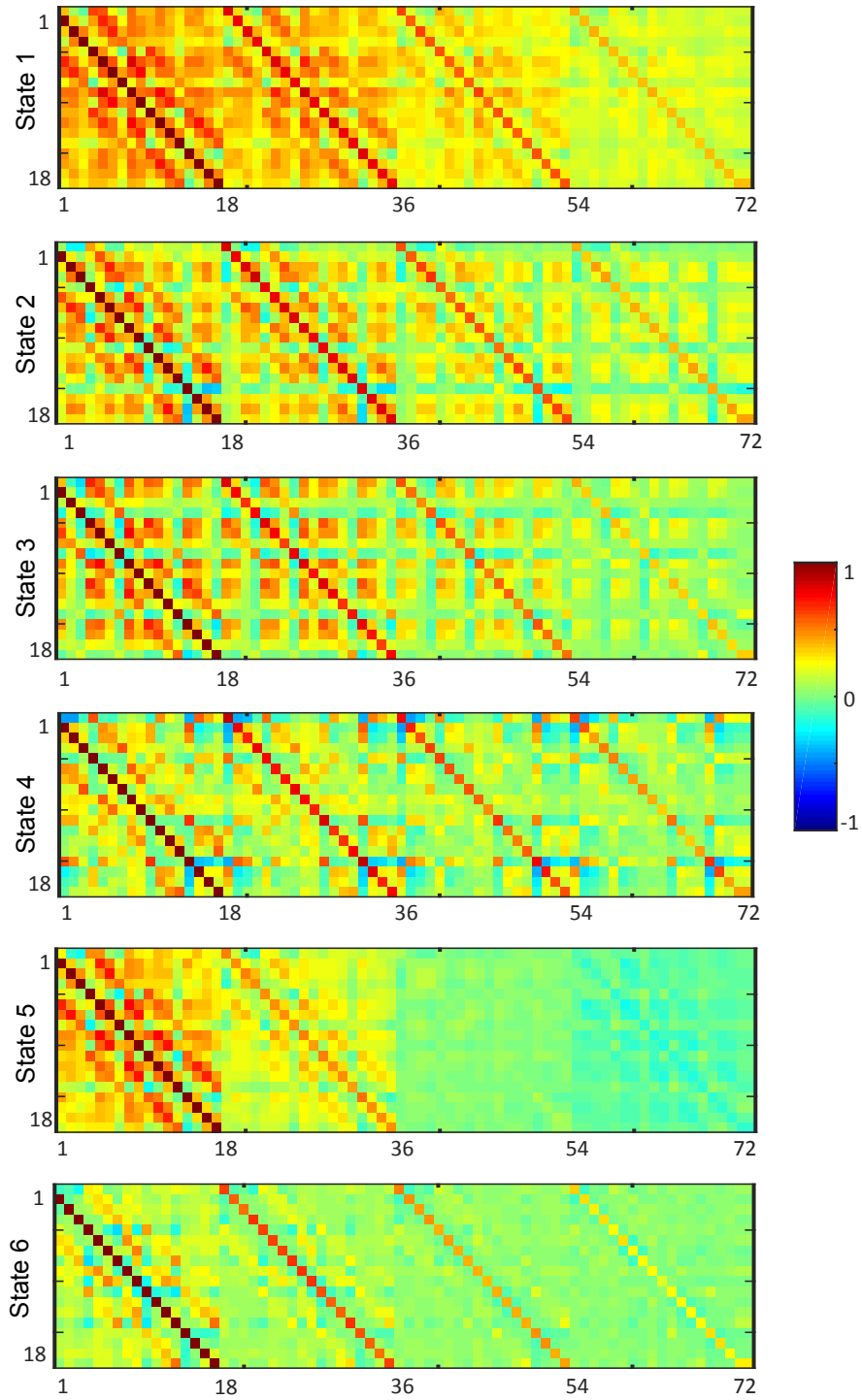
Figure C.4: Spatiotemporal microstates extracted from subjects with different coma severity and coma subtypes based on lagged cross-correlation of EEG data with k=6 microstates.

183
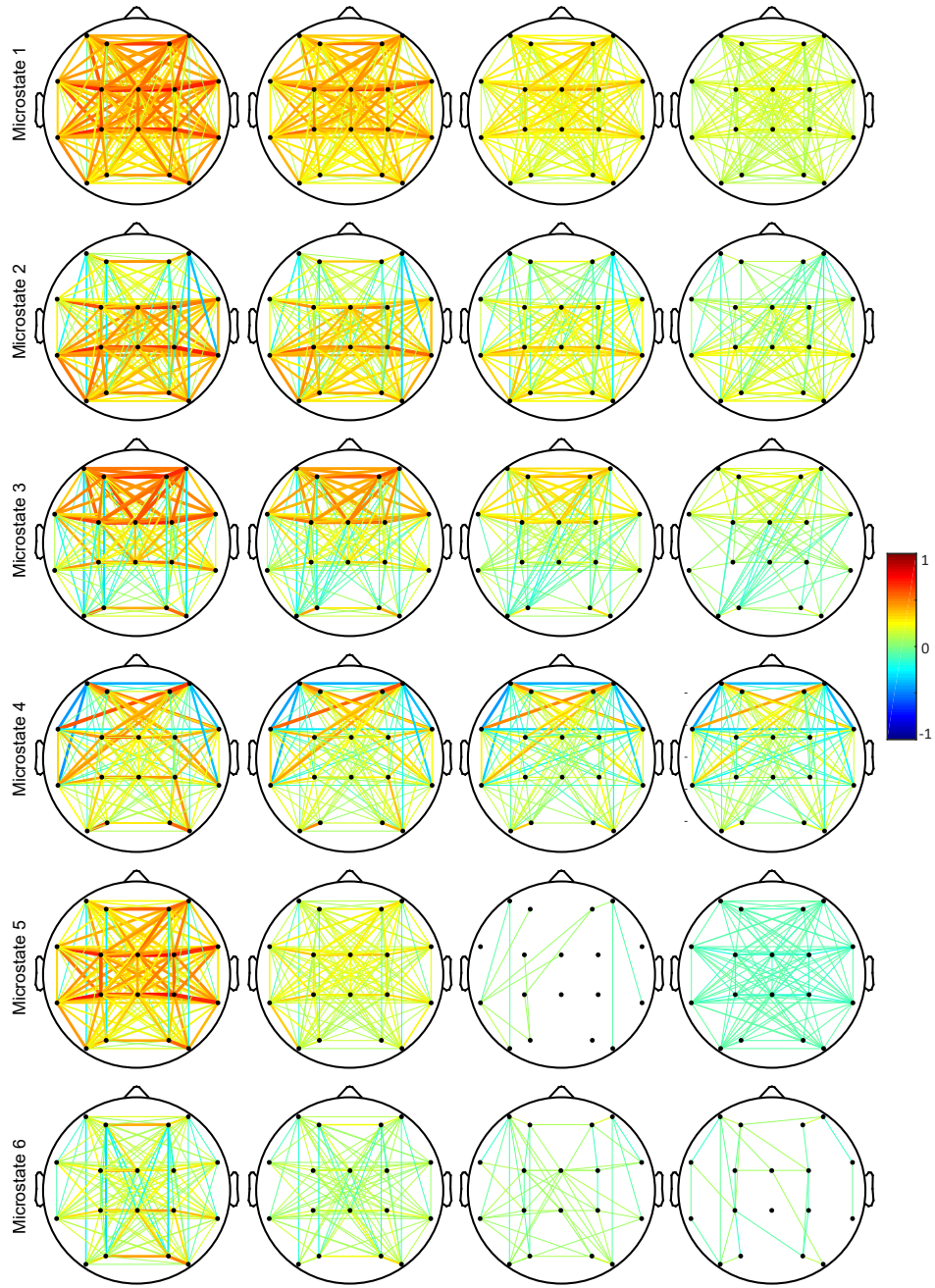
Figure C.5: Connectivity Map of all microstates with k=6. First, second, third and fourth columns are associated to zero, first, second, and third lag connectivity. All connections with correlation value less than 0.05 are set to zero.
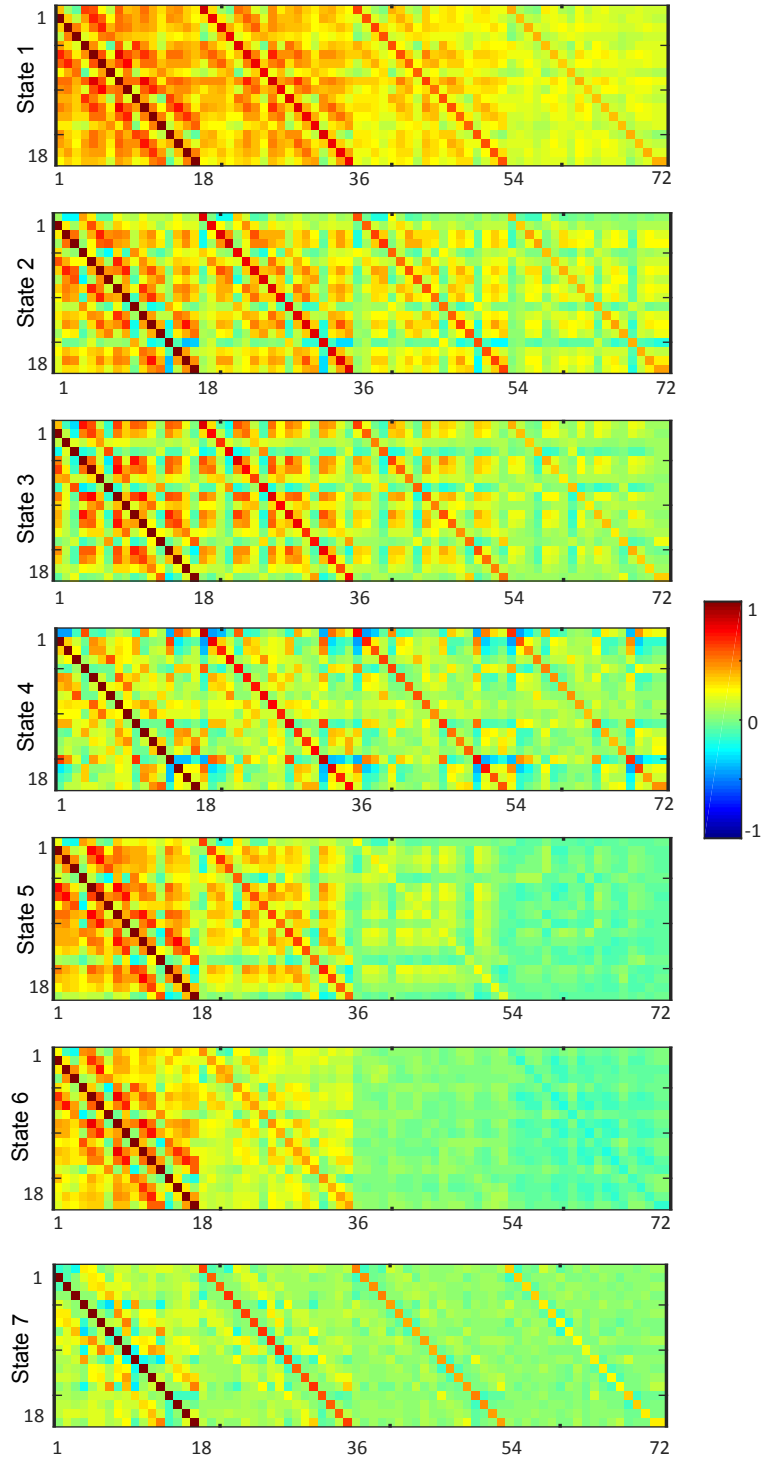
Figure C.6: Spatiotemporal microstates extracted from subjects with different coma severity and coma subtypes based on lagged cross-correlation of EEG data with k=7 microstates.

185
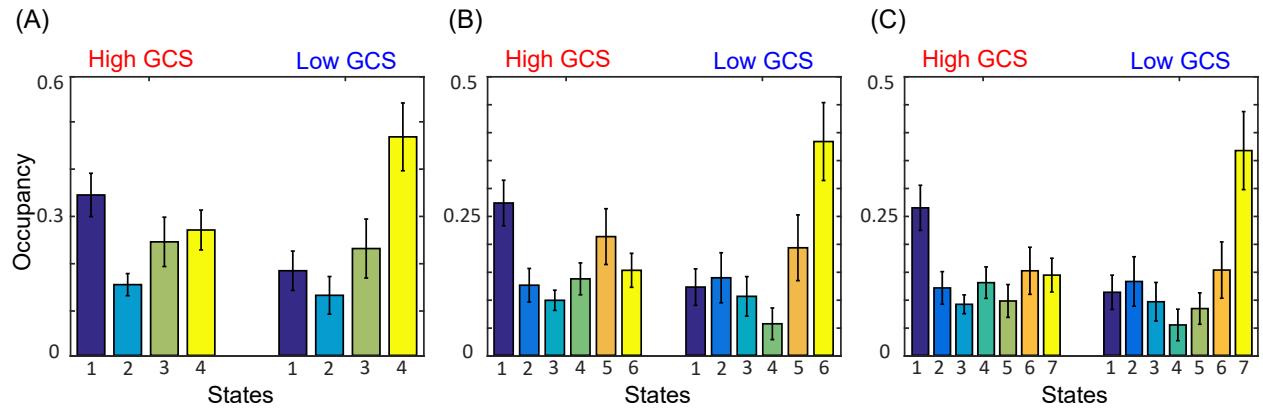
Figure C.7: Occupancy of being at a particular microstate averaged over the subjects using lagged cross-correlation to find microstates for (A) k=4, (B) k=6, and (C) k=7. Error bar shows the standard error of the mean.

# Appendix D

# Supplementary Materials for Chapter 5

## D.1  Proof of Lemma 5.1

We start the proof using contradiction. Let us assume that the sequence of $(\mathbf{u}_k)_{k=0}^{K-1}$ and $(\mathbf{r}_k)_{k=0}^{K}$ is not unique and there is another sequence of $s$-sparse $(\hat{\mathbf{u}}_k)_{k=0}^{K-1}$ and $(\hat{\mathbf{r}}_k)_{k=0}^{K}$ which satisfies the system (5.3) with noiseless measurements. Note that $\mathbf{u}_k - \hat{\mathbf{u}}_k$ has at most $2s$ nonzero elements. Similar to the depiction in Figure 5.2, we can rewrite (5.11) based on $2s$ columns of $\mathbf{B}$ corresponding to $2s$ active non-zero elements of $\mathbf{u}_k - \hat{\mathbf{u}}_k$ as

$$
\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{pmatrix} = \mathbf{O}_K \mathbf{r}_0 + \mathbf{J}_K^{2s} \begin{pmatrix} \mathbf{u}_0^{2s} \\ \mathbf{u}_1^{2s} \\ \vdots \\ \mathbf{u}_{K-1}^{2s} \end{pmatrix}, \tag{D.1}
$$

$$
\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{pmatrix} = \mathbf{O}_K \hat{\mathbf{r}}_0 + \mathbf{J}_K^{2s} \begin{pmatrix} \hat{\mathbf{u}}_0^{2s} \\ \hat{\mathbf{u}}_1^{2s} \\ \vdots \\ \hat{\mathbf{u}}_{K-1}^{2s} \end{pmatrix}. \tag{D.2}
$$

By subtracting the above equations from each other we have

$$
\mathbf{O}_K(\mathbf{r}_0 - \hat{\mathbf{r}}_0) + \mathbf{J}_K^{2s} \begin{pmatrix} \mathbf{u}_0^{2s} - \hat{\mathbf{u}}_0^{2s} \\ \mathbf{u}_1^{2s} - \hat{\mathbf{u}}_1^{2s} \\ \vdots \\ \mathbf{u}_{K-1}^{2s} - \hat{\mathbf{u}}_{K-1}^{2s} \end{pmatrix} = \mathbf{0}. \tag{D.3}
$$

Based on assumptions that $rank(\mathbf{O}_K) = n$ and $rank([\mathbf{O}_K \;\; \mathbf{J}_K^{2s}]) = n + rank(\mathbf{J}_K^{2s}), \forall \mathbf{J}_K^{2s} \in \mathcal{J}_K^{2s}$, all columns of the observability matrix must be linearly independent of each other, and of all columns of the $\mathbf{J}_K^{2s}$ matrix. Hence, the vector $\mathbf{r}_0 - \hat{\mathbf{r}}_0 = \mathbf{0}$. Having $\mathbf{r}_0 = \hat{\mathbf{r}}_0$ and the matrix $\mathbf{CB}$ satisfying the RIP condition (5.9) with isometry constant $\delta_{2s} < 1$, it is easy to see that $\mathbf{u}_k = \hat{\mathbf{u}}_k$ and therefore there exists unique state and $s$-sparse input sequences.

## D.2   Proof of Lemma 5.2

Lets assume that the the sequences $(\mathbf{r}_k^*)_{k=0}^{K}$ and $s$-sparse $(\mathbf{u}_k^*)_{k=0}^{K-1}$ are the solutions of Problem $(P2)$. In this case (D.3) can be rewritten as

$$
\mathbf{O}_K(\mathbf{r}_0^* - \bar{\mathbf{r}}_0) + \mathbf{J}_K^{2s}
\begin{pmatrix}
\mathbf{u}_0^{*2s} - \bar{\mathbf{u}}_0^{2s} \\
\mathbf{u}_1^{*2s} - \bar{\mathbf{u}}_1^{2s} \\
\vdots \\
\mathbf{u}_{K-1}^{*2s} - \bar{\mathbf{u}}_{K-1}^{2s}
\end{pmatrix}
=
\begin{pmatrix}
\mathbf{e}_0 \\
\mathbf{e}_1 \\
\vdots \\
\mathbf{e}_K
\end{pmatrix}
+
\begin{pmatrix}
\mathbf{e}_0^* \\
\mathbf{e}_1^* \\
\vdots \\
\mathbf{e}_K^*
\end{pmatrix},
\tag{D.4}
$$

where $\|\mathbf{e}_k^*\|_{\ell_2} \leq \epsilon$. Based on assumptions that $rank(\mathbf{O}_K) = n$ and $rank([\mathbf{O}_K \quad \mathbf{J}_K^{2s}]) = n + rank(\mathbf{J}_K^{2s})$, $\forall \mathbf{J}_K^{2s} \in \mathcal{J}_K^{2s}$ we can project the above equation using the projection $(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})$ where $\mathbf{P}_{\mathbf{J}_K^{2s}} = \mathbf{J}_K^{2s}(\mathbf{J}_K^{2s^T}\mathbf{J}_K^{2s})^{-1}\mathbf{J}_K^{2s^T}$. It is straightforward to verify that $(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\mathbf{J}_K^{2s} = \mathbf{0}$ and therefore we have

$$
(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\mathbf{O}_K(\mathbf{r}_0^* - \bar{\mathbf{r}}_0) = (\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})
\begin{pmatrix}
\mathbf{e}_0 \\
\mathbf{e}_1 \\
\vdots \\
\mathbf{e}_K
\end{pmatrix}
+ (\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})
\begin{pmatrix}
\mathbf{e}_0^* \\
\mathbf{e}_1^* \\
\vdots \\
\mathbf{e}_K^*
\end{pmatrix}.
\tag{D.5}
$$

Finally, we can derive the error bound for the state error from (D.5) as

$$
\|\mathbf{r}_0^* - \bar{\mathbf{r}}_0\|_{\ell_2} \leq 2\epsilon \sqrt{\frac{K\sigma_{max}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}{\sigma_{min}\left(\mathbf{O}_K^T\mathbf{O}_K\right)\sigma_{min}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}}
\tag{D.6}
$$

In a similar fashion we have an exact same bound for the state error at a different time $k$.

## D.3 Proof of Lemma 5.3

For each $j \geq 2$ and $k = 0, \cdots, K-1$ we have

$$\|\mathbf{h}_{k,T_j(k)}\|_{\ell_2} \leq s^{1/2} \|\mathbf{h}_{k,T_j(k)}\|_{l_\infty} \leq s^{-1/2} \|\mathbf{h}_{k,T_{j-1}(k)}\|_{\ell_1}, \tag{D.7}$$

and thus

$$\sum_{j \geq 2} \|\mathbf{h}_{k,T_j(k)}\|_{\ell_2} \leq s^{-1/2}(\|\mathbf{h}_{k,T_1(k)}\|_{\ell_1} + \|\mathbf{h}_{k,T_2(k)}\|_{\ell_1} + \cdots)$$
$$\leq s^{-1/2}\|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_1}. \tag{D.8}$$

Therefore, we have the following equation

$$\|\mathbf{h}_{k,T_{01}^c(k)}\|_{\ell_2} = \|\sum_{j \geq 2} \mathbf{h}_{k,T_j(k)}\|_{\ell_2} \leq \sum_{j \geq 2} \|\mathbf{h}_{k,T_j(k)}\|_{\ell_2}$$
$$\leq s^{-1/2}\|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_1}. \tag{D.9}$$

Since $(\mathbf{u}^*{}_k)_{k=0}^{K-1}$ minimizes the cost function in Problem $(P2)$,

$$\sum_{k=0}^{K-1} \|\bar{\mathbf{u}}_k\|_{\ell_1} \geq \sum_{k=0}^{K-1} \|\mathbf{u}_k^*\|_{\ell_1} = \sum_{k=0}^{K-1} \|\bar{\mathbf{u}}_k + \mathbf{h}_k\|_{\ell_1}$$
$$= \sum_{k=0}^{K-1} \left( \sum_{i \in T_0(k)} |\bar{\mathbf{u}}_{k,i} + \mathbf{h}_{k,i}| + \sum_{i \in T_0^c(k)} |\bar{\mathbf{u}}_{k,i} + \mathbf{h}_{k,i}| \right) \tag{D.10}$$
$$\geq \sum_{k=0}^{K-1} (\|\bar{\mathbf{u}}_{k,T_0(k)}\|_{\ell_1} - \|\mathbf{h}_{k,T_0(k)}\|_{\ell_1} + \|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_1} + \|\bar{\mathbf{u}}_{k,T_0^c(k)}\|_{\ell_1})$$

$\bar{\mathbf{u}}_0, \cdots, \bar{\mathbf{u}}_{K-1}$ are non-zero for $T_0(0), \cdots, T_0(K-1)$, respectively. Therefore,

$$\|\bar{\mathbf{u}}_{0,T_0^c(0)}\|_{\ell_1} = \cdots = \|\bar{\mathbf{u}}_{K-1,T_0^c(K-1)}\|_{\ell_1} = 0 \tag{D.11}$$

which gives

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_1} \leq \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_1}. \tag{D.12}$$

Considering

$$\|\mathbf{h}_{k,T_0(k)}\|_{\ell_1} \leq s^{1/2} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_2}, \tag{D.13}$$

and substituting it into (D.9) and (D.12) we have

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}^c(k)}\|_{\ell_2} \leq \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_2}. \tag{D.14}$$

## D.4 Proof of Lemma 5.4

To find the bound for $\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}$, we start with

$$\mathbf{CBh}_k = \mathbf{CBh}_{k,T_{01}(k)} + \sum_{j \geq 2} \mathbf{CBh}_{k,T_j(k)}, \tag{D.15}$$

which gives

$$\|\mathbf{CBh}_{k,T_{01}(k)}\|_{\ell_2}^2 = \langle \mathbf{CBh}_{k,T_{01}(k)}, \mathbf{CBh}_k \rangle - \langle \mathbf{CBh}_{k,T_{01}(k)}, \sum_{j \geq 2} \mathbf{CBh}_{k,T_j(k)} \rangle. \tag{D.16}$$

From (5.19) and the RIP condition for $\mathbf{CB}$,

$$\begin{aligned} |\langle \mathbf{CBh}_{k,T_{01}(k)}, \mathbf{CBh}_k \rangle| &\leq \|\mathbf{CBh}_{k,T_{01}(k)}\|_{\ell_2} \|\mathbf{CBh}_k\|_{\ell_2} \\ &\leq 2\epsilon C_0 \sqrt{1 + \delta_{2s}} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}, \end{aligned} \tag{D.17}$$

and, moreover, application of the parallelogram identity for disjoint subsets $T_0(k)$ and $T_j(k), j \geq 2$ results in

$$|\langle \mathbf{CB}\mathbf{h}_{k,T_0(k)}, \mathbf{CB}\mathbf{h}_{k,T_j(k)} \rangle| \leq \delta_{2s} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_2} \|\mathbf{h}_{k,T_j(k)}\|_{\ell_2}. \tag{D.18}$$

Inequality (D.18) holds for $T_1$ in place of $T_0$. Since $T_0$ and $T_1$ are disjoint

$$\|\mathbf{h}_{k,T_0(k)}\|_{\ell_2} + \|\mathbf{h}_{k,T_1(k)}\|_{\ell_2} \leq \sqrt{2}\|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}, \tag{D.19}$$

which results in

$$
\begin{aligned}
(1 - \delta_{2s})\|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}^2 &\leq \|\mathbf{CB}\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}^2 \\
&\leq \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}\Big(2C_0\epsilon\sqrt{1+\delta_{2s}} + \sqrt{2}\delta_{2s}\sum_{j\geq 2}\|\mathbf{h}_{k,T_j(k)}\|_{\ell_2}\Big).
\end{aligned}
\tag{D.20}
$$

It follows from (D.8) and (D.20) that

$$\|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2} \leq \alpha C_0\epsilon + \rho s^{-1/2}\|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_2}. \tag{D.21}$$

Now, using (D.12) and (D.21) we can conclude that

$$
\begin{aligned}
\sum_{k=0}^{K-1}\|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2} &\leq K\alpha C_0\epsilon + \rho s^{-1/2}\sum_{k=0}^{K-1}\mathbf{h}_{k,T_0^c(k)}\|_{\ell_2} \\
&\leq K\alpha C_0\epsilon + \rho\sum_{k=0}^{K-1}\|\mathbf{h}_{k,T_0(k)}\|_{\ell_2} \\
&\leq K\alpha C_0\epsilon + \rho\sum_{k=0}^{K-1}\|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2},
\end{aligned}
\tag{D.22}
$$

which means

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2} \leq K(1-\rho)^{-1}\alpha C_0\epsilon. \tag{D.23}$$

# Bibliography

[1] CJ Honey, O Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040, 2009.

[2] Uri Hasson, Howard C Nusbaum, and Steven L Small. Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences*, 106(26):10841–10846, 2009.

[3] Vincent Bonhomme, Audrey Vanhaudenhuyse, Athena Demertzi, Marie-Aurélie Bruno, Oceane Jaquet, Mohamed Ali Bahri, Alain Plenevaux, Melanie Boly, Pierre Boveroux, Andrea Soddu, et al. Resting-state network-specific breakdown of functional connectivity during ketamine alteration of consciousness in volunteers. *The Journal of the American Society of Anesthesiologists*, 125(5):873–888, 2016.

[4] Julián Benito-León, Elan D Louis, Juan Pablo Romero, Juan Antonio Hernández-Tamames, Eva Manzanedo, Juan Álvarez-Linera, Félix Bermejo-Pareja, Ignacio Posada, and Eduardo Rocon. Altered functional connectivity in essential tremor: A resting-state fmri study. *Medicine*, 94(49):e1936, 2015.

[5] Olaia Lucas-Jiménez, Natalia Ojeda, Javier Peña, María Díez-Cirarda, Alberto Cabrera-Zubizarreta, Juan Carlos Gómez-Esteban, María Ángeles Gómez-Beldarrain, and Naroa Ibarretxe-Bilbao. Altered functional connectivity in the default mode network is associated with cognitive impairment and brain anatomical changes in parkinson's disease. *Parkinsonism & Related Disorders*, 2016.

[6] Bruno A Olshausen et al. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[7] Panagiotis C Petrantonakis and Panayiota Poirazi. A compressed sensing perspective of hippocampal function. *Frontiers in Systems Neuroscience*, 8, 2014.

[8] Victor J Barranca, Gregor Kovačič, Douglas Zhou, and David Cai. Sparsity and compressed coding in sensory systems. *PLoS Computational Biology*, 10(8), 2014.

[9] MohammadMehdi Kafashan, Kyle Q Lepage, and ShiNung Ching. Node selection for probing connections in evoked dynamic networks. In *53rd IEEE Conference on Decision and Control*, pages 6080–6085. IEEE, 2014.

[10] MohammadMehdi Kafashan and ShiNung Ching. Optimal stimulus scheduling for active estimation of evoked brain networks. *Journal of neural engineering*, 12(6):066011, 2015.

[11] MohammadMehdi Kafashan, Ben J. Palanca, and ShiNung Ching. Bounded-observation kalman filtering of correlation in multivariate neural recordings. *36th Annu. Int. Conf. Eng. Med. Biol.*, pages 5052–5055, 2014.

[12] MohammadMehdi Kafashan, ShiNung Ching, and Ben J. Palanca. Spatiotemporal functional connectivity motifs persist during wakefulness and sevoflurane anesthesia. *Frontiers in Neural Circuit (Under Review)*, 2016.

[13] MohammadMehdi Kafashan, Shoko Ryu, Debra Roberts, Akshay Thonthakudi, Lawrence Eisenman, Terrance Kummer, and ShiNung Ching. Electroencephalographic features for classification of depressed level of consciousness: Focal versus diffuse etiologies. *Brain Informatics (Under Review)*, 2016.

[14] MohammadMehdi Kafashan, Anirban Nandi, and ShiNung Ching. Relating observability and compressed sensing of time-varying signals in recurrent linear networks. *Neural Networks*, 83:11–20, 2016.

[15] Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36, 2011.

[16] Angie Fasoula, Yohan Attal, and Denis Schwartz. Comparative performance evaluation of data-driven causality measures applied to brain networks. *J. Neurosci. Methods*, 215(2):170–189, 2013.

[17] MohammadMehdi Kafashan, Anirban Nandi, and ShiNung Ching. Recovery of sparse inputs to over-actuated linear systems. *arXiv:1408.0202 Preprint*, 2014.

[18] Mouhsin M Shafi, M Brandon Westover, Michael D Fox, and Alvaro Pascual-Leone. Exploration and modulation of brain network interactions with noninvasive brain stimulation in combination with neuroimaging. *Eur. J. Neurosci.*, 35(6):805–825, 2012.

[19] Hartwig Siebner and John Rothwell. Transcranial magnetic stimulation: new insights into representational cortical plasticity. *Exp. Brain Res.*, 148(1):1–16, 2003.

[20] Svenja Borchers, Marc Himmelbach, Nikos Logothetis, and Hans-Otto Karnath. Direct electrical stimulation of human cortex—the gold standard for mapping brain functions? *Nat. Rev. Neurosci.*, 13(1):63–70, 2012.

[21] Kyle Q. Lepage, Mark A. Kramer, and Shinung Ching. An active method for tracking connectivity in temporally changing brain networks. *Annu. Int. Conf. Eng. Med. Biol.*, *35th*, pages 4374–4377, 2013.

[22] Kyle Q. Lepage, ShiNung Ching, and Mark A. Kramer. Inferring evoked brain connectivity through adaptive perturbation. *J. Comput. Neurosci.*, 34(2):303–318, 2013.

[23] Moon-Ho Ringo Ho, Hernando Ombao, and Robert Shumway. A state-space approach to modelling brain dynamics. *Stat. Sinica*, 15:407–425, 2005.

[24] Jin Kang, Liang Wang, Chaogan Yan, Jinhui Wang, Xia Liang, and Yong He. Characterizing dynamic functional connectivity in the resting brain using variable parameter regression and kalman filtering approaches. *Neuroimage*, 56(3):1222–1234, 2011.

[25] Martin Havlicek, Jiri Jan, Milan Brazdil, and Vince D Calhoun. Dynamic granger causality based on kalman filter for evaluation of functional network connectivity in fmri data. *Neuroimage*, 53(1):65–77, 2010.

[26] Sourabh Bhattacharya, Moon-Ho Ringo Ho, and Sumitra Purkayastha. A bayesian approach to modeling dynamic effective connectivity with fmri data. *Neuroimage*, 30(3):794–812, 2006.

[27] Fei Gu, Kristopher J Preacher, and Emilio Ferrer. A state space modeling approach to mediation analysis. *J. Educ. Behav. Stat.*, 39(2):117–143, 2014.

[28] Bing Leung Patrick Cheung, Brady Alexander Riedner, Giulio Tononi, and Bart van Veen. Estimation of cortical connectivity from eeg using state-space models. *IEEE Trans. Biomed. Eng.*, 57(9):2122–2134, 2010.

[29] Kyle Q Lepage and ShiNung Ching. On the output of nonlinear systems excited by discrete prolate spheroidal sequences. *arXiv:1311.4938 Preprint*, 2013.

[30] Peter Goos and Bradley Jones. *Optimal design of experiments: a case study approach.* John Wiley & Sons, 2011.

[31] Simon S Haykin et al. *Kalman filtering and neural networks.* Wiley Online Library, 2001.

[32] Todd K Moon and Wynn C Stirling. *Mathematical methods and algorithms for signal processing.* Prentice Hall Upper Saddle River, 2000.

[33] Jouni Hartikainen, Arno Solin, and Simo Särkkä. Optimal filtering with kalman filters and smoothers. *Technical Report, Aalto University School of Science, Department of Biomedical Engineering and Computational Science*, 2011.

[34] Yilin Mo, Roberto Ambrosino, and Bruno Sinopoli. Sensor selection strategies for state estimation in energy constrained wireless sensor networks. *Automatica*, 47(7):1330–1338, 2011.

[35] James E Weimer, Bruno Sinopoli, and Bruce H Krogh. A relaxation approach to dynamic sensor selection in large-scale wireless networks. *Int. Conf. Distributed Comput. Syst.*, pages 501–506, 2008.

[36] Sijia Liu, Makan Fardad, Engin Masazade, and Pramod K Varshney. Optimal periodic sensor scheduling in large-scale dynamical networks. *arXiv:1305.5601 Preprint*, 2013.

[37] Mohammad Rezaeian. Sensor scheduling for optimal observability using estimation entropy. In *Proc. 5th Annu. IEEE Int. Conf. Pervasive Computing and Commun. Workshops (PerCom Workshops' 07)*, pages 307–312, 2007.

[38] Fang Bian, David Kempe, and Ramesh Govindan. Utility based sensor selection. In *Proc. 5th Int. Conf. Information Process. Sensor Netw.*, pages 11–18. ACM, 2006.

[39] Michael P Vitus, Wei Zhang, Alessandro Abate, Jianghai Hu, and Claire J Tomlin. On efficient sensor scheduling for linear dynamical systems. *Automatica*, 48(10):2482–2493, 2012.

[40] Peter Alriksson and Anders Rantzer. Sub-optimal sensor scheduling with error bounds. In *Proc. 16th IFAC World Congress, (Prague, Czech Republic)*, 2005.

[41] Siddharth Joshi and Stephen Boyd. Sensor selection via convex optimization. *IEEE Trans. Signal Process.*, 57(2):451–462, 2009.

[42] Vijay Gupta, Timothy H Chung, Babak Hassibi, and Richard M Murray. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42(2):251–260, 2006.

[43] Marco F Huber, Achim Kuwertz, Felix Sawo, and Uwe D Hanebeck. Distributed greedy sensor scheduling for model-based reconstruction of space-time continuous physical phenomena. *12th Int. Conf. Inf. Fusion*, pages 102–109, 2009.

[44] Manohar Shamaiah, Siddhartha Banerjee, and Haris Vikalo. Greedy sensor selection: Leveraging submodularity. In *Proc. 49th IEEE Annu. Conf. Decision and Control (CDC)*, pages 2572–2577, 2010.

[45] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294, 1978.

[46] Marshall L Fisher, George L Nemhauser, and Laurence A Wolsey. An analysis of approximations for maximizing submodular set functions—II. *Polyhedral Combinatorics*, pages 73–87, 1978.

[47] Abhimanyu Das and David Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. *arXiv Preprint arXiv:1102.3975*, 2011.

[48] Syed Talha Jawaid and Stephen L. Smith. On the submodularity of sensor scheduling for estimation of linear dynamical systems. *Am. Control Conf.*, 2014.

[49] Zoubin Ghahramani and Geoffrey E Hinton. Parameter estimation for linear dynamical systems. Technical report, Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, 1996.

[50] Robert H Shumway and David S Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.

[51] Hugh R Wilson and Jack D Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80, 1973.

[52] Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, 12(1):1–24, 1972.

[53] Yuya Maruyama, Yuta Kakimoto, and Osamu Araki. Analysis of chaotic oscillations induced in two coupled wilson–cowan models. *Biol. Cybernetics*, 108(3):355–363, 2014.

[54] Gitanjali Yadav and Suresh Babu. Nexcade: perturbation analysis for complex networks. *PLOS ONE*, 7(8):e41827, 2012.

[55] Xin Yan, Yang Wu, Xiaohui Li, Chunlin Li, and Yaogai Hu. Eigenvector perturbations of complex networks. *Phys. A*, 408:106–118, 2014.

[56] Nicholas P Franks. General anaesthesia: from molecular targets to neuronal pathways of sleep and arousal. *Nature Reviews Neuroscience*, 9(5):370–386, 2008.

[57] Robert S Schwartz, Emery N Brown, Ralph Lydic, and Nicholas D Schiff. General anesthesia, sleep, and coma. *New England Journal of Medicine*, 363(27):2638–2650, 2010.

[58] Denis Jordan, Rüdiger Ilg, Valentin Riedl, Anna Schorer, Sabine Grimberg, Susanne Neufang, Adem Omerovic, Sebastian Berger, Gisela Untergehrer, Christine Preibisch, et al. Simultaneous electroencephalographic and functional magnetic resonance imaging indicate impaired cortical top–down processing in association with anesthetic-induced unconsciousness. *The Journal of the American Society of Anesthesiologists*, 119(5):1031–1042, 2013.

[59] Oluwaseun Akeju, Kara J Pavone, M Brandon Westover, Rafael Vazquez, Michael J Prerau, Priscilla G Harrell, Katharine E Hartnack, James Rhee, Aaron L Sampson,

Kathleen Habeeb, et al. A comparison of propofol-and dexmedetomidine-induced electroencephalogram dynamics using spectral and coherence analysis. *The Journal of the American Society of Anesthesiologists*, 121(5):978–989, 2014.

[60] Oluwaseun Akeju, M Brandon Westover, Kara J Pavone, Aaron L Sampson, Katharine E Hartnack, Emery N Brown, and Patrick L Purdon. Effects of sevoflurane and propofol on frontal electroencephalogram power and coherence. *The Journal of the American Society of Anesthesiologists*, 121(5):990–998, 2014.

[61] Stefanie Blain-Moraes, U Lee, S Ku, G Noh, and George A Mashour. Electroencephalographic effects of ketamine on power, cross-frequency coupling, and connectivity in the alpha bandwidth. *Front Syst Neurosci*, 8:114, 2014.

[62] Stefanie Blain-Moraes, Vijay Tarnal, Giancarlo Vanini, Amir Alexander, Derek Rosen, Brenna Shortal, Ellen Janke, and George A Mashour. Neurophysiological correlates of sevoflurane-induced unconsciousness. *The Journal of the American Society of Anesthesiologists*, 122(2):307–316, 2015.

[63] K Kaskinoro, A Maksimow, S Georgiadis, J Långsjö, H Scheinin, P Karjalainen, and SK Jääskeläinen. Electroencephalogram reactivity to verbal command after dexmedetomidine, propofol and sevoflurane-induced unresponsiveness. *Anaesthesia*, 70(2):190–204, 2015.

[64] Patrick L Purdon, Aaron Sampson, Kara J Pavone, and Emery N Brown. Clinical electroencephalography for anesthesiologistspart i: Background and basic signatures. *The Journal of the American Society of Anesthesiologists*, 123(4):937–960, 2015.

[65] JL Vincent, GH Patel, MD Fox, AZ Snyder, JT Baker, DC Van Essen, JM Zempel, LH Snyder, M Corbetta, and ME Raichle. Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, 447(7140):83–86, 2007.

[66] Scott J Peltier, Chantal Kerssens, Stephan B Hamann, Peter S Sebel, Michael Byas-Smith, and Xiaoping Hu. Functional connectivity changes with concentration of sevoflurane anesthesia. *Neuroreport*, 16(3):285–288, 2005.

[67] Pierre Boveroux, Audrey Vanhaudenhuyse, Marie-Aurélie Bruno, Quentin Noirhomme, Séverine Lauwick, André Luxen, Christian Degueldre, Alain Plenevaux, Caroline Schnakers, Christophe Phillips, et al. Breakdown of within-and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. *The Journal of the American Society of Anesthesiologists*, 113(5):1038–1053, 2010.

[68] Ben Julian A Palanca, Anish Mitra, Linda Larson-Prior, Abraham Z Snyder, Michael S Avidan, and Marcus E Raichle. Resting-state functional magnetic resonance imaging

correlates of sevoflurane-induced unconsciousness. *The Journal of the American Society of Anesthesiologists*, 123(2):346–356, 2015.

[69] Jin Kang, Liang Wang, Chaogan Yan, Jinhui Wang, Xia Liang, and Yong He. Characterizing dynamic functional connectivity in the resting brain using variable parameter regression and kalman filtering approaches. *Neuroimage*, 56(3):1222–1234, 2011.

[70] R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Peter A Bandettini, Vince D Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H Duyn, Gary H Glover, Javier Gonzalez-Castillo, et al. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378, 2013.

[71] Vince D Calhoun, Robyn Miller, Godfrey Pearlson, and Tulay Adalı. The chronnectome: time-varying connectivity networks as the next frontier in fmri data discovery. *Neuron*, 84(2):262–274, 2014.

[72] Shella Dawn Keilholz. The neural basis of time-varying resting-state functional connectivity. *Brain connectivity*, 4(10):769–779, 2014.

[73] Pablo Barttfeld, Lynn Uhrig, Jacobo D Sitt, Mariano Sigman, Béchir Jarraya, and Stanislas Dehaene. Signature of consciousness in the dynamics of resting-state brain activity. *Proceedings of the National Academy of Sciences*, 112(3):887–892, 2015.

[74] R Matthew Hutchison, Thilo Womelsdorf, Joseph S Gati, Stefan Everling, and Ravi S Menon. Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Human brain mapping*, 34(9):2154–2177, 2013.

[75] Enrico Amico, Francisco Gomez, Carol Di Perri, Audrey Vanhaudenhuyse, Damien Lesenfants, Pierre Boveroux, Vincent Bonhomme, Jean-François Brichant, Daniele Marinazzo, and Steven Laureys. Posterior cingulate cortex-related co-activation patterns: a resting state fmri study in propofol-induced loss of consciousness. *PloS one*, 9(6):e100012, 2014.

[76] Zhifeng Liang, Xiao Liu, and Nanyin Zhang. Dynamic resting state functional connectivity in awake and anesthetized rodents. *NeuroImage*, 104:89–99, 2015.

[77] Garth John Thompson, Wen-Ju Pan, and Shella Dawn Keilholz. Different dynamic resting state fmri patterns are linked to different frequencies of neural activity. *Journal of neurophysiology*, 114(1):114–124, 2015.

[78] Jin Kang, Liang Wang, Chaogan Yan, Jinhui Wang, Xia Liang, and Yong He. Characterizing dynamic functional connectivity in the resting brain using variable parameter regression and kalman filtering approaches. *Neuroimage*, 56(3):1222–1234, 2011.

[79] Jessica S Damoiseaux and Michael D Greicius. Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Structure and Function*, 213(6):525–533, 2009.

[80] Maciej Kamiński, Mingzhou Ding, Wilson A Truccolo, and Steven L Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological cybernetics*, 85(2):145–157, 2001.

[81] Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.

[82] Anish Mitra and Marcus E Raichle. How networks communicate: propagation patterns in spontaneous brain activity. *Phil. Trans. R. Soc. B*, 371(1705):20150546, 2016.

[83] Steven L Bressler and Anil K Seth. Wiener–granger causality: a well established methodology. *Neuroimage*, 58(2):323–329, 2011.

[84] Katarzyna J Blinowska. Review of the methods of determination of directed connectivity from multichannel data. *Medical & biological engineering & computing*, 49(5):521–529, 2011.

[85] David T Jones, Prashanthi Vemuri, Matthew C Murphy, Jeffrey L Gunter, Matthew L Senjem, Mary M Machulda, Scott A Przybelski, Brian E Gregg, Kejal Kantarci, David S Knopman, et al. Non-stationarity in the "resting brain's" modular architecture. *PloS one*, 7(6):e39731, 2012.

[86] Vesa Kiviniemi, Tapani Vire, Jukka Remes, Ahmed Abou Elseoud, Tuomo Starck, Osmo Tervonen, and Juha Nikkinen. A sliding time-window ica reveals spatial variability of the default mode network in time. *Brain connectivity*, 1(4):339–347, 2011.

[87] Elena A Allen, Eswar Damaraju, Sergey M Plis, Erik B Erhardt, Tom Eichele, and Vince D Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, page bhs352, 2012.

[88] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[89] Ricardo Pio Monti, Peter Hellyer, David Sharp, Robert Leech, Christoforos Anagnostopoulos, and Giovanni Montana. Estimating time-varying brain connectivity networks from functional mri time series. *NeuroImage*, 103:427–443, 2014.

[90] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[91] MohammadMehdi Kafashan, Ben J Palanca, and ShiNung Ching. Bounded-observation kalman filtering of correlation in multivariate neural recordings. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5052–5055, 2014.

[92] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82(1):35–45, 1960.

[93] Rudolph Emi Kalman. Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5(2):102–119, 1960.

[94] Timothy O Laumann, Abraham Z Snyder, Anish Mitra, Evan M Gordon, Caterina Gratton, Babatunde Adeyemo, Adrian W Gilmore, Steven M Nelson, Jeff J Berg, Deanna J Greene, et al. On the stability of bold fmri correlations. *Cerebral Cortex*, 2016.

[95] Giulio Tononi. An information integration theory of consciousness. *BMC neuroscience*, 5(1):1, 2004.

[96] ShiNung Ching, Aylin Cimenser, Patrick L Purdon, Emery N Brown, and Nancy J Kopell. Thalamocortical model for a propofol-induced $\alpha$-rhythm associated with loss of consciousness. *Proceedings of the National Academy of Sciences*, 107(52):22665–22670, 2010.

[97] Levent Ertöz, Michael Steinbach, and Vipin Kumar. Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data. In *SDM*, pages 47–58. SIAM, 2003.

[98] Alison K Godbolt, Catharina Nygren DeBoussard, Maud Stenberg, Marie Lindgren, Trandur Ulfarsson, and Jörgen Borg. Disorders of consciousness after severe traumatic brain injury: A swedish–icelandic study of incidence, outcomes and implications for optimizing care pathways. *Journal of rehabilitation medicine*, 45(8):741–748, 2013.

[99] John Whyte, Risa Nakase-Richardson, Flora M Hammond, Shane McNamee, Joseph T Giacino, Kathleen Kalmar, Brian D Greenwald, Stuart A Yablon, and Lawrence J Horn. Functional outcomes in traumatic disorders of consciousness: 5-year outcomes from the national institute on disability and rehabilitation research traumatic brain injury model systems. *Archives of physical medicine and rehabilitation*, 94(10):1855–1860, 2013.

[100] Joseph T Giacino, Joseph J Fins, Steven Laureys, and Nicholas D Schiff. Disorders of consciousness after acquired brain injury: the state of the science. *Nature Reviews Neurology*, 10(2):99–114, 2014.

[101] Risa Nakase-Richardson, John Whyte, Joseph T Giacino, Shital Pavawalla, Scott D Barnett, Stuart A Yablon, Mark Sherer, Kathleen Kalmar, Flora M Hammond, Brian Greenwald, et al. Longitudinal outcome of patients with disordered consciousness in the nidrr tbi model systems programs. *Journal of neurotrauma*, 29(1):59–65, 2012.

[102] Steven Laureys and Mélanie Boly. The changing spectrum of coma. *Nature Clinical Practice Neurology*, 4(10):544–546, 2008.

[103] G Bryan Young. The eeg in coma. *Journal of Clinical Neurophysiology*, 17(5):473–485, 2000.

[104] S Jacobson and H Jerrier. Eeg in delirium. *Seminars in clinical neuropsychiatry*, 5(2):86–92, 2000.

[105] ShiNung Ching, Max Y Liberman, Jessica J Chemali, M Brandon Westover, Jonathan D Kenny, Ken Solt, Patrick L Purdon, and Emery N Brown. Real-time closed-loop control in a rodent model of medically induced coma using burst suppression. *The Journal of the American Society of Anesthesiologists*, 119(4):848–860, 2013.

[106] Jan Claassen, Lawrence J Hirsch, Kurt T Kreiter, Evelyn Y Du, E Sander Connolly, Ronald G Emerson, and Stephan A Mayer. Quantitative continuous eeg for detecting delayed cerebral ischemia in patients with poor-grade subarachnoid hemorrhage. *Clinical neurophysiology*, 115(12):2699–2710, 2004.

[107] Baha Şen, Musa Peker, Abdullah Çavuşoğlu, and Fatih V Çelebi. A comparative study on classification of sleep stage based on eeg signals using feature selection and classification algorithms. *Journal of medical systems*, 38(3):1–21, 2014.

[108] J Claassen, SA Mayer, RG Kowalski, RG Emerson, and LJ Hirsch. Detection of electrographic seizures with continuous eeg monitoring in critically ill patients. *Neurology*, 62(10):1743–1748, 2004.

[109] Emery N Brown, Ralph Lydic, and Nicholas D Schiff. General anesthesia, sleep, and coma. *New England Journal of Medicine*, 363(27):2638–2650, 2010.

[110] Alice D Lam, Rodrigo Zepeda, Andrew J Cole, and Sydney S Cash. Widespread changes in network activity allow non-invasive detection of mesial temporal lobe seizures. *Brain*, page aww198, 2016.

[111] Jack R Smith, William F Funke, WC Yeo, and Ralph A Ambuehl. Detection of human sleep eeg waveforms. *Electroencephalography and clinical neurophysiology*, 38(4):435–437, 1975.

[112] Takasumi Katoh, Akira Suzuki, and Kazuyuki Ikeda. Electroencephalographic derivatives as a tool for predicting the depth of sedation and anesthesia induced by sevoflurane. *The Journal of the American Society of Anesthesiologists*, 88(3):642–650, 1998.

[113] Rémy Lehembre, Olivia Gosseries, Zulay Lugo, Zayd Jedidi, Camille Chatelle, Bernard Sadzot, Steven Laureys, and Quentin Noirhomme. Electrophysiological investigations of brain function in coma, vegetative and minimally conscious patients. *Archives italiennes de biologie*, 150(2-3):122–39, 2012.

[114] Jacobo Diego Sitt, Jean-Remi King, Imen El Karoui, Benjamin Rohaut, Frederic Faugeras, Alexandre Gramfort, Laurent Cohen, Mariano Sigman, Stanislas Dehaene, and Lionel Naccache. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain*, 137(8):2258–2270, 2014.

[115] Srivas Chennu, Paola Finoia, Evelyn Kamau, Judith Allanson, Guy B Williams, Martin M Monti, Valdas Noreika, Aurina Arnatkeviciute, Andrés Canales-Johnson, Francisco Olivares, et al. Spectral signatures of reorganised brain networks in disorders of consciousness. *PLoS Comput Biol*, 10(10):e1003887, 2014.

[116] Frederic Zubler, Christa Koenig, Andreas Steimer, Stephan M Jakob, Kaspar A Schindler, and Heidemarie Gast. Prognostic and diagnostic value of eeg signal coupling measures in coma. *Clinical Neurophysiology*, 2015.

[117] Dimitris Kugiumtzis and Alkiviadis Tsimpiris. Measures of analysis of time series (mats): A matlab toolkit for computation of multiple measures on time series data bases. *arXiv preprint arXiv:1002.1940*, 2010.

[118] E Olofsen, JW Sleigh, and A Dahan. Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *British journal of anaesthesia*, 101(6):810–821, 2008.

[119] Supervised Machine Learning. A review of classification techniques. *Department of Computer Science and Technology University of Peloponnese, Greece*, 2007.

[120] Max Kuhn. A short introduction to the caret package. *R Project website. cran. r-project. org/web/packages/caret/vignettes/caret. pdf. Published August*, 6, 2015.

[121] SM Kay. Modern spectral estimation: Theory and application.

[122] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.

[123] Michel JAM van Putten, Jurriaan M Peters, Sandra M Mulder, Jan AM de Haas, Cornelis MA Bruijninckx, and Dénes LJ Tavy. A brain symmetry index (bsi) for online eeg monitoring in carotid endarterectomy. *Clinical neurophysiology*, 115(5):1189–1194, 2004.

[124] Cecile C de Vos, Susan M van Maarseveen, Paul JAM Brouwers, and Michel JAM van Putten. Continuous eeg monitoring during thrombolysis in acute hemispheric stroke patients using the brain symmetry index. *Journal of Clinical Neurophysiology*, 25(2):77–82, 2008.

[125] M Cursi, MV Meraviglia, GF Fanelli, R Chiesa, A Tirelli, G Comi, and F Minicucci. Electroencephalographic background desynchronization during cerebral blood flow reduction. *Clinical neurophysiology*, 116(11):2577–2585, 2005.

[126] Quentin Noirhomme, Rémy Lehembre, Zulay del Rosario Lugo, Damien Lesenfants, André Luxen, Steven Laureys, Mauro Oddo, and Andrea O Rossetti. Automated analysis of background eeg and reactivity during therapeutic hypothermia in comatose patients after cardiac arrest. *Clinical EEG and neuroscience*, page 1550059413509616, 2014.

[127] Cenker Eken, Mutlu Kartal, Ayse Bacanli, and Oktay Eray. Comparison of the full outline of unresponsiveness score coma scale and the glasgow coma scale in an emergency setting population. *European journal of emergency medicine*, 16(1):29–36, 2009.

[128] Joseph T Giacino, Kathleen Kalmar, and John Whyte. The jfk coma recovery scale-revised: measurement characteristics and diagnostic utility. *Archives of physical medicine and rehabilitation*, 85(12):2020–2029, 2004.

[129] Sairam Parthasarathy and Martin J Tobin. Sleep in the intensive care unit. *Intensive care medicine*, 30(2):197–206, 2004.

[130] A Temko, E Thomas, W Marnane, G Lightbody, and G Boylan. Eeg-based neonatal seizure detection with support vector machines. *Clinical Neurophysiology*, 122(3):464–473, 2011.

[131] Kaveh Samiee, Péter Kovács, and Moncef Gabbouj. Epileptic seizure classification of eeg time-series using rational discrete short-time fourier transform. *IEEE Transactions on Biomedical Engineering*, 62(2):541–552, 2015.

[132] Klaus-Robert Müller, Michael Tangermann, Guido Dornhege, Matthias Krauledat, Gabriel Curio, and Benjamin Blankertz. Machine learning for real-time single-trial eeg-analysis: from brain–computer interfacing to mental state monitoring. *Journal of neuroscience methods*, 167(1):82–90, 2008.

[133] Jennifer T Coull and Anna C Nobre. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both pet and fmri. *The Journal of Neuroscience*, 18(18):7426–7435, 1998.

[134] Anna C Nobre, GN Sebestyen, DR Gitelman, MM Mesulam, RS Frackowiak, and CD Frith. Functional localization of the system for visuospatial attention using positron emission tomography. *Brain*, 120(3):515–533, 1997.

[135] Lucina Q Uddin, AM Clare Kelly, Bharat B Biswal, F Xavier Castellanos, and Michael P Milham. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Human brain mapping*, 30(2):625–637, 2009.

[136] Peter Fransson. How default is the default mode of brain function?: Further evidence from intrinsic bold signal fluctuations. *Neuropsychologia*, 44(14):2836–2845, 2006.

[137] L Nyberg, AR Mclntosh, S Houle, L-G Nilsson, and Endel Tulving. Activation of medial temporal structures during episodic memory retrieval. *Nature*, 380(6576):715–717, 1996.

[138] Md Nurujjaman, Ramesh Narayanan, and AN Sekar Iyengar. Comparative study of nonlinear properties of eeg signals of normal persons and epileptic patients. *Nonlinear biomedical physics*, 3(1):1, 2009.

[139] Mandyam V Srinivasan, Simon B Laughlin, and Andreas Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B: Biological Sciences*, 216(1205):427–459, 1982.

[140] Roland Baddeley, Larry F Abbott, Michael CA Booth, Frank Sengpiel, Tobe Freeman, Edward A Wakeman, and Edmund T Rolls. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London B: Biological Sciences*, 264(1389):1775–1783, 1997.

[141] Daniel J Graham, Damon M Chandler, and David J Field. Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, 46(18):2901–2913, 2006.

[142] Florian Blättler and Richard HR Hahnloser. An efficient coding hypothesis links sparsity and selectivity of neural responses. *PloS one*, 6(10):e25506, 2011.

[143] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696, 2009.

[144] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, 11:19–60, 2010.

[145] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems*, pages 801–808, 2006.

[146] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.

[147] Yonina C Eldar and Gitta Kutyniok. *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[148] Jarvis Haupt, Waheed U Bajwa, Michael Rabbat, and Robert Nowak. Compressed sensing for networked data. *IEEE Signal Process. Mag.*, 25(2):92–101, 2008.

[149] Surya Ganguli and Haim Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual review of neuroscience*, 35:485–508, 2012.

[150] Victor J Barranca, Gregor Kovačič, Douglas Zhou, and David Cai. Network dynamics for optimal compressive-sensing input-signal recovery. *Phys. Rev. E*, 90(4):042908, 2014.

[151] David L Donoho. For most large underdetermined systems of linear equations the minimal $l_1$ norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59(6):797–829, 2006.

[152] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[153] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005.

[154] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[155] Namrata Vaswani. Kalman filtered compressed sensing. In *Proc. 15th IEEE International Conference on Image Processing*, pages 893–896, 2008.

[156] Adam Charles, M Salman Asif, Justin Romberg, and Christopher Rozell. Sparsity penalties in dynamical system estimation. In *Proc. 45th IEEE Annual Conference on Information Sciences and Systems (CISS),*, pages 1–6, 2011.

[157] Michael B Wakin, Borhan M Sanandaji, and Tyrone L Vincent. On the observability of linear systems from random, compressive measurements. In *Proc. 49th IEEE Conference on Decision and Control (CDC)*, pages 4447–4454, 2010.

[158] Borhan M Sanandaji, Tyrone L Vincent, Michael B Wakin, Roland Tóth, and Kameshwar Poolla. Compressive system identification of lti and ltv arx models. In *Proc. 50th*

*IEEE Conf. on Decision and Control and European Control Conference (CDC-ECC)*, pages 791–798, 2011.

[159] Domenico Napoletani and Timothy D Sauer. Reconstructing the topology of sparsely connected dynamical networks. *Phys. Rev. E*, 77(2):026103, 2008.

[160] Yasser Shoukry, Pierluigi Nuzzo, Alberto Puggelli, Alberto L Sangiovanni-Vincentelli, Sanjit A Seshia, and Paulo Tabuada. Secure state estimation under sensor attacks: A satisfiability modulo theory approach. *arXiv preprint arXiv:1412.4324*, 2014.

[161] Hamza Fawzi, Paulo Tabuada, and Suhas Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Trans. Automatic Control*, 59(6):1454–1467, 2014.

[162] Demba E Ba, Behtash Babadi, Patrick L Purdon, and Emery N Brown. Exact and stable recovery of sequences of signals with sparse increments via differential $l_1$ minimization. In *Proc. Advances in Neural Information Processing Systems (NIPS)*, pages 2636–2644, 2012.

[163] Shahin Sefati, Noah J Cowan, and René Vidal. Linear systems with sparse inputs: Observability and input recovery. In *Proc. IEEE American Control Conference (ACC)*, pages 5251–5257, 2015.

[164] H Jaeger. Short term memory in echo state networks. *GMD report, German National Research Center for Information Technology*, 2001.

[165] Olivia L White, Daniel D Lee, and Haim Sompolinsky. Short-term memory in orthogonal neural networks. *Phys. Rev. Lett.*, 92(14):148102, 2004.

[166] Surya Ganguli, Dongsung Huh, and Haim Sompolinsky. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, 105(48):18970–18975, 2008.

[167] Michiel Hermans and Benjamin Schrauwen. Memory in linear recurrent neural networks in continuous time. *Neural Networks*, 23(3):341–355, 2010.

[168] Edward Wallace, Hamid Reza Maei, and Peter E Latham. Randomly connected networks have short temporal memory. *Neural Comput.*, 25(6):1408–1439, 2013.

[169] Surya Ganguli and Haim Sompolinsky. Short-term memory in neuronal networks through dynamical compressed sensing. In *Proc. Advances in neural information processing systems*, pages 667–675, 2010.

[170] Adam S Charles, Han Lun Yap, and Christopher J Rozell. Short-term memory capacity in networks via the restricted isometry property. *Neural Comput.*, 26(6):1198–1235, 2014.

[171] Iori Ito, Rose Chik-Ying Ong, Baranidharan Raman, and Mark Stopfer. Sparse odor representation and olfactory learning. *Nat. Neurosci.*, 11(10):1177–1184, Oct 2008.

[172] Baranidharan Raman, Joby Joseph, Jeff Tang, and Mark Stopfer. Temporally diverse firing patterns in olfactory receptor neurons underlie spatiotemporal neural codes for odors. *J. Neurosci.*, 30(6):1994–2006, Feb 2010.

[173] Srdjan Ostojic, Nicolas Brunel, and Vincent Hakim. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J. Neurosci.*, 29(33):10234–10253, 2009.

[174] Larry F Abbott. Lapicqueś introduction of the integrate-and-fire model neuron (1907). *Brain Res. Bull.*, 50(5):303–304, 1999.

[175] P. Dayan and L.F. Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Comput. Neurosci. Cambridge, MA, USA: Massachusetts Institute of Technology Press, 2005.

[176] Eugene M Izhikevich et al. Simple model of spiking neurons. *IEEE Trans. Neural Netw.*, 14(6):1569–1572, 2003.

[177] Rodney J Douglas and Kevan AC Martin. Recurrent neuronal circuits in the neocortex. *Curr. Biol.*, 17(13):R496–R500, 2007.

[178] Teuvo Kohonen and Erkki Oja. Fast adaptive formation of orthogonalizing filters and associative memory in recurrent networks of neuron-like elements. *Biol. Cybern.*, 21(2):85–95, 1976.

[179] H Sebastian Seung, Daniel D Lee, Ben Y Reis, and David W Tank. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron*, 26(1):259–271, 2000.

[180] Nihal Fatma Güler, Elif Derya Übeyli, and İnan Güler. Recurrent neural networks employing lyapunov exponents for EEG signals classification. *Expert Syst. Appl.*, 29(3):506–514, 2005.

[181] Christian W Omlin and C Lee Giles. Extraction of rules from discrete-time recurrent neural networks. *Neural Netw.*, 9(1):41–52, 1996.

[182] Bernard Ghanem and Narendra Ahuja. Sparse coding of linear dynamical systems with an application to dynamic texture recognition. In *Proc. 20th IEEE Int. Conf. Pattern Recog. (ICPR)*, pages 987–990, 2010.

[183] Xian Wei, Hao Shen, and Martin Kleinsteuber. An adaptive dictionary learning approach for modeling dynamical textures. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pages 3567–3571. IEEE, 2014.

[184] Lei Li. Fast algorithms for mining co-evolving time series. Technical Report CMU-CS-11-127., Carnegie Inst. Tech., Dept. Computer Science, 2011.

[185] Yufei Tao, Christos Faloutsos, Dimitris Papadias, and Bin Liu. Prediction and indexing of moving objects with unknown motion patterns. In *Proc. ACM SIGMOD Int. Conf. Management of data*, pages 611–622, 2004.

[186] Shanmugavelayutham Muthukrishnan. *Data streams: Algorithms and applications.* New Brunswick, NJ, USA: Now Publishers Inc, 2005.

[187] R Gray Parker and Ronald L Rardin. *Discrete optimization.* San Diego, CA, USA: Academic Press Professional, Inc., 1988.

[188] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

[189] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, 2006.

[190] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[191] Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Uniform uncertainty principle for bernoulli and subgaussian ensembles. *Constr. Approx.*, 28(3):277–289, 2008.

[192] Emmanuel Candes, Laurent Demanet, David Donoho, and Lexing Ying. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899, 2006.

[193] Emmanuel J Candès and David L Donoho. New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities. *Commun. Pure Appl. Math.*, 57(2):219–266, 2004.

[194] Inc. CVX Research. CVX: Matlab software for disciplined convex programming, version 2.0, August 2012.

[195] Michael C Grant and Stephen P Boyd. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, pages 95–110. Springer, 2008.

[196] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[197] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. *Image Processing, IEEE Transactions on*, 16(12):2992–3004, 2007.

[198] Alison L Barth and James FA Poulet. Experimental evidence for sparse firing in the neocortex. *Trends in Neurosci.*, 35(6):345–355, 2012.

[199] Mahmood S Hoseini and Ralf Wessel. Coherent and intermittent ensemble oscillations emerge from networks of irregular spiking neurons. *Journal of neurophysiology*, 115(1):457–469, 2016.

[200] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440–442, 1998.

[201] Paul D King, Joel Zylberberg, and Michael R DeWeese. Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of v1. *The Journal of Neuroscience*, 33(13):5475–5485, 2013.

[202] Hong Jung and Jong Chul Ye. Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques. *International Journal of Imaging Systems and Technology*, 20(2):81–98, 2010.

[203] Fred Attneave. Some informational aspects of visual perception. *Psychological Review*, 61(3):183, 1954.

[204] Simon B Laughlin. Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology*, 11(4):475–480, 2001.

[205] William Bialek, Rob R de Ruyter Van Steveninck, and Naftali Tishby. Efficient representation as a design principle for neural coding and computation. In *2006 IEEE International Symposium on Information Theory*, pages 659–663, 2006.

[206] Pietro Vertechi, Wieland Brendel, and Christian K Machens. Unsupervised learning of an efficient short-term memory network. In *Advances in Neural Information Processing Systems*, pages 3653–3661, 2014.

[207] Mandyam V Srinivasan, Simon B Laughlin, and Andreas Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B: Biological Sciences*, 216(1205):427–459, 1982.

[208] Daniel J Graham, Damon M Chandler, and David J Field. Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, 46(18):2901–2913, 2006.

[209] Odelia Schwartz and Eero P Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825, 2001.

[210] Evan C Smith and Michael S Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–982, 2006.

[211] Yuan Sophie Liu, Charles F Stevens, and Tatyana O Sharpee. Predictable irregularities in retinal receptive fields. *Proceedings of the National Academy of Sciences*, 106(38):16499–16504, 2009.

[212] Guy Major and David Tank. Persistent neural activity: prevalence and mechanisms. *Current Opinion in Neurobiology*, 14(6):675–684, 2004.

[213] Eizaburo Doi, Jeffrey L Gauthier, Greg D Field, Jonathon Shlens, Alexander Sher, Martin Greschner, Timothy A Machado, Lauren H Jepson, Keith Mathieson, Deborah E Gunning, et al. Efficient coding of spatial information in the primate retina. *The Journal of Neuroscience*, 32(46):16256–16264, 2012.

[214] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696, 2009.

[215] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, 11:19–60, 2010.

[216] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems*, pages 801–808, 2006.

[217] Bruno A Olshausen and David J Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997.

[218] Bruno A Olshausen and David J Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, 2004.

[219] Ralph Linsker. Local synaptic learning rules suffice to maximize mutual information in a linear network. *Neural Computation*, 4(5):691–702, 1992.

[220] John J Hopfield, DI Feinstein, and RG Palmer. 'unlearning' has a stabilizing effect in collective memories. *Nature*, 304:158–159, 1983.

[221] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):123–231, 2013.

[222] Heinz H Bauschke, Rafal Goebel, Yves Lucet, and Xianfu Wang. The proximal average: basic theory. *SIAM Journal on Optimization*, 19(2):766–785, 2008.

[223] Xi Chen, Qihang Lin, Seyoung Kim, Jaime G Carbonell, Eric P Xing, et al. Smoothing proximal gradient method for general structured sparse regression. *The Annals of Applied Statistics*, 6(2):719–752, 2012.

[224] Robert L Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, 1953.

[225] Gholamreza Salimi-Khorshidi, Thomas E Nichols, Stephen M Smith, and Mark W Woolrich. Using gaussian-process regression for meta-analytic neuroimaging inference based on sparse observations. *Medical Imaging, IEEE Transactions on*, 30(7):1401–1416, 2011.

[226] Zhe Chen. Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69, 2003.

# Vita

MohammadMehdi Kafashan

| | |
|---|---|
| **Degrees** | Ph.D. Electrical Engineering, Washington University in St. Louis, Missouri, USA, December 2016 |
| | M.Sc. Electrical Engineering, Sharif University of Technology, Tehran, Iran, May 2011 |
| | B.Sc. Electrical Engineering, Amirkabir University of Technology, Tehran, Iran, September 2008 |
| | |
| **Professional Membership** | The Institute of Electrical and Electronics Engineers (IEEE) |
| | Society for Neuroscience (SFN) |
| | |
| **Publications** | Journal Publications: |

**M. Kafashan**, and S. Ching, "Optimal Stimulus Scheduling for Active Estimation of Evoked Brain Networks," *Journal of Neural Engineering*, 12(6), 066011, 2015.

**M. Kafashan**, A. Nandi, and S. Ching, "Relating Observability and Compressed Sensing of Time-Varying Signals in Recurrent Linear Networks," *Neural Networks*, 83, 11-20, 2016.

**M. Kafashan**, S. Ching, and B. Palanca, "Spatiotemporal Functional Connectivity Motifs Persist during Wakefulness and Sevoflurane Anesthesia," under review in *Frontiers in Neural Circuit*, 2016.

**M. Kafashan**, S. Ryu, D. Roberts, A. Thonthakudi, L. Eisenman, T. Kummer, and S. Ching, "Electroencephalographic Features for Classification of Depressed Level of Consciousness: Focal versus Diffuse Etiologies," under review in *Brain Informatics*, 2016.

A. Nandi, **M. Kafashan**, and S. Ching, "Control Analysis and Design for Statistical Models of Spiking Networks," under review in *IEEE Transactions on Control of Network Systems*, 2016.

Conference Publications:

**M. Kafashan**, P. Ben, and S. Ching, "Bounded-Observation Kalman Filtering of Correlation in Multivariate Neural Recordings," *IEEE Engineering in Medicine and Biology Society (EMBS)*, Chicago, USA, Aug. 2014.

**M. Kafashan**, K. Lepage, and S. Ching, "Node Selection for Probing Connections in Evoked Dynamic Networks," *IEEE Conference on Decision and Control (CDC)*, LA, USA, Dec. 2014.

G. Kumar, D. Menolascino, **M. Kafashan**, and S. Ching, "Controlling Linear Networks with Minimally Novel Inputs," *IEEE American Control Conference (ACC)*, Chicago, USA, Jul. 2015.

A. Nandi, **M. Kafashan**, and S. Ching, "Controlling Point Process Generalized Linear Models of Neural Spiking," *IEEE American Control Conference (ACC)*, Boston, USA, Jul. 2016.

Abstracts:

**M. Kafashan**, S. Ching, and B. Palanca, "Spatiotemporal Motifs of Correlated Interactions across Resting State Networks," *Society for Neuroscience (SFN)*, Chicago, USA, Oct. 2015.

**M. Kafashan**, S. Ryu, L. Eisenmann, T. Kummer, and S. Ching, "Disambiguating Structural and Non-Structural Coma using Network Reachability Analysis," *Society for Neuroscience (SFN)*, Chicago, USA, Oct. 2015.

December 2016