Understanding Neuronal Synchronisation in High-Dimensions

Carla J. Pinkney, BSc (Hons), MRes

Submitted for the degree of Doctor of Philosophy at Lancaster University.

December 2025

Abstract

Advances in neural recording technologies allow for the exciting opportunity to study large scale neural dynamics in the living brain. The focus of this thesis is to infer connectivity in the brain network from modern high-dimensional spike train data. These data, which represent the firing times of individual neurons, provide the gold standard for measuring localised neural activity. Statistically, neural spike trains can be modelled in continuous time, using the raw spikes, i.e., the exact time of neural firing, or in discrete time, when the spikes are transformed into binary data.

The contributions of this thesis are threefold, each addressing a distinct challenge in neural spike train analysis. In the first instance, we focus on advancing techniques in spectral analysis, to infer the dependence structure of a multivariate and high-dimensional point process. We propose novel methodology that combines a Whittle pseudo-likelihood with ridge or Lasso style penalties, improving the efficiency of spectral estimation in the point process framework. We establish both asymptotic and large sample properties for our proposed estimators, and assess their performance on synthetic data simulated from a multivariate Hawkes process. Finally, we apply our methodology to neural spike train data, demonstrating its ability to infer connectivity in the brain network.

As a second focus of this thesis, we propose a novel method to identify the effective (directional) connectivity of a population of neurons, under a binary time series representation of the neural data. In this framework, we directly account for non-stationary firing rates and propose an inference procedure to quantify the uncertainty associated with our estimated networks of neural interactions. We empirically validate the performance of our model and the inferential procedure, illustrating its ability to detect both excitatory and inhibitory neural interactions.

Lastly, we extend our discussion to the multi-subject setting, exploring the variability of functional and effective connectivity estimates across experimental subjects. We also make comparisons between our proposed methods and existing approaches for neuronal network estimation, in both the single and multi-subject setting.

Ultimately, this thesis contributes to the field of high-dimensional statistics, for the analysis of neural spiking data. In a world where neural recording technologies continue to advance, it is imperative that the statistics literature evolves in parallel, to enable the best possible insights from these complex datasets.

Acknowledgements

First and foremost, I would like to thank my academic supervisors Alex Gibberd and Carolina Euán. I am exceptionally grateful for all of your support, patience and kindness throughout the PhD process. I have learned a lot from you both, and feel very fortunate to have had such knowledgeable mentors during my time as a PhD student. From my first ever conference in Oaxaca, to our weekly catch-ups in the PSC, I can honestly say there has never been a dull moment. Thank you for being there when I needed you, I could not have asked for better supervisors.

I count myself extremely lucky to have completed my PhD at the EPSRC-funded STOR-i Centre for Doctoral Training, a place full of great staff, students and opportunities. I would like to thank the great friends I have made at the centre, especially Matt, Eleanor, Lídia, Robert and Thomas. My PhD experience would not have been as enjoyable without you all. I would also like to thank the STOR-i leadership team, namely, Jon, Idris, Anna and Rachel. A special thank you to Jonathan Tawn, who is not only academically brilliant but hilariously funny. My time in STOR-i would not have been the same without you. I would also like to thank Dan and Bash for the IT support. Thanks also to Kim, Nicky and Wendy for all of their help over the years.

During my PhD, I have been fortunate to collaborate with some excellent researchers at the University of Washington, as part of my strategic project partnership. In particular, I would like to thank Ali Shojaie for his expert guidance and mentorship, especially regarding the third chapter of this thesis. I would also like to thank Nick Steinmetz for the fascinating discussions during my visits to Seattle.

Last but not least, I would like to thank my family. To the Famous Five: Mum, Dad, Nadia and Jay, thank you for always being there for me, and for your constant support and encouragement. To Granny and Grandad (Mr and Mrs Sup G), thank you for the weekly catch-ups over FaceTime, and the seemingly infinite supply of home-made soup, apple pies and Bourbon biscuits. To Auntie Pat, thank you for your unwavering support and encouraging messages, your words of wisdom are truly unmatched. I would also like to thank my late Grandad Pinkney, who championed all of my academic endeavours. I think I just about have an answer to your favourite question: *'How's your education?'*. Finally, to Sophie, thank you for brightening up the last two years of my life. Your kindness, love and support have been, and continue to be, a great source of comfort.

To Mum and Dad, for teaching me the core values of the Famous Five.

"You is kind, you is smart, you is important."

- Kathryn Stockett

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Chapter 3 has been accepted for publication as Pinkney, C., Euan, C., Gibberd, A., & Shojaie, A. (2026). Regularised Spectral Estimation for High-Dimensional Point Processes. *Biometrika.*

Chapter 4 has been submitted for publication as Pinkney, C., Euan & C., Gibberd, A. (2025). Identifying Neural Connectivity using Bernoulli Autoregressive Partially Linear Additive Models.

The word count for this thesis is approximately 44,000 words.

Carla J. Pinkney

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The brain is a highly complex structure which processes information via an interconnected network of neurons. In recent years, knowledge of the brain's structure and function has significantly improved, thanks to the development of new technologies which allow for the exciting opportunity to study large scale neural activity in the living brain (Jun et al., 2017; Steinmetz et al., 2019; Ye et al., 2024). However, of equal importance, has been the development of new statistical and computational tools, capable of extracting meaningful information from the complex and high-dimensional data generated by these modern technologies. Indeed, there are several neuroscientific objectives associated with the analysis of neural data. For example, one might be interested in understanding how neurons interact, or how the brain handles certain cognitive processes such as learning, perception and memory.

Understanding the intricate neural processes which drive brain function has captured the attention of many scientists, and is a major goal of international brain science programs, such as the BRAIN Initiative in the US (Miller et al., 2024), the Human Brain Project in Europe (Amunts et al., 2019), and the Brain/MINDS project in Japan

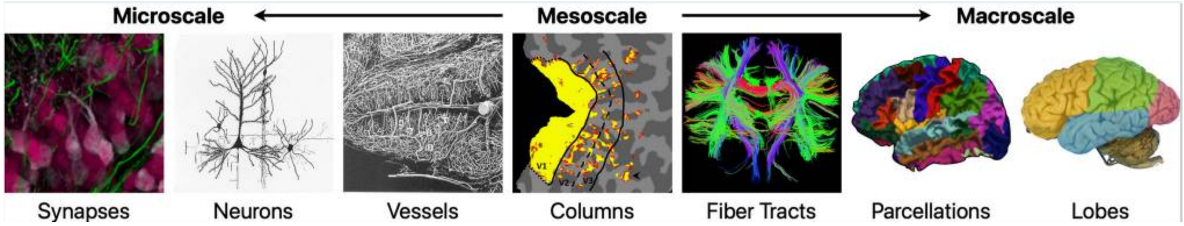


Figure 1.1.1: Hierarchical scales of neuroanatomical topology (Source: Marchant et al., 2024).

(Okano et al., 2016). In this thesis, our primary focus lies in characterising the connectivity between populations of neurons. In particular, we will represent relationships between neurons as a network, and will develop novel statistical methodologies capable of inferring this network structure in a variety of experimental settings. This is an active research problem in both statistics and neuroscience (Keeley et al., 2020; Ren et al., 2020; Vareberg et al., 2024; El-Yaagoubi et al., 2025) and one which is essential to our continued understanding of the brain.

The concept of brain connectivity spans multiple scales, from individual synapses on a neuron, to large functionally connected regions (Marchant et al., 2024). Researchers have studied the brain network at a variety of spatial scales, including; the macroscopic level; to infer connections between anatomical brain regions, the mesoscopic level; to infer connections between individual neurons and the microscopic level; to understand the operational mechanisms of single neurons (Abril et al., 2018).

Constrained by the technology available in the past, neuroscientists have traditionally studied brain connectivity at the macroscopic level, using techniques such as magnetoencephalography (MEG), electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). Indeed, many statistical methodologies have been developed for the analysis of these data, primarily in the domain of time series and network analysis (see the book Ombao et al., 2016, for a review).

Some progress has also been made towards identifying neural interactions at the mesoscopic level, using a small number of simultaneously recorded spike trains (Brillinger, 1988; Dahlhaus et al., 1997; Eichler et al., 2003). However, these existing statistical techniques are not suitable for the complexity and high-dimensionality of modern neuroscience spike train data. Related work for the identification of mesoscopic interactions in *in vitro* neuronal assemblies is discussed in Poli et al. (2015, 2016).

In this thesis, we study the brain network at the mesoscopic level, and focus on the development of novel statistical methodology to infer connections between individual neurons, both locally, within brain regions, and globally, across regions. To do so, we will utilise data obtained from technologies such as the multi-electrode array and Neuropixels probes (Jun et al., 2017; Steinmetz et al., 2019, 2021), which give rise to complex high-dimensional datasets.

As stated, the primary objective of this thesis is to characterise the connectivity amongst neural populations. It is therefore important to distinguish exactly what we mean 'brain connectivity'; a concept which can typically be categorised as either functional, effective or structural. Functional connectivity refers to the statistical association or dependence among measurements of neuronal activity, and is classically measured using Pearson correlations (Lang et al., 2012). Effective connectivity is defined as the direct interaction between pairs of neurons after accounting for any indirect effects. In this case, it is often assumed that the neural data have been generated from a particular statistical model. Then, estimates of effective connectivity can be inferred by optimising the model parameters. By contrast, structural connectivity (also sometimes referred to as anatomical connectivity) represents the existence of actual synapses (Poli et al., 2016). Importantly, the identification of a functional or effective interaction does not necessarily guarantee the existence of a structural connection (Abril et al., 2018).

In this thesis, our aim is to develop novel statistical methodology that can accurately recover the latent connectivity structure in the brain network. While we primarily focus on methods for the identification of functional and effective interactions, references to structural connectivity will be mentioned in the final chapter. Interestingly, there does not yet exist a set of universally agreed measures to characterise brain connectivity (Hu et al., 2019). For example, connectivity can be described at varying levels of detail including, but not limited to, existence, direction, sign, magnitude and temporal dynamics (Abril et al., 2018). As such, a secondary aim of this thesis is to assess the consistency of different neuronal network estimation methods, both within and across experimental subjects.

1.2 Overview of the Thesis

As stated, the primary aim of this thesis is to develop statistical models that can accurately represent mesoscopic interactions in the brain network, from modern high-dimensional spike train data. While a rich body of literature exists for brain connectivity analysis at the macroscopic scale, there remains significant potential to develop new methods for connectivity inference at the individual neuron level. Moreover, given the variety of methods developed for this purpose in recent years, we subsequently explore the sources of variability in estimated functional and effective networks, using real data from the DANDI archive. The remainder of the thesis is organised as follows:

Chapter 2 provides an introduction to key concepts and methods used in the analysis of neural data. We begin by discussing the data collection and pre-processing pipeline, before introducing key statistical concepts used in subsequent chapters of the thesis.

In Chapter 3, we represent neural spike train data as a multivariate point process. We characterise the second order structure of this process via the spectral density matrix, a frequency domain equivalent of the covariance matrix. In the context of neuronal analysis, statistics based on the spectral density matrix can be used to infer connectivity in the brain network between individual neurons. However, the high-dimensional nature of spike train data mean that it is often difficult, or at times impossible, to compute these statistics. To improve the efficiency of spectral estimation for point processes, we propose methodology that combines a Whittle pseudo-likelihood with ridge or Lasso style penalties. We establish asymptotic and large sample properties for our proposed estimators and evaluate their performance on synthetic data simulated from multivariate Hawkes processes. Finally, we apply our methodology to neuroscience spike train data in order to illustrate its ability to infer brain connectivity.

In Chapter 4, we take a different approach to the problem of inferring brain connectivity from neural spike train data. More specifically, under a binned spike count representation of these data, we propose a Bernoulli autoregressive partially linear additive (BAPLA) model to identify the effective connectivity of a population of neurons. Estimates of the model parameters are obtained using a regularised maximum likelihood estimator, where an ℓ_1 penalty is used to find sparse and interpretable estimates of neuronal interactions. We also account for non-stationary firing rates by adding a non-parametric trend to the model and propose an inference procedure to quantify the uncertainty associated with our estimated networks of neuronal interactions. We empirically validate the performance of the BAPLA method and the inferential procedure, highlighting its ability to detect both excitatory and inhibitory interactions in various settings. Finally, we apply our method to a neural spiking dataset from the DANDI archive, where we study the interactions of neural processes in reaction to various stimulus-response type neuroscience experiments.

In Chapter 5, we extend the work developed in Chapter 3 and Chapter 4 to the multi-session and multi-subject setting, exploring the variability of functional connectivity estimates across experimental subjects. We also compare the proposed methods to existing approaches for neuronal network estimation. Using the spike train data from Steinmetz et al. (2019, 2024), we discuss the similarities and indeed differences between classical and modern techniques, offering insights into the advantages and disadvantages of each method. In doing so, our aim is to better understand the ongoing challenges associated with the estimation and understanding of functional brain networks, at the individual neuron level.

Finally, in Chapter 6, we conclude by summarising the main contributions of this thesis and discussing possible avenues for future work.

Chapter 2

Preliminaries

2.1 Introduction

This chapter details foundational concepts in statistics and computational neuroscience that are relevant to the discussion in the main contributory chapters of this thesis. We begin, in Section 2.2, by summarising some notation which is used throughout the chapter. In Section 2.3, we describe the data collection process and preprocessing steps which give rise to the high-dimensional neural spike train data that underpin our statistical analyses. In Section 2.4, we describe univariate methods for the analysis of a single spike train, before extending our discussion to the multivariate setting in Section 2.5. Lastly, in Section 2.6, we review relevant literature for the graphical modelling of multivariate data. In particular, we focus on existing methods for functional connectivity analyses in the statistics and neuroscience literature, for the identification of both macroscopic and mesoscopic neural interactions.

The ideas developed within this thesis lie within the realm of high-dimensional statistics for the graphical modelling of multivariate spike train data, in both discrete and continuous time. While we focus our efforts on inferring neuronal connectivity in the brain network, multivariate processes appear in a wide variety of disciplines outside of neuroscience. Therefore, the contributions of this work may also be of interest in other fields, including, but not limited to, finance, health and social networks.

2.2 Notation

For the convenience of the reader, we summarise here some notation to be used throughout this chapter.

We use the standard notation $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ to denote the set of natural numbers which includes 0, whereas $\mathbb{N}^* = \{1, 2, 3, \dots\}$ is the set excluding zero. Similarly, we write $\mathbb{R}_{\geq 0}$ to denote the set of non-negative real numbers.

Given a matrix $U \in \mathbb{R}^{p\times p}$, $\phi_{\min}(U)$ and $\phi_{\max}(U)$ denote the minimum and maximum eigenvalues of U, respectively. Additionally, $\phi_j(U)$ represents the j^{th} eigenvalue of U. Furthermore, we recall the standard matrix notation $\succ$ and $\succeq$. Specifically, for a $p \times p$ matrix U we have $U \succ 0$ denoting that U is positive definite, while $U \succeq 0$ means that U is positive semi-definite. The ℓ_∞-operator norm is defined as

$$|||U|||_\infty := \max_{q=1,\dots,p} \sum_{r=1}^{p} |U_{qr}|,$$

and we use $\|U\|_\infty$ to denote the element-wise maximum $\max_{q,r} |U_{qr}|$. The Frobenius norm is defined as $\|U\|_F := \sqrt{\sum_{qr} U_{qr}^2}$ and ℓ_1-norm is defined as $\|U\|_1 := \sum_{qr} |U_{qr}|$.

For asymptotics, we use the following standard notation: we write $f(n) = \mathcal{O}(g(n))$ if $f(n) \leq cg(n)$ for some constant $c < \infty$ and $f(n) = \Omega(g(n))$ if $f(n) \geq c'g(n)$ for some constant $c' > 0$. The notation $f(n) \asymp g(n)$ means that $f(n) = \mathcal{O}(g(n))$ and $f(n) = \Omega(g(n))$.

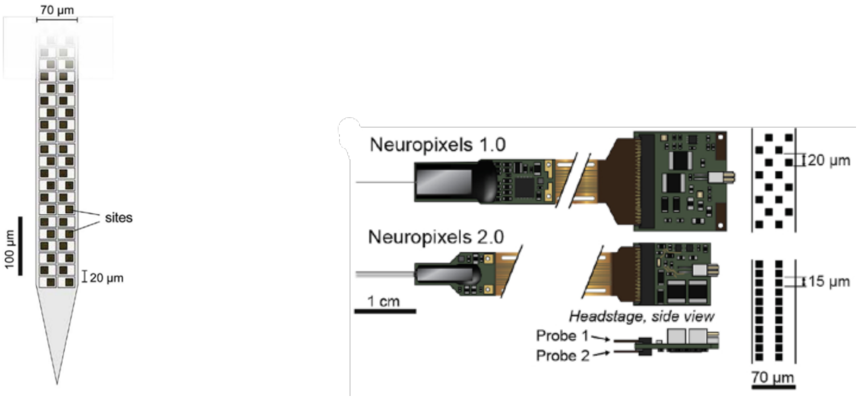
2.3 Neural Spike Train Data

2.3.1 Data Collection and Preprocessing

High-dimensional measurements of localised neural activity can be captured by modern technologies such as calcium fluorescence imaging methods or via multiple electrodes implanted directly into the brain. The choice between these different recording modalities typically depends on the research question of interest. For example, calcium imaging methods provide a broader spatial coverage, and are therefore better suited for the investigation of large-scale population dynamics, while electrodes offer higher temporal resolution and precise measurements of individual neuronal spiking activity, albeit for a sparser subset of neurons (Siegle et al., 2021). In this thesis, we will focus on the latter, analysing data obtained from recent technologies such as the Neuropixels probe, whose direct measurements provide the gold standard for measuring localised neural activity (Jun et al., 2017; Steinmetz et al., 2021; Ye et al., 2024).

The first generation of Neuropixels probes (Neuropixels 1.0) were developed in 2017 by an international group of researchers at the Howard Hughes Medical Institute's Janelia Research Campus, the Allen Institute for Brain Science and University College London (Jun et al., 2017). These probes were designed with the intention of overcoming the limitations of existing technologies, such as the extracellular microelectrode (Buzsáki, 2004) capable of recording from only a few dozen neurons per shank, or calcium imaging methods which offered larger coverage but lacked the temporal resolution required to accurately distinguish single action potentials (spikes).

Each Neuropixels 1.0 probe has 384 recording channels, containing 960 TIN (titanium nitride) electrodes, on a single 10 mm long, 70×20 μm cross section silicon shank. This unprecedented combination of dense recording sites and high channel count, enabled the simultaneous recording of hundreds of neurons from multiple brain structures in freely moving animals, transforming the study of the brain's structure and function.



(a) (b)

Figure 2.3.1: (a) Diagram of the Neuropixels 1.0 tip showing checkerboard site layout (source: Jun et al., 2017). (b) Comparison of the Neuropixels 1.0 and Neuropixels 2.0 devices (source: Steinmetz et al., 2021).

More recently, scientists have developed the Neuropixels 2.0 (Steinmetz et al., 2021) and Neuropixels Ultra probes (Ye et al., 2024), which are even smaller and lighter than Neuropixels 1.0 probes, and enable recordings from even larger neuronal populations via a more sophisticated multi-shank design (see Figure 2.3.1). In either design, each channel can record action potentials (the individual firing times of single neurons) and local field potentials (LFPs) which measure the collective, low-frequency electrical activity of the local neural circuit. While the statical methodologies in this thesis have been developed for the estimation of neural interactions from single spiking data, analysing brain connectivity using LFP measurements is also an active area of research (Wang et al., 2019; El-Yaagoubi et al., 2025).

In summary, variations of these state-of-the-art technologies are used in labs across the world to collect data from a range of species, including fish, rodents, monkeys, reptiles, and humans (Ye et al., 2024). Thus, there is an ever increasing volume of data, for which the methods proposed in this thesis are relevant.

Spike Sorting

Extracting the firing times of single neurons from extracellular recordings is achieved via a process known as ‘spike sorting’. This procedure involves a number of algorithmic steps to identify and separate electrical signals from individual neurons recorded by electrodes implanted in the the brain, and is considered a classical technique in neuroscience, with literature spanning several decades (Carlson and Carin, 2019). The main steps in the spike sorting process (discussed in Rey et al., 2015) are summarised below, alongside a diagrammatic description in Figure 2.3.2. However, the spike sorting literature is continually evolving, driven by new recording technologies and data challenges (International Brain Laboratory, 2024).

1. **Preprocessing:** A bandpass filter is applied, e.g., between 300 Hz and 3000 Hz, to separate spikes from noise and low-frequency signals like LFPs.

2. **Spike Detection:** An amplitude threshold is applied to the filtered data to identify the time points at which spikes occur.

3. **Spike Alignment:** The detected spikes are aligned to a consistent feature (e.g. peak or trough) allowing for consistent comparison of waveforms.

4. **Feature Extraction:** Each detected spike waveform is transformed into a point in a lower-dimensional feature space. Common techniques include principal component analysis (PCA), wavelet transforms or template features (e.g., amplitude, width, energy).

5. **Post-processing and Validation:** Manual curation of clusters to assess cluster quality metrics such as waveform stability and firing rate.

Indeed, a number of spike sorting algorithms (e.g., kilosort, MountainSort, IronClust, SpyKING, CIRCUS) have been developed in recent years, and each method is tailored to different experimental setups, such as the choice of recording hardware

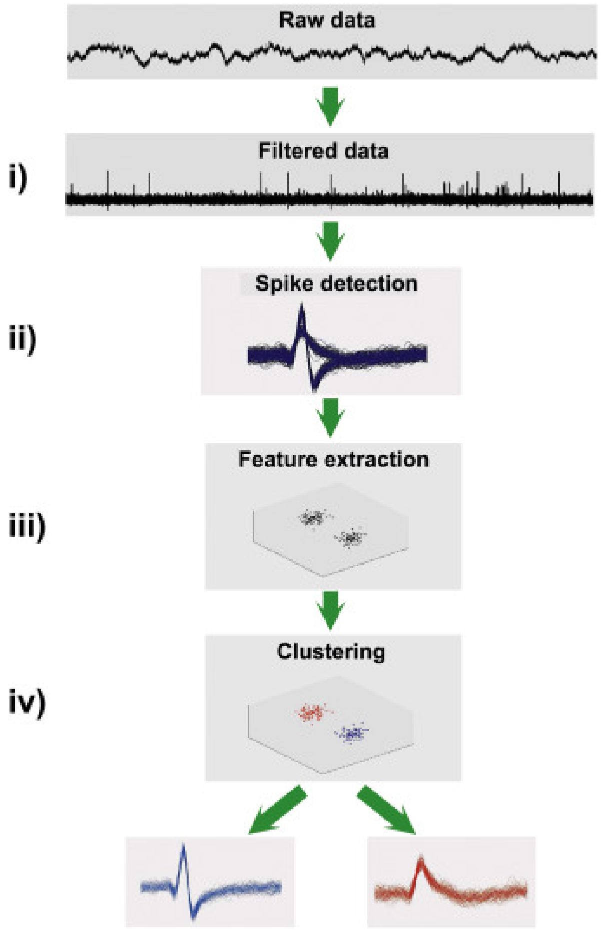
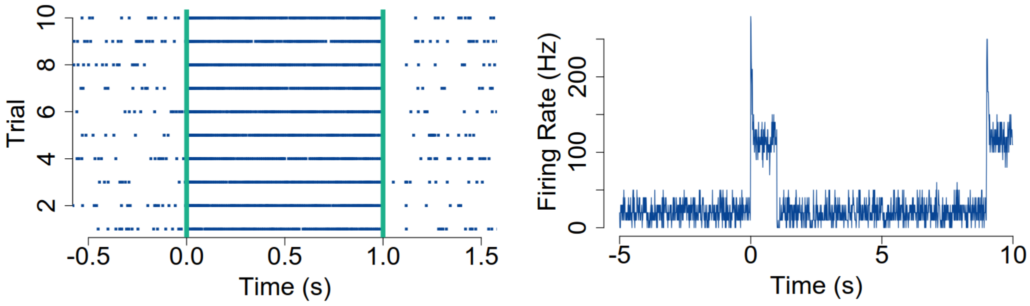


Figure 2.3.2: Overview of the spike sorting process. (Source: Rey et al., 2015).

(e.g., Neuropixels, tetrodes, MEAs). To aid comparison between the different spike sorting approaches, open source software such as SpikeForest (Magland et al., 2020) and SpikeInterface (Buccino et al., 2020) were developed to systemically compare the performance of these algorithms using real and simulated datasets with a known ground truth. However, despite these advances, the spike sorting process remains error-prone. Consequently, spike trains should be considered as *approximate* estimates of the true neural firing patterns, rather than exact measurements.

Overall, the spike sorting process is somewhat complicated, and further discussion is beyond the scope of this thesis. However, it is important to acknowledge the complexity of this task, prior to the discussion of methods which can be used to infer connectivity from the preprocessed (spike sorted) data.



(a) (b)

Figure 2.3.3: Spike train data from Bolding and Franks (2018) with (a) raster plot and (b) trial-averaged firing rate.

2.3.2 Modelling Spike Train Data

After adequate preprocessing of the raw electrode data, measurements known as spike trains are obtained, which represent the firing times of individual neurons in the brain. Figure 2.3.3 shows an example of spike train data obtained from the experiments of Bolding and Franks (2018). The figure on the left hand-side is known as a raster plot, and depicts the firing times (indicated by a dot in the plot) of an example neuron in the dataset across a series of ten experimental trials. Comparatively, the right hand-side shows the firing rate (interpreted as the number of spikes per second) of this neuron averaged across the experimental trials. In both plots, we observe an obvious increase in firing rate[1], which is due to the onset of a stimulus at time 0. For further discussion of these data, see Chapter 3.

Biologically, action potentials (or spikes) are produced as a result of the interactions between individual neurons. In particular, when a neuron sends a message it produces a spike, corresponding to a substantial increase of its membrane potential. Statistically, these data are most commonly represented in either one of two ways: (1) in continuous-time via a multivariate point process, or (2) in discrete-time typically taking the form of a multivariate binary time series. In this thesis, we will develop novel statistical methodology for use in both settings.

---

[1] There are indeed spaces between the points in the high firing (0-1s) period.

Remark 2.3.1. *Ultimately, we will be interested in inferring a form of network structure to summarise the interactions between neural processes. Of particular interest, are networks in which nodes (representing neurons) are characterised either as point processes or as binary time series. However, it is first useful to introduce concepts related to the analysis of a single spike train, i.e., the univariate setting, prior to an extension to multivariate methods in Section 2.5.*

A useful way in which to characterise a set of neuronal spiking observations is in terms of the number of spikes observed over a given time interval. Specifically, we define the counting process $N(t)$ with $t \geq 0$ as the number of spikes that have occurred in the time interval $(0, T], t \leq T$, whose realisations are non-decreasing and take a unit step up whenever an event occurs.

Definition 2.3.2 (Counting Process)**.** *A* counting processes *is a stochastic process* $\{N(t), t \geq 0\}$ *that takes values in* $\mathbb{N}_0$ *that satisfies* $N(0) = 0$*, is finite, and is a right-continuous step function with jumps of size* $+1$*.*

Alternatively, we might also consider the sequence of random spike times $t = \{t_1, t_2, \dots\}$ at which the counting process $N(\cdot)$ has jumped. The process defined in terms of the spike times is known as a point process.

Definition 2.3.3 (Point Process)**.** *Let* $\{t_i \in \mathbb{R}_{\geq 0}, i \in \mathbb{N}_0\}$ *be a sequence of non-negative random variables such that* $t_i < t_{i+1}, \forall i \in \mathbb{N}_0$*. We call* $\{t_i, i \in \mathbb{N}_0\}$ *a simple point process on* $\mathbb{R}_{\geq 0}$*, and assume that* $t_0 = 0$*.*

However, since recording the spike times is equivalent to recording the times at which the counting processes increases, it is easy to see that the definitions given above are equivalent in terms of their ability to characterise neural processes. As such, the point and counting process terminology will be used interchangeably throughout this thesis.

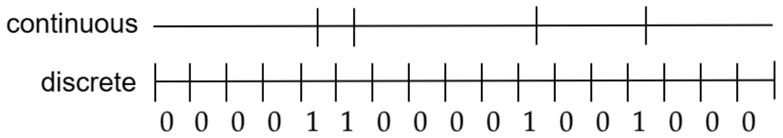


Figure 2.3.4: Comparison of the continuous and discrete time representation of neural spike train data. In each time bin, a 1 indicates that a spike has occurred.

A popular alternative to the continuous time representation of neural data is that of the binned spike count, which refers to a discretised version of the continuous time process (Kass et al., 2011). The so-called binned spike count is constructed by decomposing the time axis into bins of equal width δ, resulting in a discrete integer-valued time series where each value indicates the number of action potentials (spikes) within a specified time bin, i.e. $Y_t = N(t+\delta) - N(t), t \leq T$. Typically, $\delta = 1$ms is chosen to ensure that at most one spike per neuron can occur within a time bin, resulting in a binary time series representation of the neural data, which takes value 1 in a time bin where there is a spiking event, and 0 everywhere else (Figure 2.3.4). Importantly, it can be shown (Appendix A) that under the usual regularity condition, where no two events occur at the same time, that the likelihood function of the binary time series approximates the likelihood of the point process as $\delta \to 0$ (Chen et al., 2010; Kass et al., 2011).

The discretisation of neural spike train data is a standard technique in computational neuroscience, as evidenced by its extensive use in the literature (Chen et al., 2010; Kass et al., 2011; Ito et al., 2011; Zhao et al., 2012; Song et al., 2013; Leen and Shea-Brown, 2015; Kadirvelu et al., 2017; Zoltowski and Pillow, 2018; Kobayashi et al., 2019; Liew et al., 2021; Vareberg et al., 2024). However, there are indeed conflicting opinions regarding the suitability of this approach. While the arguments for discretisation are both biologically motivated and computationally efficient, continuous-time methods are often considered more sophisticated since they model the exact timing of the spikes without binning.

2.4 Univariate Point Processes

2.4.1 The Conditional Intensity Function

Any point process can be characterised by its conditional intensity function (Daley et al., 2003)

$$\Lambda(t|\mathcal{H}_t) = \lim_{dt \to 0} \frac{\mathrm{pr}\left(dN(t) = 1|\mathcal{H}_t\right)}{dt},$$

where $dN(t) = N(t+dt) - N(t)$ denotes the number of spikes in some small time interval dt (Bartlett, 1963a,b) and $\mathcal{H}_t$ is the past spiking history of the process up to time t. Thus, for sufficiently small dt, $\Lambda(t|\mathcal{H}_t)dt$ is approximately the probability of observing a spike in the time interval $(t, t+dt]$ given the spiking history. Hence, one could equivalently write the conditional intensity function as the expected rate of spikes conditioned on the past, i.e.,

$$\Lambda(t|\mathcal{H}_t) = \lim_{dt \to 0} \frac{\mathbb{E}(dN(t)|\mathcal{H}_t)}{dt}.$$

Going forward, we do not continue to write the conditions on the histories $\mathcal{H}(\cdot)$ explicitly, and for ease of notation will now refer to $\Lambda(t|\mathcal{H}_t)$ as $\Lambda(t)$.

If a spike event causes $\Lambda(t)$ to increase, then the process is called self-exciting. This type of behaviour causes temporal clustering of the spike times $t = \{t_1, t_2, \dots\}$, meaning that we are likely to exhibit collections of spikes firing in quick succession. In this case, it is important to choose $\Lambda(t)$ appropriately, in order to avoid explosion, i.e., $N(t) = \infty$ for finite t with non-zero probability (Laub et al., 2015).

However, it also possible to observe a reduction in the conditional intensity function after an arrival. Processes of this nature are known as self-regulating, and exhibit the opposite behaviour of self-exciting processes. In the context of our neuronal analysis, one might consider the electrophysiological properties of neural membranes, which often limit how fast a neuron can fire immediately after a spiking event. Specifically, neurons experience an absolute refractory period, whereby the neuron physically cannot fire another spike, regardless of the stimulus. Additionally, they might experience a relative refractory period in which the neuron can fire another spike, but in order to do so it will require a more effective stimulus (Iremonger and Bains, 2016).

Estimation of the instantaneous firing rate $\Lambda(t)$ has received significant attention in the literature (Eden et al., 2012; Zhou et al., 2015; Sarmashghi et al., 2021). With applications in neural encoding, decoding[2] and brain-machine interfaces, modelling the firing patterns of individual neurons is fundamental to our understanding of how the brain processes information.

Point process generalised linear models (GLMs) are often used for the estimation of $\Lambda(t)$. These models assume that $\Lambda(t)$ can be modelled by a linear combination of covariates that influence neural firing, through a non-linear link function e.g.,

$$\Lambda(t|\theta, \mathcal{H}_t) = \exp\left\{\sum_{i=1}^{d} \theta_i v_i(t)\right\}$$

where $v_i(t)$ is the i^{th} covariate and d is the dimension of the parameter θ. Examples of covariates include external stimuli (e.g., visual or auditory cues) and the spiking history of individual neurons (e.g. for refractory effects or bursting) derived based on $\mathcal{H}_t$. In general, these models are easy to fit and interpret (Sarmashghi et al., 2021). As such, they have been used throughout the literature to better understand the coding properties of individual neurons within wider populations (see for example Truccolo et al., 2005; Huang et al., 2009; Tripathy et al., 2013; Park et al., 2014; Steinmetz et al., 2019). Here, coding refers to how external stimuli and a neuron's spiking history, influences its firing rate. More generally, neural coding describes how external stimuli

---

[2] Encoding refers to how information (e.g. external stimuli) is represented in neural activity whereas decoding describes how information can be inferred from that activity (Shimazaki, 2025).

are translated into neural activities that ultimately drive behaviour (Shimazaki, 2025).

When the relationship between the instantaneous firing rate and the covariates are non-linear, spline functions can be used within the point process GLM. For example, splines are often used to model stimulus or task-related modulations of the firing rate (Eden et al., 2012; Zhou et al., 2015; Sarmashghi et al., 2021).

2.4.2 The Poisson Process

The simplest example of a point process model is that of the Poisson process, which assumes that the number of points (or spiking events in our case) in disjoint intervals are independent and have a Poisson distribution. In the context of our neuronal analysis, this amounts to an assumption that the probability of a neuron firing in some small time interval is independent of its previous firing history. Generally, Poisson processes are not recommended for modelling spike train data, since neural firing is not entirely independent, as evidenced by features such as refractory periods and bursting (Bair et al., 1994; Dostrovsky et al., 2000).

However, it is useful to introduce the Poisson process prior to the discussion of history dependent processes in later sections.

Definition 2.4.1 (Inhomogeneous Poisson Process)**.** *Consider a point process $N(t)$ with time-varying intensity function $\Lambda(t)$ such that*

$$\begin{aligned} &\operatorname{pr}\{dN(t) = 1 | N(s)(s \leq t)\} = \Lambda(t)dt + o(dt), \\ &\operatorname{pr}\{dN(t) > 1 | N(s)(s \leq t)\} = o(dt). \end{aligned}$$

Furthermore, if $N(0) = 0$ and $N(t)$ has independent increments, i.e., for all choices of $t_1 < t_2 \leq t_3 < t_4$, the random variables $N(t_2) - N(t_1)$ and $N(t_4) - N(t_3)$ are independent. Then, $N(t)$ is an inhomogeneous Poisson process.

The notation $o(dt)$ is used for any dt-dependent quantity satisfying $\frac{o(dt)}{dt} \to 0, dt \to 0$.

Hence, the above definition says that in a Poisson process, the probability of two or more events occurring in a small interval dt is negligible, thereby aligning with the definition of a simple point process (Daley et al., 2003).

Definition 2.4.1 describes how to characterise the so called inhomogeneous Poisson process. In this instance, we are considering the case in which the probability of a neuron firing in some small time interval dt varies over time. Importantly, if $\Lambda(t) = \Lambda$ for all $t \geq 0$ in the above definition, then $N(t)$ is known as a homogeneous Poisson process and has the added property of stationary increments. In this case, the spiking probability models do not depend on time. For completeness, we say that $N(t)$ has stationary increments if,

$$N(t_1 + s) - N(t_1) \overset{d}{=} N(t_2 + s) - N(t_2) \quad \text{for all } t_1, t_2, s \geq 0.$$

That is, the distribution of the number of points in any time interval depends only on the length of the interval, not on its starting position.

2.4.3 The Hawkes Process

As mentioned, the assumption that all events are stochastically independent of each other is unrealistic in the context of our neuronal analyses. Indeed, the history dependent nature of neural data is an important part of virtually all neural spiking activity. Hence, accurate models which incorporate this history dependence are essential in fully describing these data.

One particular point process model which is thought to adequately capture the spiking activity of neurons is the Hawkes process (Hawkes, 1971a,b). This process can be used to model a sequence of spikes over time, where each spike is said to “excite” the process. In the univariate case, this means that there is some time period after an initial spiking event during which the likelihood of the occurrence of future spikes is increased. Consequently, Hawkes processes are often referred to as "self-exciting" processes.

Definition 2.4.2 (The Univariate Hawkes Process). *Consider a point process* $N(t)$ *such that*

$$\begin{aligned}\mathrm{pr}\left\{dN(t)=1|N(s)(s\leq t)\right\}&=\Lambda(t)dt+o(dt),\\ \mathrm{pr}\left\{dN(t)>1|N(s)(s\leq t)\right\}&=o(dt).\end{aligned}$$

Now, assume that the process is self-exciting in the sense that the conditional intensity function may be written as

$$\Lambda(t)=\nu+\int_{-\infty}^{t}g(t-u)\ dN(u), \tag{2.4.1}$$

where $\nu\geq 0$ *and* $g:(0,\infty)\rightarrow[0,\infty)$ *are known as the background intensity and excitation function respectively. Then,* $N(t)$ *is a Hawkes process.*

The self-exciting nature of the univariate Hawkes process is made explicit in (2.4.1) where the current intensity of events is determined by events in the past, i.e., $\mathcal{H}_t$ (Hawkes, 1971b). While the definition given in (2.4.1) has been used throughout the literature, it is often convenient to rewrite the conditional intensity function in terms of past spiking events. Specifically, if we consider $\{t_1,\ldots,t_k\}$ to be the observed sequence of past spiking events up to time t, then the Hawkes conditional intensity function can be written as

$$\Lambda(t)=\nu+\sum_{t_i<t}g(t-t_i). \tag{2.4.2}$$

The quantity $\nu\geq 0$ in (2.4.2) describes the arrival of spiking events triggered by external sources, and provides a Poisson base level for the process (Hawkes, 2018). The arrival of events according to this quantity are therefore independent of previous spiking events within the process. The self-exciting nature of the process therefore arises through the summation part of (2.4.2), where the excitation function $g(t - t_i)$ adjusts the intensity function at time t to account for the effect of an event occurring at time t_i. Typically, the function $g(\cdot)$ is taken to be monotonically decreasing so that more recent events have a stronger influence on the current event intensity, compared to events that have occurred previously and further away in time (Rizoiu et al., 2017). As noted in Hawkes (1971b), if $g(\cdot)$ decays rapidly, then there will be a largely local effect, whereas if $g(\cdot)$ has a hump, then the process might contain longer term effects (Hawkes, 1971b).

The structure of the conditional intensity function requires careful specification of both the background intensity ν and excitation function $g(\cdot)$. A common choice for $g(\cdot)$ is that of exponential decay. In this case, the conditional intensity function $\Lambda(t)$ is specified by constants $\alpha, \beta > 0$ such that

$$\Lambda(t) = \nu + \sum_{t_i < t} \alpha e^{-\beta(t - t_i)}. \tag{2.4.3}$$

The above definition leads an intuitive interpretation of the constants (α, β) whereby the arrival of each spike event instantaneously increases the intensity function by α, and over time this event's influence decays at rate β.

Other specifications of the excitation function $g(\cdot)$ exist, including that of the power law function. In this case, the conditional intensity function is

$$\Lambda(t) = \nu + \sum_{t_i < t} \frac{\alpha}{(c + (t - t_i))^p},$$

where α, c and p are positive scalars (Lesage et al., 2022). This excitation function is often used in geophysical applications, and was popularised by a geological model called Omori's law (see Ogata, 1999, for details). While many other specifications exist, the remainder of this thesis will focus on Hawkes processes with exponentially decaying intensity functions.

Cluster Representation of the Hawkes Process

An alternative view of the Hawkes process is given by its cluster (or immigration-birth) representation. In this framework, spikes (or events) are seen as forming branching structures, whereby immigrants arrive according to a homogeneous Poisson process with rate ν. Then, each event (immigrant or descendent) independently generates offspring events, and the arrival of births forms an an inhomogeneous Poisson process. More specifically, an event occurring at time $t_i \in \mathbb{R}$ will produce events at future times $t > t_i$ at the rate $g(t - t_i)$. Direct offspring of the event at t_i are known as the first generation, and their offspring are known as the second-generation, and so on. Collectively, these generations are called the descendants of the event at t_i.

Figure 2.4.1 provides a diagrammatic description of this process, generated using the `hawkesbow` package in R. The plot on the left shows the cluster representation, where the immigrants are represented by squares and the offspring/descendents are represented by circles. Circles in pink, green, blue and turquoise denote the first, second, third and fourth generations, respectively, while the crosses denote the resulting point process. The plot on the right hand side shows the corresponding intensity function for the simulated Hawkes process.

Simulating from the Univariate Hawkes Process

In this section, we outline a simulation algorithm that can be used to generate realisations from the univariate Hawkes process. This is particularly important for Chapter 3, where the Hawkes process is used as an illustrative example for our spectral estimation methods. While there are multiple ways to simulate from a univariate Hawkes process, we will limit our discussion to Ogata's modified thinning algorithm (Ogata, 1981). This method is an adaptation of a pre-existing algorithm (Lewis and Shedler, 1979) for generating realisations of an inhomogeneous Poisson process. As such, we will first discuss Lewis's algorithm prior to a discussion of Ogata's method.

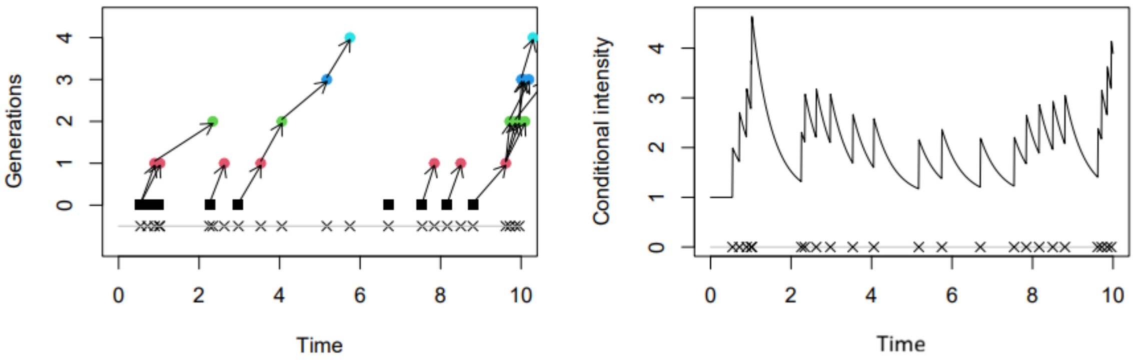


Figure 2.4.1: Cluster representation of the Hawkes process (left) and corresponding intensity function (right). Squares denote immigrants, circles denote offspring/descendents and the crosses denote the resulting point process.

In Lewis's algorithm, points from an inhomogeneous Poisson process are simulated sequentially, by generating a "faster" homogeneous Poisson process. Then, points are removed probabilistically (thinned) so that those remaining satisfy $\Lambda(t)$. This idea is formalised in the following result, a proof of which can be found in Lewis and Shedler (1979).

Theorem 2.4.3. *Consider a homogeneous Poisson process* $\tilde{N}(t)$ *with intensity function* $\tilde{\Lambda}$. *Let* $\tilde{t}_t, \ldots, \tilde{t}_{\tilde{N}(T)}$ *be the points of the process in the interval* $(0, T]$. *Suppose that for* $0 \leq t \leq T, 0 \leq \Lambda(t) \leq \tilde{\Lambda}$, *where* $\Lambda(t)$ *is the intensity function of the inhomogeneous Poisson process. For* $k = 1, \ldots, \tilde{N}(T)$, *delete the point* $\tilde{t}_k$ *with probability* $1 - \Lambda(\tilde{t}_k)/\tilde{\Lambda}$; *then the remaining points form a point process* $N(t)$ *with intensity function* $\Lambda(t)$.

It was shown in Ogata (1981) that Lewis's algorithm (Lewis and Shedler, 1979) could be modified to obtain realisations from Hawkes processes. In particular, we see from (2.4.3) that given the information of the first i points in the process, $t_1, \ldots, t_i$, the intensity $\Lambda(t)$ is deterministic on $[t_i, t_{i+1}]$, where t_{i+1} is the time of the next point, which is stochastic. Hence, generating the next point in a Hawkes process can also be seen as generating the first point in an inhomogeneous Poisson process.

Algorithm 1 (Ogata, 1981) Simulation of a univariate Hawkes process with exponential excitation function $g(u) = \alpha e^{-\beta u}$ on $(0, T]$.

```
Require: $\nu, \alpha, \beta, T$
 1: Initialise $\mathcal{T} = \emptyset, t = 0, n = 0$;
 2: while $t < T$ do
 3:     Set $\tilde{\Lambda} = \Lambda(t^+) = \nu + \sum_{\tau \in \mathcal{T}} \alpha e^{-\beta(t-\tau)}$      ▷ Find new upper bound
 4:     Let $w \sim \mathrm{Exp}(\tilde{\Lambda})$;
 5:     Set $t = t + w$;                                  ▷ $t$ is the next candidate point
 6:     Generate $D \sim \mathrm{Unif}(0, 1)$;            ▷ Keep candidate point with some probability
 7:     if $D\tilde{\lambda} < \Lambda(t) = \nu + \sum_{\tau \in \mathcal{T}} \alpha e^{-\beta(t-\tau)}$ then      ▷ Accept with probability $\Lambda(t)/\tilde{\Lambda}$
 8:         $n = n + 1$;                                  ▷ Update number of accepted points
 9:         $t_n = t$;                                    ▷ Name it $t_n$
10:         $\mathcal{T} = \mathcal{T} \cup \{t_n\}$;     ▷ Add $t_n$ to the ordered set $\mathcal{T}$
11:     end if
12: end while
13: if $t_n \leq T$ then
14:     return $\{t_i\}_{i=1,...,n}$
15: else
16:     return $\{t_i\}_{i=1,...,n-1}$
17: end if
```

More specifically, for any bounded $\Lambda(t)$ we can find a constant $\tilde{\Lambda}$ such that $\Lambda(t) \leq \tilde{\Lambda}$ in a given time interval (Rizoiu et al., 2017). Having already simulated spike times $t_1, \ldots, t_i$, the algorithm obtains a new spike time t_{i+1} in the following way. A candidate point w is drawn from an exponential distribution with rate $\tilde{\Lambda}$, and the counter t is updated (steps 3-5 in Algorithm 1). This point is then accepted or rejected according to the ratio of the true event rate $\Lambda(t)$ to the thinning rate $\tilde{\Lambda}$. If the candidate point is accepted, we record the spike time $i + 1$ and add it to the ordered set $\mathcal{T}$. Otherwise, we repeat the procedure of sampling a candidate point until one is accepted. Importantly, even if a candidate point is rejected, the time counter t is still updated. This demonstrates the principle of thinning a homogeneous Poisson process with a higher intensity value (Rizoiu et al., 2017). Finally, we check (line 13) if the last accepted point is within the range $(0, T]$.

In summary, Ogata's method simulates a homogeneous Poisson process with intentionally high intensities $\tilde{\Lambda}$. The conditional intensity function is then used to thin out the points that are too high, and the points which remain satisfy $\Lambda(t)$.

Figure 2.4.2 illustrates the use of Algorithm 1 to generate realisations from a uni-

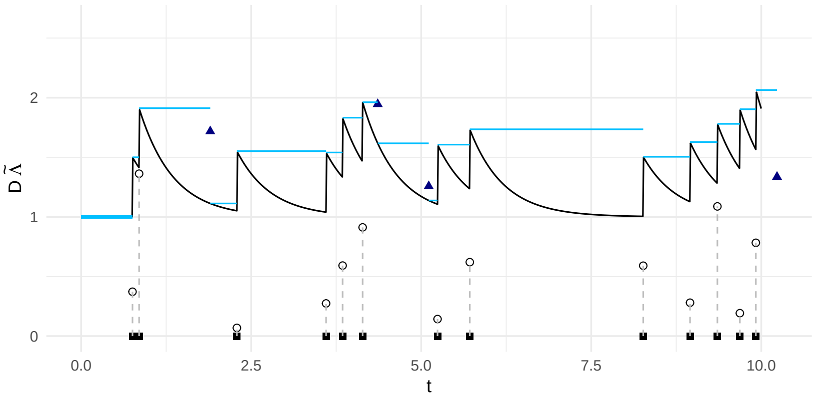


Figure 2.4.2: Illustration of Ogata's method to simulate a univariate Hawkes process with parameters $(\nu, \alpha, \beta) = (1, 0.5, 2)$

variate Hawkes process with exponentially decaying excitation function with parameters $(\nu, \alpha, \beta) = (1, 0.5, 2)$. The conditional intensity $\Lambda(t)$ is plotted as the solid black curve. The left-continuous piece-wise constant $\tilde{\Lambda}$, which is used in each iteration of the while loop at line 2 of Algorithm 1, is plotted as the blue lines. Furthermore, the values of $D\tilde{\Lambda}$ at line 7 of Algorithm 1 are plotted, and are either labelled as circles for those accepted (when $D\tilde{\Lambda} < \Lambda(\cdot)$) or with dark blue triangles for those rejected. Finally, vertical lines connect the value of $D\tilde{\Lambda}$ at the accepted points to black squares on the x-axis, highlighting the resulting point process.

2.4.4 The Univariate Spectrum

In Chapter 3, we will characterise the second order structure of the multivariate point process in the frequency domain. However, it is first useful to understand the spectral properties of point processes in the univariate setting.

Recall that $dN(t)$ denotes the number of events observed in some small time interval dt, where $dN(t) = N(t + dt) - N(t)$. As previously discussed, the first order properties of $N(t)$ are characterised by the intensity function $\Lambda(t)$, which in the context of our neuronal analysis denotes the expected number of spiking events to occur in some small time interval $t + dt$. In the event that $\Lambda(t)$ is constant for all t, i.e., $\Lambda(t) = \Lambda$, and the covariance density function,

$$\mu(\tau) = \frac{\mathbb{E}\{dN(t+\tau)dN(t)\}}{dt^2} - \Lambda^2 \quad (\tau > 0), \tag{2.4.4}$$

does not depend on t, we refer to process $N(t)$ as being second-order stationary. Moreover, for $\tau < 0$ we have $\mu(\tau) = \mu(-\tau)$. However, for $\tau = 0$, $\mathbb{E}\{[dN(t)]^2\} = \mathbb{E}\{dN(t)\}$ since $dN(t) \in \{0, 1\}$. This is a direct result of the fundamental assumption that the occurrence of two or more spikes arriving simultaneously is prohibited. Consequently, Bartlett (1963a) defined the complete covariance density, for all real τ, as

$$\mu_c(\tau) = \Lambda\delta(\tau) + \mu(\tau), \tag{2.4.5}$$

where $\delta(\tau)$ is the Dirac delta-function, and $\mu(\tau)$ is continuous at $\tau = 0$. The importance of the Dirac delta function in (2.4.5) can be demonstrated with particular reference to the Poisson process. For example, it is easy to show that when $N(t)$ is a Poisson process, $\mu(0) = 0$ as per definition (2.4.4). Hence, the inclusion of the delta function is necessary to ensure that the variance of the process at lag $\tau = 0$ is equal to Λ rather than 0.

The spectral density function for $N(t)$ is subsequently defined as the Fourier transform of the complete covariance density function

$$\begin{aligned} S(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\tau\omega}\mu_c(\tau)d\tau \\ &= \frac{1}{2\pi}\left\{\Lambda + \int_{-\infty}^{\infty} e^{-i\tau\omega}\mu(\tau)d\tau\right\}, \end{aligned}$$

and describes the frequencies at which events occur within a process. For example, consider the Poisson process which exhibits complete randomness, meaning that all frequencies contribute equally to the spacing of events. Consequently, the Poisson process has a flat spectral density function of value $S(\omega) = \Lambda/2\pi$.

The Spectral Density of a Univariate Hawkes Process

Recall from (2.4.1) the definition of the conditional intensity function $\Lambda(t)$ of a univariate Hawkes process. In this context, the rate function is random, and therefore the stationarity condition is such that $\mathbb{E}\{\Lambda(t)\} = \Lambda$. Under this stationarity assumption, we can obtain

$$\begin{aligned}\Lambda = \mathbb{E}\{\Lambda(t)\} &= \nu + \int_{-\infty}^{t} g(t-u)\mathbb{E}\left\{\frac{dN(u)}{du}\right\} du \\ &= \nu + \Lambda \int_{-\infty}^{t} g(t-u) du\end{aligned}$$

or equivalently (by substitution of $v = t - u$)

$$\Lambda = \nu \bigg/ \left\{1 - \int_{0}^{\infty} g(v) dv\right\}.$$

Thus, we must have $\nu > 0$ and

$$\int_{0}^{\infty} g(v)\ dv < 1, \tag{2.4.6}$$

to result in a stationary process. In the specific case of exponential decay, we have $\int_0^\infty \alpha e^{-\beta v} dv = \alpha/\beta$ and therefore

$$\Lambda = \mathbb{E}\{\Lambda(t)\} = \frac{\beta\nu}{\beta - \alpha}.$$

Remark 2.4.4. *For a Hawkes process with exponential decay, the condition in* (2.4.6) *amounts to the assumption that* $\alpha < \beta$.

To obtain an equation for the covariance density function, we observe (Hawkes, 1971b) that for $\tau > 0$,

$$\begin{aligned}\mu(\tau) &= \frac{\mathbb{E}\{dN(t+\tau)dN(t)\}}{dt^2} - \Lambda^2 \\ &= \mathbb{E}\left[\frac{dN(t)}{dt}\left\{\nu + \int_{-\infty}^{t+\tau} g(t+\tau-u)dN(u)\right\}\right] - \Lambda^2 \\ &= \int_{-\infty}^{\tau} g(\tau - v)\mu_c(v)\ dv.\end{aligned}$$

Applying (2.4.5), we find that for $\tau > 0$,

$$\mu(\tau) = \Lambda g(\tau) + \int_{-\infty}^{\tau} g(\tau - v)\mu(v)\ dv. \tag{2.4.7}$$

Due to the symmetry of $\mu(\tau)$, equation (2.4.7) can be rewritten as

$$\mu(\tau) = \Lambda g(\tau) + \int_0^{\infty} g(\tau + v)\mu(v)\ dv + \int_0^{\tau} g(\tau - v)\mu(v)\ dv \quad (\tau > 0). \tag{2.4.8}$$

In general, the integral in (2.4.8) is difficult to solve. However, an analytical solution is obtainable when $g(\cdot)$ decays exponentially (2.4.2). In particular, Hawkes (1971b) showed that in the exponential case, the covariance density function and spectral density are given by

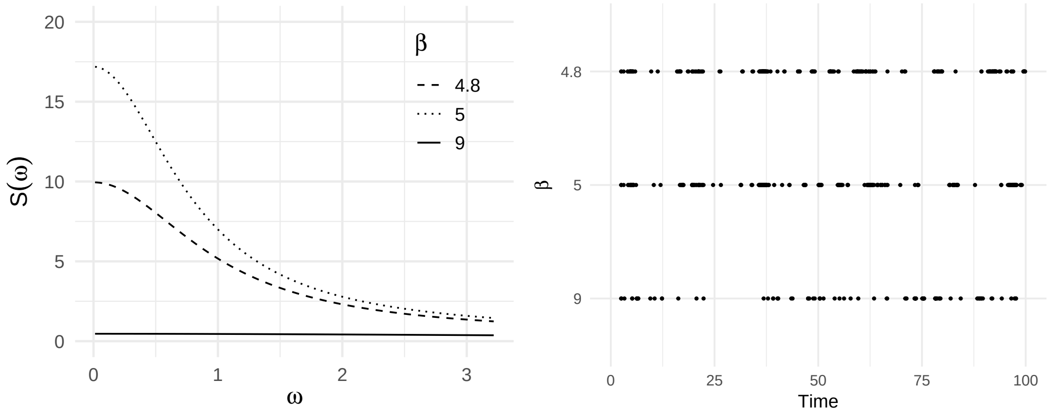
$$\mu(\tau) = \frac{\alpha\nu\beta(2\beta - \alpha)}{2(\beta - \alpha)^2} e^{-(\beta-\alpha)\tau} \quad (\tau > 0),$$

and

$$S(\omega) = \frac{\nu\beta}{2\pi(\beta - \alpha)}\left\{1 + \frac{\alpha(2\beta - \alpha)}{(\beta - \alpha)^2 + \omega^2}\right\},$$

respectively.

Figure 2.4.3 shows three examples of theoretical spectra for the univariate Hawkes process with exponentially decay. Importantly, we highlight the difference in shape of the spectral densities as determined by their parameters. Specifically, when β is large in comparison to α, there is less dependence in the process. In fact, for β large enough, the Hawkes process begins to act more like a Poisson process as evidenced by the flat



(a) (b)

Figure 2.4.3: Example (a) theoretical spectra and (b) temporal point process for a simulated Hawkes processes with exponentially decaying intensities specified by $(\nu, \alpha) = (0.5, 4)$ and varying values of β.

spectra shown by the solid black line. However, when α and β take similar values, there is more dependence in the process, and this is apparent in the spectral density as evidenced by the exponential decay at low frequencies. To aid comparison between the frequency and time-domain representations, Figure 2.4.3b shows the generated point pattern of the simulated Hawkes process with $(\nu, \alpha) = (0.5, 4)$ and varying values of β. Here, we see that concentration in low frequencies of the spectrum coincides with temporal clustering of the point pattern in the time domain.

2.5 Multivariate Point Processes

In this section, we extend our discussion to the multivariate setting whereby we consider p point processes $N_q(t)$ $(q = 1, \dots, p)$ forming a vector process $N(t)$. That is, we consider a p-dimensional point process $N(t)$ whose q^{th} component gives the number of events of type q that have occurred in the time interval $(0, t]$, $t \leq T$. Similar to the univariate setting, we denote by $dN_q(t)$ the number of events observed in process q in some small time interval dt, where $dN_q(t) = N_q(t + dt) - N_q(t)$. Moreover, we characterise the first order properties of $N(t)$ via the intensity function $\Lambda(t) \in \mathbb{R}^p_{\geq 0}$, defined for each unit q as $\Lambda_q(t) := \mathbb{E}\{dN_q(t)\}/dt$.

2.5.1 The Multivariate Hawkes Process

The univariate Hawkes process can be easily extended to the multivariate setting (Hawkes, 1971a). While the univariate case focussed on the self-excitation of an individual process, the main objective in the multivariate setting is to investigate both self-excitation and cross-excitation among different processes. In particular, we will examine processes that are collections of one-dimensional Hawkes processes which excite both themselves and each other. As such, multivariate Hawkes processes are often referred to as "mutually exciting" processes.

Definition 2.5.1 (The Multivariate Hawkes Process)**.** *Consider the multivariate point processes $N_q(t)$ $(q = 1, \ldots, p)$ forming a vector process $N(t)$ such that*

$$\begin{aligned}\mathrm{pr}\left\{dN_q(t) = 1 | N(s)(s \leq t)\right\} &= \Lambda_q(t)dt + o(dt),\\ \mathrm{pr}\left\{dN_q(t) > 1 | N(s)(s \leq t)\right\} &= o(dt),\end{aligned}$$

independently for each q. Now, assume that the stochastic intensity function of the marginal point process $N_q(t)$ is defined as

$$\Lambda_q(t) = \nu_q + \sum_{r=1}^{p} \int_{-\infty}^{t} g_{qr}(t-u)dN_r(u),$$

where $\nu_q > 0$ and $g_{qr} : (0, \infty) \to [0, \infty)$ are the background intensity and excitation functions respectively. Then $N(t)$ is a multivariate Hawkes process.

Alternatively, the intensity function for a multivariate Hawkes process can be represented in vector notation as

$$\Lambda(t) = \nu + \int_{-\infty}^{t} g(t-u)dN(u), \tag{2.5.1}$$

where $\nu \in \mathbb{R}^p$ and g is a $p \times p$ matrix. Motivated by applications in neuroscience, we focus our attention on multivariate Hawkes processes with exponentially decaying intensity functions. In this case, the intensity function for each unit q takes the form

$$\Lambda_q(t) = \nu_q + \sum_{r=1}^{p} \sum_{j:t_j^r<t} \alpha_{qr} e^{-\beta_{qr}(t-t_j^r)}, \tag{2.5.2}$$

where $\nu_q > 0, \alpha_{qr} \geq 0$ and $\beta_{qr} \geq 0$ for $q, r = 1, \ldots, p$. Intuitively, the parameters have the following interpretation: α_{qr} determines the size of the jump in $\Lambda_q(t)$ caused by an event in $N_r(\cdot)$ and β_{qr} determines the decay exhibited in $\Lambda_q(t)$ caused by an event in $N_r(\cdot)$. Hence, the parameter α_{qr} determines the level of self-excitation when $q = r$ and cross excitation when $q \neq r$.

2.5.2 Simulating from the Multivariate Hawkes Process

Realisations from a multivariate Hawkes processes can also be obtained using a thinning algorithm (Ogata, 1981). The main difference compared to the univariate setting is that there is an additional step which decides which dimension an accepted point belongs to. In short, the simulation of multivariate point processes can be achieved by distributing accepted points to each dimension with probabilities proportional to their intensities.

Proposition 2.5.2 (Ogata (1981), p.24, Proposition 1)**.** *Let* $N(t) = \{N_q(t)\}_{q\in\{1,\ldots,p\}}$ *be a multivariate point process on the interval* $(0, T]$ *with stochastic intensities* $\Lambda_q(t)$ *for* $q = 1, \ldots, p$. *Now, suppose there exists a predictable process* $\tilde{\Lambda}(t)$ *which is defined path-wisely satisfying*

$$\sum_{q=1}^{p} \Lambda_q(t) \leq \tilde{\Lambda}(t), \quad 0 < t \leq T,$$

and set $\Lambda_0(t) = \tilde{\Lambda}(t) - \sum_{q=1}^{p} \Lambda_q(t)$. *Let* $\tilde{t}_1, \ldots, \tilde{t}_{\tilde{N}(t)} \in (0, T]$ *be the points of the process* $\tilde{N}(t)$ *with stochastic intensity* $\tilde{\Lambda}(t)$. *For each of the points* $\tilde{t}_k$, *where* $k = 1, \ldots, \tilde{N}(t)$, *we attach a mark* $q = 1, \ldots, p$ *with probability* $\Lambda_q(\tilde{t}_k)/\tilde{\Lambda}(\tilde{t}_k)$ *respectively. Then, the points*

Algorithm 2 (Ogata, 1981) Simulation of a multivariate Hawkes process with exponential excitation function $g_{qr}(u) = \alpha_{qr} e^{-\beta_{qr} u}$ for $q, r = 1, \dots, p$ on $(0, T]$.

```
Require: $\nu \in \mathbb{R}^p, \alpha \in \mathbb{R}^{p\times p}, \beta \in \mathbb{R}^{p\times p}, T$
 1: Initialise $\mathcal{T}^1 = \cdots = \mathcal{T}^p = \emptyset, n^1 = \cdots = n^p = 0, t = 0$;
 2: while $t < T$ do
 3:     Set $\tilde{\Lambda} = \sum_{q=1}^{p} \Lambda_q(t^+) = \sum_{q=1}^{p} \left(\nu_q + \sum_{r=1}^{p} \sum_{\tau\in\mathcal{T}^r} \alpha_{qr} e^{-\beta_{qr}(t-\tau)}\right)$    ▷ Find upper bound
 4:     Let $w \sim \mathrm{Exp}(\tilde{\Lambda})$;
 5:     Set $t = t + w$;    ▷ $t$ is the next candidate point
 6:     Generate $D \sim \mathrm{Unif}(0, 1)$;
 7:     if $D\tilde{\lambda} \le \sum_{q=1}^{p} \Lambda_q(t) = \sum_{q=1}^{p} \left(\nu_q + \sum_{r=1}^{p} \sum_{\tau\in\mathcal{T}^r} \alpha_{qr} e^{-\beta_{qr}(t-\tau)}\right)$ then
 8:         $p = 1$;
 9:         while $D\tilde{\Lambda} > \sum_{q=1}^{p} \Lambda_q(t)$ do    ▷ Find dimension that the point belongs to
10:             $p = p + 1$
11:         end while
12:         $n^p = n^p + 1$;    ▷ Update number of accepted points in dimension $p$
13:         $t^p_{n^p} = t$;    ▷ Name it $t^p_{n^p}$
14:         $\mathcal{T}^p = \mathcal{T}^p \cup \{t^p_{n^p}\}$;    ▷ Add $t^p_{n^p}$ to the ordered set $\mathcal{T}^p$
15:     end if
16: end while
17: if $t^p_{n^p} \le T$ then
18:     return $\mathcal{T}^q$ for $q = 1, \dots, p$
19: else
20:     return $\mathcal{T}^1, \dots, \mathcal{T}^p \backslash \{t^p_{n^p}\}, \dots, \mathcal{T}^p$
21: end if
```

with marks $q = 1, \dots, p$ *provide a multivariate point process with stochastic intensities* $\Lambda_q(t)$ *for* $q = 1, \dots, p$.

Algorithm 2 outlines the pseudocode for the simulation procedure of a multivariate Hawkes process with exponential decays. When compared to Algorithm 1, we can see that main difference in the multivariate setting is the additional step which determines which dimension an accepted point belongs to. More specifically, according to Proposition 2.5.2, we know that given a point is accepted at time t, it should be assigned to dimension q with probability $\Lambda_q(t) / \sum_{q=1}^{p} \Lambda_q(t)$. This is implemented in line 9 of Algorithm 2, where a 'while' loop is used to find the first dimension p such that $D\tilde{\lambda} \le \sum_{q=1}^{p} \Lambda_q(t)$.

2.5.3 The Multivariate Spectrum

Typically, the first and second order properties of the multivariate process $N(t)$ are characterised in the time domain via the intensity function $\Lambda(t) \in \mathbb{R}^p_{\geq 0}$ and covariance density matrix, defined at times t and u as

$$\mu(t, u) = \frac{\mathbb{E}\{dN(u)dN^\top(t)\}}{dtdu} - \Lambda(u)\Lambda^\top(t).$$

However, as mentioned, a large part of this thesis will be dedicated to the discussion of methodological and theoretical advances in spectral analysis for multivariate point processes, building on the frequency domain representation of the stochastic process (Bartlett, 1963a; Brillinger, 1972). Throughout, we will assume that $N(t)$ is second-order stationary, i.e., $\Lambda(t)$ is constant for all t, and $\mu(t, u)$ depends only on lag $\tau = u - t$. In this case, we denote the intensity function as $\Lambda(t) = \Lambda$ and the covariance density matrix simply as $\mu(\tau)$.

The spectral density matrix of a stationary point process is subsequently defined as the Fourier transform of its complete covariance density matrix, namely

$$S(\omega) = \frac{1}{2\pi}\left\{\text{diag}(\Lambda) + \int_{-\infty}^{\infty} \exp(-i\omega\tau)\mu(\tau)d\tau\right\}, \tag{2.5.3}$$

where $S(\omega)$ is a $p \times p$ Hermitian positive definite matrix. Analogous to the covariance matrix in the time domain, the spectral density matrix captures both the within and between type dynamics of the multivariate point process. More specifically, it describes the variance in each process or the covariance between processes that is attributable to oscillations in the data at a particular frequency $\omega \in \mathbb{R}$.

Given the complex nature of the spectrum, a common measure of association between two processes $N_q(t)$ and $N_r(t)$ is the squared spectral coherence,

$$R^2_{qr}(\omega) := \frac{|S_{qr}(\omega)|^2}{S_{qq}(\omega)S_{rr}(\omega)} \in [0, 1]. \tag{2.5.4}$$

If $R^2_{qr}(\omega) = 0$ for all ω, then the two processes are linearly independent. Conversely, a value of 1 indicates a perfect linear relationship. More specifically, $R^2_{qr}(\omega)$ provides a measure of the correlation structure between pairs of process in the frequency domain, and has been used in the literature for the analysis of neuronal spike train data (Jarvis and Mitra, 2001; Cohen and Gibberd, 2021).

The Spectral Density Matrix of a Multivariate Hawkes Process

Assuming stationarity, we have from (Hawkes, 1971a) and (2.5.1) that

$$\begin{aligned}\Lambda &= \mathbb{E}\{\Lambda(t)\}\\ &= \nu + \int_{-\infty}^{\infty} g(u)\Lambda du\end{aligned}$$

so that

$$\Lambda = (I - G(0))^{-1}\nu,$$

where ν is the vector of background intensities, $\mathbb{I}$ is the identity matrix and G is the Fourier Transform of the excitation function. In particular, for the case of exponential decay, we have

$$\begin{aligned}G_{qr}(\omega) &= \int_{-\infty}^{\infty} \exp(-i\omega\tau)\alpha_{qr}\exp(-\beta_{qr}\tau)d\tau\\ &= \frac{\alpha_{qr}}{\beta_{qr} + i\omega}.\end{aligned}$$

Furthermore, for stationarity, we require $\max_j |\phi_j(G(0))| < 1$ where $\phi_j(\cdot)$ represents the j^{th} eigenvalue. The spectral density matrix of a stationary multivariate Hawkes process is given in Hawkes (1971a) as

$$S(\omega) = \frac{1}{2\pi}\{I - G(\omega)\}^{-1} D \left\{I - G^{\top}(-\omega)\right\}^{-1},$$

where $D = \mathrm{diag}(\Lambda)$.

2.5.4 Estimation of the Multivariate Spectrum

A classical non-parametric estimator of the spectral density matrix for a stationary point process is based on the periodogram, that is a (Hermitian) outer-product of the Fourier transform of the data (Bartlett, 1963a). This estimate is often improved upon by applying a taper to the data prior to taking the Fourier transform (Percival et al., 1993). The main purpose of the taper is to smooth the ends of the point process and remove any high-frequency artefacts, arising from an abrupt start or end to the point process, which in turn reduces bias in the periodogram.

In the sequel, we propose to utilise a variant of the multitaper periodogram (3.2.2) to estimate the spectral density matrix. However, it is first useful to understand some properties of this object in relation to the point process (continuous time) setting.

The Tapered Fourier Transform for Point Processes

As in regular time series analysis, different choices of taper functions may be useful for different tasks, each having their own concentration properties in the Fourier domain. For the purposes of our discussion, we focus on the canonical choice of $k = 1, \ldots, m$ non-overlapping tapers that act to segment the process $N(t)$ into m intervals of equal length n. This aligns with the experimental motivation for our methodology, where we can attribute a single taper to each of the m trials in our neuronal-spiking experiment (see Figure 2.3.3). For simplicity, we choose $h_k(t/T) = (T/2\pi z)^{1/2}$ where $\mathrm{supp}(h_k) \cap \mathrm{supp}(h_l) = \varnothing, \forall k \neq l$ and $|\mathrm{supp}(h_k)| = z$ for all $k = 1, \ldots, m$, where $z := n/T$. That is, the tapers are supported on a subset of $[0, 1]$, i.e., can be considered as being defined in re-scaled time. However, the framework developed in Chapter 3 can be extended to different choices of taper function and/or trial length.

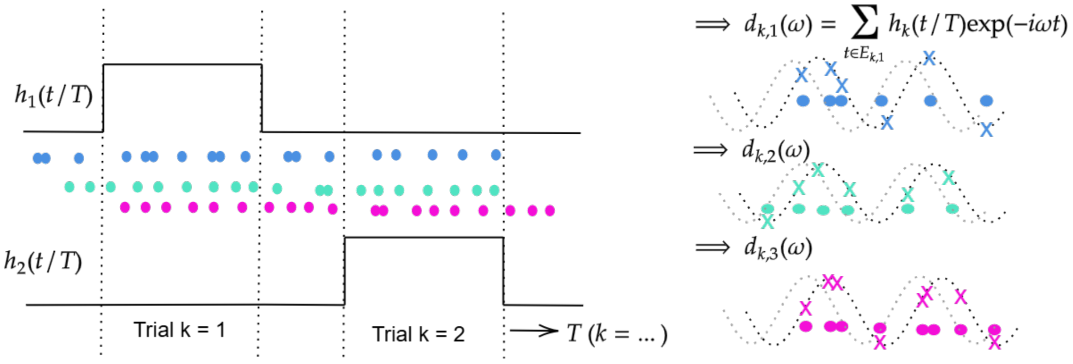


Figure 2.5.1: Diagrammatic description of the chosen taper functions and resulting Fourier transforms.

Definition 2.5.3. *The tapered Fourier Transform of the process $N_q(t)$ for $t \in (0, T]$ is defined as*

$$d_{T;k;q}(\omega) := \int_0^T h_k(t/T)\exp(-i\omega t)dN_q(t).$$

An alternative view of the Fourier transform is given by the random set of events induced by the point process. That is, instead of the counting process representation, we can consider the events $E_{k,q} = \{t \mid dN_q(t) = 1\ ,\ t/T \in \mathrm{supp}(h_k)\}$. In this case, we can write

$$d_{T;k;q}(\omega) = \sum_{t \in E_{k,q}} h_k(t/T)\exp(-i\omega t).$$

As the rate of the point process is non-zero and positive, it is important to remove this mean prior to obtaining estimates of the spectrum (Jarvis and Mitra, 2001; Rajala et al., 2023). Therefore, we will work with these mean-corrected variates to obtain estimates of the spectrum.

Definition 2.5.4. *The mean-corrected tapered Fourier Transform is defined as*

$$\bar{d}_{T;k;q}(\omega) := d_{T;k;q}(\omega) - d_{T;k;q}(0)\frac{H_k(T\omega)}{H_k(0)}, \tag{2.5.5}$$

where $H_k(\omega)$ denotes the Fourier transform of $h_k(z)$, i.e. $H_k(\omega) = \int h_k(z)\exp(-i\omega z)dz$.

At this point, it is useful to introduce further conditions on $N(t)$, enabling further exploration and understanding of the asymptotic ($T \to \infty$) behaviour of the Fourier transform.

Assumption 2.5.5. *(Brillinger, 1972). Assume the p-dimensional point process $N(t)$ is strictly stationary, i.e. the j^{th}-order cumulant*

$$C_{q_1,\ldots,q_j}(t_1,\ldots,t_j)dt_1\ldots dt_j = \operatorname{cum}\{dN_{q_1}(t_1),\ldots,dN_{q_j}(t_j)\},$$

is invariant to shifts in time. Let us define the reduced cumulant density $r_{q_1,\ldots,q_j}(u_1,\ldots,u_{j-1}) = C_{q_1,\ldots,q_j}(u_1,\ldots,u_{j-1},0)$, then we further assume all moments exist, and that

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left|r_{q_1,\ldots,q_j}(u_1,\ldots,u_{j-1})\right|\mathrm{d}u_1\cdots\ \mathrm{d}u_{j-1} < \infty\ ,$$

holds for any $q_1,\ldots,q_j = 1,\ldots,p$ and all $j \geq 2$. Finally, we assume the covariance density satisfies

$$\int_{-\infty}^{\infty}|u|\,|r_{q,r}(u)|\,\mathrm{d}u < \infty\ ,$$

for all $q, r = 1,\ldots,p$.

The above conditions amount to a form of mixing condition (Assumption 2.2 Brillinger, 1972) which ensures that the dependence of the point process decays at a sufficient rate to enable the following central limit-theorem on the finite Fourier transform.

Proposition 2.5.6 (Theorem 4.2, Brillinger (1972))**.** *Let $N(t)$ be a p-dimensional point process satisfying Assumption 2.5.5. Furthermore, let f_T be an integer with $\omega_T = 2\pi f_T/T \to \omega$ as $T \to \infty$ and*

$$d_{T;k}(\omega) = \int_0^T \exp(-i\omega t)dN(t),$$

for $-\infty < \omega < \infty$. Then, $d_{T;k}(\omega_T), k = 1,\ldots,m$, are asymptotically independent

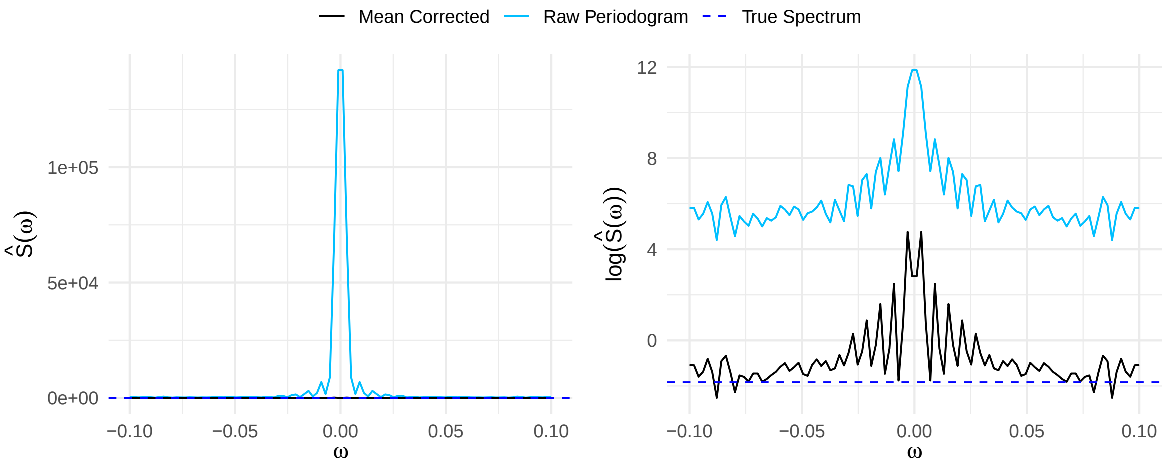


Figure 2.5.2: Comparison of the raw periodogram vs the mean corrected periodogram.

$N_p^c(0, 2\pi T S(\omega))$ *variates, respectively. Also,* $d_{T;k}(0) = N(0, T]$ *is asymptotically normal independently of the previous variates.*

Remark 2.5.7. *In Chapter 3, we will show that for our specific choice of taper function, the mean corrected coefficients* (2.5.5) *are also asymptotically complex Gaussian variates.*

The Multi-Trial Periodogram

We now use this classical result to motivate studying a multi-trial estimate of the spectrum,

$$\hat{S}_m(\omega) := \frac{1}{m} \sum_{k=1}^{m} \bar{d}_{T;k}(\omega) \bar{d}_{T;k}^H(\omega), \tag{2.5.6}$$

where $k = 1, \dots, m$ represent distinct trials (Fig 2.5.1). We demonstrate the practical importance of using the mean-corrected variates in (2.5.6) using synthetic data simulated from a homogeneous Poisson process with rate $\lambda = 1$. Estimates of the spectrum were obtained at 100 equally spaced frequencies in the range $\omega \in [-0.1, 0.1]$ via the raw and mean-corrected periodogram estimators. Figure 2.5.2 shows that frequencies close to $\omega = 0$ should be avoided when the mean-correction is not applied. Moreover, even when using the mean-corrected variates, there is indeed deviation from the true spectrum around $\omega = 0$, albeit very localised to the zero frequency. Thus, our findings align with those in Jarvis and Mitra (2001) and Rajala et al. (2020) who highlight that the periodogram is biased unless mean-corrected.

Asymptotic Distributional Results

Under Assumption 2.5.5, the estimator in (2.5.6) is an asymptotically, p fixed, $T \to \infty$ unbiased estimator of the spectrum. Moreover, due to the asymptotic complex normality of $\bar{d}_{T;k}(\omega)$, we have that the estimator in (2.5.6) is distributed as a Complex Wishart distribution, i.e.,

$$\hat{S}_m(\omega) \sim m^{-1} \mathcal{W}_p^c(m, S(\omega)),$$

with m degrees of freedom and centrality matrix $S(\omega)$ (Brillinger, 1972). As $m \to \infty$, $\hat{S}_m(\omega)$ is a consistent estimator of the spectrum. As a further corollary of the asymptotic complex normality, we know from Goodman (1963) that if $p < m$, then the estimated squared spectral coherence (2.5.4) has the density function

$$f_{\hat{R}^2_{qr}}(x) = (m-1)\big(1 - R^2_{qr}\big)\big(1 - x^2\big)^{(m-2)} \; {}_2F_1\big(m, m; 1; R^2_{qr} x\big), \tag{2.5.7}$$

where ${}_2F_1(\alpha_1, \alpha_2; \beta_1; z)$ denotes the hypergeometric function with 2 and 1 parameters α_1, α_2 and β_1 and scalar argument z.

To demonstrate how these asymptotic properties hold in practice, we give an example of the empirical distribution of the coherence estimated from synthetic data simulated from a multivariate Hawkes process. In particular, we simulate data from a

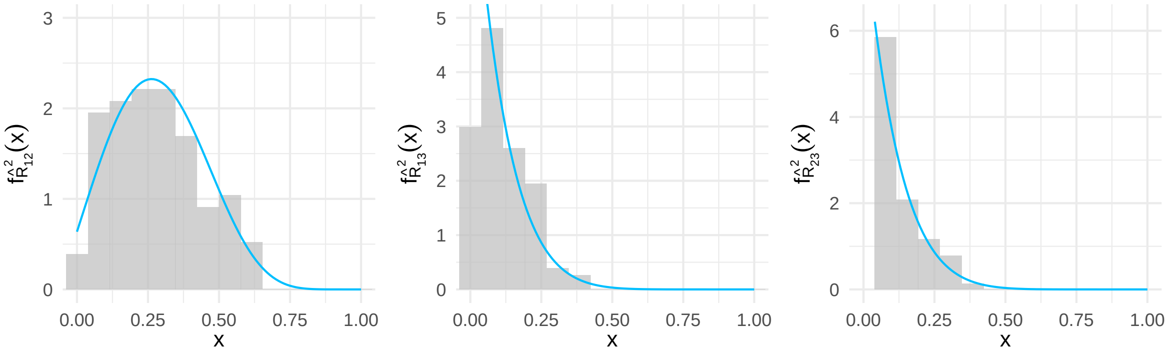


Figure 2.5.3: Distribution of different entries in the estimated squared coherence matrix using synthetic data simulated from multivariate Hawkes process. Fit to the asymptotic distribution shown in blue.

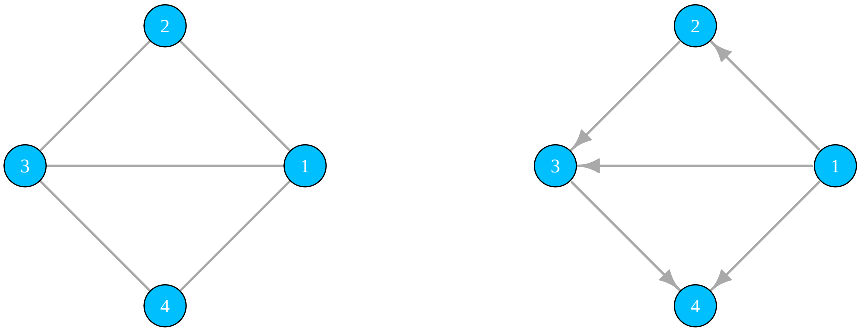
3-dimensional Hawkes process with parameters

$$\nu = (0.2, 0.2, 0.2), \quad \alpha = \begin{pmatrix} 0.00 & 0.25 & 0.00 \\ 0.00 & 0.25 & 0.00 \\ 0.00 & 0.00 & 0.25 \end{pmatrix}, \quad \beta = \begin{pmatrix} 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 \end{pmatrix}.$$

By construction, this is a multivariate Hawkes process in which there is cross-excitation between processes $N_1(t)$ and $N_2(t)$, $N_3(t)$ is subject only to self-excitation.

Data was simulated on the interval $(0, 1000]$ via Ogata's modified thinning algorithm, using the `hawkes` package in R. Estimates of the multivariate spectrum were obtained over 100 Monte Carlo samples, using the multi-trial periodogram estimator (2.5.6) with $m = 10$ trials.

The results, shown in Figure 2.5.3, demonstrate the appropriateness of the asymptotic distribution (2.5.7) for large T, when $m > p$. Moreover, the dependence between processes $N_1(t)$ and $N_2(t)$ is evident in the distribution of the coherence estimate $\hat{R}^2_{12}(\omega)$. Comparatively, the distributions of $\hat{R}^2_{13}(\omega)$ and $\hat{R}^2_{23}(\omega)$ are skewed much closer to zero, and values of $\hat{R}^2_{i,3}(\omega) > 0.4$, for $i = 1, 2$, have little to no probability of being observed.



(a) Undirected graph (b) Directed graph

Figure 2.6.1: Examples of different graph types, with $V = \{1, \ldots, 4\}$ and five edges.

2.6 Graphical Models

2.6.1 Directed and Undirected Graphical Models

As mentioned, our primary interest lies in inferring a network structure to summarise the interactions between neural processes. Mathematically, this translates to constructing a graph $G = (V, E)$ consisting of a node set $V = \{1, \ldots, p\}$ representing each of the $q = 1, \ldots, p$ neurons and an edge set $E \subseteq V \times V$ where an edge $(q, r) \in E$ for $q \in V$ and $r \in V$ encodes the presence of a relation between neurons q and r. Thus, in this thesis, vertices represent neurons and edges represent interactions.

Generally, graphs are categorised as either undirected (Figure 2.6.1a) or directed (Figure 2.6.1b). If undirected, then $(q, r) \in E \iff (r, q) \in E$. However, this is not necessarily the case in a directed graph, where edges have a direction. In the case of our neural spike train example, a directed graph would capture causal or time-ordered interactions, representing effective connections (e.g., Zhao et al., 2012; Wang et al., 2025; Wang and Shojaie, 2021). On the other hand, an undirected graph models symmetric interactions, without implying causality, and is often used in functional connectivity analyses (e.g., Pinkney et al., 2024; Deb et al., 2024).

Often, graphs are represented in terms of their adjacency matrix A. Specifically, for a non-weighted graph, we write $A \in \{0,1\}^{p \times p}$ where

$$A_{qr} = \begin{cases} 1 & \text{if } (q,r) \in E \\ 0 & \text{if } (q,r) \notin E. \end{cases}$$

That is, a non-zero $(q,r)^{th}$ entry in A indicates the presence of an edge between nodes q and r in the corresponding graph. For weighted graphs, where each edge $e \in E$ has an associated weight w_e, the non-zero entries are set equal to the edge weights, i.e.,

$$A_{qr} = \begin{cases} w_e & \text{if } (q,r) \in E \\ 0 & \text{if } (q,r) \notin E. \end{cases}$$

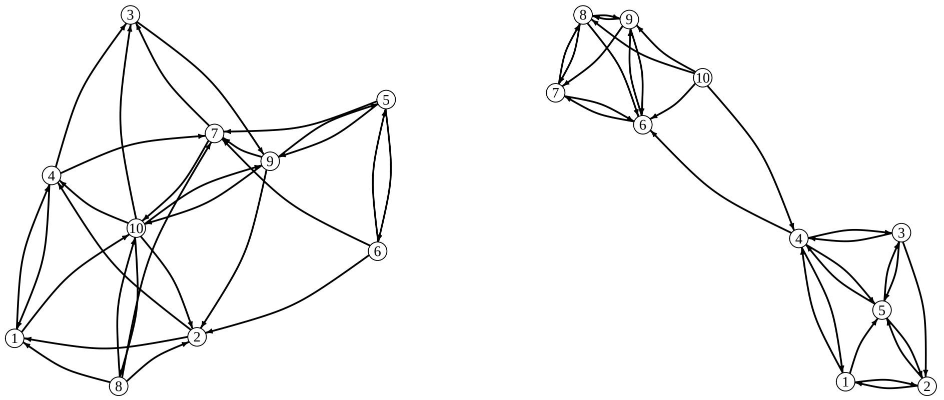
In the case of undirected graphs, the adjacency matrix A is symmetric.

2.6.2 Random Graph Models

In Chapters 3 and 4, we will be interested in evaluating the performance of our proposed network estimation methods on simulated data whose 'ground truth' is known. In particular, we will use random graph models to specify the interactions between neural processes, thus enabling the comparison between our estimates of neural connectivity the true underlying network structure.

Random graph models are networks in which edges between nodes are determined probabilistically rather than deterministically. A classic example is the Erdős-Rényi (ER) model (Erdős, 1960), which assumes that an edge $(q,r) \in E$ is included in the graph with some fixed probability, independent of all the other edges. That is, one can sample the entries of the adjacency matrix independently as

$$A_{qr} \sim \text{Bernoulli}(\tilde{p}),$$



(a) (b)

Figure 2.6.2: Illustrations of the different network structures generated by (a) the Erdős-Rényi random graph and (b) the Stochastic block model with $d = 10$ nodes.

where $\tilde{p} \in (0, 1)$ denotes the probability of an edge appearing between any two nodes. While the simplicity of the ER model makes it an attractive candidate for use in many elementary understanding of basic network properties, its main limitation is that it fails to capture key features of many real-world systems, e.g., clustering or community structure.

A popular alternative, which addresses some of these limitations, is the stochastic block model (SBM), which extends the idea of random graphs to incorporate community structures. In particular, the node set V is partitioned into K disjoint sets, known as blocks, and the probability between two nodes depends on the blocks to which they belong. More specifically, consider the matrix $Z = (z_1, \ldots, z_p)^\top$ with $z_q = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ and $z_{qk} = 1$ indicating that the q^{th} node belongs to the k^{th} block. Given a $K \times K$ matrix B with $B_{kl} \in (0, 1)$ representing the probability of an edge between blocks k and l, a graph G can be sampled via the adjacency matrix

$$A_{qr} | B, Z \sim \text{Bernoulli}(z_q^\top B z_r).$$

Typically, the SBM model is parametrised such that nodes within the same block have a higher chance of being connected, compared to nodes in different blocks. In the context of our neuronal spike train analysis, one could imagine the blocks as being different brain regions, whereby connections within a region are (perhaps) more likely than those between regions.

While the ER and SB models will be used in the synthetic experiments of Chapter 4, analysing the properties of these random graphs is not the focus of this thesis. Comparatively, the proposed methods align better with the time-series and Gaussian graphical modelling literature, which we will discuss in the remainder of this chapter.

2.6.3 Gaussian Graphical Models

Building on the discussion of directed and undirected graphical models, we now turn our attention to an important subgroup of the latter, called Gaussian Graphical models (GGMs). These models capture the conditional independence structure among a set of normally distributed random variables.

Let $X = (X_1, \ldots, X_p)^\top$ denote a zero-mean p-dimensional Gaussian random vector, with density

$$f(x_1, \ldots, x_p; \Theta) = \frac{1}{\sqrt{(2\pi)^p \det(\Theta^{-1})}} \exp\left\{-\frac{1}{2}x^\top \Theta x\right\},$$

parametrised by the inverse covariance matrix $\Theta = (\Sigma)^{-1} \succ 0$.

GGMs model the conditional dependence structure among the variables $(X_1, \ldots, X_p)$ via an undirected graph $G = (V, E)$ where each node $q \in V$ corresponds to a variable X_q and an edge between nodes q and r, i.e., $(q, r) \in E$, exists if and only if X_q and X_r are conditionally dependent given all other variables. That is, the pairwise independence $X_q \perp X_r | X_{V\setminus\{q,r\}}$ for an edge (q, r) implies its exclusion from the set E. In the Gaussian case, these pairwise partial-covariances are encoded in the inverse covariance matrix Θ, i.e., $(X_q \perp X_r | X_{V\setminus\{q,r\}}) \iff \Theta_{qr} = 0$ (Lauritzen, 1996).

Accurate estimation of Θ is therefore central to revealing the intricate dependencies between variables, and the resulting graphical model. Given n i.i.d. observations $\{X^{(k)}\}_{k=1}^{n}$ of the zero mean random vector $X \in \mathbb{R}^p$, our interest lies in the estimation of Θ, to infer the edge structure of the so-called partial correlation graph. Under the normal assumption, the negative log-likelihood (up to a constant) can be written in terms of the inverse covariance matrix Θ as

$$\ell(X^{(1)}, \dots, X^{(n)}; \Theta) = \operatorname{tr}(\Theta\hat{\Sigma}) - \log\det(\Theta),$$

where

$$\hat{\Sigma} := \frac{1}{n}\sum_{k=1}^{n} X^{(k)}(X^{(k)})^\top,$$

is the sample covariance matrix. The resulting maximum likelihood estimator for the inverse covariance matrix is given as

$$\hat{\Theta}_{\mathrm{ML}} := \operatorname{argmin}_{\Theta\succeq 0}\left\{-\log\det(\Theta) + \operatorname{tr}(\Theta\hat{\Sigma})\right\}.$$

In general, the estimator $\hat{\Theta}_{\mathrm{ML}}$ is dense and therefore leads to an almost surely complete graph, i.e., where all nodes are connected. Moreover, in high-dimensional settings (i.e., when $p > n$) the covariance estimator $\hat{\Sigma}$ is rank deficient and there is no unique inverse for $\hat{\Sigma}$. In these settings, it is common to add a penalty e.g. to encourage sparsity, and indeed interpretability of the GGM. In particular, the graphical Lasso estimator is defined

$$\hat{\Theta}_{GL} := \operatorname{argmin}_{\Theta\succeq 0}\left\{-\log\det(\Theta) + \operatorname{tr}(\Theta\hat{\Sigma}) + \lambda\|\Theta\|_1\right\}, \tag{2.6.1}$$

where λ is a non-negative tuning parameter, $\|\Theta\|_1 = \sum_{q\neq r}\Theta_{qr}$ is the ℓ_1 norm of Θ, and the minimisation is taken over symmetric positive definite matrices. Typically, the optimisation problem in (2.6.1) is solved using coordinate descent (Friedman et al., 2008; Mazumder and Hastie, 2012) or via the alternating direction method of multipliers algorithm (Boyd et al., 2011).

Alternative approaches for sparse estimation of the inverse covariance matrix in high-dimensional settings include the CLIME (constrained ℓ_1 minimisation for inverse matrix estimation) estimator of (Cai et al., 2011) and neighbourhood selection (Meinshausen and Bühlmann, 2006).

2.6.4 Theory for Gaussian Graphical Models

An important part of the GGM literature concerns the analysis of the graphical Lasso estimator (2.6.1) in high-dimensional settings, i.e., when $p > n$. In particular, it is often of interest to distinguish how close the estimated precision matrix $\hat{\Theta}_{GL}$ is to the ground truth Θ. In this section, we will introduce error bounds for the graphical Lasso estimator, with particular reference to works of Rothman et al. (2008) and Ravikumar et al. (2011).

Deviation bound for the Sample Covariance Matrix

Since $\hat{\Sigma}$ is used as a surrogate for the unknown covariance in (2.6.1), any type of consistency for $\hat{\Theta}_{GL}$ requires a bound on the distance $\hat{\Sigma} - \Sigma$. Thus, we will first discuss a deviation bounds for the sample covariance matrix, prior to introducing results for the graphical Lasso estimator.

Below, we state a Lemma from Ravikumar et al. (2011) showing that the entries of $\hat{\Sigma} - \Sigma$ satisfy an exponential-type tail bound, under a sub-Gaussian assumption of the re-scaled variates $(X_q/\sqrt{\Sigma_{qq}})_{q=1}^{p}$.

Definition 2.6.1. *A zero mean random variable Z is sub-Gaussian if there exists a constant $\sigma \in (0, \infty)$ such that*

$$\mathbb{E}\{\exp(tZ)\} \leq \exp(\sigma^2 t^2/2) \quad \text{for all } t \in \mathbb{R}.$$

Proposition 2.6.2 (Lemma 1, Ravikumar et al. (2011))**.** *Consider a zero-mean random* *vector* $(X_1, \ldots, X_p)$ *with covariance* Σ *such that each* $X_q/\sqrt{\Sigma_{qq}}$ *is sub-Gaussian with parameter* σ. *Given* n *i.i.d. samples, the associated sample covariance* $\hat{\Sigma}$ *satisfies the tail bound*

$$\mathrm{pr}\left(\left|\hat{\Sigma}_{qr} - \Sigma_{qr}\right| > \delta\right) \leq 4 \exp\left\{-\frac{n\delta^2}{128(1+4\sigma^2)^2 \max_q(\Sigma_{qq})^2}\right\}$$

for all $\delta \in (0, \max_q(\Sigma_{qq})8(1+4\sigma^2))$.

Remark 2.6.3. *The proof of Proposition 2.6.2, which is omitted here for brevity, relies on a sub-Gaussianity assumption of the* $(X_q/\sqrt{\Sigma_{qq}})_{q=1}^p$ *variates. In Chapter 3, we extend this result to the complex setting, and provide a deviation bound for our multi-trial periodogram estimate of the spectrum* (2.5.6) *based on the asymptotic complex normality of our mean-corrected coefficients* (2.5.5). *Then, we derive a large (but finite) sample deviation bound under an assumption of weak dependence (Doukhan and Louhichi, 1999; Doukhan and Neumann, 2007) to study the finite sample properties of our spectral estimators.*

Lemma 2.6.2 ensures that each entry of $\hat{\Sigma}$ concentrates tightly around its expectation, even when $p > n$. This result is essential for the finite sample analysis of the graphical Lasso estimator $\hat{\Theta}_{GL}$.

Remark 2.6.4. *A similar result is given in Rothman et al. (2008), where the authors directly apply Lemma* 3 *from Bickel and Levina (2008) under the assumption that the minimum and maximum eigenvalues of* Σ *are bounded.*

Error bounds for the graphical Lasso (Rothman et al., 2008)

Foundational analysis for the graphical Lasso estimator was described in Rothman et al. (2008), who provided rates in Frobenius norm. Their result required the following assumptions

A1. Let the set $E(\Theta) = \{(q,r) : \Theta_{qr} \neq 0, q \neq r\}$. Then $\mathrm{card}\{E(\Theta)\} \leq s$.

A2. $\phi_{\min}(\Sigma) \geq \underline{k} > 0$, or equivalently $\phi_{\max}(\Theta) \leq 1/\underline{k}$.

A3. $\phi_{\max}(\Sigma) \leq \bar{k}$.

Proposition 2.6.5 (Theorem 1, Rothman et al. (2008)). *Let $\hat{\Theta}_{GL}$ be the minimiser defined by* (2.6.1)*. Under $A1, A2, A3$ if $\lambda \asymp \sqrt{\frac{\log p}{n}}$*

$$\|\hat{\Theta}_{GL} - \Theta\|_F = \mathcal{O}_p\left(\sqrt{\frac{(p+s)\log p}{n}}\right).$$

Remark 2.6.6. *The worst part of the rate, $\sqrt{p \log p/n}$, comes from estimating the diagonal. Therefore, it is possible to reduce the error rate from $\sqrt{(p+s)\log p/n}$ to $\sqrt{s \log p/n}$ by using the sample correlation matrix instead of the sample covariance matrix. See Rothman et al. (2008) for further details.*

Thus, Rothman et al. (2008) established consistency of the graphical Lasso estimator via Proposition 2.6.5. However, it remained to show that with high probability, $\hat{\Theta}_{GL}$ could correctly identify the zero pattern of the true inverse covariance matrix. This result was achieved in later work by Ravikumar et al. (2011) who provided stronger results which guaranteed model selection consistency of the graphical Lasso estimator.

Rates in Elementwise ℓ_∞-norm and Model Selection Consistency (Ravikumar et al., 2011)

Recall from assumption A1 that the set of off-diagonal entries in Θ used to define the edge set of the graph is denoted by $E(\Theta)$. Ravikumar et al. (2011) define the augmented set $S(\Theta) = \{E(\Theta) \cup (1,1), \ldots, (p,p)\}$ which includes the diagonal elements, and $S^c(\Theta)$ denotes the complement of $S(\Theta)$ corresponding to all pairs (q,r) for which $\Theta_{qr} = 0$. The shorthand S and S^c are adopted for ease of notation. Moreover, the so-called sparsity index $s := \text{card}\{E(\Theta)\}$ denotes the total number of non-zero elements in off-diagonal positions of Θ, i.e., twice the number of edges in a GGM. Moreover, the maximum node degree is defined as

$$d := \max_{q=1,\dots p} \left|\{r \in \{1,\dots,p\} : \Theta_{qr} \neq 0\}\right|,$$

which denotes the maximum number of non-zeros in any row of Θ.

Let the Hessian of the multivariate Gaussian log-likelihood evaluated at the true concentration matrix be given by $\Gamma = \Sigma \otimes \Sigma$, where $\otimes$ denotes the Kronecker matrix product. By definition, Γ is the Fisher information of the model, and we have that $\Gamma_{(q,r)(s,t)} = \mathrm{cov}(X_q X_r, X_s X_t)$. Hence, Γ can be viewed as an edge-based counterpart to the usual covariance matrix Σ.

The main results in Ravikumar et al. (2011) require an application of the ℓ_∞-operator norm to the covariance matrix Σ and to the inverse of a sub-block of the Fisher information Γ. More specifically, the ℓ_∞-operator norm of the true covariance matrix is defined as

$$\kappa_\Sigma := |||\Sigma|||_\infty = \left(\max_{q=1,\dots,p} \sum_{r=1}^{p} |\Sigma_{qr}|\right). \tag{2.6.2}$$

Similarly, we consider the subset of the Hessian relevant to the true model subset

$$\Gamma_{SS} = (\Sigma \otimes \Sigma)_{SS} \in \mathbb{R}^{(s+p)\times(s+p)}$$

and let

$$\kappa_\Gamma := |||(\Gamma_{SS}^{-1})|||_\infty. \tag{2.6.3}$$

The quantity in (2.6.3) captures the magnitude of the Hessian restricted on the true support set, whereas the quantity in (2.6.2) captures the size of the covariance matrix Σ, by looking at the largest row. These quantities play an important role in describing the behaviour of the graphical Lasso estimator, and are used in the final result (Proposition 2.6.8). Finally, the Hessian is assumed to satisfy the following irrepresentability condition.

Assumption 2.6.7. *There exists some* $\alpha \in (0, 1]$ *such that*

$$\max_{e \in S^c} \|\Gamma_{eS}(\Gamma_{SS})^{-1}\|_1 \le (1 - \alpha). \tag{2.6.4}$$

The intuition behind assumption 2.6.7 is that it limits the influence that the non-edge terms in S^c can have on the edge terms in S. More specifically, if we define the zero-mean edge random variables by

$$Y_{(q,r)} := X_q X_r - \mathbb{E}(X_q X_r)$$

for all $q, r \in \{1, \ldots, p\}$ and note that $\Gamma_{(q,r),(s,t)} = \mathbb{E}\{Y_{(q,r)} Y_{(s,t)}\}$. Then, letting the vector $Y_S := \{Y_{(q,r)}, (q, r) \in S\}$, the irrepresentability condition reduces to

$$\max_{e \in S^c} \|\mathbb{E}(Y_e Y_S^\top)\mathbb{E}(Y_S Y_S^\top)^{-1}\|_1 \le (1 - \alpha).$$

Therefore, the irrepresentability condition (2.6.4) can be interpreted as enforcing the the requirement that there should be no edge variable, (e.g., $Y_{(q,r)}$) that is not included in the graph (i.e., $(q, r) \in S^c$) that is highly correlated with variables within the true edge-set.

The following result establishes an elementwise deviation bound for the graphical Lasso estimator which can be used to establish consistent estimation of the GGM.

Proposition 2.6.8. *(Ravikumar et al., 2011). Consider a zero-mean random vector* $(X_1, \ldots, X_p)$ *with covariance* Σ *such that each* $\tilde{X}_q = X_q/\sqrt{\Sigma_{qq}}$ *is sub-Gaussian with parameter* σ *and the samples are drawn independently. Suppose also that* $(\tilde{X}_1, \ldots, \tilde{X}_q)$ *satisfies the the incoherence assumption* (2.6.7) *with parameter* $\alpha \in (0, 1]$*. Then, if the* *sample size is lower bounded as*

$$n > C_1 d^2 \left(1 + \frac{8}{\alpha}\right)^2 (\tau \log p + \log 4),$$

where $C_1 := \{48\sqrt{2}(1+4\sigma^2)\max_q\{\Sigma_{qq}\}\max\{\kappa_\Sigma, \kappa_\Gamma, \kappa_\Sigma^3, \kappa_\Gamma^2\}\}^2$, *then, under sufficient regularisation, i.e.,* $\lambda \asymp \sqrt{\tau \log p/n}$, *with probability* $1 - 1/p^{\tau-2} \to 1$, *we have:*

(a) The estimate $\hat{\Theta}_{GL}$ *satisfies the elementwise* ℓ_∞*-bound:*

$$\|\hat{\Theta}_{GL} - \Theta\|_\infty \leq \{16\sqrt{2}(1+4\sigma^2)\max_q\{\Sigma_{qq}\}(1+8\alpha^{-1})\kappa_\Gamma\}\sqrt{\frac{\tau \log p + \log 4}{n}},$$

(b) It specifies an edge set $E(\hat{\Theta}_{GL})$ *that is a subset of the true edge set* $E(\Theta)$ *and it includes all edges* (q, r) *with*

$$|\Theta_{qr}| > \{2(1+8\alpha^{-1})\kappa_\Gamma\}\sqrt{128(1+4\sigma^2)^2\max_q(\Sigma_{qq})^2}\sqrt{\frac{\tau \log p + \log 4}{n}}.$$

Remark 2.6.9. *The above result is a corollary of Theorem 1 in Ravikumar et al. (2011) specific to the sub-Gaussian setting. The parameter* $\tau > 2$ *reflects the rate of convergence in probability, where a high value results in high-probability claims. However,* τ *also appears in the setting of the sample size* n*, whereby a larger* τ *will result in an increased lower bound on the sample size. Moreover, it appears in the choice of the regularisation parameter, i.e.,* $\lambda = c\{\tau \log p/n\}^{1/2}$ *for some constant* $c > 0$*, ensuring that the maximum deviation of the sample covariance from the true covariance is bounded.*

Part (b) of Proposition 2.6.8 states that the edge set $E(\hat{\Theta}_{GL})$ returned by the graphical Lasso estimator is contained within the true edge set $E(\Theta)$. However, sign consistency can also be achieved by modifying Proposition 2.6.8 to provide sufficient conditions linking the lower bound on the sample size with the minimum value

$$\theta_{\min} := \min_{(q,r)\in E(\Theta)} |\Theta_{qr}|.$$

In this case, the graphical Lasso estimator will not only have the same edge set as Θ, but it will also recover the correct signs on these edges. That is, if we define the event

$$\mathcal{M}(\hat{\Theta}_{GL};\Theta) := \Big\{\text{sign}(\hat{\Theta}_{qr}) = \text{sign}(\Theta)\ \forall(q,r) \in E(\Theta)\Big\},$$

then with sufficient modification to the sample size $n = \Omega((d^2 + \theta_{\min}^{-2})\tau \log p$, the graphical Lasso estimator is model selection consistent with high probability as $p \to \infty$,

$$\text{pr}\Big\{\mathcal{M}(\hat{\Theta}_{GL};\Theta)\Big\} \geq 1 - 1/p^{\tau-2} \to 1.$$

Remark 2.6.10. *A proof of the above result, using the primal-dual witness approach, can be found in Theorem* 2 *in Ravikumar et al. (2011).*

2.6.5 Time Series Graphical Models

In this section, we extend the graphical modelling framework to the multivariate time series setting. In particular, we discuss the estimation and structure learning of the inverse spectral density matrix, a frequency domain analogue of the inverse covariance matrix. Analogous to the GGM framework, partial coherence, which measures the partial correlation structure between pairs of time series in the frequency domain, is encoded in the inverse spectral density matrix (Dahlhaus, 2000). Thus, efficient estimation of the inverse spectrum has received significant attention in the statistics literature (Jung et al., 2015; Fiecas et al., 2019; Tugnait, 2022; Dallakyan et al., 2022; Deb et al., 2024) and the resulting estimates of partial coherence have practical importance in a number of application areas, including, but not limited to, neuroscience (Salvador et al., 2005), oceanography (Koutitonsky et al., 2002) and climatology (Song et al., 2020).

Let $X_t = (X_{t,1}, \ldots, X_{t,p})^\top, t \in \mathbb{Z}$ denote the p-dimensional weakly stationary zero-mean real-valued time series with autocovariance matrix function $\Gamma(\tau) = \text{cov}(X_t, X_{t-\tau}) = E(X_t X_{t-\tau}^\top)$. The spectral density matrix is defined as the Fourier transform of the autocovariance function

$$S(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \Gamma(\tau) \exp(-i\omega\tau), \quad \omega \in [-\pi, \pi]. \tag{2.6.5}$$

Remark 2.6.11. *Note that $S(\omega)$ is defined on the interval $\omega \in [-\pi, \pi]$ in the discrete time setting.*

Similar to the continuous time setting, a classic non-parametric estimator of (2.6.5) is based on the periodogram. That is, given n consecutive observations $(X_t)_{t=1}^n$, the vector-valued discrete Fourier transform at frequency ω is defined as

$$d_n(\omega) := \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \exp(-i\omega t),$$

where $d_n(\omega) = (d_{n,1}(\omega), \ldots, d_{n,p}(\omega))^\top$ and the $p \times p$ periodogram matrix is defined as

$$I_n(\omega) := d_n(\omega) d_n^H(\omega), \tag{2.6.6}$$

where H denotes the Hermitian transpose. However, the periodogram (2.6.6) is not a consistent estimator of the spectrum, and is only asymptotically (p fixed, $n \to \infty$) unbiased (Böhm and von Sachs, 2009). Therefore, it is customary to work with a smoothed version of the periodogram, i.e., smoothing over frequency, to guarantee consistency and asymptotic unbiasedness of spectral estimators. Smoothed periodogram estimators typically take the form

$$\tilde{S}(\omega) := \frac{1}{m_n} \sum_{j=-(m_n-1)/2}^{(m_n-1)/2} I_n(\omega + \omega_j) \tag{2.6.7}$$

where ω_j denotes the Fourier frequency $2\pi j/n$. In the classical framework, where the dimension p is fixed, the conditions $m_n/n \to 0$ and $m_n \to \infty$ as $n \to \infty$ guarantee consistency and asymptotic unbiasedness of $\tilde{S}(\omega)$ at a fixed frequency $\omega \in [-\pi, \pi]$.

Remark 2.6.12. *A similar result could also be relevant for point process data. More specifically, if we had spiking data from a single (large) trial, then averaging over frequencies could be an alternative to the multi-trial estimator* (2.5.6)*. However, extending our spectral estimators to the more general multi-tapering framework of Walden (2000) awaits future work.*

As mentioned, estimation of the inverse spectrum $\hat{\Theta}(\omega) = \hat{S}(\omega)^{-1}$ is of particular interest to enable graphical modelling of the multivariate time series. However, when the dimension of the time series p is large compared to the observed number of time points n, this can be a challenging task due to the large condition numbers exhibited by standard smoothed periodogram estimates (2.6.7). As such, the resulting estimates of the inverse spectrum are numerically unstable (Böhm and von Sachs, 2009).

Shrinkage Estimators

A potential solution lies in shrinkage estimation, whereby estimators of the spectrum are constructed as a convex combination of a 'classical estimator' (e.g., smoothed periodogram) and a so-called 'target matrix'. For example, Böhm and von Sachs (2009) proposed a shrinkage estimator which is a linear combination of the smoothed periodogram estimator $\tilde{S}(\omega)$ and the identity matrix,

$$\hat{S}(\omega) := a(\omega)\mathbb{I} + b(\omega)\tilde{S}(\omega),$$

where the factors $a(\omega)$ and $b(\omega)$ are chosen in a data-driven way. The authors show that the above shrinkage estimator has a strictly smaller mean-squared error than the averaged periodogram $\tilde{S}(\omega)$ in a double asymptotic framework, i.e., when $p, n \to \infty$.

While the above method improves the condition number of the resulting spectral estimator (leading to numerically stable estimates of $\hat{\Theta}(\omega)$) it is limited in the sense that the target matrix is too simplistic and may under represent the true dependence structure. An alternative solution is therefore to shrink the smoothed periodogram estimate toward a more general shrinkage target. In particular, estimators of the form

$$\hat{S}(\omega) = W_n(\omega)\tilde{V}(\omega) + (1 - W_n(\omega))\tilde{S}(\omega),$$

where $\tilde{S}(\omega)$ is the smoothed periodogram estimator, $\tilde{V}(\omega)$ is the shrinkage target and $W_n(\omega)$ is the shrinkage weight, were considered in Fiecas and Ombao (2011) and Fiecas and von Sachs (2014). The former, built on the work of Böhm and von Sachs (2009) to obtain a better fit to the spectral density matrix. In particular, the authors used the class of vector autoregressive (VAR) models as the shrinkage target $\tilde{V}(\omega)$ and the shrinkage weights were determined at each frequency using a data-driven approach. The resulting estimate had good frequency resolution, adequately capturing the frequencies driving the dynamics of the process. However, the authors did not analyse the performance of this estimator in high-dimensional settings, due to the large samples and multiple traces of data available for their application to EEG recordings.

The work of Fiecas and von Sachs (2014) essentially tried to bridge the gap between the methods of Böhm and von Sachs (2009) and Fiecas and Ombao (2011) to provide an estimator which was balanced in terms of both regularisation and fit, for use in high-dimensional settings. They proposed to use a diagonal matrix as the shrinkage target, whose $(q, q)^{th}$ element was given by the estimated parametric spectral density function of the autoregressive model fit to $X_{t,q}$. That is, each dimension of X_t was modelled independently, using the class of AR models, whose order was selected via the Bayesian Information Criterion (BIC).

Estimators based on the Whittle Likelihood

An important drawback of the above shrinkage estimators is that they do not produce sparse estimates of the inverse spectrum. As such, a popular alternative is to estimate the inverse spectrum directly by maximising a Whittle likelihood with appropriate (ℓ_1)

regularisation (Jung et al., 2015; Dallakyan et al., 2022; Tugnait, 2022; Deb et al., 2024). These estimators rely on the asymptotic complex normality of the discrete Fourier transforms (Brillinger, 2001) allowing standard methods in the GGM framework to be transported across to the spectral setting. Generally, an estimator of the inverse spectrum based on the Whittle likelihood is obtained by solving

$$\hat{\Theta}(\omega) := \text{argmin}_{\Theta(\omega)\in\mathcal{C}} \left\{ -\log\det\{\Theta(\omega)\} + \text{tr}\{\hat{S}(\omega)\Theta(\omega)\} + \lambda\|\Theta(\omega)\|_1 \right\}, \tag{2.6.8}$$

where $\lambda > 0$ is the penalty parameter, $\|\Theta(\omega)\|_1 = \sum_{q\neq r} |\Theta_{qr}(\omega)|$ and the optimisation is over the constraint set $\mathcal{C} := \{\Theta(\omega) \in \mathbb{C}^{p\times p} : \Theta(\omega) \succ 0\}$.

The estimator obtained from (2.6.8) is reminiscent of the estimator in Deb et al. (2024) for estimation of the inverse spectrum at a single frequency of interest ω. However, many works (Jung et al., 2015; Dallakyan et al., 2022; Tugnait, 2022) estimate the inverse spectrum at multiple frequencies, for example, via application of a group-Lasso penalty (Danaher et al., 2014) to capture shared zero patterns across frequencies. In this case, the goal is to solve

$$\hat{\Theta}(\cdot) := \text{argmin}_{\Theta(\cdot)\in\mathcal{C}} \left\{ \sum_{\omega\in\mathcal{W}} \Big[-\log\det\{\Theta(\omega)\} + \text{tr}\{\hat{S}(\omega)\Theta(\omega)\} \Big] + \lambda\|\Theta(\cdot)\|_1 \right\},$$

where $\Theta(\cdot) = \{\Theta(\omega)\}_{\omega\in\mathcal{W}}$ defined for frequencies ω in the set $\mathcal{W} = \{\omega_1, \omega_2, \omega_3, \dots\}$ and $\|\Theta(\cdot)\|_1 = \sum_{q\neq r} \sqrt{\sum_{\omega\in\mathcal{W}} |\Theta_{qr}(\omega)|^2}$ is the group-Lasso penalty.

Remark 2.6.13. *The theoretical properties of these estimators have also been well studied, see for example Sun et al. (2018); Fiecas et al. (2019); Deb et al. (2024).*

2.6.6 Point Process Graphical Models

Spectral Methods

The graphical modelling framework can also be extended to the point process setting, where an undirected graph $G = (V, E)$ can be used to visualise conditional relationships between dimensions of the multivariate point process $N(t) := \{N_q(t)\}_{q=1}^{p}$. Similar to the time series setting, the $(q, r)^{th}$ entry of the inverse spectral density matrix $\Theta(\omega) := S(\omega)^{-1}$ encodes a measure of association between two processes $N_q(t)$ and $N_r(t)$ after the linear effects of the multivariate process $N_{\backslash qr}(t)$ have been subtracted (Dahlhaus, 2000; Eichler et al., 2003). In the sequel, we will be interested in obtaining estimates of the squared *partial* coherence

$$\rho_{qr}^2(\omega) := \frac{|\Theta_{qr}(\omega)|^2}{\Theta_{qq}(\omega)\Theta_{rr}(\omega)} \in [0, 1], \tag{2.6.9}$$

to infer brain connectivity from neural spike train data, at specified frequencies of interest. The quantity in (2.6.9) provides a normalised measure of the partial correlation between pairs of processes in the frequency domain.

Related work regarding point process graphical models in the spectral domain has been discussed in Dahlhaus et al. (1997) and Eichler et al. (2003). However, these existing spectral estimation methods for multivariate point processes are not suitable for modern high-dimensional spike train data. A key contribution of this thesis (discussed in Chapter 3) therefore concerns the regularised estimation of the spectrum, and its inverse, in the context of multivariate and high-dimensional point processes.

Graphical Hawkes Process Models

A popular alternative to frequency domain methods, is to model spike train data using multivariate Hawkes processes (Chen et al., 2017a; Eichler et al., 2017; Wang et al., 2025). In this case, the latent connectivity structure of $N(t)$ is represented with a graph $G = (V, E)$ whose nodes represent components of the multivariate point process $N(t)$ and each directed edge, $(q \to r) \in E$, indicates that the probability of future events in $N_r(t)$ depends on the history of the events in $N_q(t)$.

For example, consider the multivariate Hawkes process with exponentially decaying intensity functions (2.5.2). In this case, the matrix of α_{qr} values forms the adjacency matrix of the resulting graphical model, where an entry $\alpha_{qr} > 0$ indicates a directed edge from node r to node q. More generally, if we define the intensity function of the q^{th} dimension of $N(t)$ as

$$\Lambda(t) = \nu_q(t) + \sum_{r=1}^{p} \int_0^t g_{qr}(u) dN_j(t-u),$$

where ν_q is the baseline intensity of the q^{th} component and the excitation functions g_{qr} satisfy $g_{qr}(u) = 0$ for $u \leq 0$ and $\int_0^\infty \|g_{qr}(u)\| du \leq 1$. Then N_r does not Granger cause N_q if an only if $g_{qr}(u) = 0$ for all $u \in \mathbb{R}$ (Eichler et al., 2017). This formulation has been used throughout the literature to infer Granger causal interactions, and estimators with sparsity inducing penalties have received significant attention (Zhou et al., 2013; Hansen et al., 2015; Xu et al., 2016).

Traditionally, the Hawkes process was known as a mutually-exciting process, where an event in one dimension of the process could trigger future events in any other dimension, including its own (Hawkes, 1971a,b). However, it is often desirable to allow for inhibitory effects, where events can suppress future events in different dimensions. This is particularly important for our neuroscience application, where a spike in one neuron may limit how fast another neuron, or itself, can subsequently fire.

To overcome this limitation, Hawkes process models allowing for both excitatory and inhibitory effects have been developed in both the univariate (Costa et al., 2018) and multivariate (Chen et al., 2017a) settings. Here, we limit our discussion to graphical models as described in Wang et al. (2025), where the intensity function of the Hawkes process for each unit $q \in \{1, \dots, p\}$ takes the form

$$\Lambda_q(t) = \max\left(0, \nu_q + \sum_{r=1}^{p} \sum_{j:t_j^r} \tilde{g}_{qr}(t - t_j^r)\right) \tag{2.6.10}$$

where ν_q is the background intensity of unit q, and $\tilde{g}_{qr}(\cdot) : \mathbb{R}^+ \to \mathbb{R}$ is the transfer function. Specifically, $\tilde{g}_{qr}(t - t_j^r)$ represents the influence of the j^{th} event in process $N_r(\cdot)$ on the intensity of unit q at time t. The use of the link function $\max(0, \cdot)$ in (2.6.10) ensures that the intensity $\Lambda_q(t)$ is always non-negative. Comparatively, the traditional definition of a Hawkes process (2.5.2) intensity function states that the excitation function $g_{qr}(\cdot)$ is always non-negative (Hawkes, 1971a,b).

Wang et al. (2025) consider a parametric transfer function of the form $\tilde{g}_{qr}(t) = \varphi_{qr}\kappa_r(t)$ with a transition kernel $\kappa_r(\cdot) : \mathbb{R}^+ \to \mathbb{R}$ which captures the decay of the dependence on the history of the process. This leads to $\sum_{j:t_j^r<t} \tilde{g}_{qr}(t - t_j^r) = \varphi_{qr}x_r(t)$ where

$$x_r(t) = \int_0^{t-} \kappa_r(t-u)dN_r(u),$$

summarises the entire history of unit r of the multivariate process.

In this model, the so-called connectivity coefficient φ_{qr} determines the underlying structure of the network, where φ_{qr} represents the influence of unit r's past events on unit q's intensity. Therefore, the connectivity coefficient provides an insight into the Granger causal interactions of the multivariate process (Shojaie and Fox, 2022). More specifically, a positive φ_{qr} implies that the past events of unit r excite future events of unit q, whereas a negative φ_{qr} represents an inhibitory influence of one units past events on the other. Thus, graphical models can be constructed using the connectivity coefficients α_{qr} for $q, r = 1, \dots p$.

2.6.7 Graphical Models for Binary Time Series

Chapter 4 of this thesis concerns the graphical modelling of multivariate binary time series, obtained under the binned spike count representation of the spike train data (Fig 2.3.4). In this section, we review previous work in the computational neuroscience literature for the identification and modelling of neural connectivity in this setting (Zhao et al., 2012; Song et al., 2013) alongside existing work in the statistics literature (Hall et al., 2016).

An ℓ_1 Regularised Logistic Model (Zhao et al., 2012)

Generalised linear models (GLMs) are a popular tool in neuroscience research for the analysis of neural firing rates. Popularised by Brillinger (1988) who studied small networks of three neurons, GLMs have since been used for a variety of problems in computational neuroscience including encoding (Paninski, 2004; Truccolo et al., 2005) and decoding (Gao et al., 2003) analyses [3].

However, GLMs also offer a flexible framework to analyse the interactions between hundreds of simultaneously recorded neurons. For example, Zhao et al. (2012) showed how they could be used to infer directional interactions between simultaneously recorded neural spike trains.

In their so-called L_1-regularised logistic regression model (L_1L), neural spike trains were represented as a binary time series under the binned spike count representation, with $\delta = 1$ms. They denote by $N^q_{1:t}$ the spike train of neuron q within the first t bins; δN^q_t denotes the number of spikes observed for neuron q within the t^{th} time bin and Λ^q_t denotes the conditional firing rate (i.e., number of spikes per second) at bin t, for $q = 1 \ldots, p$ neurons. With this notation, the authors define the following logistic

---

[3] encoding models aim to predict neural activity from stimuli whereas decoding models predict stimuli or behaviour from neural activity.

regression model for the conditional firing rate

$$\log\left(\frac{\Lambda_t^q\delta}{1-\Lambda_t^q\delta}\right) = \beta_q + \underbrace{\sum_{j=1}^{J}\beta_{qj}\delta N_{t-j}^q}_{I} + \underbrace{\sum_{q\neq r}\beta_{qr}\left(\sum_{k=1}^{K}\delta N_{t-k}^r\right)}_{II} + \underbrace{I(\alpha_q)}_{III}. \tag{2.6.11}$$

The first term β_q in (2.6.11) denotes the baseline firing rate. Term I models the spiking history of neuron q with the coefficient β_{qj} indicating the magnitude of the effect at lag j, up to a $J\delta$ms lag. Term II captures the interactions between other neural processes and term III denotes a function of extrinsic covariate effects, which is linear in parameters α (Truccolo et al., 2005). Despite including the last term (III) in definition of their model, the authors remove it completely for both their simulation study and real data section of the paper, arguing that their main focus is that of the detection/extraction of the interactions between neurons. In Chapter 4, we discuss the importance of including this term for accurate recovery of neural interactions, and model task-related modulations in firing rate directly, via smoothing spines.

The main parameter of interest (β_{qr}) represents the short-term interaction between neurons q and r, given the activity of all of the other neurons. Specifically, $\beta_{qr} > 0$ means that neuron q is excited by neuron r and $\beta_{qr} < 0$ is representative of an inhibitory interaction. Intuitively, $\beta_{qr} = 0$ denotes a lack of interaction.

The model in (2.6.11) is fitted to each of the $q = 1, \ldots, p$ spike trains separately, and parameter estimates are obtained via a regularised maximum likelihood procedure, where an ℓ_1 penalty is used to obtain sparse estimates of the interaction parameters.

The authors show that their proposed L_1L method can detect excitatory and inhibitory interactions on both simulated and real electrophysiological datasets. However, they only consider small networks of $p = 30$ neurons and their model assumes that neural firing rates remain stationary over time. In Chapter 4, we seek to extend the generalised linear model framework to account for non-stationary dynamics that are often observed in stimulus-response type neuroscience experiments.

The Generalised Functional Additive Model (Song et al., 2013)

Related work on the identification of sparse neural connectivity using spike train data was discussed in Song et al. (2013). There, the authors formulate the so-called generalised functional additive model (GFAM) which consists of a set of basis functions convolving the input signals, and a link function generating the firing probability of the output neuron from the summation of the convolutions weighed by the sought model coefficients.

More specifically, to identify interactions between the neural population, they model the causal relationship between the spike train recorded from one neuron (i.e., single output denoted $y(t)$) and the spike trains recorded from the whole neuron population denoted $X(t) = (x_1(t), \ldots, x_p(t))$ for a total of p neurons. That is, the vector X can contain the single output y to account for autoregressive effects. Similar to the approach in Zhao et al. (2012), both x and y are assumed to take binary values, under the binned-spike count representation of neural data.

A particular quantity of interest in their analysis is

$$\theta(t) = \mathrm{pr}\left(y(t) = 1 | X(t-\tau), 0 \leq \tau \leq M\right), \tag{2.6.12}$$

i.e., the probability of a spike being observed at time t given the input signals at time $t - \tau$, where M is the memory length of the system. Similar to Zhao et al. (2012), the interactions between neural processes are determined only on the basis of neural spiking activities and the derived model is behavioural task-independent. In Chapter 4, we discuss the importance of including additional model components to incorporate this task-dependence, particularly in the case of stimulus-response type neuroscience experiments.

While there are many approaches to solve the modelling problem in (2.6.12), Song et al. (2013) proposed the generalised function additive model (GFAM)

$$g(\theta(t)) = k_0 + \sum_{q=1}^{p} \sum_{\tau=0}^{M} k^{(q)}(\tau) x_q(t-\tau), \tag{2.6.13}$$

where k_0 is a scalar and $k^{(q)}$ are first-order kernel functions describing the relationship between the output spike probability and the q^{th} input. The term $k^{(q)}(\tau)$ in (2.6.13) models the influence of neuron q's past spikes on the current firing probability of a target (output) neuron at time t. Thus $k^{(q)}(\tau) > 0$ implies that a spike in neuron q at time $t - \tau$ increases the likelihood of the target neuron firing at time t; $k^{(q)}(\tau) < 0$ means that a spike in neuron q at time $t - \tau$ suppresses the target neuron's firing at that lag; and finally $k^{(q)}(\tau) = 0$ indicates that neuron q has no effect on the target neuron at that lag. Thus, the GFAM model is different to the L_1L method in the way that it captures interactions between pairs of neurons.

The authors use the probit link function for $g(\cdot)$ and approximate the kernel function as a weighted summation of a set of basis functions,

$$g(\theta(t)) \approx c_0 + \sum_{q=1}^{p} \sum_{j=1}^{J} c^{(q)}(j) v_j^{(q)}(t),$$

where $v_j^{(q)}(t) = \sum_{\tau=0}^{M} b_j(\tau) x_q(t-\tau)$. Two variations of the GFAM are considered: (1) using a global basis (e.g., Laguerre basis) and group Lasso estimation, and (2) using a local basis (e.g., B-spline basis) and group bridge estimation. Estimates of the model parameters are obtained using a regularised maximum likelihood approach.

Generally the GFAM model appears to perform well, with simulation and experimental results showing that both methods can capture complex dynamics between neural processes. However, the numerical experiments and real data analysis are conducted on small datasets with only $p = 16$ neurons, therefore it is unclear whether and how well this method might scale to larger datasets, more reflective of those obtained in practice. In Chapter 4, we seek to uncover interactions for a larger set of neurons, up to $p = 100$, and we also incorporate task-related dynamics in our model.

The Generalised Linear Autoregressive Model

In the statistics literature, Hall et al. (2018) studied a broad class of generalised linear autoregressive (GLAR) models for use in high-dimensional and multivariate settings. Specifically, the GLAR model is defined as

$$X_{t+1,q}|X_t \sim p(\nu_q + a_q^{*\top} X_t), \tag{2.6.14}$$

where $X_{t+1,q}$ is the q^{th} variate of X_{t+1}, where $1 \leq q \leq p$, $(X_t)_{t=0}^{\infty}$ are p-variate vectors and a^* is an unknown parameter vector, v is a constant offset parameter and p is an exponential family probability distribution. The formulation in (2.6.14) follows a component wise structure, whereby one can infer that time series x_r does not Granger cause series x_q if and only if $a_{qr} = 0$. For binary data, the GLAR model reduces to

$$X_{t+1,q}\Big|X_t \sim \text{Bernoulli}\left(\frac{1}{1 + \exp\left(-\nu - a_q^{*\top} X_t\right)}\right).$$

Hall et al. (2018) employ ℓ_1 regularisation of A (a $p \times p$ matrix constructed row-wise from a_q). They also derive statistical guarantees, including mean squared error bounds for the ℓ_1-regularised estimator.

Generally, their work focusses on the statistical and theoretical properties of these estimators, compared to the more practical and application driven papers discussed in the previous sections. However, neural spike trains are given as a motivating example for their methodology, alongside other problems in finance and epidemiology.

2.7 Summary

As shown in this brief literature review, a variety of statistical tools have been developed for the analysis of neural spike train data. However, emerging datasets present new statistical challenges, requiring novel contributions to the field of high-dimensional statistics, for the important task of neuronal network estimation. The remainder of this thesis is dedicated to the discussion of two novel statistical approaches, and the application of these methods to modern spike train datasets.

In particular, Chapter 3 introduces novel methodology to estimate the inverse spectrum of a multivariate point process. Building on existing work (Section 2.6.6) we show how regularisation-based methods can be used to infer brain connectivity at the individual neuron level, in high-dimensional settings. We establish asymptotic and large sample properties of our proposed estimators, and evaluate their performance on synthetic data simulated from multivariate Hawkes processes.

In Chapter 4, we build on the literature presented in Section 2.6.7 for the graphical modelling of binary time series. More specifically, we propose novel methodology to identify the effective connectivity of a population of neurons in the presence of non-stationary trends. Our work extends the current literature for identifying neural connectivity in two ways: 1) by accounting for non-stationary firing rates, and 2) by providing confidence intervals for entries in the interaction matrix. We apply our method to a neural spiking dataset, showcasing its ability to detect interactions of neural processes in response to various neuroscience experiments.

In Chapter 5, we extend the work presented in Chapter 3 and Chapter 4 to the multi-session and multi-subject setting, exploring the variability of functional brain networks across experimental subjects.

Finally, in Chapter 6, we conclude by summarising the main contributions of this thesis and discuss possible avenues for future work.

Chapter 3

Regularised Spectral Estimation for High-Dimensional Point Processes

3.1 Introduction

We consider the p-dimensional point process $N(t) := \{N_q(t)\}_{q \in \{1,\dots,p\}}$ whose q^{th} component gives the number of events of type q that have occurred in the time interval $(0, t]$, $t \leq T$. As in Bartlett (1963a), we denote by $dN_q(t)$ the number of events observed in process q in some small time interval dt, whereby $dN_q(t) = N_q(t + dt) - N_q(t)$. The first order properties of $N(t)$ are characterised via the intensity function $\Lambda(t) \in \mathbb{R}^p_{\geq 0}$ defined as $\Lambda(t) := \mathbb{E}\{dN(t)\}/dt$.

Hawkes (1971a) introduced an important class of point process models in which past events of $N(t)$ can affect the evolution of the process. In particular, the intensity function for each unit q takes the form

$$\Lambda_q(t) = \nu_q + \sum_{r=1}^{p} \int_{-\infty}^{t} \gamma_{qr}(t - u) dN_r(u),$$

where $\nu_q > 0$ and $\gamma_{qr}(\cdot)$ represent the background intensity and excitation functions respectively. The excitation function, $\gamma_{qr}(t - u)$ represents the effect on the q^{th} type at time t due to an event in the r^{th} type at a previous time point, u. Motivated by neuroscience applications, the multivariate Hawkes process will serve as an illustrative throughout this chapter.

The second order structure of $N(\cdot)$ at times t and u is given by the covariance density matrix,

$$\mu(t, u) = \frac{\mathbb{E}\{dN(u)dN^\top(t)\}}{dtdu} - \Lambda(u)\Lambda^\top(t),$$

where the superscript $(\cdot)^\top$ denotes transpose. When $\Lambda(t)$ is constant for all t, and $\mu(t, u)$ depends only on lag $\tau = u - t$, we refer to process $N(t)$ as being second-order stationary. In this case, we denote the intensity function as $\Lambda(t) = \Lambda$ and the covariance density matrix simply as $\mu(\tau)$.

In this work, we characterise the second order structure of $N(t)$ via the spectral density matrix, building on the frequency domain representation of the stochastic process (Bartlett, 1963a; Brillinger, 1972). The spectral density matrix $S(\omega)$ of a stationary point process is defined as the Fourier transform of its complete covariance density matrix, namely

$$S(\omega) = \frac{1}{2\pi}\left\{\text{diag}(\Lambda) + \int_{-\infty}^{\infty} \exp(-i\omega\tau)\mu(\tau)d\tau\right\},$$

where $S(\omega)$ is a $p \times p$ Hermitian positive definite matrix and $\text{diag}(\Lambda)$ is a $p \times p$ diagonal matrix with the elements of Λ on the diagonal. Analogous to the covariance matrix in the time domain, the spectral density matrix captures both the within and between type dynamics of the multivariate point process. More specifically, it describes the variance in each process or the covariance between processes that is attributable to oscillations in the data at a particular frequency $\omega \in \mathbb{R}$.

Given the complex nature of the spectrum, a common measure of association between two processes $N_q(t)$ and $N_r(t)$ is the squared spectral coherence,

$$R^2_{qr}(\omega) := \frac{|S_{qr}(\omega)|^2}{S_{qq}(\omega)S_{rr}(\omega)} \in [0,1].$$

If $R^2_{qr}(\omega) = 0$ for all ω, then the two processes are linearly independent. Conversely, a value of 1 indicates a perfect linear relationship. However, in the context of high-dimensional point processes, it is important to assess whether the association between $N_q(t)$ and $N_r(t)$ is due to a direct connection between them, or whether their correlation can be attributed to an indirect connection involving other processes, i.e. whether this correlation could potentially be explained away through conditioning on other streams. In this work, we address this problem by studying the inverse spectral density matrix $\Theta(\omega) := S(\omega)^{-1}$, whose $(q,r)^{th}$ entry encodes a measure of association between $N_q(t)$ and $N_r(t)$ after the linear effects of the multivariate process $N_{\backslash qr}(t)$ have been subtracted (Dahlhaus, 2000; Eichler et al., 2003). Specifically, the squared partial coherence between two processes $N_q(t)$ and $N_r(t)$ at a particular frequency ω

$$\rho^2_{qr}(\omega) := \frac{|\Theta_{qr}(\omega)|^2}{\Theta_{qq}(\omega)\Theta_{rr}(\omega)},$$

provides a normalised measure on $[0,1]$ of the partial correlation between pairs of processes in the frequency domain.

In the time series literature, partial coherence has been used in a variety of application areas, from medicine and oceanography to climatology (Kocsis et al., 1999; Koutitonsky et al., 2002; Song et al., 2020). In the point process setting, partial coherence has been used to identify functional neuronal connectivity from simultaneously recorded neural spike trains (see, for example, Dahlhaus et al., 1997; Eichler et al., 2003). However, these existing spectral estimation methods for multivariate point processes are not suitable for modern high-dimensional spike train data. In particular, the number of recorded neurons p is often large in comparison to the number of experimental trials, which we consider as a monotonically increasing sequence m_T. In this setting (see Figure 3.2.1c as an example) classical periodogram based estimators exhibit a very large condition number and thus the estimation of the inverse can be highly susceptible to sampling noise. Our main goal in this chapter is to better understand the behaviour of the periodogram in high-dimensional settings (for point processes), and then apply regularisation methods to strike a bias-variance trade-off enabling more efficient estimation of $\Theta(\omega)$, even when $p \geq m_T$.

The estimation of high-dimensional (inverse) spectral density matrices for time series data has received significant attention in the statistics literature. For example, Böhm and von Sachs (2009); Fiecas and Ombao (2011) and Fiecas and von Sachs (2014) explore shrinkage estimators for the spectral density matrix, leading to numerically stable and therefore invertible estimates. Comparatively, Jung et al. (2015); Tugnait (2022); Dallakyan et al. (2022) obtained estimates of the inverse spectral density matrix directly by optimising a penalised Whittle likelihood. There, a group Lasso penalty is used to capture shared zero patterns across neighbouring frequencies in the spectral domain. More recently, Deb et al. (2024) used a penalised Whittle likelihood to estimate the inverse spectral density matrix at a single frequency, and developed a fast path-wise coordinate descent algorithm for use in the time series setting. The Whittle approximation has also been studied in the Bayesian framework; see for example Tank et al. (2015).

In this work, we also estimate the inverse spectrum directly using a penalised Whittle likelihood. However, a key challenge in our point process framework lies in controlling the error for a preliminary multivariate estimate of the spectral density matrix (e.g. the periodogram), which is used as input to the Whittle likelihood. In particular, we require a bound on the distance $\hat{S}(\omega) - S(\omega)$ prior to establishing consistency of our proposed estimators of the inverse spectrum. In short, we show that the entries of $\hat{S}(\omega) - S(\omega)$ satisfy exponential-type tail bounds, in both asymptotic and large sample settings. We establish asymptotic properties of a mean-corrected version of the periodogram under classical mixing conditions (Brillinger, 1972) in a multi-trial design setting relevant to our neuroscience application. To enable a truly high-dimensional analysis of our proposed estimators, we then study the finite sample properties of the periodogram for the multivariate point process under a weakly-mixing framework, as developed in Doukhan and Louhichi (1999); Doukhan and Neumann (2007). In terms of regularisation, we consider both a ridge type penalty (shrinking all elements of the inverse spectra) and a Lasso type penalty. The latter imposes a sparsity assumption on the inverse spectral density matrix, easing interpretation of the spectrum, and enabling association of the point process with a form of graphical model. To the best of our knowledge, we believe this work is the first to consider regularisation of the Whittle likelihood in the context of multivariate point processes.

3.2 Preliminaries

3.2.1 The Tapered Fourier Transform for Point Processes

A classical non-parametric estimator of the spectral density matrix for a stationary point process is based on the periodogram, that is a (Hermitian) outer-product of the Fourier transform of the data (Bartlett, 1963a). This estimate is often improved upon by applying a taper to the data prior to taking the Fourier transform (Percival et al., 1993). Here, we propose to utilise a variant of the multitaper periodogram (3.2.2) to estimate the spectral density matrix. However, as the periodogram is based on the Fourier transform, it is first useful to understand some properties of this object in relation to the (continuous time) point process setting.

As in regular time series analysis, different choices of taper functions may be useful for different tasks, each having their own concentration properties in the Fourier domain. For the purposes of our discussion, we will focus on the canonical choice of $k = 1, \dots, m_T$

non-overlapping tapers that act to segment the process $N(t)$ into m_T intervals of equal length n_T, where m_T is a monotonically increasing sequence in T. This aligns with the experimental motivation for our methodology, where we can attribute a single taper to each of the m_T trials in our neuronal-spiking experiment. For simplicity, we choose $h_k(t/T) = (T/2\pi z_T)^{1/2}$ where $\text{supp}(h_k) \cap \text{supp}(h_l) = \varnothing, \forall k \neq l$ and $|\text{supp}(h_k)| = z_T$ for all $k = 1, \ldots, m_T$, where $z_T := n_T/T$. However, the framework is extensible to both other choices of taper function, and differing trial lengths.

Definition 3.2.1. *The tapered Fourier Transform of the process $N_q(t)$ for $t \in (0, T]$ is defined as*

$$d_{T;k;q}(\omega) := \int_0^T h_k(t/T)\exp(-i\omega t)dN_q(t).$$

An alternative view of the Fourier transform is given by the random set of events induced by the point process. That is, instead of the counting process representation, we can consider the events $E_{k,q} = \{t \mid dN_q(t) = 1\ ,\ t/T \in \text{supp}(h_k)\}$. In this case, we can write

$$d_{T;k;q}(\omega) = \sum_{t \in E_{k,q}} h_k(t/T)\exp(-i\omega t).$$

As the rate of the point process is non-zero and positive, it is important to remove this *mean* prior to obtaining estimates of the spectrum (Jarvis and Mitra, 2001; Rajala et al., 2023). Therefore, we will work with the mean-corrected variates (3.2.1) for the remainder of the paper.

Definition 3.2.2. *The mean-corrected tapered Fourier Transform is defined as*

$$\bar{d}_{T;k;q}(\omega) := d_{T;k;q}(\omega) - d_{T;k;q}(0)\frac{H_k(T\omega)}{H_k(0)}, \tag{3.2.1}$$

where $H_k(\omega)$ denotes the Fourier transform of $h_k(z)$, i.e. $H_k(\omega) = \int h_k(z)\exp(-i\omega z)dz$.

At this point, it is useful to introduce further conditions on $N(t)$, enabling further exploration and understanding of the asymptotic ($T \to \infty$) behaviour of the Fourier transform.

Assumption 3.2.3. *Assume the p-dimensional point process $N(t)$ is strictly stationary, i.e. the j^{th}-order cumulant*

$$C_{q_1,\ldots,q_j}(t_1,\ldots,t_j)dt_1\ldots dt_j = \mathrm{cum}\{dN_{q_1}(t_1),\ldots,dN_{q_j}(t_j)\},$$

is invariant to shifts in time. Let us define the reduced cumulant density

$$r_{q_1,\ldots,q_j}\left(u_1,\ldots,u_{j-1}\right) = C_{q_1,\ldots,q_j}\left(u_1,\ldots,u_{j-1},0\right),$$

then we further assume all moments exist, and that

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left|r_{q_1,\ldots,q_j}\left(u_1,\ldots,u_{j-1}\right)\right|\mathrm{d}u_1\cdots\ \mathrm{d}u_{j-1} < \infty\ ,$$

holds for any $q_1,\ldots,q_j = 1,\ldots,p$ and all $j \geq 2$. Finally, we assume the covariance density satisfies

$$\int_{-\infty}^{\infty}|u|\left|r_{q,r}(u)\right|\mathrm{d}u < \infty\ ,$$

for all $q,r = 1,\ldots,p$.

The above conditions amount to a form of mixing condition (Assumption 2.2 Brillinger, 1972) which ensures that the dependence of the point process decays at a sufficient rate to enable a central limit-theorem on the mean corrected Fourier transform.

3.2.2 The Multi-Trial Periodogram

We now use this classic result to motivate studying the multi-trial periodogram

$$\hat{S}_{m_T}(\omega) := \frac{1}{m_T}\sum_{k=1}^{m_T}\bar{d}_{T;k}(\omega)\bar{d}_{T;k}^{H}(\omega)\ . \tag{3.2.2}$$

Remark 3.2.4. *Here, we consider the number of experimental trials as a monotonically increasing sequence m_T, compared to a fixed m discussed in Chapter 2, i.e., (2.5.6).*

Under Assumption 3.2.3, the estimator in (3.2.2) is an asymptotically, p fixed, $T \to \infty$, consistent estimator for the spectrum. In the sequel we will be interested in the finite sample properties of (3.2.2). However, we first recall some properties of $\hat{S}_{m_T}(\omega)$ as $T \to \infty$. In the context of our neuronal-spiking experiment, this corresponds to a setting in which we have a fixed number of infinitely long trials, i.e., $m_T = m$ and $n_T \to \infty$ as $T \to \infty$.

Using Assumption 3.2.3 and Theorem 4.2 of Brillinger (1972), one can show that $\{\bar{d}_{T;k}(\omega)\}_{k=1}^{m}$ are asymptotically independent complex Gaussian, and thus the multi-trial periodogram is distributed as a Complex Wishart distribution, i.e., $\hat{S}_m(\omega) \sim m^{-1}\mathcal{W}_p^C(m, S(\omega))$ with m degrees of freedom and centrality matrix $S(\omega)$. As a further corollary of the asymptotic complex normality of $\{\bar{d}_k(\omega)\}_{k=1}^{m}$, we have (Goodman, 1963) that if $p < m$, then the estimated squared spectral coherence $\hat{R}^2_{qr}(\omega) = |\hat{S}_{qr}(\omega)|^2/\{\hat{S}_{qq}(\omega)\hat{S}_{rr}(\omega)\}$ has the density function

$$f_{\hat{R}^2_{qr}}(x) = (m-1)\big(1-R^2_{qr}\big)\big(1-x^2\big)^{(m-2)}\ {}_2F_1\big(m,m;1;R^2_{qr}x\big), \tag{3.2.3}$$

where ${}_2F_1(\alpha_1,\alpha_2;\beta_1;z)$ denote the hypergeometric function with 2 and 1 parameters α_1, α_2 and β_1 and scalar argument z.

As discussed in Chapter 2, the above properties are well known; however, it is nowadays of interest to examine the statistical properties of our multi-trial periodogram in high-dimensional, large $p \geq m$ settings. In these situations, the Goodman distribution (3.2.3) will not apply, however, the errors of the periodogram may still be bounded. As an alternative to the classic asymptotic framework (Brillinger, 1972), we propose to study finite-sample deviation bounds, i.e. for finite m, that can control the behaviour (error) of the multi-trial periodogram. We begin by discussing the asymptotic setting, where $T \to \infty$, before stating a deviation bound which holds for $T < \infty$ and where we allow m_T to be a monotonically increasing sequence in T. We later harness this result as a lemma when deriving bounds on our regularised estimators estimates of the inverse spectrum.

Let us define the Lipschitz modulus of a function g as $\operatorname{Lip} g = \sup_{x \neq y} |g(x) - g(y)|/\|x-y\|_1$, where $\|z\|_1 = \sum_i |z_i|$. Additionally, for a function $f : \mathbb{C}^u \to \mathbb{R}$ let us define $\|f\|_\infty := \sup_{x \in \mathbb{C}^u} |f(x)|$. Finally, we denote by $\mathbb{N}^* = \{1, 2, 3, \dots\}$ the set of natural numbers excluding zero. Then, the following weak dependence condition (Doukhan and Louhichi, 1999) is sufficient for the finite sample analysis which follows.

Assumption 3.2.5. *The variates* $\{\bar{d}_{T;k}(\omega)\}_{k=1}^{m_T}$ *are so-called* η*-weakly dependent. That is, there exists a sequence* $\{\eta(\tau)\}_{\tau \in \mathbb{N}}$ *converging to 0, such that for any* u*-tuple* $(s_1, \dots, s_u) \in \mathbb{Z}^u$ *and any* v*-tuple* $(t_1, \dots, t_v) \in \mathbb{Z}^v$ *with* $s_1 \le \cdots \le s_u \le s_u + \tau = t_1 \cdots \le t_v$ *and arbitrary measurable functions* $f : \mathbb{C}^{up} \to \mathbb{R}, g : \mathbb{C}^{vp} \to \mathbb{R}$ *with* $\|f\|_\infty \le 1$, $\|g\|_\infty \le 1$, $\operatorname{Lip} f < \infty$ *and* $\operatorname{Lip} g < \infty$, *such that*

$$\begin{aligned} &\left|\operatorname{cov}\left[f\left\{\bar{d}_{T;s_1}(\omega), \dots, \bar{d}_{T;s_u}(\omega)\right\}, g\left\{\bar{d}_{T;t_1}(\omega), \dots, \bar{d}_{T;t_v}(\omega)\right\}\right]\right| \\ &\qquad\qquad \le \left\{u(\text{Lip } \ f)\|g\|_\infty + v(\text{Lip } \ g)\|f\|_\infty\right\}\eta(\tau), \end{aligned}$$

where $(u, v) \in \mathbb{N}^* \times \mathbb{N}^*$.

Remark 3.2.6. *Assumption 3.2.5 controls dependence between the trial specific Fourier transforms via the covariance of transformation functions* f, g, *for a specific number of trials. Intuitively, this assumption is more restrictive than Assumption 3.2.3 which bounds the cumulant function over unbounded intervals. Related work on the notion of weak-dependence for point processes as an alternative to more traditional mixing properties can be found in Doukhan and Lang (2016).*

In the sequel, we introduce a useful result for the periodogram (3.2.2) which provides a deviation bound, firstly in the asymptotic setting, and then in the setting where m_T

and T are both finite. Prior to stating this bound, we highlight that the periodogram is asymptotically unbiased. However, for finite samples there can be a trade-off between bias and variance, based on the choice of taper function. The bias of the periodogram is given by the following result.

Lemma 3.2.7. *Define* $Z_T(\omega,\omega') := H_k(T\omega - T\omega') - H_k(T\omega)H_k^*(T\omega')/H_k(0)$ *and let* $\tilde{S}_{qr}(\omega) := \mathbb{E}\{\hat{S}_{m_T;qr}(\omega)\}$ *, then*

$$\tilde{S}_{qr}(\omega) = \int |Z_T(\omega,\omega')|^2 S(\omega')d\omega'.$$

Let us track the bias using $\delta_{b;T}(\omega) := \max_{qr} |\tilde{S}_{qr}(\omega) - S_{qr}(\omega)|$*. If* $S_{qr}(\omega)$ *is smooth and has uniformly bounded second derivatives* $S''_{qr}(\omega) \leq M_2$*, then choosing* ω_* *such that* $H_k(T\omega_*) = 0$*, and assuming* $f_{W_T(\omega)}(x) = |H_k(T(\omega - x))|^2$ *forms a valid probability density, we have*

$$\delta_{b;T}(\omega_*) \leq S'(\omega_*)\big(\mathbb{E}[W_T(\omega_*)] - \omega_*\big) + \frac{M_2}{2}\mathrm{Var}[W_T(\omega_*)]\ ,$$

where $W_T(\omega)$ *is a random variable with density* $f_{W_T(\omega)}(x)$*.*

Remark 3.2.8. *In the discrete time setting, the bias properties of the periodogram depend on the choice of taper function* $h(\cdot)$ *(see for example Walden (2000) for detailed discussion). In our case, we find* $\mathbb{E}[W_T(\omega)] = \omega$*, and* $\mathrm{Var}[W_T(\omega)] = \mathcal{O}(T^{-2})$ *(see supplementary material for details). It is worth noting that the minimum allowable spacing between frequencies* ω_* *decreases as* $1/T$*. Consequently, while it is not possible to estimate the spectrum at all frequencies without bias, the bias diminishes as the observation interval* T *increases. Therefore, asymptotically, the estimator becomes unbiased. The sidelobe behaviour around* ω_* *will depend on the form of the chosen taper. A generalised understanding of the multi-taper periodogram in the point process setting awaits future work. However, in our empirical experiments the bias appears to have minimal impact, and the main issue is the high variance of our periodogram estimates.*

Proposition 3.2.9. *Deviation Bound for the Multi-Trial Periodogram.*

(a) Let $N(t)$ satisfy Assumption 3.2.3 and let $\hat{S}_m(\omega)$ be the multi-trial periodogram estimator defined in (3.2.2) with $m_T = m$ and $n_T = T/m$, then

$$\mathrm{pr}\left\{|\hat{S}_{m;qr}(\omega) - S_{qr}(\omega)| \geq \delta\right\} \leq 8\exp\left\{\frac{-m\delta^2}{c_1 S_{qq}(\omega) S_{rr}(\omega)}\right\},$$

for some $c_1 > 0$ and all $\delta \in [0, 12\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}]$ as $T \to \infty$.

(b) Let $\hat{S}_{m_T}(\omega)$ be the multi-trial periodogram estimator defined in (3.2.2) satisfying Assumption 3.2.5. Furthermore, assume that the bias is bounded uniformly by $\delta_b(\omega) \leq (1-\varphi)\delta$ for some $\varphi \in (0,1), \delta > 0$, then

$$\mathrm{pr}\left(\left|\hat{S}_{m_T;qr}(\omega) - S_{qr}(\omega)\right| \geq \delta\right) \leq 8\exp\left\{-\frac{m_T(\varphi\delta)^2}{c_2 \tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)}\right\},$$

for some $c_2 > 0$ and all $\delta \in [0, \{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}/\varphi]$.

Proof. The proof, which is based on results in Bickel and Levina (2008) and Doukhan and Neumann (2007), is given in Section 3.7.2. □

To demonstrate how these properties hold in practice, and the challenges when $p \geq m_T$, we give an example for the empirical distribution of, and error incurred, by $\hat{S}_{m_T}(\omega)$. To illustrate the behaviour, we consider a simple homogeneous Poisson point process with rate $\lambda = 1$ and let $T = 1000$, and obtain 1000 Monte-Carlo replicates of the multi-trial spectrum with $m_T = 10$ trials. The frequency $\omega = 0.0628$ is considered without loss of generality since the process is Poisson[1]. The results, depicted in Figure 3.2.1, confirm the appropriateness of the asymptotic distribution (3.2.3), and how the deviations grow as a function of $\log(p)$. Empirical 95% confidence intervals, obtained from the 0.025 and 0.975 quantiles of the 1000 replicate estimates, are shown in red.

---

[1] As the spectrum of a Poisson process is constant across all frequencies, we could in theory run our experiments for any $\omega \in \mathbb{R}$ and our conclusions would be the same.

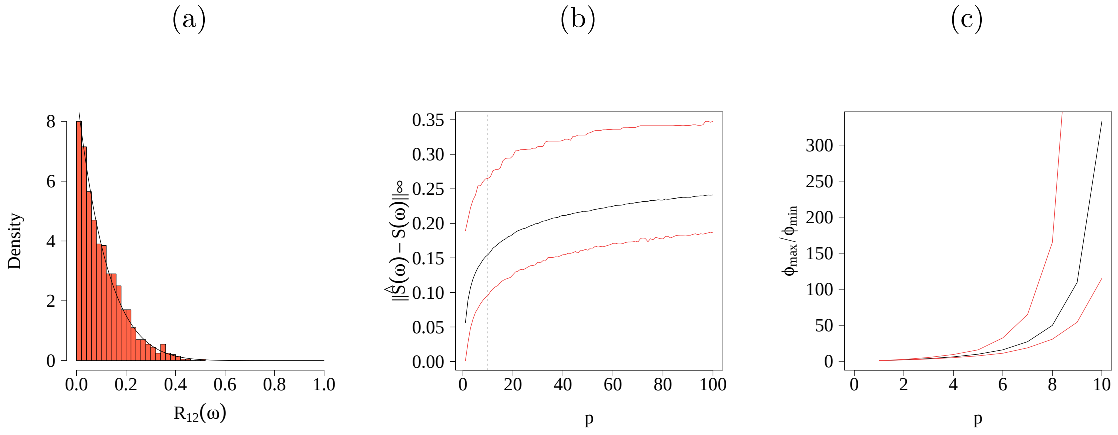


Figure 3.2.1: (a) Empirical (histogram) vs theoretical (solid line) distribution of coherence for $p = 7$. (b) Plot of the l_∞ norm of spectral matrix estimation error, the dashed line indicates when $p = m$. (c) Plot of the condition number of the multivariate spectrum as a function of dimensionality p.

To highlight the issue of instability of the periodogram as one approaches the high-dimensional setting, we plot the condition number of the multivariate spectrum, defined as the ratio $\phi_{\max}/\phi_{\min}$ of its largest to its smallest eigenvalue, in Fig 3.2.1c. We observe that the variability of the eigenvalues of $\hat{S}_{m_T}(\omega)$ is not well-controlled when $p \approx m_T$. Therefore, estimates of the inverse spectrum obtained by simply inverting $\hat{S}_{m_T}(\omega)$ would be highly inaccurate (Böhm and von Sachs, 2009). It is these scenarios, i.e., when $p \approx m_T$, and $p > m_T$ where the methods proposed in this chapter will be of most utility.

3.3 Regularised Spectral Estimation

3.3.1 Methodology

We propose a novel methodology to estimate the inverse spectral density matrix for a high-dimensional multivariate point process at a specified frequency of interest. While estimates of a similar nature have undergone significant development for time series data (Jung et al., 2015; Nadkarni et al., 2016; Tugnait, 2022; Dallakyan et al., 2022; Deb et al., 2024) to the best of our knowledge, this is the first examination of related estimators in the point process framework.

Remark 3.3.1. *Going forward, we will focus on regularised spectral estimators for high-dimensional settings. We will therefore work under Assumption 3.2.5 and let* $m_T > 1$ *be a monotonically increasing sequence in* T*. The results presented here hold for all frequencies at which the bias can be bounded* $\delta_b(\omega) \leq (1-\varphi)\delta$ *and consider the doubly asymptotic setting where both* p, m_T *grow, and where it is possible that* $p > m_T$*. The results can be applied in the setting* $\delta \asymp \{\log(p)/m_T\}^{1/2}$ *and when* T *is finite.*

Our estimator is motivated by the asymptotic complex normality of our multi-trial Fourier coefficients (Theorem 4.2 from Brillinger (1972)), where we appeal to the following pseudo (negative) log-likelihood (Whittle, 1953)

$$\mathcal{L}(\Theta(\omega)) := -\log\det(\Theta(\omega)) + \operatorname{tr}\left\{\hat{S}_{m_T}(\omega)\Theta(\omega)\right\},$$

where $\det(\cdot)$ and $\operatorname{tr}\{\cdot\}$ denote the determinant and trace of a matrix respectively.

If $m_T > p$, the negative log-likelihood achieves its minimum for $\hat{\Theta}(\omega) = \hat{S}_{m_T}(\omega)^{-1}$. However, as discussed in Section 3.2.2, in the high-dimensional setting where $p > m_T$, the inverse of the multi-trial periodogram is not uniquely defined. To overcome this, it is common to add a penalty to $\mathcal{L}(\omega)$ in order to constrain the maximum likelihood estimate $\hat{\Theta}(\omega)$ to a feasible region.

Definition 3.3.2. *Let* $\lambda(\omega) > 0$*, the regularised spectral estimator for* $\Theta(\omega)$ *is defined as the M-estimator*

$$\hat{\Theta}(\omega) := \arg\min_{\Theta(\omega)\in\mathcal{C}} \left[\mathcal{L}\{\Theta(\omega)\} + \lambda(\omega)P\left\{\Theta(\omega)\right\}\right], \tag{3.3.1}$$

with constraint set $\mathcal{C} := \{\Theta(\omega) \in \mathbb{C}^{p\times p} : \Theta(\omega) \succ 0\}$*. We consider two distinct penalties* *as below, each possessing their unique advantages*

$$P_1\{\Theta(\omega)\} = \|\Theta(\omega)\|_1 := \sum_{qr} \left[\mathrm{Re}^2\{\Theta_{qr}(\omega)\} + \mathrm{Im}^2\{\Theta_{qr}(\omega)\}\right]^{1/2} \tag{3.3.2}$$

$$P_2\{\Theta(\omega)\} = \mathrm{tr}\{\Theta(\omega)\}. \tag{3.3.3}$$

Throughout this chapter, we refer to the regularised spectral estimator with penalties (3.3.2) *and* (3.3.3) *as the Lasso and ridge estimators, respectively.*

The penalty in (3.3.2) encourages sparsity in the inverse spectrum by jointly penalising both the real and imaginary parts of the entries in $\Theta(\omega)$. This can be seen an example of the group-Lasso penalty (Huang and Zhang, 2010). In cases where the inverse spectrum is known to be sparse, it is beneficial to use the Lasso-type penalty. By contrast, the penalty in (3.3.3) proportionally shrinks elements of $\Theta(\omega)$ (without inducing sparsity) and can therefore be viewed as a ridge type estimator. Similar estimators for the estimation of high-dimensional covariance matrices have been discussed in Friedman (1989) and Warton (2008).

An advantage of the ridge estimator is that solutions to (3.3.1) can be found analytically. Specifically, $\hat{\Theta}(\omega) = \{\hat{S}_{m_T}(\omega) + \lambda(\omega)\mathbb{I}\}^{-1}$, where $\hat{S}_{m_T}(\omega)$ is an estimate of the spectrum and $\mathbb{I}$ is the identity matrix. When the Lasso penalty in (3.3.2) is used, the convex optimisation problem (3.3.1) can be solved numerically using the alternating direction method of multipliers algorithm, or via the complex graphical Lasso (Deb et al., 2024). In this chapter, we adopt the former approach due to its simplicity. This approach has been used throughout the time series literature to solve optimisation problems similar to those considered (Jung et al., 2015; Nadkarni et al., 2016; Dallakyan et al., 2022). A full description of the algorithm is given in the Supplementary Material.

3.3.2 Error Bounds for the Regularised-Spectral Estimator

In this section, we establish asymptotic properties of the Lasso and ridge estimators, providing rates in Frobenius norm. We begin by specifying the following assumptions about the true $\Theta(\omega)$ and $S(\omega)$.

A1. Let the set $E\{\Theta(\omega)\} := \{(q,r) | \Theta_{qr}(\omega) \neq 0, q \neq r\}$. Then $\text{card}[E\{\Theta(\omega)\}] \leq s$.

A2. $\phi_{\min}(S(\omega)) \geq \underline{k} > 0$.

The second assumption provides a lower bound on the eigenvalues of the spectral density matrix, and thus ensures positive definiteness of $S(\omega)$. This guarantees the existence of $\Theta(\omega)$.

Proposition 3.3.3. *Let $\hat{\Theta}_R(\omega)$ be the regularised spectral estimator defined in* (3.3.1) *with penalty* (3.3.3)*, and regularisation parameter $\lambda = c(p\tau \log p / m_T)^{1/2}$ for any $\tau > 2$ and some frequency-specific constant $c > 0$. Then, under A2, there exists a constant $c_2 > 0$ such that using sufficient trials,*

$$m_T > \tau \log p \max\left[c_2, 64c^2\underline{k}^{-2}p^2\right],$$

we have, with probability at least $1 - 8/p^{\tau-2}$,

$$\left\| \hat{\Theta}_R(\omega) - \Theta(\omega) \right\|_F \leq \frac{8c\tau^{1/2}}{\underline{k}^2} \left(\frac{p^2 \log p}{m_T} \right)^{1/2}.$$

Proof. The proof is given in Section 3.7.5. □

The above result is based on lower-bounding the eigenvalues of the estimated spectrum, alongside bounding the deviations of the periodogram based on Proposition 3.2.9. This leads to the stated bound, holding for $m_T = \Omega(p^2 \log p)$ which ensures that $\|\hat{\Theta}_R(\omega)\|_2$ is bounded. However, as we will later demonstrate in our simulation study, the ridge estimator can accurately recover the inverse spectrum, even in settings where $m_T < p^2$. Thus, while the condition on m_T is required for the theoretical analysis, we do not recommend enforcing it in practice.

While the bound for the ridge estimator is not particularly tight, it is feasible that tighter bounds can be obtained if the true spectrum maintained a specific structure, e.g., if it were approximately low rank then it may be possible to extend the results of Bunea and Xiao (2015) to the spectral setting. As an alternative, we now present a result on the Lasso spectral estimator, which can make use of the structure within the edge set $E\{\Theta(\omega)\}$ by selecting certain entries of the inverse spectral density matrix to be exactly zero.

Proposition 3.3.4. *Let $\hat{\Theta}_G(\omega)$ be the regularised spectral estimator defined in (3.3.1) with penalty (3.3.2) and regularisation parameter $\lambda = c(\tau \log p/m_T)^{1/2}$ for any $\tau > 2$ and some frequency-specific constant $c > 0$. Then, under A1-2, there exists a constant $c_2 > 0$ such that using sufficient trials,*

$$m_T > \tau \log p \max\left[c_2, 64c^2\underline{k}^{-2}(p+s)\right],$$

we have, with probability at least $1 - 8/p^{\tau-2}$,

$$\left\|\hat{\Theta}_G(\omega) - \Theta(\omega)\right\|_F \leq \frac{8c\tau^{1/2}}{\underline{k}^2}\left\{\frac{(p+s)\log p}{m_T}\right\}^{1/2}. \tag{3.3.4}$$

Proof. The proof is given in Section 3.7.6. □

The stated bound for the Lasso estimator (3.3.4) is similar to that in Rothman et al. (2008) for sparse estimates of the inverse covariance matrix (c.f. Chapter 2). The result for the Lasso estimator sharpens the rate by order $p^{1/2}$ compared with the ridge. The order $(p+s)^{1/2}$ term in the bound originates as we need to estimate both the diagonal and off-diagonal structure in the inverse spectral density. In comparison to Proposition 3.3.3, this bound will hold as long as $m = \Omega((p+s)\log p)$, whilst this allows a relatively high-dimensional setting we still need sufficient samples to ensure $\|\hat{\Theta}_G(\omega)\|_2$ is bounded. In the following section we further sharpen these bounds by introducing a more restrictive incoherence condition on the spectrum.

3.3.3 Consistent Estimation of Partial Coherence Graphs

The Lasso regularised spectral estimator in (3.3.1) can be used to construct a frequency specific partial coherence graph $\mathcal{G}(\omega) = (V, E)$ which can be used to visualise conditional relationships between dimensions of the multivariate point process (Dahlhaus, 2000). In this graph, the nodes $V = \{1, \ldots, p\}$ represent the $q = 1, \ldots, p$ processes, $N_q(t)$, and the edge set is characterised by the squared partial coherence. If the squared partial spectral coherence $\rho^2_{qr}(\omega) = 0$ for all frequencies ω, then the edge between nodes q and r would be excluded from the corresponding conditional independence graph (Eichler et al., 2003).

Whilst we have shown consistency of the Lasso estimator in (3.3.4), it remains to demonstrate that, with high probability, this estimate can correctly identify the zero pattern of the matrix $\Theta(\omega)$. In this section, we state a result on the model selection consistency of the Lasso estimator, which builds on the work presented in Ravikumar et al. (2011) and Deb et al. (2024). Before stating the result, we first set out some quantities required in our analysis.

Recall from Assumption A1 that the set of off-diagonal entries in $\Theta(\omega)$ used to define the edge set of the graph is denoted by $E\{\Theta(\omega)\}$, we define the maximum degree of the graph associated with these edges as $d = \max \deg\{E\{\Theta(\omega)\}$, or equivalently the maximum number of non-zeros on any row of $\Theta(\omega)$. We denote the supported elements of the inverse spectral density by the set $\mathcal{M}(\omega) := E(\Theta(\omega)) \cup \{(1,1), \ldots, (p,p)\}$. Furthermore, we use $\mathcal{M}^{\perp}(\omega)$ to denote the complement of $\mathcal{M}(\omega)$, corresponding to all pairs (q, r) for which $\Theta_{qr} = 0$. For convenience, we adopt the shorthand $\mathcal{M}$ and $\mathcal{M}^{\perp}$ respectively, but these sets may still remain frequency specific.

Let the Hessian of the Whittle likelihood at frequency ω be given by $\Gamma(\omega) = S(\omega) \otimes S(\omega) \in \mathbb{C}^{p^2 \times p^2}$. We define the term

$$\kappa_S := \|\!|S(\omega)\|\!|_\infty = \max_{q=1,\ldots,p} \sum_{r=1}^{p} |S_{qr}(\omega)| \,, \tag{3.3.5}$$

corresponding to the ℓ_∞ norm of the true spectral density matrix. Similarly, we consider the subset of the Hessian relevant to the true model subset $\Gamma_{\mathcal{M}\mathcal{M}}(\omega)$ and let

$$\kappa_\Gamma = \|\!| \left\{\Gamma_{\mathcal{M}\mathcal{M}}(\omega)\right\}^{-1} \|\!|_\infty. \tag{3.3.6}$$

The two quantities (3.3.5) and (3.3.6) play an important role in describing the behaviour of the Lasso estimator, and are used in the final result.

We now introduce a frequency specific incoherence condition, that holds across all frequencies, which is required for the model selection consistency of the Lasso estimator.

A3. Let $\Gamma := \Gamma(\omega)$. Then for frequency $\omega \in \mathbb{R}$ there exists some $\alpha \in (0, 1]$ such that

$$\max_{e \in \mathcal{M}^\perp} \left\| \Gamma_{e\mathcal{M}} (\Gamma_{\mathcal{M}\mathcal{M}})^{-1} \right\| \leq (1 - \alpha). \tag{3.3.7}$$

The intuition behind this assumption is that it limits the influence of the non-edge terms, indexed by $\mathcal{M}^\perp$, on the edge terms, indexed by $\mathcal{M}$; see Saegusa and Shojaie (2016) for further details.

The next result establishes an elementwise deviation bound for the Lasso estimator which can be used to establish consistent estimation of the resulting partial coherence graphs.

Proposition 3.3.5. *Consistent Estimation of Partial Coherence Graphs. Consider a distribution satisfying the incoherence assumption* (3.3.7) *with parameter* $\alpha \in (0, 1]$. *Let* $\hat{\Theta}_G(\omega)$ *be the regularised spectral estimator defined by* (3.3.1) *with penalty* (3.3.2) *and regularisation parameter* $\lambda = (8c/\alpha)(\log p^\tau / m_T)^{1/2}$ *for any* $\tau > 2$ *and frequency-specific* *constant $c > 0$. Then, there exists a constant $c_2 > 0$ such that under assumptions A1-A3 with $p > 3$ and sufficient trials,*

$$m_T > \tau \log p \max \left[c_2, \left\{6c(1+8/\alpha)^2 d\kappa_S^3 \kappa_\Gamma^2\right\}^2, \left\{6c(1+8/\alpha) d\kappa_S^2 \kappa_\Gamma \max(1, \kappa_S \kappa_\Gamma)\right\}^2\right],$$

with probability greater than $1 - 8/p^{\tau-2}$, we have:

(a) The estimate $\hat{\Theta}_G(\omega)$ satisfies the element-wise ℓ_∞-bound:

$$\left\|\hat{\Theta}_G(\omega) - \Theta(\omega)\right\|_\infty \leq 2c\kappa_\Gamma \left(1 + \frac{8}{\alpha}\right) \left(\frac{\tau \log p}{m_T}\right)^{1/2}.$$

(b) The estimate $\hat{\Theta}_G(\omega)$ specifies an edge set $E(\hat{\Theta}_G(\omega))$ that is a subset of the true edge set $E(\Theta(\omega))$, and includes all edges (q, r) with

$$|\Theta_{qr}(\omega)| > 2c\kappa_\Gamma \left(1 + 8\alpha^{-1}\right) (\tau \log p / m_T)^{1/2}.$$

Proof. The proof, similar to Deb et al. (2024), adopts the primal dual witness approach (Ravikumar et al., 2011), and is given in Appendix B.1.

□

Part (b) of Proposition 3.3.5 states that the estimated edge set is a subset of the true edge set. More specifically, it asserts that $E(\hat{\Theta}_G(\omega))$ correctly excludes all non-edges and only includes edges that are 'large' relative to decay of the error $c(\tau \log p/m)^{1/2}$.

Sign consistency can also be achieved by placing a lower bound on the entries in the inverse spectral density matrix. In this case, not only will the estimator have the same edge set as $\Theta(\omega)$, but it will also be able to recover the correct signs on these edges. Mathematically, this translates to the event

$$\mathcal{K}(\hat{\Theta}(\omega); \Theta(\omega)) := \left\{\text{sign}(\hat{\Theta}_{qr}(\omega)) = \text{sign}(\Theta_{qr}(\omega)) \ \forall (q, r) \in \mathcal{M}(\omega)\right\}.$$

Using this notation, and with a sufficiently modified bound on the number of trials, the estimator $\hat{\Theta}_G(\omega)$ is model selection consistent with high probability as $p \to \infty$,

$$\mathrm{pr}(\mathcal{K}(\hat{\Theta}(\omega); \Theta(\omega))) \geq 1 - 8/p^{\tau-2} \to 1.$$

The sign consistency can also be achieved without requiring Assumption A3 and instead using a thresholded or adaptive Lasso penalty (Bühlmann and Van De Geer, 2011).

3.4 Synthetic Experiments

3.4.1 Simulation Setting

We use synthetic data to compare the performance of the Lasso and ridge estimators to existing statistical approaches where applicable. The goal of the study is two-fold: (1) to assess the accuracy of the recovered inverse spectrum and (2) to investigate the conditional relationships of high-dimensional multivariate point processes. Throughout, we use simulated data from different parameterisations of the multivariate Hawkes process. This particular point process has been used extensively to model neuronal interactions in the brain, and as such will serve as an illustrative example in our discussion (see for example Pernice et al., 2011; Chen et al., 2017b; Lambert et al., 2018; Wang and Shojaie, 2021).

In the simulation study, we consider three different parameterisations of stationary multivariate Hawkes processes. Motivated by applications in neuroscience, we consider processes that are governed by exponential excitation functions. More specifically, the conditional intensity function of the process can be written as $\Lambda = (\mathbb{I} - G(0))^{-1}\nu$, where ν is a vector of background intensities, $\mathbb{I}$ is the identity matrix and G is the Fourier transform of the excitation function defined elementwise in the exponential case as

$$G_{qr}(\omega) = \int_{-\infty}^{\infty} \exp(-i\omega\tau)\alpha_{qr}\exp(-\beta_{qr}\tau)d\tau = \frac{\alpha_{qr}}{\beta_{qr} + i\omega}.$$

The spectral density matrix of a stationary multivariate Hawkes process is subsequently given as

$$S(\omega) = \frac{1}{2\pi}\left[\{\mathbb{I} - G(\omega)\}^{-1} D \{\mathbb{I} - G^{\top}(-\omega)\}^{-1}\right]$$

where $D = \text{diag}(\Lambda)$. Further details regarding the multivariate Hawkes process used in this section can be found in the Supplementary Material.

We consider multivariate Hawkes processes of dimension $p = 12, 48$ and 96 and simulate $m = 10$ and $m = 50$ independent trials of synthetic data using Ogata's method (Ogata, 1981), via the `hawkes`[2] package in R. $T = 200$ seconds were generated for each process.

Multiple scenarios are considered in this study. In the first instance, we simulate from a p-dimensional Hawkes process whose dependence structure is captured by a block diagonal excitation matrix α. Each block consists of a 3×3 matrix α^* designed to ensure both self- and cross-excitation between the dimensions in a particular block. Thus, dependence between dimensions of the process is null between the blocks and non-null within each block. We consider two models of this nature: model (a) in which there are interactions between two of the processes within a block, and an additional independent process and model (b) where all three processes in a particular block interact with each other. The excitation matrix for model (a) is low-rank compared to model (b) which is full rank. In the final setting (c) we simulate from a p-dimensional Hawkes process whose dependence structure is driven by an arbitrarily sparse excitation matrix α, which is also low-rank. All parameters were chosen in such a way to ensure both stability and stationarity of the multivariate process. For further details on the

---

[2] https://cran.r-project.org/web/packages/hawkes/index.html

specification of the Hawkes process used in this study, see Appendix B.

3.4.2 Parameter Tuning and Performance Metrics

The ridge and Lasso estimators require careful specification of a tuning parameter λ (where we have dropped the dependence on ω for ease of notation). To reduce the computational complexity of our experiments, a group of independently simulated samples are used to select sensible values of the regularisation parameter for use in our simulation study. Going forward, we will refer to this independent set of samples as the 'tuning set'. More specifically, estimates of $\Theta(\omega)$ are obtained over a fixed grid of λ values, and an optimal value of λ is selected according to an appropriate performance metric. Often, one may consider a measure of accuracy (i.e. precision of the recovered inverse spectrum), or edge selection (i.e. estimation of the correct sparsity pattern). For the former, we use the notion of mean squared error (MSE) defined in (3.4.2) and for the latter, we consider the F_1 score defined as

$$\mathrm{F}_1 = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})}$$

where TP is the number of correctly identified edges in the estimate graph, while FP and FN are the number of false positives and false negatives respectively. In the absence of a ground truth comparison, we recommend the use of the Extended Bayesian information criterion (eBIC) for tuning parameter selection, which is defined in Foygel and Drton (2010) as

$$\mathrm{eBIC}_{\gamma}(\hat{\Theta}_{\lambda}(\omega)) = 2\ell_m(\hat{\Theta}_{\lambda}(\omega)) + |E|\log(m) + 4\gamma|E|\log(p), \tag{3.4.1}$$

where E is the edge set of a candidate graph, $\ell_m(\hat{\Theta}(\omega))$ denotes the maximised log-likelihood of the associated model and γ is a tuning parameter. For each sample in the tuning set, an optimal choice of λ is chosen either by maximising the F_1 score or minimising the MSE and eBIC. These values are then averaged across the number of samples in the tuning set to arrive at a final optimal parameter λ^*.

We estimate the inverse spectral density matrix across $N = 100$ Monte Carlo samples. Then, we quantify the accuracy of our estimation using the mean squared error (MSE) defined as

$$\text{MSE}\left\{\hat{\Theta}^{(n)}(\omega), \Theta(\omega)\right\} = \frac{2}{Np(p-1)} \sum_{n=1}^{N} \sum_{q<r} \left(\hat{\Theta}_{qr}^{(n)} - \Theta_{qr}\right)^2 \tag{3.4.2}$$

where N denotes the number of Monte Carlo samples in our simulation study. We also report results for the recovery of the sparsity pattern for the Lasso estimate, and calculate the area under receiver operating characteristic (AUROC) curve. Since the ridge and periodogram estimators do not produce sparse results, we only report AUROC scores for the Lasso estimate. Numerical results averaged across 100 repetitions are presented in Table 3.4.1.

3.4.3 Simulation Results

Table 3.4.1 reports simulation results for Models (a)-(c), respectively. Numerical results are averaged across 100 Monte Carlo samples, and each entry of the table is of the form mean (standard error). We omit standard errors of $\leq 10^{-2}$ for brevity. Lasso_1, Lasso_2 and Lasso_3 refer to the Lasso estimator tuned using the MSE, F_1 score and eBIC, respectively. Results are obtained at a particular frequency $\omega = 0.0268$. As shown in Figure 2.4.3, temporal clustering of the point process results in increased power in the spectrum at lower frequencies. Thus, we are interested in assessing the performance of our spectral estimators at these low frequencies. In neuroscience, these are more commonly known as frequencies in the delta, $(0, 4)$Hz, band (Buzsáki, 2004).

The results across all simulation settings reveal that the multi-trial periodogram matrix cannot be inverted in settings where $p > m_T$, thus motivating the need for regularisation or shrinkage.

Table 3.4.1: Simulation results over 100 replications for estimating the inverse spectral density matrix, with mean squared error (MSE) and AUROC. All results are recorded at a particular frequency $\omega = 0.0628$ and are in the form of mean (standard error). Standard errors of $< 10^{-2}$ are omitted for brevity. Hyphenated entries (-) denote that the multi-trial periodogram matrix could not be inverted.

p	m	Mean Squared Error - ISDM					AUROC		
		IP	Ridge	Lasso$_1$	Lasso$_2$	Lasso$_3$	Lasso$_1$	Lasso$_2$	Lasso$_3$
					Model (a)				
12	10	-	3.67 (0.04)	2.50 (0.05)	5.26 (0.01)	5.20 (0.02)	1.00	0.91 (0.01)	0.92
	50	5.14	1.36 (0.02)	0.86 (0.01)	4.85 (0.04)	2.97 (0.02)	1.00	1.00	1.00
48	10	-	1.18	0.88 (0.01)	1.27	1.26	0.99	0.85	0.86
	50	38522.09	0.81 (0.01)	0.34	1.19	0.95	1.00	1.00	1.00
96	10	-	0.61	0.48	0.63	0.63	0.99	0.84	0.84
	50	-	0.51	0.21	0.59	0.48	1.00	1.00	1.00
					Model (b)				
12	10	-	0.34 (0.03)	0.18	0.87	0.44 (0.01)	0.98	0.92 (0.01)	0.96
	50	0.94	0.18 (0.02)	0.07	0.81	0.09	0.99	1.00	1.00
48	10	-	0.14	0.07	0.21	0.21	0.99	0.80	0.74
	50	6411.92	0.08	0.02	0.20	0.14	1.00	0.99	1.00
96	10	-	0.09	0.04	0.10	0.10	1.00	0.74	0.72
	50	-	0.04	0.01	0.10	0.09	1.00	0.99	1.00
					Model (c)				
12	10	-	3.35 (0.01)	2.24 (0.05)	4.50	3.15 (0.04)	0.98	0.95	0.98
	50	26.07	2.33 (0.01)	1.56 (0.02)	4.28	1.87 (0.02)	0.97	0.95	0.97
48	10	-	0.93	0.69 (0.01)	1.09	1.03	0.98	0.93	0.95
	50	214203.50	0.77	0.30	1.08	0.71	0.97	0.95	0.98
96	10	-	0.48	0.42	0.54	0.52	0.97	0.88	0.95
	50	-	0.42	0.17	0.54	0.49	0.97	0.95	0.95

In terms of MSE, the ridge and Lasso estimators are comparable across all experimental settings, though generally the Lasso estimator yields (slightly) more favourable results. In particular, Lasso$_1$ outperforms all estimators in every scenario, while the ridge estimator is most competitive with the Lasso$_3$ estimator in settings (a) and (b). This highlights a potential drawback of the ridge estimator in that it is biased towards a scaled identity matrix, of which the sparsity pattern of models (a) and (b) somewhat resemble (Fiecas et al., 2019). It also suggests the use of the Lasso estimator when the sparsity pattern of the inverse spectrum is unspecified rather than containing some sort of banded structure. Notably, the inverted periodogram (IP) estimate is somewhat competitive in terms of MSE in scenarios where $p = 12$ and $m = 50$. This highlights the competence of the multi-trial estimate in scenarios where $p < m$, while also emphasising the importance for appropriate regularisation in high-dimensional settings.

The AUROC results highlight that the Lasso estimator generally yields favourable results in terms of sparse recovery of the inverse spectrum. However, since the Lasso_3 estimator provides favourable results both in terms of MSE and sparse recovery of the spectrum, we recommend its use in practical settings.

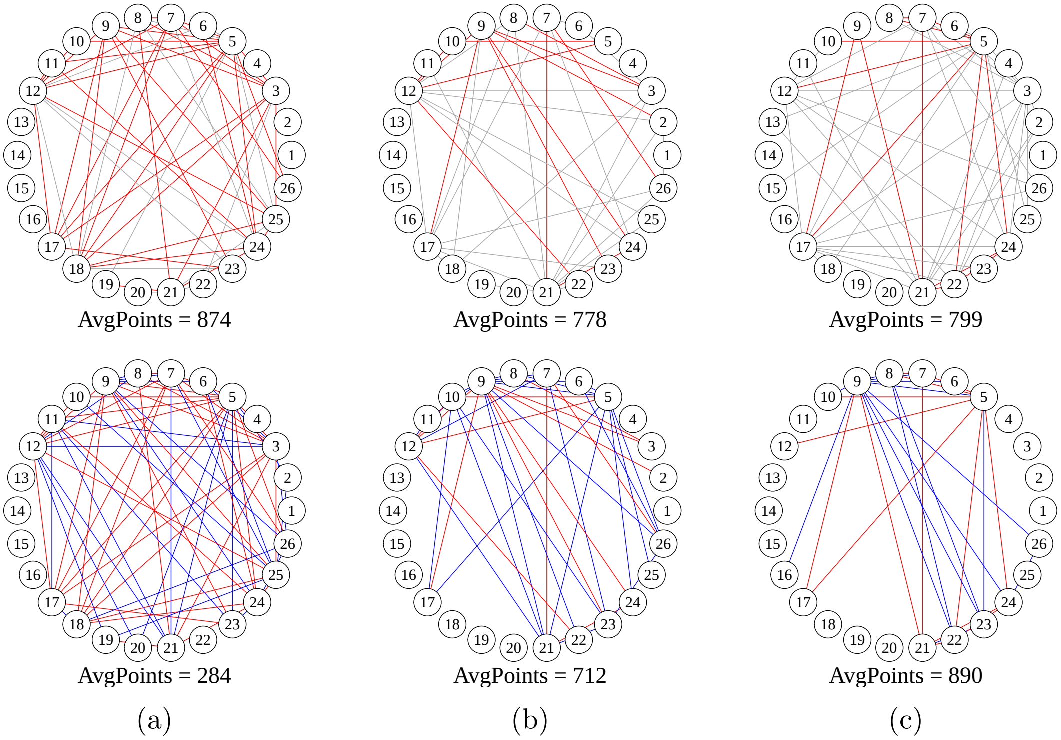
Overall, the Lasso estimator is preferable when the goal is to learn the sparsity pattern of the inverse spectrum. However, the ridge estimator yields competitive results in terms of the accuracy of entries in the recovered estimate. Furthermore, the results presented here highlight that the proposed methodology improves upon existing estimators, such as the multi-trial periodogram, which break down in high-dimensional settings.

3.5 Neuronal Synchronicity and Partial Coherence Networks

We illustrate how our proposed methodology can be used to infer a network of interactions between a population of neurons, using the spike train data from Bolding and Franks (2018). In this experiment, an optogenetic approach was used to directly stimulate the olfactory bulb (OB) region of the mouse brain. In particular, 1 second light pulses were presented above the OB region at increasing intensity levels from 0 to 50mW/mm^2 over a total of $m = 10$ experimental trials.

The dataset consists of spike times for multiple mice, collected at 8 intensities of laser stimuli. As an illustrative example, we will consider data collected at three stimuli ($0, 10$ and 50mW/mm^2) from a single mouse with $p = 26$ neurons measured in the OB region.

For each of the considered levels of stimuli, data is collected during a 'laser on' and 'laser off' condition. For the laser on condition, we have 1 second worth of data,



(a) (b) (c)

Figure 3.5.1: Estimated networks of neural interactions on the delta band using the spike train data from Bolding and Franks (2018). Each column shows the estimated networks for the laser off (top) and laser on (bottom) condition for laser stimuli (a) 0mW/mm^2, (b) 10mW/mm^2 and (c) 50mW/mm^2. Common edges between the on and off conditions for each level of intensity are shown in red, and edges unique to the laser on condition are shown in blue. All other edges are shown in grey.

compared to a total of 3 seconds for the laser off condition. We use our Lasso estimator to visualise functional connectivity for each condition, under the differing levels of laser stimuli. Since a laser stimulus of intensity 0mW/mm^2 is equivalent to the laser off condition, we use this setting as a plausibility check for the proposed methodology.

Initial exploratory analysis showed that each of the 26 point processes are driven by lower frequencies in the spectrum, which in neuroscience are named as frequencies in the delta, $(0, 4)$ Hz, and theta, $(4, 8)$ Hz, bands (Buzsaki and Draguhn, 2004). Consequently, we will apply our method to these lower frequency bands. The overall aim of our analysis is to use the regularised estimator of the partial coherence matrix to determine if and how estimates of neuronal connectivity differ in response to the changing stimuli.

To estimate connectivity in a particular frequency band, we consider an adaptation of the Lasso estimator whereby we use both a trial and frequency smoothed periodogram estimate as input to optimisation problem (3.3.1). Specifically, we define the trial-frequency smoothed estimator as

$$\hat{S}(\omega_*) = \frac{1}{\mathrm{card}(F)} \sum_{f \in F} \frac{1}{m} \sum_{k=1}^{m} \bar{d}_k(\omega_f) \bar{d}_k^H(\omega_f), \tag{3.5.1}$$

where the frequencies f lie in a specified range. This modified version of the Lasso estimator is tuned using the eBIC in (3.4.1) with $\gamma = 0.5$, as suggested in Foygel and Drton (2010). The degrees of freedom are determined by the number of non-zero elements in our estimate $\hat{\Theta}(\cdot)$.

Remark 3.5.1. *The additional smoothing across frequencies in* (3.5.1) *further reduces the variability of estimated spectrum, compared to the previously discussed multi-trial estimator* (3.2.2)*. While a full theoretical analysis of* (3.5.1) *awaits future work, this modified estimator enables estimation of the inverse spectrum across specified frequency bands of interest. Alternatively, one could estimate the inverse spectrum at multiple frequencies directly, utilising the multi-trial estimator alongside a group-Lasso penalty (c.f. Chapter 2). However, this is left as future work.*

Figure 3.5.1 shows the estimated partial coherence graphs obtained for the delta band under the laser on and laser off conditions. That is, we present estimates of neuronal interactions for frequencies in the range $(0, 4)$ Hz that are specific to each experimental condition. The nodes in the graphs represent neurons and edges indicate non-zero values of partial coherence estimated using the Lasso estimator. Each plot is annotated with the average number of spikes that are observed under each experimental condition. We highlight that for the laser on condition, the number of spikes observed increases with the increasing level of laser stimuli.

In general, the network structure changes in response to the external stimuli. Less edges are detected when the laser is applied (40 under $10mW/mm^2$ and 29 under $50mW/mm^2$ for the laser on setting versus 40 under $10mW/mm^2$ and 50 under $50mW/mm^2$ for the laser off setting). For the $0mW/mm^2$ setting, 56 edges are detected in the 'laser off' condition versus 70 for the 'laser on' condition. We hypothesise that this difference in network structure is a consequence of the difference in the observed number of points. Overall, the results presented here align with the findings in Bolding and Franks (2018) where it is highlighted by neuroscientists that the OB response is sensitive to the application of external stimuli.

Remark 3.5.2. *For further analysis of a different neural spiking dataset using the proposed regularised spectral estimator, see Chapter 5.*

3.6 Discussion

We have proposed novel statistical methodology for the estimation of high-dimensional inverse spectral density matrices in the point process setting. The main advantage of our proposal is the ability to estimate the inverse spectrum when the dimension of the multivariate process is large, i.e. when $p > m_T$, unlike existing methods which break down in high-dimensional settings. Importantly, our Lasso estimator yields sparse estimates of the inverse spectrum, easing interpretation of the resulting partial coherence networks. Moreover, under the stated incoherence condition, this estimator can correctly identify the sparse structure of the inverse spectrum. In general, we would recommend the Lasso-based estimator due to this interpretability; however, in situations where the sparsity assumption may not hold, we demonstrate that a ridge-based regulariser can also provide estimates of the spectral density in high-dimensional scenarios.

There are many possible and interesting avenues for future work. Whilst we have demonstrated consistent estimation of the inverse spectrum, it remains to provide a way in which to quantify uncertainty around the resulting partial coherence graphs.

Existing work on statistical inference has been discussed for high-dimensional inverse covariance estimation in Janková and van de Geer (2015). Furthermore, extending the methodology to non-stationary settings is also of interest, where one could easily extend the localised estimation methods provided in Roueff and Sachs (2019),which may find utility both in neuroscience, but also in high-frequency econometric applications. Moreover, one may consider further extensions beyond the current assumption of non-overlapping trials; for example, it would be relatively straightforward to apply our regularisation methods (and analysis) in the more general multi-tapering framework of Walden (2000); Cohen and Gibberd (2021).

3.7 Proofs

3.7.1 Proof of Lemma 3.2.7

Recall the form of the taper, corresponding to trial k, as

$$h_k(z) = \begin{cases} (T/2\pi z_T)^{1/2} & \text{for } z \in (s, s+z_T] \\ 0 & \text{otherwise.} \end{cases}$$

The Fourier transform of $h(z)$ for our rectangular taper is given by

$$\begin{aligned} H_k(\omega) &:= \int_0^1 h(z) e^{-i\omega z} dz \\ &= (T/2\pi z_T)^{1/2} \int_0^{z_T} e^{-i\omega z} dz \\ &= (T/2\pi z_T)^{1/2} \left[\frac{e^{-i\omega z}}{-i\omega} \right]_0^{z_T} \\ &= (T/2\pi z_T)^{1/2} \left[\frac{e^{-i\omega z_T} - 1}{-i\omega} \right] \\ &= 2(T/2\pi z_T)^{1/2} \left[\frac{e^{-i\omega z_T/2}}{\omega} \left(\frac{e^{i\omega z_T/2} - e^{-i\omega z_T/2}}{2i} \right) \right] \\ &= (T z_T/2\pi)^{1/2} e^{-i\omega z_T/2} \mathrm{sinc}(\omega z_T/2) \ . \end{aligned}$$

In particular, we will be interested in the size of this function $|H_k(\omega)|^2 = \frac{z_T T}{2\pi} \mathrm{sinc}^2(\omega z_T/2)$.

Consider the definition of the mean-corrected finite Fourier transform

$$\begin{aligned} \bar{d}_k(\omega) &:= \int_0^T h_k(t/T) e^{-i\omega t} dN(t) - \frac{H_k(T\omega)}{H_k(0)} \int_0^T h_k(t/T) dN(t) \\ &\sim \int_0^{n_T} h_k(t/T) e^{-i\omega t} dN(t) - \frac{H_k(T\omega)}{H_k(0)} \int_0^{n_T} h_k(t/T) dN(t) \ , \end{aligned}$$

where the second line is correct up to a phase shift. The multi-trial periodogram is given as

$$\hat{S}_{m_T}(\omega) := \frac{1}{m_T} \sum_{k=1}^{m_T} \bar{d}_{T;k}(\omega) \bar{d}^H_{T;k}(\omega) \ .$$

Taking expectations we have

$$\mathbb{E}[\hat{S}_{m_T}(\omega)] = \phi(\omega)\mathbb{E}[\bar{d}_{T;k}(\omega)\bar{d}^H_{T;k}(\omega)] \ ,$$

for any $k \in \{1, \dots, m_T\}$, due to the stationarity of $N(t)$, and where $\phi(\omega) = m_T^{-1}\sum_{k=1}^{m_T} e^{-i\omega s_k}$ and $|\phi(m)| = 1$. Thus without loss of generality we use $k = 1, s_1 = 0$, $\mathrm{supp}(h_1) = (0, z_T]$.

We now study

$$\begin{aligned}
\mathbb{E}[\bar{d}_{T;k}(\omega)\bar{d}^H_{T;k}(\omega)] &= \mathbb{E}\Bigg[\bigg(\int_0^{n_T} h_k(t/T)e^{-i\omega t}dN(t) - \frac{H_k(T\omega)}{H_k(0)}\int_0^{n_T} h_k(t/T)dN(t)\bigg) \\
&\qquad \times \bigg(\int_0^{n_T} h_k(t/T)e^{-i\omega t}dN(t) - \frac{H_k(T\omega)}{H_k(0)}\int_0^{n_T} h_k(t/T)dN(t)\bigg)^H\Bigg] \\
(I) &= \int_0^T\int_0^T h_k(t/T)h_k(u/T)e^{-i\omega(t-u)}\mu_c(t-u)dtdu \\
(II) &- \frac{H_k^*(T\omega)}{H_k(0)}\int_0^T\int_0^T h_k(t/T)h_k(u/T)e^{-i\omega t}\mu_c(t-u)dtdu \\
&- \frac{H_k(T\omega)}{H_k(0)}\int_0^T\int_0^T h_k(t/T)h_k(u/T)e^{i\omega u}\mu_c(t-u)dtdu \\
(III) &+ \frac{|H_k(T\omega)|^2}{H_k^2(0)}\int_0^T\int_0^T h_k(t/T)h_k(u/T)\mu_c(t-u)dtdu \ ,
\end{aligned}$$

where we have substituted $\mathbb{E}[dN(t)dN(u)] = \mu_c(t,u)dtdu + \Lambda\Lambda^\top dtdu$, (to get to the second equality) where the complete auto-covariance density is given by $\mu_c(\tau) = \mu(\tau) + \mathrm{diag}(\lambda)\delta(\tau)$, see e.g. Hawkes (1971b).

Note, the terms involving $\Lambda\Lambda^\top$ cancel out across the three lines (I, II, III). Using the time-scale relation for the Fourier transform, alongside the convolution theorem, the above integrals can be expressed as integrals with respect to the Fourier transforms

$$\begin{aligned}
(I) &= \int |H_k(T(\omega - \omega'))|^2 S(\omega')d\omega' \\
(II) &= -2\int \mathrm{Re}\bigg[\frac{H_k^*(T\omega)}{H_k(0)}H_k(T\omega - T\omega')H_k(T\omega')S(\omega')\bigg]d\omega' \\
(III) &= \frac{|H_k(T\omega)|^2}{H_k^2(0)}\int |H_k(T\omega')|^2 S(\omega')d\omega'
\end{aligned}$$

which taken together gives the stated form for the expectation, i.e.

$$\begin{aligned}\mathbb{E}[\hat{S}_{m_T}(\omega)] &= \phi(\omega)\int\left|H_k(T(\omega-\omega')) - \frac{H_k(T\omega)H_k^*(T\omega')}{H_k(0)}\right|^2 S(\omega')d\omega' \\ &\equiv \phi(\omega)\int |Z_T(\omega,\omega')|^2 S(\omega')d\omega' \ .\end{aligned}$$

For the bound on the bias, we consider the first order Taylor series approximation of $S(\omega')$ about ω_*, recalling that for ω_* we have $H(T\omega_*) = 0$, thus we have the simplification $|Z_T(\omega_*,\omega')|^2 = |H_k(T(\omega_* - \omega'))|^2$. Substituting the series expansion (with remainder) into (3.7.1) (with this simplification) gives

$$\begin{aligned}\mathbb{E}[\hat{S}_{m_T}(\omega_*)] &= \phi(\omega_*)\int |H_k(T(\omega_*-\omega'))|^2 S(\omega')d\omega' \\ &= \phi(\omega_*)\int |H_k(T(\omega_*-\omega'))|^2\left[S(\omega_*) + S'(\omega_*)(\omega'-\omega_*) + \frac{S^{''}(\bar{\omega})}{2}(\omega'-\omega_*)^2\right]d\omega' \ ,\end{aligned}$$

where $\bar{\omega}$ is in the interval formed by ω' and ω_*. Under our assumption we have $S^{''}(\bar{\omega}) \leq M_2$, and that for all ω we have $\int |H_k(T(\omega-\omega'))|^2 d\omega' = 1$, and indeed $f_{W_T(\omega)}(x) := |H_k(T(\omega - x))|^2$ can be treated like a density for an associated random variable $W_T(\omega)$. As such we have

$$\begin{aligned}\left|\mathbb{E}[\hat{S}_{m_T}(\omega_*)] - \phi(\omega_*)S(\omega_*)\right| &\leq \left|S'(\omega_*)\int f_{W_T(\omega_*)}(\omega')(\omega'-\omega_*)d\omega'\right| \\ &\quad + \frac{M_2}{2}\left|\int f_{W_T(\omega_*)}(\omega')(\omega'-\omega_*)^2 d\omega'\right| \\ &\leq |S'(\omega_*)|(\mathbb{E}[W_T(\omega_*)] - \omega_*) + \frac{M_2}{2}\mathrm{Var}(W_T(\omega_*)) \ .\end{aligned}$$

As a corollary, for our specific tapers we can explicitly take the expectations to find

$$
\begin{aligned}
\mathbb{E}[W_T(\omega)] &= \int f_{W_T(\omega_*)}(\omega')\omega' d\omega' \\
&= \frac{z_T T}{2\pi} \int \operatorname{sinc}^2([T(\omega_* - \omega')]z_T/2)\omega' d\omega' \\
&= \frac{1}{\pi} \int \operatorname{sinc}^2(x_* - x')\big(\frac{2x'}{Tz_T}\big)dx' \\
&= \frac{1}{\pi}\frac{2}{Tz_T} \int \operatorname{sinc}^2(x_* - x')x' dx' \\
&= \frac{2}{\pi z_T T} \int u \operatorname{sinc}^2(u)du + x^* \int \operatorname{sinc}^2(u)du \\
&= \frac{2x_*}{Tz_T} = \omega_*
\end{aligned}
$$

where we have used substitution $x_* = T\omega_* z_T/2$, $x' = T\omega' z_T/2 \implies dx'/d\omega' = Tz_T/2$ on line 3, and the substitution $u = x_* - x$ on line 5.

Now for the variance. First use $x_* = Tz_T\omega_*/2$ and $x = Tz_T\omega/2$ thus

$$
\begin{aligned}
\int f_{W_T(\omega_*)}(\omega)\omega^2 d\omega &= \frac{z_T T}{2\pi} \int \operatorname{sinc}^2([T(\omega_* - \omega)]z_T/2)\omega^2 d\omega \\
&= \frac{1}{\pi} \int \operatorname{sinc}^2(x_* - x)\frac{4x^2}{T^2 z_T^2}dx \\
&= -\frac{1}{\pi} \int \operatorname{sinc}^2(u)\frac{4(x_* - u)^2}{T^2 z_T^2}du \\
&= -\frac{4}{\pi T^2 z_T^2} \int \operatorname{sinc}^2(u)(x_* - u)^2 du \\
&= -\frac{4}{\pi T^2 z_T^2}\left[x_*^2 \int \operatorname{sinc}^2(u)du - 2x_* \int \operatorname{sinc}^2(u)udu + \int \operatorname{sinc}^2(u)u^2 du\right] \\
&= \omega_*^2 - 0 - \frac{4}{\pi T^2 z_T^2} \int \operatorname{sinc}^2(u)u^2 du
\end{aligned}
$$

where on the second line we use $u = x_* - x$, and the middle term on the last line is zero due to symmetry. Using Plancherel's theorem on the remaining term, we find

$$
\int \operatorname{sinc}^2(u)u^2 du = \int |\mathcal{F}\{u\operatorname{sinc}(u)\}|^2(\omega)d\omega = \int |\mathcal{F}\{\sin(u)\}\{\omega\}|^2 d\omega \,.
$$

Now note $\mathcal{F}\{\sin(u)\}\{\omega\} = -i\pi[\delta(\omega-1) - \delta(\omega+1)]$, and thus we have

$$\mathrm{Var}(W_T(\omega)) = \int f_{W_T(\omega_*)}(\omega')(\omega' - \omega_*)^2 d\omega' = \mathcal{O}(T^{-2}) \ .$$

3.7.2 Proof of Proposition 3.2.9

Proof. We begin the proof by recalling the form of the trial averaged periodogram

$$\hat{S}_{m_T}(\omega) = \frac{1}{m_T}\sum_{k=1}^{m_T} \bar{d}_{T;k}(\omega)\bar{d}^H_{T;k}(\omega).$$

Now, for each pair (q,r) and $\delta > 0$, we define the event

$$\mathcal{A}_{T;qr}(\delta,\omega) = \left\{ \left| \frac{1}{m_T}\sum_{k=1}^{m_T} \bar{d}_{T;k;q}(\omega)\bar{d}^H_{T;k;r}(\omega) - S_{qr}(\omega) \right| > \delta \right\}.$$

Furthermore, each $\bar{d}_{T;k}(\omega) = \{\bar{d}_{T;k;1}(\omega), \ldots, \bar{d}_{T;k;p}(\omega)\}^\top$ can be described by the $2p$ dimensional real-valued random vector $z_k = (\bar{x}_k^\top, \bar{y}_k^\top)^\top$ where (for ease of notation we drop the dependence on ω and T)

$$\bar{x}_k = \mathrm{Re}\{\bar{d}_{T;k}(\omega)\}$$
$$\bar{y}_k = \mathrm{Im}\{\bar{d}_{T;k}(\omega)\}$$

so that $\bar{d}_{T;k}(\omega) = \bar{x}_k + i\bar{y}_k$. More specifically, $z_k = (\bar{x}_k^\top, \bar{y}_k^\top)^\top$ encodes the fFT for the k^{th} segment of the point process. Our proof will proceed by following the de-coupling argument of Bickel and Levina (2008) to study the normalised variates

$$\tilde{x}_{k;q} = \bar{x}_{k;q}/\{S_{qq}(\omega)\}^{1/2} \quad \text{and} \quad \bar{y}_{k;q}/\{S_{qq}(\omega)\}^{1/2}.$$

However, we first state the following lemma:

Lemma 3.7.1. *A complex random variable* $\mathrm{supp}(X) \in \mathbb{C}$ *can be bound according to*

$$|X - \mathbb{E}[X]| \leq |\operatorname{Re}(X) - \mathbb{E}[\operatorname{Re}(X)]| + |\operatorname{Im}(X) - \mathbb{E}[\operatorname{Im}(X)]|$$

via the triangle inequality. Moreover, if $\mathrm{supp}(X) \in \mathbb{C}^p$ *then*

$$\|X - \mathbb{E}(X)\| \leq \|\operatorname{Re}(X) - \mathbb{E}\{\operatorname{Re}(X)\}\| + \|\operatorname{Im}(X) - \mathbb{E}\{\operatorname{Im}(X)\}\|$$

for any norm $\|\cdot\| : \mathbb{R}^p \mapsto \mathbb{R}$ *(or* $\mathbb{C}^p \mapsto \mathbb{R}$*).*

Applying Lemma 3.7.1, we consider bounds on the real and imaginary components of $\bar{d}_{T;k}(\omega)$ separately. In particular, we note that $\mathrm{pr}\left\{\left|\hat{S}_{m_T;qr}(\omega) - S_{qr}(\omega)\right| \geq \delta\right\}$ is at most

$$\mathrm{pr}\underbrace{\left[\left|\operatorname{Re}\{\hat{S}_{m_T;qr}(\omega) - S_{qr}(\omega)\}\right| \geq \delta/2\right]}_{\mathcal{A}'_{T;qr}(\delta,\omega)} + \mathrm{pr}\underbrace{\left[\left|\operatorname{Im}\{\hat{S}_{m_T;qr}(\omega) - S_{qr}(\omega)\}\right| \geq \delta/2\right]}_{\mathcal{A}''_{T;qr}(\delta,\omega)}.$$

3.7.3 Proof of Proposition 3.2.9 (a)

We begin our proof by showing that, under Assumption 3.2.3, the mean corrected variates are asymptotically complex Gaussian, i.e., $\bar{d}_{T;k}(\omega) \sim \mathcal{N}_c(0, S(\omega))$ as $T \to \infty$.

Lemma 3.7.2. *Brillinger (1972) Under Assumption 3.2.3, we consider the set of* $k = 1, \ldots, m$ *trials. Now, let* f_T *be an integer sequence with* $\omega_T = 2\pi f_T / z_T T \to \omega$ *as* $T \to \infty$*. The tapered and mean corrected Fourier coefficients are distributed as*

$$\bar{d}_k(\omega_T) \sim \mathcal{N}_C(0, S(\omega)) \text{ as } T \to \infty.$$

Proof. Consider first the form of the mean corrected tapered Fourier transform defined in Brillinger (1972) for a particular frequency ω as

$$\bar{d}_{T;k}(\omega) = d_{T;k}(\omega) - \underbrace{\frac{d_{T;k}(0)H_k(T\omega)}{H_k(0)}}_{b(\omega)},$$

where we refer to $b(\omega)$ as the so called "mean correction" term. We proceed by considering the mean and variance of $b(\omega)$ before an application of Theorem 4.2 Brillinger (1972) to complete the proof. We compute the mean of $b(\omega)$ as

$$\begin{aligned}\mathbb{E}\{b(\omega)\} &= \frac{H_k(T\omega)}{H_k(0)}\mathbb{E}\left\{d_{T;k}(0)\right\} \\ &= \operatorname{sinc}(T\omega z_T/2)\exp\{-iT\omega z_T/2\}\mathbb{E}\{d_{T;k}(0)\}.\end{aligned}$$

At the sampled frequencies $\omega_T = 2\pi f_T/z_T T \to \omega$ as $T \to \infty$ we have that $\operatorname{sinc}(T\omega z_T/2) = \frac{\sin\{\pi f_{T'}\}}{\pi f_{T'}} = 0$ and thus $\mathbb{E}\{b(\omega)\} = 0$. The variance of $b(\omega)$ is

$$\begin{aligned}\operatorname{var}\{b(\omega)\} &= \left\{\frac{H_k(T\omega)}{H_k(0)}\right\}^2 \operatorname{var}\{d_k(\omega)\} \\ &= \left[\operatorname{sinc}(T\omega z_T/2)\exp\{-iT\omega z_T/2\}\right]^2 \operatorname{var}\{d_{T;k}(0)\}.\end{aligned}$$

As above, we have that $\operatorname{sinc}(T\omega z_T/2) = 0$ at the sample frequencies ω_T. Therefore, $\operatorname{var}\{b(\omega)\} = 0$. Thus at the sampled frequencies, Theorem 4.2 Brillinger (1972) also applies to the mean corrected tapered Fourier transform. Asymptotically, we therefore have that

$$\bar{d}_{T;k}(\omega) \sim \mathcal{N}_C(0, S(\omega)),$$

and $\bar{d}_{T;k}(0) = 0$. $\square$

From Brillinger (1972) we therefore have that each $\bar{d}_{k;q} \sim \mathcal{N}_c(0, S_{qq}(\omega))$, as $T \to \infty$. Using the fact that $\bar{d}_{k;q} = \bar{x}_{k;q} + i\bar{y}_{k;q}$, we have that $\bar{x}_{k;q} \sim \mathcal{N}(0, S_{qq}(\omega)/2)$ and $\bar{y}_{k;q} \sim \mathcal{N}(0, S_{qq}(\omega)/2)$, and therefore the normalised variates $\tilde{x}_{k;q} \sim N(0, 1/2)$ as $T \to \infty$. We proceed by defining the auxiliary random variables

$$u_{k;qr} = \tilde{x}_{k;q} + \tilde{x}_{k;r} \quad \text{and} \quad v_{k;qr} = \tilde{x}_{k;q} - \tilde{x}_{k;r} \ .$$

With this notation, we then claim:

Lemma 3.7.3. *For all $\delta > 0$, pr$\left\{\mathcal{A}'_{T;qr}(\delta, \omega)\right\}$ is upper bounded by*

$$pr\left(\left|\sum_{k=1}^{m}\left[u_{k;qr}^2 - \{1 + 2R_{qr}(\omega)\}\right]\right| \geq \frac{\delta m}{\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}}\right) \tag{3.7.1}$$

$$+\, pr\left(\left|\sum_{k=1}^{m}\left\{v_{k;qr}^2 - \{1 - 2R_{qr}(\omega)\}\right]\right| \geq \frac{\delta m}{\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}}\right). \tag{3.7.2}$$

where $R_{qr}(\omega) = \mathrm{Re}\{S_{qr}(\omega)\}/\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}$.

Proof. For all $q = 1, \ldots, p$ we have that the normalised variates $\tilde{x}_{k;q} \sim N(0, 1/2)$ as $T \to \infty$. Thus,

$$u_{k;qr} \sim N\left(0, 1 + 2R_{qr}(\omega)\right) \quad \text{and} \quad z_{k;qr} \sim N(0, 1 - 2R_{qr}(\omega))$$

where $R_{qr}(\omega) = \mathrm{cov}(\tilde{x}_{k;q}, \tilde{x}_{k;r}) = \mathrm{Re}[R_{qr}(\omega)/\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}]$.

Now, using straightforward algebra it is easy to show that

$$\sum_{k=1}^{m}\{\tilde{x}_{k;q}\tilde{x}_{k;r} - R_{qr}(\omega)\} = \left(\frac{1}{4}\sum_{k=1}^{m}\left[\{\tilde{x}_{k;q} + \tilde{x}_{k;r}\}^2 - \{1 + 2R_{qr}(\omega)\}\right]\right) - \left(\frac{1}{4}\sum_{k=1}^{m}\left[\{\tilde{x}_{k;q} - \tilde{x}_{k;r}\}^2 - \{1 - 2R_{qr}(\omega)\}\right]\right).$$

Thus, we have that pr $\left[\sum_{k=1}^{m}|\{\tilde{x}_{k;q}\tilde{x}_{k;r} - R_{qr}(\omega)\}| \geq \delta/2\right]$ is at most

$$\mathrm{pr}\left(\left|\frac{1}{4}\sum_{k=1}^{m}\left[u_{k;qr}^2 - \{1 + 2R_{qr}(\omega)\}\right]\right| \geq \frac{\delta}{4}\right) + \mathrm{pr}\left(\left|\frac{1}{4}\sum_{k=1}^{m}\left[v_{k;qr}^2 - \{1 - 2R_{qr}(\omega)\}\right]\right| \geq \frac{\delta}{4}\right)$$

and therefore $\mathrm{pr}\left\{\mathcal{A}'_{T;qr}(\delta,\omega)\right\}$ is upper bounded by

$$\begin{aligned}&\mathrm{pr}\left(\left|\sum_{k=1}^{m}\left[u_{k;qr}^2-\{1+2R_{qr}(\omega)\}\right]\right|\geq\frac{\delta m}{\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}}\right)\\&+\mathrm{pr}\left(\left|\sum_{k=1}^{m}\left\{v_{k;qr}^2-\{1-2R_{qr}(\omega)\}\right]\right|\geq\frac{\delta m}{\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}}\right).\end{aligned}$$

which completes the proof. $\square$

To complete the proof of Proposition 3.2.9(a), it remains to upper bound the tail probabilities (3.7.1) and (3.7.2).

Remark 3.7.4. *Lemma 3.7.3 is reminiscent of Lemma* 9 *part* (b) *in Ravikumar et al. (2011). There, the authors proceed to bound* (3.7.1) *and* (3.7.2) *under a sub-Gaussianity assumption of their equivalent* $u_{k;qr}$ *and* $v_{k;qr}$ *variates. By contrast, we will proceed by exploiting the asymptotic complex normality of our mean-corrected coefficients* $\bar{d}_{T;k}(\omega)$*.*

From standard properties of Gaussian random variables, we have that $u_{k;qr}^2\sim(1+2R_{qr})\chi_1^2$ and $\sum_{k=1}^{m}u_{k;qr}^2\sim(1+2R_{qr})\chi_m^2$. To proceed, we recall the form of the standard Laurent and Massart (2000) bounds for a random variable $X\sim\chi_k^2$

$$\mathrm{pr}(X-k\geq2\sqrt{kx}+2x)\leq e^{-x}\quad\text{and}\quad\mathrm{pr}(k-X\geq2\sqrt{kx})\leq e^{-x},$$

and using the union bound, we obtain $\mathrm{pr}(|X-k|\geq2\sqrt{kx}+2x)\leq2e^{-x}$. Thus, for our $u_{k;qr}^2$ variates we have

$$\mathrm{pr}\left\{\left|\sum_{k=1}^{m}u_{k;qr}^2-m(1+2R_{qr})\right|\geq(1+2R_{qr})(2\sqrt{mx}+2x)\right\}\leq2\exp(-x),$$

or equivalently,

$$\mathrm{pr}\left\{\left|\sum_{k=1}^{m_T}u_{k;qr}^2-m(1+2R_{qr})\right|\geq(1+2R_{qr})(2\sqrt{m(cm\delta^2)}+2cm\delta^2)\right\}\leq2\exp(-cm\delta^2),$$

where $x = cm\delta^2$. Letting $a = (1 + 2R_{qr})$, then

$$\begin{aligned} 2a\{(cm^2\delta^2)^{1/2} + cm\delta^2\} &= 2am(c^{1/2}\delta + c\delta^2) \\ &\leq 2am(2c^{1/2}\delta) \end{aligned}$$

if $c\delta^2 \leq c^{1/2}\delta \implies \delta \leq c^{-1/2}$. Then,

$$\text{pr}\left(\left|\sum_{k=1}^{m}\left[u_{k;qr}^2 - \{1 + 2R_{qr}(\omega)\}\right]\right| \geq 4amc^{1/2}\delta\right) \leq 2\exp(-cm\delta^2),$$

or equivalently, with $c = 1/[16\{1 + 2R_{qr}(\omega)\}^2]$, we have

$$\begin{aligned} \text{pr}\left(\left|\sum_{k=1}^{m}\left[u_{k;qr}^2 - \{1 + 2R_{qr}(\omega)\}\right]\right| \geq m\delta\right) &\leq 2\exp\left[-\frac{m\delta^2}{16\{1 + 2R_{qr}(\omega)\}^2 S_{qq}(\omega)S_{rr}(\omega)}\right] \\ &\leq 2\exp\left[-\frac{m\delta^2}{16(3)^2 S_{qq}(\omega)S_{rr}(\omega)}\right] \end{aligned}$$

valid for $0 \leq \delta \leq 12\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}$, where the last line (and condition on δ) hold since

$R_{qr}(\omega) \leq 1$. A similar argument yields the same tail bound for the deviation involving $\ddot{v}_{k;qr}$. Therefore, using lemma 3.7.3 we conclude that

$$\text{pr}\left\{\mathcal{A}'_{T;qr}(\delta,\omega)\right\} \leq 4\exp\left[\frac{-m\delta^2}{144 S_{qq}(\omega)S_{rr}(\omega)}\right],$$

valid for $0 \leq \delta \leq 12\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}$. Applying, a similar argument to bound the term $\mathcal{A}''(\delta,\omega)$ we arrive at the final result

$$\text{pr}\{|\hat{S}_{T;m;qr(\omega)} - S_{qr}(\omega)| \geq \delta\} \leq 8\exp\left[\frac{-m\delta^2}{144 S_{qq}(\omega)S_{rr}(\omega)}\right],$$

valid for $0 \leq \delta \leq 12\{S_{qq}(\omega)S_{rr}(\omega)\}^{1/2}$.

$\square$

3.7.4 Proof of Proposition 3.2.9 (b)

The key difference between part (a) and (b) from Proposition 3.2.9 is that T is finite. We will build the proof based on two elements: 1) assuming weak dependence on Fourier coefficients of the point processes and 2) controlling the bias term that will appear when estimating the spectrum (as in classical time series analysis). For each pair (q, r) and $\delta > 0$, we define the event

$$\mathcal{B}_{m_T;qr}(\delta,\omega) = \left\{\left|\hat{S}_{m_T;qr}(\omega) - \mathbb{E}\{\hat{S}_{m_T;qr}(\omega)\} + \mathbb{E}\{\hat{S}_{m_T;qr}(\omega)\} - S_{qr}(\omega)\right| \geq \delta\right\}$$

Let us define $\tilde{S}_{qr}(\omega) = \mathbb{E}\{\hat{S}_{m_T;qr}(\omega)\}$, then we have that

$$\begin{aligned}
\mathrm{pr}\{\mathcal{B}_{m_T;qr}(\delta,\omega)\} &\leq \mathrm{pr}\left(|\hat{S}_{m_T;qr}(\omega) - \tilde{S}_{qr}(\omega)| \geq \varphi\delta\right) + \mathrm{pr}\left(|\tilde{S}_{qr}(\omega) - S_{qr}(\omega)| \geq (1-\varphi)\delta\right) \\
&\leq \mathrm{pr}\left(|\hat{S}_{m_T;qr}(\omega) - \tilde{S}_{qr}(\omega)| \geq \varphi\delta\right) \quad \text{if } \delta_{\mathrm{b}} < (1-\varphi)\delta \\
&\leq \underbrace{\mathrm{pr}\left(\left|\mathrm{Re}[\hat{S}_{T;m:qr}(\omega) - \tilde{S}_{qr}(\omega)\}]\right| \geq \frac{\varphi\delta}{2}\right)}_{\mathcal{B}'_{m_T;qr}(\delta,\omega)} \\
&\quad + \underbrace{\mathrm{pr}\left(\left|\mathrm{Im}[\hat{S}_{T;m:qr}(\omega) - \tilde{S}_{qr}(\omega)]\right| \geq \frac{\varphi\delta}{2}\right)}_{\mathcal{B}''_{m_T;qr}(\delta,\omega)}.
\end{aligned}$$

Similar to the proof of Proposition 3.2.9 (a), we define the auxiliary random variables $u_{k;qr} = \tilde{x}_{k;q} + \tilde{x}_{k;r}$,$v_{k;qr} = \tilde{x}_{k;q} - \tilde{x}_{k;r}$ where $\tilde{x}_{k;q} = \bar{x}_{k;q}/\{\tilde{S}_{qq}(\omega)\}^{1/2}$. With this notation, we then claim:

Lemma 3.7.5. *For all $\delta > 0$, pr$\left\{\mathcal{B}'_{m_T;qr}(\delta,\omega)\right\}$ is upper bounded by*

$$pr\left(\left|\sum_{k=1}^{m}\left[\underbrace{u_{k;qr}^2 - 2\{1+\tilde{R}_{qr}(\omega)\}}_{\ddot{u}_{k;qr}}\right]\right| \geq \frac{m_T\varphi\delta}{\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}}\right) \tag{3.7.3}$$

$$+\, pr\left(\left|\sum_{k=1}^{m}\left[\underbrace{v_{k;qr}^2 - 2\{1-\tilde{R}_{qr}(\omega)\}}_{\ddot{v}_{k;qr}}\right]\right| \geq \frac{m_T\varphi\delta}{\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}}\right), \tag{3.7.4}$$

where $\tilde{R}_{qr} = \mathrm{Re}\{\tilde{S}_{qr}(\omega)/\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}\}$ *and* $\mathbb{E}(\ddot{u}_{k;qr}) = 0$ *and* $\mathbb{E}(\ddot{v}_{k;qr}) = 0$.

Proof. The proof is immediate from the proof of Lemma 3.7.3 with the exception of calculating the expectations $\mathbb{E}(u^2_{k;qr})$ and $\mathbb{E}(v^2_{k;qr})$. Consider first

$$
\begin{aligned}
\mathbb{E}\left(\sum_{k=1}^{m_T} u^2_{k;qr}\right) &= \sum_{k=1}^{m_T} \mathbb{E}\{(\tilde{x}_{k;q} + \tilde{x}_{k;r})^2\} \\
&= \sum_{k=1}^{m_T} \left\{\mathbb{E}(\tilde{x}^2_{k;q}) + 2\mathbb{E}(\tilde{x}_{k;q}\tilde{x}_{k;r}) + \mathbb{E}(\tilde{x}^2_{k;r})\right\} \\
&= \sum_{k=1}^{m_T} \left\{\mathbb{E}\left(\frac{\bar{x}^2_{k;q}}{\tilde{S}_{qq}(\omega)}\right) + 2\tilde{R}_{qr}(\omega) + \mathbb{E}\left(\frac{\bar{x}^2_{k;r}}{\tilde{S}_{rr}(\omega)}\right)\right\} \\
&= \sum_{k=1}^{m_T} \left\{\frac{1}{\tilde{S}_{qq}(\omega)}\mathbb{E}(\bar{x}^2_{k;q}) + 2\tilde{R}_{qr}(\omega) + \frac{1}{\tilde{S}_{rr}(\omega)}\mathbb{E}(\bar{x}^2_{k;r})\right\} \\
&= \frac{\sum_{k=1}^{m_T} \mathbb{E}\{\bar{d}_{T;k;q}(\omega)\bar{d}_{T;k;q}(\omega)\}}{\frac{1}{m_T}\sum_{k=1}^{m_T} \mathbb{E}\{\bar{d}_{T;k;q}(\omega)\bar{d}_{T;k;q}(\omega)\}} + 2m_T\tilde{R}_{qr}(\omega) + \frac{\sum_{k=1}^{m_T} \mathbb{E}\{\bar{d}_{T;k;r}(\omega)\bar{d}_{T;k;r}(\omega)\}}{\frac{1}{m_T}\sum_{k=1}^{m_T} \mathbb{E}\{\bar{d}_{T;k;r}(\omega)\bar{d}_{T;k;r}(\omega)\}} \\
&= 2m_T\{1 + \tilde{R}_{qr}(\omega)\}
\end{aligned}
$$

A similar argument yields $\mathbb{E}\left(\sum_{k=1}^{m_T} v^2_{k;qr}\right) = 2m_T\{1 - \tilde{R}_{qr}(\omega)\}$. Thus, $\mathbb{E}(\ddot{u}_{k;qr} = 0)$ and $\mathbb{E}(\ddot{v}_{k;qr}) = 0$. $\square$

Remark 3.7.6. *To bound the probabilities* (3.7.3) *and* (3.7.4), *we will use an argument from Doukhan and Neumann (2007) under an assumption of weak dependence (Doukhan and Louhichi, 1999) of our mean-corrected Fourier coefficients. More specifically, we will make use of the following Bernstein-type inequality which holds for sums of weakly dependent random variables.*

Theorem 3.7.7 (Theorem 1, Doukhan and Neumann (2007))**.** *Assume that* $X_1, \ldots, X_n$ *are zero mean random variables which fulfil appropriate weak dependence conditions. Then, for all* $\delta > 0$, *there exist constants* $K, M, L_1, L_2 < \infty, \mu, \nu \geq 0$ *such that,*

$$
pr\left(\sum_{i=1}^{n} X_i \geq \delta\right) \leq \exp\left\{-\frac{\delta^2/2}{A_n + B_n^{1/(\mu+\nu+2)}\delta^{(2\mu+2\nu+3)/(\mu+\nu+2)}}\right\},
$$

where A_n *can be chosen as any number greater than or equal to* $\sigma_n^2 = \mathrm{var}(\sum_{i=1}^n X_i)$ *and*

$$B_n = 2\max(K, M) L_2 \max\left\{ \frac{2^{4+\mu+\nu} n K^2 L_1}{A_n}, 1 \right\}.$$

The proof will proceed by showing that under Assumption 3.2.5, the variates $(\ddot{u}_{k;qr})_{k=1}^{m_T}$ and $(\ddot{v}_{k;qr})_{k=1}^{m_T}$ are weakly dependent. Then, we will utilise Theorem 3.7.7 to derive the the stated deviation bound.

The following Lemma and subsequent corollary are required to show that, under Assumption 3.2.5, the $(\ddot{u}_{k;qr})_{k=1}^{m_T}$ variates are weakly dependent.

Lemma 3.7.8 (Bardet et al. (2008))**.** *Let* $(X_i)_{i\in\mathbb{Z}}$ *be an* $\mathbb{R}^p$*-valued and* $\mathbb{L}^n$*-stationary time series with* $n > 1, h : \mathbb{R}^p \to \mathbb{R}$ *be a function such that* $|h(x)| \leq c\|x\|^a$ *and* $|h(x) - h(y)| \leq c\|x - y\|(\|x\|^{a-1} + \|y\|^{a-1})$ *for* $(x, y) \in (\mathbb{R}^p)^2, \|\cdot\|$ *a norm on* $\mathbb{R}^p$*, with* $0 < c$ *and* $1 \leq a < n$*. Let* $(Y_i)_{i\in\mathbb{Z}}$ *be the stationary time series defined by* $Y_i = h(X_i)$ *for* $i \in \mathbb{Z}$*. Then, if* (X_i) *is an* $\eta(\tau)$*-weakly dependent time series, then* $(Y_i)_{i\in\mathbb{Z}}$ *is a stationary* $\eta(\tau)^y$*-weakly dependent time series, such that* $\forall \tau \in \mathbb{N}$

$$\eta(\tau)^y = C \cdot \eta(\tau)^{\frac{n-a}{n-1}},$$

with a constant $C > 0$*.*

Corollary 3.7.9. *Let* $h_{qr}(x) = (c_1 x_q + c_2 x_r)^2$*, for* $x \in \mathbb{R}^p$*. If* $(x_k)_{k=1}^{m_T}$ *is* $\eta(\tau)$*-weakly dependent, then* $\{(c_1 x_{k;q} + c_2 x_{k;r})^2\}_{k=1}^{m_T}$ *is* $C \cdot \eta(\tau)^{\frac{n-2}{n-1}}$*-weakly dependent.*

Proof. We find that $|h_{qr}(x)| = |(c_1 x_q + c_2 x_r)^2| \leq 4\max^2(c_1, c_2)\|x\|_\infty$ where $\|x\|_\infty =$

$\max_{q\in\{1,\ldots,p\}} |x_q|$. Now, letting $\Delta = x - y$, we find

$$\begin{aligned}
|h_{qr}(x) - h_{qr}(x+\Delta)| &= |(c_1 x_q + c_2 x_r)^2 - (c_1 x_q + \Delta_q + c_2 x_r + \Delta_r)^2| \\
&= |c_1^2 x_q^2 + 2c_1 c_2 x_q x_r + c_2^2 x_r^2 - (c_1 x_q + \Delta_q + c_2 x_r + \Delta_r)^2| \\
&= |-2c_1 x_q \Delta_q - 2\Delta_q \Delta_r - 2c_2 \Delta_r x_r + \Delta_r^2 - \Delta_q^2 - 2c_1 \Delta_q x_r - 2c_2 \Delta_r x_q| \\
&= |-\Delta_q(2c_1 x_q + 2\Delta_r + \Delta_q + 2c_2 x_r) - \Delta_r(2c_2 x_r + \Delta_r + 2c_1 x_q)| \\
&\leq \|\Delta\|_\infty (|4c_1 x_q + 4c_2 x_r + 3\Delta_r + \Delta_q|) \\
&\leq 4\max(c_1, c2) \cdot \|\Delta\|(\|x\|_\infty + \|x+\Delta\|_\infty)
\end{aligned}$$

where the last line holds since $\|x\|_\infty + \|x+\Delta\|_\infty \geq \|\Delta\|_\infty$ via the reverse triangle inequality. The result is then immediate under Lemma 3.7.8 with a choice of $a = 2$ and $c = 4\max^2(c_1, c_2)$. □

Lemma 3.7.10. *Under Assumption 3.2.5, the variates* $(\ddot{u}_{k;qr})_{k=1}^{m_T}$ *are weakly dependent across experimental trials* $k = 1, \ldots, m_T$.

Proof. Under Assumption 3.2.5, we know that $(\bar{x}_k)_{k=1}^{m_T}$ are η-weakly dependent. Thus, using corollary 3.7.9, we have that $(u_{k;qr}^2)_{k=1}^{m^T}$ is $\eta(\tau)^{\frac{n-2}{n-1}}$-weakly dependent for all $q, r = 1, \ldots, p$.

Moreover, the variates $(\ddot{u}_{k;qr})_{k=1}^{m_T} = (u_{k;qr}^2 - c_{qr})_{k=1}^{m_T}$, where $c_{qr} = 2m_T\{1 + \tilde{R}_{qr}(\omega)\}$ are also $\eta(\tau)^{\frac{n-2}{n-1}}$ weakly dependent since the Lipschitz constant of a function is invariant under translations of its input. That is, if $f : \mathbb{R}^{ud} \to \mathbb{R}$ and $f_c(u_{s1;qr}^2, \ldots, u_{s_u;qr}^2) = f(u_{s1;qr}^2 - c, \ldots, u_{s_u;qr}^2)$, then $\operatorname{Lip} f_c = \operatorname{Lip} f$. □

Therefore, from Lemma 3.7.5 and Lemma 3.7.10 we have that the variates $(\ddot{u}_{k;qr})_{k=1}^{m}$ are $\eta(\tau)^{\frac{n-2}{n-1}}$-weakly dependent random variables with zero mean. Thus, to bound the probability in (3.7.3) we can directly apply Theorem 1 in Doukhan and Neumann (2007) which yields

$$\begin{aligned}
&\operatorname{pr}\left[\left|\sum_{k=1}^{m_T} \ddot{u}_{k;qr}\right| \geq \frac{m_T\varphi\delta}{\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}}\right] \\
&\leq 2\exp\left[-\frac{m_T^2(\varphi\delta)^2}{4\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\left\{A_{m_T} + B_{m_T}^{\{1/(\mu+\nu+2)\}}\left[\frac{m_T(\varphi\delta)}{\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}}\right]^{(2\mu+2\nu+3)/(\mu+\nu+2)}\right\}}\right].
\end{aligned}$$

Now, letting $A_m = c_1' m_T$ and $B_m = c_2'$ yields

$$\begin{aligned}
&\operatorname{pr}\left[\left|\sum_{k=1}^{m_T} \ddot{u}_{k;qr}\right| \geq \frac{m_T\varphi\delta}{\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}}\right] \\
&\leq 2\exp\left[-\frac{m_T^2(\varphi\delta)^2}{4m_T\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\left\{c_1 + c_2^{\{1/(\mu+\nu+2)\}}\left[\frac{\varphi\delta}{\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}}\right]^{(2\mu+2\nu+3)/(\mu+\nu+2)}\right\}}\right] \\
&\leq 2\exp\left(-\frac{-m_T(\varphi\delta)^2}{4\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\left[c_1 + c_2^{\{1/(\mu+\nu+2)\}}\right]}\right)
\end{aligned}$$

for $\varphi\delta/\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2} \leq 1$ and $m_T \geq 1$. Finally, we have

$$\operatorname{pr}\left[\left|\sum_{k=1}^{m_T} \ddot{u}_{k;qr}\right| \geq \frac{m_T\varphi\delta}{\{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}}\right] \leq 2\exp\left[-\frac{m_T(\varphi\delta)^2}{c_3\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)}\right]$$

where $c_3 = \left(4[c_1' + (c_2')^{\{1/(\mu+\nu+2)\}}]\right)^{-1}$ with $c_1' \geq \max\{\sigma_m^2/m, 2^{4+\mu+\nu}K^2L_1\}$ and $c_2' = 2L_2\max(K, M)$. A similar argument will yield the same tail bound for (3.7.4). Consequently, using Lemmas 3.7.1 and 3.7.3, we conclude that for some constant $c_3 > 0$,

$$\operatorname{pr}(\mathcal{B}_{m_T;qr}) \leq 8\exp\left[-\frac{m_T(\varphi\delta)^2}{c_3\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)}\right]$$

valid for $\varphi\delta \leq \{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}$ and $T < \infty$. Lastly, as per Lemma 3.2.7, the bias can be controlled such that $\delta_b \leq (1-\varphi)\delta$ occurs for sufficiently large (but finite) T. This is relevant for the condition on δ to hold and the error bounds in Propositions 3.3.3 and

3.3.4.

3.7.5 Proof of Proposition 3.3.3

Proof. In the proof, we will need the following lemma which is a direct consequence of the deviation bound given in Proposition 3.2.9(b).

Lemma 3.7.11. *For any $q, r = 1, \dots, p$ and $\tau > 2$ there exists a frequency specific constant c such that*

$$pr\left(\max_{qr} |\hat{S}_{qr}(\omega) - S_{qr}(\omega)| \geq c\sqrt{\frac{\log p^{\tau}}{m}}\right) \leq \frac{8}{p^{\tau-2}} \to 0 \text{ as } p \to \infty.$$

Proof. Recall the tail bound given in Proposition 3.2.9. Applying a union bound argument yields

$$\text{pr}\left\{\max_{qr} |\hat{S}_{qr}(\omega) - S_{qr}(\omega)| \geq c\delta\right\} \leq 8p^2 \exp\left\{\frac{-m_T c^2(\varphi\delta)^2}{c_2 \max_q \{\tilde{S}_{qq}(\omega)\}^2}\right\},$$

where we define

$$c = \left[\frac{c_2 \max_q \{\tilde{S}_{qq}(\omega)\}^2}{\varphi^2}\right]^{1/2}.$$

Hence,

$$\begin{aligned}\text{pr}\left\{\max_{qr} |\hat{S}_{qr}(\omega) - S_{qr}(\omega)| \geq c\delta\right\} &\leq 8p^2 \exp\{-\delta^2 m_T\} \\ &= 8\exp\{\log p^2 - \delta^2 m_T\}.\end{aligned}$$

Letting $\delta = \sqrt{\frac{\log p^{\tau}}{m_T}}$ yields

$$\begin{aligned}
\mathrm{pr}\left\{\max_{qr} |\hat{S}_{qr}(\omega) - S_{qr}(\omega)| \geq c\delta\right\} &\leq 8\exp\{\log p^2 - \tau \log p\} \\
&= 8\exp\{\log p^2 - \log p^{\tau-2} - \log p^2\} \\
&= 8\exp\{-\log p^{\tau-2}\} \\
&= \frac{8}{p^{\tau-2}} \to 0 \text{ as } p \to \infty.
\end{aligned}$$

□

We begin the proof of Proposition 3.3.4 by defining the event

$$\mathcal{E} := \left\{\|\hat{S}_{qr}(\omega) - S_{qr}(\omega)\|_{\infty} \leq c\sqrt{\frac{\log p^{\tau}}{m_T}}\right\}. \tag{3.7.5}$$

We have shown in Lemma 3.7.11 that $\mathrm{pr}(\mathcal{E}^c) \leq 8/p^{\tau-2}$, and therefore $\mathrm{pr}(\mathcal{E}) \geq 1-8/p^{\tau-2}$. We condition on the event $\mathcal{E}$ in the following analysis.

Going forward, we acknowledge a change in notation for the true inverse spectral density matrix, which we will now denote by Θ_0 rather than Θ. We will also drop the dependence on ω for ease of notation.

Consider the function

$$\begin{aligned}
F(\hat{\Theta}) :&= \mathcal{L}(\hat{\Theta}) - \mathcal{L}(\Theta_0) + \lambda(P_2\left\{\hat{\Theta}\right\} - P_2\left\{\Theta_0\right\}) \\
&= \underbrace{\mathrm{tr}\{(\hat{\Theta} - \Theta_0)(\hat{S} - S)\}}_{(i)} - \underbrace{(\log\det(\hat{\Theta}) - \log\det(\Theta_0))}_{(ii)} \\
&\quad + \underbrace{\mathrm{tr}\{(\hat{\Theta} - \Theta_0)S\}}_{(iii)} + \underbrace{\lambda(\mathrm{tr}\left\{\hat{\Theta}\right\} - \mathrm{tr}\left\{\Theta_0\right\})}_{(iv)}
\end{aligned} \tag{3.7.6}$$

where $\Theta_0 = S(\omega)^{-1}$.

Let $\hat{\Theta} = \Theta_0 + \Delta$. Our aim is to bound the size of Δ such that $F(\hat{\Theta}) \leq F(\Theta_0) = 0$. Since the ridge estimator $\hat{\Theta}$ is the solution to a minimisation problem, we know that it will always attain a lower loss than the true value Θ_0.

The main idea of the proof is based on an adaptation of the argument presented in Rothman et al. (2008) where one can consider the shell of errors

$$\mathcal{B}_m(G) = \{\Delta \mid \Delta = \Delta^H,\ \|\Delta\|_F = Gr_m\}, \tag{3.7.7}$$

i.e. all possible symmetric errors of a given radius e.g. $Gr_m, G > 0$. We can then adjust the radius r_m such that

$$F(\Theta_0 + \Delta) > 0,$$

and when this occurs we know that our shell is too big, and therefore the error must lie within $\mathcal{B}_m(G)$ and thus $\|\Delta\|_F \leq Gr_m$.

We now consider each term in (3.7.6) separately letting $\Delta = \hat{\Theta} - \Theta_0$. We start by simplifying term (ii) in (3.7.6).

Lemma 3.7.12. *Taking the first order Taylor expansion of* $f(t) = \log\det(\Theta_0 + t\Delta)$ *with the integral form of the remainder yields*

$$\begin{aligned}
&\log\det(\Theta_0 + \Delta) - \log\det(\Theta_0) = \\
&\qquad \operatorname{tr}\{S\Delta\} - \vec{\Delta}^H \left[\int_0^1 (1-v)((\Theta_0 + t\Delta)^{-1})^H \otimes (\Theta_0 + t\Delta)^{-1})dv\right] \vec{\Delta}
\end{aligned}$$

where $\otimes$ *denotes the Kronecker product.*

Proof. The proof of Lemma 3.7.12 begins by computing the derivatives

$$\begin{aligned}
\frac{\partial f(t)}{\partial t} &= \frac{1}{\det(\Theta_0 + t\Delta)} \left\{\frac{\partial}{\partial t} \det(\Theta_0 + t\Delta)\right\} \\
&= \frac{1}{\det(\Theta_0 + t\Delta)} \det(\Theta_0 + t\Delta) \operatorname{tr}\left\{(\Theta_0 + t\Delta)^{-1} \frac{\partial}{\partial t}(\Theta_0 + t\Delta)\right\} \quad \text{(Jacobi's formula)} \\
&= \operatorname{tr}\{(\Theta_0 + t\Delta)^{-1}\Delta\}.
\end{aligned}$$

Computing the second derivative

$$\frac{\partial^2 f(t)}{\partial t^2} = \frac{\partial}{\partial t}\left(\operatorname{tr}\{(\Theta_0 + t\Delta)^{-1}\Delta\}\right) = \frac{\partial}{\partial t}\underbrace{\operatorname{vec}(\Delta^H)}_{f_1(t)}\underbrace{\operatorname{vec}((\Theta_0 + t\Delta)^{-1})}_{f_2(t)}$$

where by the product rule

$$\frac{\partial^2 f(t)}{\partial t^2} = f_2(t)\frac{\partial f_1(t)}{\partial t} + f_1(t)\frac{\partial f_2(t)}{\partial t}.$$

Since $\frac{\partial f_1(t)}{\partial t} = 0$ we proceed by considering

$$\begin{aligned}\frac{\partial f_2(t)}{\partial t} &= \frac{\partial}{\partial t}\operatorname{vec}((\Theta_0 + t\Delta)^{-1})\\ &= \operatorname{vec}\left(\frac{\partial}{\partial t}(\Theta_0 + t\Delta)^{-1}\right)\\ &= \operatorname{vec}(-(\Theta_0 + t\Delta)^{-1}\Delta(\Theta_0 + t\Delta)^{-1})\\ &= -((\Theta_0 + t\Delta)^{-1})^H \otimes (\Theta_0 + t\Delta)^{-1}\operatorname{vec}(\Delta)\end{aligned}$$

Hence,

$$\frac{\partial^2 f(t)}{\partial t^2} = -\vec{\Delta}^H[((\Theta_0 + t\Delta)^{-1})^H \otimes (\Theta_0 + t\Delta)^{-1}]\vec{\Delta}.$$

The first order Taylor expansion with integral remainder term is defined as

$$f(t) = f(a) + \frac{\partial f(a)}{\partial t}(t-a) + \int_a^t \frac{\partial^2 f(a)}{\partial t}(t-v)dv.$$

Applying the above Taylor expansion with $t = 1, a = 0$ and re-arranging yields,

$$\begin{aligned}&\log\det(\Theta_0 + \Delta) - \log\det(\Theta_0) =\\ &\qquad \operatorname{tr}\{S\Delta\} - \vec{\Delta}^H\left[\int_0^1 (1-v)((\Theta_0 + v\Delta)^{-1})^H \otimes (\Theta_0 + v\Delta)^{-1})dv\right]\vec{\Delta}\end{aligned}$$

$\square$

Therefore, we can write (3.7.6) as

$$F(\Theta_0 + \Delta) = \underbrace{\operatorname{tr}\{\Delta(\hat{S} - S)\}}_{(i)} + \underbrace{\vec{\Delta}^H \left[\int_0^1 (1-v)((\Theta_0 + v\Delta)^{-1})^H \otimes (\Theta_0 + v\Delta)^{-1}) dv\right] \vec{\Delta}}_{(ii)}$$

$$+ \underbrace{\lambda(\operatorname{tr}\left\{\hat{\Theta}\right\} - \operatorname{tr}\left\{\Theta_0\right\})}_{(iii)}. \tag{3.7.8}$$

Noting that

$$(iii) = \lambda(\operatorname{tr}\{\Theta_0 + \Delta\} - \operatorname{tr}\{\Theta_0\}) = \lambda(\operatorname{tr}\{\Delta\}),$$

we can rewrite (3.7.8) as

$$F(\Theta_0 + \Delta) = \underbrace{\operatorname{tr}\{\Delta(\hat{S} - S + \lambda\mathbb{I})\}}_{(i')} + \underbrace{\vec{\Delta}^H \left[\int_0^1 (1-v)((\Theta_0 + v\Delta)^{-1})^H \otimes (\Theta_0 + v\Delta)^{-1}) dv\right] \vec{\Delta},}_{(ii)} \tag{3.7.9}$$

and proceed by bounding each term separately.

We begin by writing

$$\begin{aligned}(i') \le |\operatorname{tr}\{\Delta(\hat{S} - S + \lambda\mathbb{I})\}| &\le \left|\operatorname{tr}\{\Delta(\hat{S} - S)\}\right| + |\lambda \operatorname{tr}\{\Delta\mathbb{I}\}|. \\ &\le \|\hat{S} - S\|_F \|\Delta\|_F + \lambda\|\Delta\|_F \|\mathbb{I}\|_F \\ &\le p\|\hat{S} - S\|_\infty \|\Delta\|_F + \lambda\sqrt{p}\|\Delta\|_F \\ &\le \left(c\sqrt{\frac{p^2 \log p^\tau}{m_T}} + \lambda\sqrt{p}\right)\|\Delta\|_F.\end{aligned}$$

To bound the integral term (ii) in (3.7.9), we recall that for a symmetric matrix M we have that the minimum eigenvalue is $\phi_{\min}(M) = \min_{\|x\|=1} x^H M x$. Hence, after factoring out the norm of $\vec{\Delta}$, we have for $\Delta \in \mathcal{B}_m(G)$

$$\begin{aligned}
&\phi_{\min}\left(\int_0^1 (1-v)((\Theta_0 + v\Delta)^{-1})^H \otimes (\Theta_0 + v\Delta)^{-1})dv\right) \\
&\quad \geq \int_0^1 (1-v)\phi_{min}^2(\Theta_0 + v\Delta)^{-1}dv \\
&\quad \geq \frac{1}{2}\min_{0\leq v\leq 1} \phi_{\min}^2(\Theta_0 + v\Delta)^{-1} \\
&\quad \geq \frac{1}{2}\min\{\phi_{\min}^2(\Theta_0 + \Delta)^{-1} : \|\Delta\|_F \leq Gr_m\}.
\end{aligned}$$

The first inequality uses the property that eigenvalues of Kronecker products of Hermitian matrices are equal to the products of the eigenvalues of their factors, i.e., $\text{eig}(A \otimes B) = eig(A)eig(B)$.

Now,

$$\phi_{\min}^2(\Theta_0 + \Delta)^{-1} = \phi_{\max}^{-2}(\Theta_0 + \Delta) = \frac{1}{\|\Theta_0 + \Delta\|_2^2} \geq \frac{1}{\|\Theta_0\|_2^2 + \|\Delta\|_2^2}$$

for all $\|\Delta\|_F \leq Gr_m$. Noting that $\|\Delta\|_F \leq Gr_m \implies \|\Delta\|_2 \leq Gr_m$ we have that

$$\begin{aligned}
\frac{1}{\|\Theta_0\|_2^2 + \|\Delta\|_2^2} &\geq \frac{1}{\|\Theta_0\|_2^2 + G^2r_m^2} \\
&\geq \frac{1}{2\|\Theta_0\|_2^2} = \frac{1}{2}\underline{k}^2.
\end{aligned}$$

Thus,

$$\phi_{\min}\left(\int_0^1 (1-v)((\Theta_0 + v\Delta)^{-1})^H \otimes (\Theta_0 + v\Delta)^{-1})dv\right) \geq \frac{1}{4}\underline{k}^2, \tag{3.7.10}$$

and we have the lower bound

$$F(\Theta_0 + \Delta) \geq \frac{1}{4}\underline{k}^2\|\Delta\|_F^2 - 2c\sqrt{\frac{p^2\log p^\tau}{m_T}}\|\Delta\|_F,$$

where we have substituted $\lambda = c(p\log p^\tau/m_T)^{1/2}$.

Solving this quadratic in $\|\Delta\|_F$ yields one positive root

$$\|\Delta\|_F = \frac{8c}{\underline{k}^2}\sqrt{\frac{p^2 \log p^\tau}{m_T}},$$

which equating with Gr_m gives the stated error bound. Importantly, (ii) in (3.7.11) was bounded according to (3.7.10) under the assumption that we have sufficient trials

$$m_T \geq 16\tau \underline{k}^{-2}(2c)^2 p^2 \log p$$

to ensure that $Gr_m \leq \|\Theta_0\|_2 = \underline{k}^{-1}$.

Since the above analysis was conditioned on the event $\mathcal{E}$, Proposition 3.3.3 holds with probability $1 - \frac{8}{p^{\tau-2}} \to 1$, for any $\tau > 2$ and sufficient trials as stated. □

3.7.6 Proof of Proposition 3.3.4

Proof. In the proof, we adopt the notation $M^+ = \mathrm{diag}(M)$ for a diagonal matrix with the same diagonal as matrix M, and $M^- = M - M^+$.

In the same manner as the proof of Proposition 3.3.3, we condition on the event $\mathcal{E}$ and seek to find the value Gr_m which defines the shell of errors (3.7.7). The main difference is that of the regularisation term, therefore (3.7.8) becomes

$$F(\Theta_0 + \Delta) = \underbrace{\mathrm{tr}\{\Delta(\hat{S} - S)\}}_{(i)} + \underbrace{\vec{\Delta}^H \left[\int_0^1 (1-v)((\Theta_0 + v\Delta)^{-1})^H \otimes (\Theta_0 + v\Delta)^{-1}) dv\right] \vec{\Delta}}_{(ii)}$$
$$+ \underbrace{\lambda(\|(\Theta_0 + \Delta)\|_1 - \|\Theta_0\|_1)}_{(iv)} \tag{3.7.11}$$

where $\Theta_0 = S^{-1}(\omega)$.

We consider bounding each term in (3.7.11) separately, and begin by writing

$$(i) \leq |\operatorname{tr}\{\Delta(\hat{S}-S)\}| \leq \underbrace{\left|\sum_{q\neq r}(\hat{S}_{qr}-S_{qr})\Delta_{qr}\right|}_{\mathrm{I}} + \underbrace{\left|\sum_{q}(\hat{S}_{qq}-S_{qq})\Delta_{qq}\right|}_{\mathrm{II}} = \mathrm{I}+\mathrm{II}.$$

To bound I, we appeal to our earlier result in Lemma 3.7.11 and argue that with probability tending to 1,

$$\max_{q\neq r}|\hat{S}_{qr}-S_{qr}| \leq c\sqrt{\frac{\log p^{\tau}}{m}}$$

therefore term I is bounded by

$$\mathrm{I} \leq c\sqrt{\frac{\log p^{\tau}}{m}}\|\Delta^{-}\|_1.$$

To bound the second term, we consider an application of the Cauchy-Schwartz inequality combined with our result from Lemma 3.7.11, namely,

$$\begin{aligned}
\mathrm{II} \leq \|\hat{S}_{qq}-S_{qq}\|_F \ \|\Delta^{+}\|_F &= \sqrt{\sum_{q=1}^{p}(\hat{S}_{qq}-S_{qq})^2}\ \|\Delta^{+}\|_F \\
&\leq \sqrt{p}\|\hat{S}_{qq}-S_{qq}\|_\infty \ \|\Delta^{+}\|_F \\
&\leq c\sqrt{\frac{p\log p^{\tau}}{m}}\|\Delta^{+}\|_F \\
&\leq c\sqrt{\frac{(p+s)\log p^{\tau}}{m}}\|\Delta^{+}\|_F.
\end{aligned}$$

To bound the penalty term (iii) we apply the decomposability argument of Negahban et al. (2012), arguing that the norm is linearly decomposable over the on and off support of the true model.

Importantly, we note the applicability of this argument to the group regulariser $\|\Theta\|_1$ defined for a complex valued matrix $\Theta \in \mathbb{C}^{p\times p}$ since the group norm is specified at the level of elements in $\mathcal{M}$.

Hence, by this decomposability argument we have that

$$\begin{aligned}
&\lambda(\|\Theta_0+\Delta\|_1-\|\Theta_0\|_1)\\
=&\lambda(\|\Theta^-_{0\mathcal{M}}+\Theta^-_{0\mathcal{M}^\perp}+\Theta^+_{0\mathcal{M}}+\Delta^-_{\mathcal{M}}+\Delta^-_{\mathcal{M}^\perp}+\Delta^+_{\mathcal{M}}\|_1-\|\Theta^-_{0\mathcal{M}}+\Theta^-_{0\mathcal{M}^\perp}+\Theta^+_{0\mathcal{M}}\|_1)\\
=&\lambda(\underbrace{\|\Theta^-_{0\mathcal{M}}+\Delta^-_{\mathcal{M}}\|_1-\|\Theta^-_{0\mathcal{M}}\|_1}_{(a)}+\underbrace{\|\Theta^-_{0\mathcal{M}^\perp}+\Delta^-_{\mathcal{M}^\perp}\|_1-\|\Theta^-_{0\mathcal{M}^\perp}\|_1}_{(b)}+\underbrace{\|\Theta^+_{0\mathcal{M}}+\Delta^+_{\mathcal{M}}\|_1-\|\Theta^+_{0\mathcal{M}}\|_1)}_{(c)}.
\end{aligned}$$

Bounding term separately, we use the reverse triangle inequality to write

$$\begin{aligned}
(a)=\|\Theta^-_{0\mathcal{M}}-(-\Delta^-_{\mathcal{M}})\|_1-\|\Theta^-_{0\mathcal{M}}\|_1 &\geq \big|\|\Theta^-_{0\mathcal{M}}\|_1-\|-\Delta^-_{\mathcal{M}}\|_1\big|-\|\Theta^-_{0\mathcal{M}}\|_1\\
&\geq \|\Theta^-_{0\mathcal{M}}\|_1-\|\Delta^-_{\mathcal{M}}\|_1-\|\Theta^-_{0\mathcal{M}}\|_1=-\|\Delta^-_{\mathcal{M}}\|_1.
\end{aligned}$$

A similar argument yields $(c)\geq-\|\Delta^+_{0\mathcal{M}}\|_1$. Finally, under the true model $\|\Theta^-_{0\mathcal{M}^\perp}\|_1=0$. Hence, $(b)=\|\Delta^-_{\mathcal{M}^\perp}\|_1$.

The above implies that

$$\lambda(\|\Theta_0+\Delta\|_1-\|\Theta_0\|_1)\geq\lambda(\|\Delta^-_{\mathcal{M}^\perp}\|_1-\|\Delta^-_{\mathcal{M}}\|_1-\|\Delta^+_{\mathcal{M}}\|_1).$$

and the lower bound

$$\begin{aligned}
F(\Theta_0+\Delta)\geq\frac{1}{4}\underline{k}^2\|\Delta\|_F^2-c\sqrt{\frac{\tau\log p}{m}}\|\Delta^-\|_1-c\sqrt{\frac{(p+s)\tau\log p}{m}}\|\Delta^+\|_F\\
+\lambda(\|\Delta^-_{\mathcal{M}^\perp}\|_1-\|\Delta^-_{\mathcal{M}}\|_1-\|\Delta^+_{\mathcal{M}}\|_1).
\end{aligned}$$

Now, substituting $\lambda=c\sqrt{\frac{\tau\log p}{m}}$ yields

$$F(\Theta_0+\Delta)\geq\frac{1}{4}\underline{k}^2\|\Delta\|_F^2-2c\sqrt{\frac{\tau\log p}{m}}\|\Delta^-_{\mathcal{M}}\|_1-c\sqrt{\frac{\tau\log p}{m}}\|\Delta^+_{\mathcal{M}}\|_1-c\sqrt{\frac{(p+s)\log p}{m}}\|\Delta^+\|_F.$$

Now,

$$\|\Delta^-_{\mathcal{M}}\|_1 \le \sqrt{s}\|\Delta^-_{\mathcal{M}}\|_F \le \sqrt{s}\|\Delta^-\|_F \le \sqrt{p+s}\|\Delta^-\|_F,$$

and similarly, $\|\Delta^+_{\mathcal{M}}\|_1 \le \sqrt{p+s}\|\Delta^+\|_F$. Thus, we have

$$\begin{aligned} F(\Theta_0+\Delta) \ge \|\Delta^-\|_F^2 &\left[\frac{1}{4}\underline{k}^2 - 2c\sqrt{\frac{(p+s)\log p^\tau}{m}}\|\Delta^-\|_F^{-1}\right] \\ &+ \|\Delta^+\|_F^2\left[\frac{1}{4}\underline{k}^2 - 2c\sqrt{\frac{(p+s)\log p^\tau}{m}}\|\Delta^+\|_F^{-1}\right] \end{aligned}$$

where using $\|\Delta\|_F = Gr_m$ and letting $r_m = \sqrt{\frac{(p+s)\log p^\tau}{m}}$ we have

$$F(\Theta_0+\Delta) \ge \|\Delta\|_F^2\left[\frac{1}{4}\underline{k}^2 - \frac{2c}{G}\right] > 0$$

for G sufficiently large i.e.,

$$G \ge \frac{8c}{\underline{k}^2}.$$

Hence,

$$\|\Delta\|_F \le Gr_m = \frac{8c}{\underline{k}^2}\sqrt{\frac{(p+s)\log p^\tau}{m}} \tag{3.7.12}$$

In the above, we note that (ii) in (3.7.11) was bounded according to (3.7.10) under the assumption that we have sufficient trials

$$m > \frac{8^2c^2}{\underline{k}^2}(p+s)\log p^\tau,$$

which ensures that $Gr_m \le \underline{k}^{-1}$. Finally, as before, the above analysis was conditioned on the event $\mathcal{E}$, statement (3.7.12) holds with probability $1-\frac{8}{p^{\tau-2}} \to 1$, for any $\tau > 2$.

□

Chapter 4

Identifying Neural Connectivity Using Bernoulli Autoregressive Partially Linear Additive Models

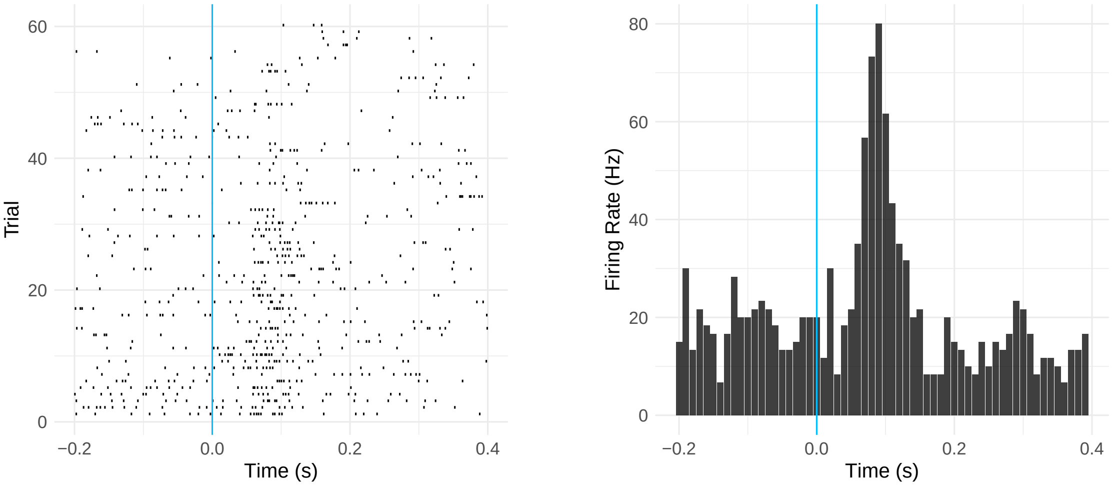
4.1 Introduction

Characterising the synchronisation of spiking neurons from recorded signals is a challenging and active research problem in neuroscience (Keeley et al., 2020; Ren et al., 2020; Vareberg et al., 2024). In the past decade, there have been significant advances in neural recording technologies, allowing for the exciting opportunity to study large scale neural activity in the living brain. For example, the first generation of Neuropixels probes enabled the simultaneous recording of hundreds of neurons distributed across the mouse brain, during a single experimental trial (Jun et al., 2017). More recent advances include the development of the Neuropixels 2.0 and Neuropixels Ultra probes, which allow for recordings from even larger neuronal populations in small brains, during behaviour and free movement (Steinmetz et al., 2021; Ye et al., 2024). The use of high-throughput neural probes result in recordings of noisy, high-dimensional data, which present interesting challenges for both neuroscientists and statisticians. In this paper, we seek to understand how populations of neurons interact, and present a statistical methodology which can be used to estimate networks of neural interactions. The development of such a tool, which is capable of extracting accurate and interpretable descriptions of interactions from multi-region neural recordings, is imperative to our continued understanding of the brain's processes.

In neuroscience, a common and widely accepted measure of functional interactions between neurons is that of cross-correlation (Narayanan and Laubach, 2009). Traditionally, neuroscientists have studied neural correlations using a variant of the cross-correlation histogram (Perkel et al., 1967) including, but not limited to, the joint peri-stimulus time histogram (JPSTH) (Gerstein and Perkel, 1972), the snowflake plot (Perkel et al., 1975; Czanner et al., 2005) or the shuffle-corrected cross-correlogram (Aertsen et al., 1989; Brody, 1999). These methods, which are readily available in popular neuroscience data analysis packages such as Pynapple and SpikeInterface, are regularly used for the analysis of electrophysiological data. However, the above approaches are limited in the sense that they only study a handful of neurons at a time, and are therefore ignore likely contributions from other neurons. Consequently, detected correlations could be ambiguous as a result of the possible, and yet excluded, influence of other neurons (Zhao et al., 2012).

Generalised linear models (GLMs) are also routinely used by neuroscientists for the analysis of simultaneously recorded neural data. Popularised by Brillinger (1988), who studied small networks of 3 neurons, GLMs offer a highly interpretable framework for the analysis of firing rates, and have also been used for a variety of other problems in computational neuroscience including encoding (Paninski, 2004; Truccolo et al., 2005), decoding (Gao et al., 2003) and for the assessment of neural interactions (Zhao et al., 2012).

In this paper, we extend the popular and widely used GLM framework in order



(a) (b)

Figure 4.1.1: Example neuron from the spike train dataset of Steinmetz et al. (2019): (a) Raster plot showing single neuron response to 60 repeats of a visual stimulus and (b) peri-stimulus time histogram (PSTH) for the same data.

to account for non-stationary dynamics that are often observed in stimulus-response type neuroscience experiments. Take, as example, the spike train data from Steinmetz et al. (2019) shown in Figure 4.1.1. In this experiment, neural activity was recorded over a series of trials, while mice performed a visual discrimination task (see section 4.5 for full details). Figure 4.1.1 shows a raster plot and Peristimulus time histogram (PSTH) aligned to the onset of a visual stimulus, at time $t = 0$, for an example neuron in the dataset. Each row in the raster plot is representative of a single spike train, which represents the neural response, or more specifically the firing times of the neuron in a particular trial. By contrast, the PSTH describes the trial-averaged activity of the neuron, by pooling data across trials. It is clear to the see that the firing rate of the neuron increases in response to the visual stimulus, thus motivating the need for methods which account for inherent non-stationarities commonly observed in neural data. It is in these stimulus-response type experiments where the methodology outlined in this paper would be of most use.

Often, neural spike trains are represented in continuous time as a multivariate point process $\mathbf{N}(t) = \{N_i(t)\}_{i\in\{1,\ldots,d\}}$ whose i^{th} component gives the number of spikes observed in neuron i in the time interval $(0,t], t \leq n$. In the statistics literature, the multivariate Hawkes (Hawkes, 1971a) process has received significant attention, and has become a popular tool for the identification and modelling of functional connectivity from recorded neural spike trains (Lambert et al., 2018; Wang and Shojaie, 2021; Wang et al., 2024). Conversely, in the computational neuroscience literature, neural spike trains are often represented in terms of their binned spike count, which refers to a discretised version of the continuous time process (Kass et al., 2011). The so-called binned spike count is constructed by decomposing the time axis into bins of equal width δ, resulting in a discrete integer-valued time series where each value indicates the number of action potentials (spikes) within a specified time bin, i.e., $Y_{t,i} = N_i(t+\delta) - N_i(t), t \leq n$, for each $i = 1,\ldots,d$.

Here, in line with many other related works (Kadirvelu et al., 2017; Zoltowski and Pillow, 2018; Kobayashi et al., 2019; Vareberg et al., 2024) we take the discrete-time approach, and choose δ to be small enough to ensure that at most one spike per neuron can occur within a time bin, e.g., $\delta = 1$ms. Therefore, we consider a binary time series representation of the neural spiking data, which takes value 1 in a time bin where there is a spiking event, and 0 everywhere else. Importantly, it can be shown that under the usual regularity condition, where no two events occur at the same time, that the likelihood function of the binary time series approximates the likelihood of the point process as $\delta \to 0$ (Chen et al., 2010; Kass et al., 2011).

In this work, we build on the regularised logistic model proposed in Zhao et al. (2012) for the detection of neuronal interactions. There, the authors studied a network with $d = 30$ neurons, and constructed a model assuming that neural firing rates remained stationary over time. In our approach, we account for non-stationary firing rates by adding a non-parametric component to the logistic model, and consider larger networks with more neurons. Alongside other groups in statistics and neuroscience (Eden et al., 2012; Zhou et al., 2015; Sarmashghi et al., 2021) we approximate the non-parametric component as a weighted summation of a set of basis functions. Thus, the novelty of our approach lies in incorporating this smooth function within the Bernoulli autoregressive model, to detect neuronal interactions in the presence of non-stationary trends, that are commonly observed in stimulus-response type neuroscience experiments (Fig 4.1.1). Related work on the identification of sparse neural functional connectivity using penalised likelihood estimation and basis functions is discussed in Song et al. (2013). There, the authors formulate a generalised functional additive model (GFAM) which consists of a set of basis functions convolving the input signals, and a link function generating the firing probability of the output neuron from the summation of the convolutions weighted by the sought model coefficients. Comparatively, we use basis functions to model the non-stationary firing rate of individual neurons, and infer a network of interactions using an approach which is more similar in nature to that of Zhao et al. (2012). For further details on these alternative approaches, see Chapter 2.

Admittedly, it is difficult to determine a 'ground truth' for network interactions among populations of real neurons. Therefore, in addition to calibrating our estimator using synthetic data, we also propose an empirically validated inference procedure to quantify the uncertainty associated with our estimated networks. To the best of our knowledge, our approach is the first to: 1) account for non stationary firing rates in the regularised logistic model for the estimation of effective connectivity between neurons, and 2) provide confidence intervals for these estimates.

The rest of this paper is organised as follows. In Section 4.2, we describe the methodology used for the detection of neural interactions, including the required estimation procedure, before giving details of our inference procedure in Section 4.3. In Section 4.4, we conduct a simulation study to demonstrate the satisfactory performance of our estimator. In Section 4.5, we apply our model to electrophysiology data obtained from the DANDI archive, where we study the interactions of neural processes in reaction to the stimulus-response type neuroscience experiments described in Steinmetz et al. (2019). We conclude with a discussion in Section 4.6.

4.2 Bernoulli Autoregressive Partially Linear Additive Models

Consider a multivariate binary time series $\mathbf{Y}_t = (Y_{t,1}, \ldots, Y_{t,d})^\top$ such that the series at time t are conditionally independent of one another given past realisations, i.e.,

$$\Pr(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{<t}) = \prod_{i=1}^{d} \Pr(Y_{t,i} = y_{t,i} | \mathbf{Y}_{<t}),$$

where $\mathbf{Y}_{<t}$ denotes the entire history of the multivariate series observed up to, but not including, time t.

Definition 4.2.1. *(Shojaie and Fox, 2022). Time series Y_j is Granger non-causal for time series Y_i if and only if $\forall$ t,*

$$\Pr(Y_{t,i} = 1 | Y_{<t,1}, \ldots Y_{<t,j}, \ldots, Y_{<t,d}) = \Pr(Y_{t,i} = 1 | Y_{<t,1}, \ldots, Y_{<t,(j-1)}, Y_{<t,(j+1)}, \ldots, Y_{<t,d}).$$

Granger causality has proven to be an effective method for the analysis of neural spike train data, due to its ability to detect directional interactions amongst populations of neurons (Kim et al., 2011; Lintas et al., 2018; Casile et al., 2021). We wish to develop a class of models that allow for easy elicitation of such Granger causal relations.

4.2.1 Model Formulation

Consider a set of spike trains, with non-stationary firing rates, recorded from a population of neurons. Under a binned spike count representation of these data, we propose to use the following Bernoulli autoregressive partially linear additive (BAPLA) model to identify the effective connectivity of this neural population.

Let $\{Y_{t,i}\}_{t=1}^{n} \in \{0,1\}$ be the i^{th} variate of the d-variate random vector $\mathbf{Y}_t$ for $1 \leq i \leq d$, i.e., $Y_{t,i}$ denotes the state of neuron i at time t. Then, the BAPLA model is defined as

$$Y_{t,i}|\mathbf{Y}_{t-1} \sim \text{Bernoulli}\left(\frac{1}{1+\exp\{-\beta_i - \boldsymbol{\gamma}_i^{\top}\mathbf{Y}_{t-1} - f_i(t/n)\}}\right), \tag{4.2.1}$$

where $\mathbf{Y}_{t-1} = (Y_{t-1,1}, Y_{t-1,2}, \ldots, Y_{t-1,d})^{\top}$ denotes the previous values of the multivariate process, β_i is a constant offset parameter, $\boldsymbol{\gamma}_i^{\top} \in \mathbb{R}^{1\times d}$ is an unknown parameter vector, and $f_i(t/n) : [0,1] \to \mathbb{R}$ is an unknown smooth one-dimensional and differentiable function. To ensure identifiability of the intercept term β_i and the smooth function $f_i(t/n)$, we impose the constraint that $\int_0^1 f_i(u)du = 0$ for $i = 1, \ldots, d$.

We are interested in using the BAPLA model to investigate Granger casual interactions in high-dimensional non-stationary multivariate binary time series. In the context of our neuronal analysis, the components of the BAPLA model have the following interpretations: 1) β_i models the base-line firing rate of neuron i; 2) $\boldsymbol{\gamma}_i^{\top}$ provides an insight into the interactions between the spike train recorded from neuron i and the spike trains from the whole neural population; and 3) $f_i(t/n)$ models the non-stationary firing rate of neuron i. Admittedly, our proposed method can be considered as an extension to a particular case of the generalised linear autoregressive (GLAR) model previously studied in Hall et al. (2018). However, a key methodological advancement of the BAPLA model is its ability to capture non-linear dynamics via the time varying function $f_i(t/n)$. The inclusion of such a component is essential in order to account for the non-stationary dynamics commonly observed in real world data (Figure 4.1.1).

The model in (4.2.1) can be written in terms of the conditional distribution of observing a spike at time t

$$\Pr(Y_{t,i} = 1 | \mathbf{Y}_{t-1} = \boldsymbol{y}_{t-1}) = \mathrm{logit}^{-1}\left(\beta_i + \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} + f_i(t/n)\right), \tag{4.2.2}$$

where $\mathrm{logit}^{-1}(a) = 1/1 + \exp(-a)$. Using this representation, equation (4.2.2) can be used to infer a network of interactions $\mathbf{\Gamma} \in \mathbb{R}^{d \times d}$ which is constructed row-wise from $\boldsymbol{\gamma}_i^\top$ (i.e. $\boldsymbol{\gamma}_i^\top$ is the i^{th} row of $\mathbf{\Gamma}$). Therefore, given observed data $\{y_{t,i}\}_{t=1}^n$, for $i = 1, \ldots, d$, one can identify and model connectivity via an estimate of the matrix $\mathbf{\Gamma}$. In particular, by leveraging the component-wise structure of (4.2.2), we can infer that time series Y_j does not Granger cause series Y_i if and only if $\Gamma_{ij} = 0$.

In general, it is reasonable to assume that interactions among a population of real neurons are sparse. For example, a cortical neuron receives between $10^3 - 10^4$ inputs, which is a particularly small fraction of the 10^{10} neurons in the cerebral cortex (Zhao et al., 2012). Consequently, we now introduce the following notation for the true, and assumed sparse, interaction matrix $\mathbf{\Gamma}$. Firstly, we define the edge set

$$E(\mathbf{\Gamma}) := \{(i, j) : \Gamma_{ij} \neq 0, i \neq j\} \tag{4.2.3}$$

and let $\mathrm{card}(E(\mathbf{\Gamma})) \leq s$. That is, the set of off-diagonal entries in $\mathbf{\Gamma}$ which determine the edge set of the graph, is at most s. Going forward, we will refer to the true matrix $\mathbf{\Gamma}$ as being s-sparse, i.e., the corresponding graph has at most s edges. In addition, we denote by $\tilde{d} = \max_{i=1,\ldots,d} |\{j \in \{1, \ldots, d\} : \Gamma_{ij} \neq 0\}|$, the maximum degree of a node, which stipulates the maximum number of non-zero entries in any row of $\mathbf{\Gamma}$. Finally, we will assume that entries in the interaction matrix are bounded, i.e., $\mathbf{\Gamma} \in [\gamma_{\min}, \gamma_{\max}]^{d \times d}$.

It is also necessary to place certain assumptions on the time varying function $f_i(t/n)$. Firstly, we assume that each $f_i(t/n)$ satisfies some smoothness conditions, such as being differentiable up to a certain order. Taking advantage of this smoothness, we approximate each $f_i(t/n)$ as a weighted summation of a set of basis functions

$$f_i(t/n) \approx \sum_{k=1}^{m} c_{ik}\phi_{ik}(t/n) \quad \text{for } t = 1, 2, \ldots, n,$$

where $\phi_{ik}(\cdot)$ and c_{ik} are the basis functions and their coefficients respectively. Consequently, we can write a parametric form of model (4.2.2) as

$$\Pr(Y_{t,i} = 1 | \mathbf{Y}_{t-1} = \boldsymbol{y}_{t-1}) = \text{logit}^{-1}\left(\beta_i + \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} + \sum_{k=1}^{m} c_{ik}\phi_{ik}(t/n)\right).$$

4.2.2 Parameter Estimation and Model Selection

Alongside other groups in statistics and neuroscience (Eden et al., 2012; Zhou et al., 2015; Sarmashghi et al., 2021) we model the non-stationary firing rate $f_i(t/n)$ of each neural spike train as a weighted summation of a set of basis functions. While there are a wide variety of basis functions that could be used, we choose the B-spline basis due to its simplicity and numerical properties. Indeed, the B-spline basis is a popular modelling choice for neural spike train data, offering a flexible and computationally efficient way to estimate complex neural dynamics (Song et al., 2013; Rodríguez-Álvarez et al., 2023). That said, the BAPLA model is extensible to different choices of basis functions. For example, if we suspected a piecewise constant trend, then we could utilise the Haar family of wavelets Vidakovic (2009).

Generally, the number of B-spline basis functions m is chosen to be large enough to ensure that the fitted curve is representative of the data. In our estimation procedure, we specify a fixed number of basis functions over a given sequence of knots, chosen to be equally spaced for $t \in [0, n]$. Specifically, let $\boldsymbol{\phi}_i(t/n) = (\phi_{i,1}(t/n), \ldots, \phi_{i,m}(t/n))$ be the associated B-spline basis for each $i = 1, \ldots, d$. Then, we estimate $\hat{f}_i(t/n) = \hat{\mathbf{c}}_i^\top \boldsymbol{\phi}(t/n)$ where $\hat{\mathbf{c}}_i$ is the estimated vector of coefficients for each neuron i.

An obvious extension to our model would be the inclusion of an additional penalty term to automatically select the number of basis functions, rather than specifying a fixed m. However, we emphasise that the main goal of this work is to get an accurate estimate of the interaction matrix $\boldsymbol{\Gamma}$, for which a reasonably good estimate of $f_i(t/n)$ is sufficient. Indeed, we will show later on in the simulation studies that there exists an $a > 0 \in \mathbb{Z}^+$ such that for $m > a$, the estimation error of the interaction matrix $\boldsymbol{\Gamma}$ does not improve. Thus, selecting a 'large enough' m is sufficient for the purposes of this work.

We obtain estimates of the model parameters using a **regularised maximum likelihood estimator**. Given data $\{y_{t,i}\}_{t=1}^n$ for $i = 1, \ldots, d$, we estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}_i, \hat{\beta}_i, \hat{c}_{i,1}, \ldots, \hat{c}_{i,m})$ by maximising the following penalised log-likelihood

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\phi}_i(t/n), \lambda) := \Bigg[& \sum_{t=1}^{n} y_{t,i} \Bigg(\beta_i + \sum_{k=1}^{m} c_{i,k}\phi_{i,k}(t/n) + \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} \Bigg) \\ & - \log \left(1 + \exp \left\{ \beta_i + \sum_{k=1}^{m} c_{i,k}\phi_{i,k}(t/n) + \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} \right\} \right) \\ & - \lambda \|\boldsymbol{\gamma}_i\|_1 \Bigg], \end{aligned} \tag{4.2.4}$$

where $\|\boldsymbol{\gamma}_i\|_1 = \sum_{j=1}^d |\Gamma_{ij}|$ denotes the ℓ_1 norm and $\lambda > 0$ is a regularisation parameter, such that

$$\begin{aligned} \hat{\boldsymbol{\theta}} := \arg\max_{\boldsymbol{\theta}} \quad & L(\boldsymbol{\theta}, \boldsymbol{\phi}_i(t/n), \lambda) \\ \text{such that} \quad & \sum_{k=1}^{m} c_{ik}\phi_{i,k}(t/n) = 0 \quad \text{for } t = 1, \ldots, n \,. \end{aligned} \tag{4.2.5}$$

The above regularised maximum likelihood estimator attempts to find a sparse estimate of the interaction matrix $\boldsymbol{\Gamma}$ whilst also ensuring a good fit to the data by modelling the non-linear interactions via $f_i(t/n)$. The constraint in (4.2.5) is the sample analogue of the assumption that $\int f_i(u)du = 0$, which ensures identifiability of the intercept term β_i and the smooth non-linear function $f_i(t/n)$.

Following the work of Guo et al. (2013), we convert (4.2.4) and (4.2.5) into an unconstrained problem by centering the basis functions as follows. Let

$$\bar{\phi}_{i,k}(t/n) = \frac{1}{n}\sum_{t=1}^{n} \phi_{i,k}(t/n), \quad \text{and} \quad \tilde{\phi}_{i,k}(t/n) = \phi_{i,k}(t/n) - \bar{\phi}_{i,k}(t/n).$$

Then, estimates of the model parameters $\boldsymbol{\theta} = (\hat{\boldsymbol{\gamma}}_i, \hat{\beta}_i, \hat{c}_{i,1}, \ldots, \hat{c}_{i,m})$ are obtained by maximising the unconstrained problem, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \tilde{\boldsymbol{\phi}}_i(t/n), \lambda) \ , \tag{4.2.6}$$

and the corresponding estimator of the unknown smooth function is $\hat{f}_i(t) = \sum_{k=1}^{m} \hat{c}_{i,k}\tilde{\phi}_{i,k}(t/n)$. Note, for simplicity, in the rest of the paper we will assume that $\phi_{i,k}(t/n) = \phi_k(t/n)$, i.e. the basis functions are chosen to be the same for all neurons $i = 1, \ldots, d$.

4.2.3 Computational Choices

The estimator in (4.2.6) requires careful specification of a tuning parameter λ. Several approaches have been considered in the literature, including the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and cross-validation procedures. In this work, the regularisation parameter is selected using a BIC-type criterion due to its computational efficiency over cross-validation procedures which require additional model fittings. Moreover, BIC tends to select larger values of λ compared to AIC, thereby yielding a more interpretable representation of effective connectivities between neurons.

Using a training group of simulated data, we obtain estimates of $\boldsymbol{\gamma}_i$ over a fixed search grid of λ values and choose the tuning parameter which minimises

$$\mathrm{BIC}_\lambda(\hat{\boldsymbol{\gamma}}_i) = -2\ell(\hat{\boldsymbol{\gamma}}_i) + \log(n)(k + 1 + m),$$

where $\ell(\hat{\boldsymbol{\gamma}}_i)$ denotes the maximised log-likelihood of the associated model and k denotes the number of non-zero entries in $\hat{\boldsymbol{\gamma}}_i$. These values are then averaged across the number of samples in the training set to arrive at a final optimal parameter λ^*.

We solve the optimisation problem in (4.2.6) using a coordinate descent algorithm (Friedman et al., 2010). However, consider first the *unpenalised* equivalent of (4.2.6) which can be solved via the Newton algorithm, which amounts to iteratively re-weighted least squares (Friedman et al., 2023). That is, if we have current estimates of the parameters $(\tilde{\beta}_i, \tilde{\boldsymbol{\gamma}}_i, \tilde{c}_{i,1}, \ldots, \tilde{c}_{i,m})$, then we can use a Taylor expansion around these estimates to obtain a quadratic approximation to the log-likelihood

$$\begin{aligned}\ell_Q(\beta_i, \boldsymbol{\gamma}_i, c_{i,1}, \ldots, c_{i,m}) = -\frac{1}{2}\sum_{t=1}^{n}\left\{w_{t,i}\left(z_{t,i} - \beta_i - \sum_{k=1}^{m} c_{i,k}\tilde{\phi}_k(t/n) - \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1}\right)^2\right\} \\ + C\left(\tilde{\beta}_i, \tilde{\boldsymbol{\gamma}}_i, \tilde{c}_{i1}, \ldots, \tilde{c}_{im}\right)^2,\end{aligned}$$

where $C(\cdot)$ is a constant with respect to the new parameter positions, and we have

$$\begin{aligned} z_{t,i} &= \tilde{\beta}_i + \tilde{\boldsymbol{\gamma}}_i^\top \boldsymbol{y}_{t-1} + \sum_{k=1}^{m} \tilde{c}_{i,k}\tilde{\phi}_k(t/n) + \frac{y_{t,i} - \tilde{p}(\boldsymbol{y}_{t-1})}{w_{t,i}}, && \text{(working response)} \\ w_{t,i} &= \tilde{p}(\boldsymbol{y}_{t-1})(1 - \tilde{p}(\boldsymbol{y}_{t-1})), && \text{(weights)}\end{aligned}$$

where $\tilde{p}(\boldsymbol{y}_{t-1})$ is evaluated at the current parameters, i.e., $\tilde{p}(\boldsymbol{y}_{t-1}) = 1/1 + \exp\left\{-\tilde{\beta}_i - \sum_{k=1}^{m} \tilde{c}_{i,k}\tilde{\phi}_k(t/n) - \tilde{\boldsymbol{\gamma}}_i^\top \boldsymbol{y}_{t-1}\right\}$. The Newton update is then obtained by minimising ℓ_Q. The approach we take to maximise the regularised estimator (4.2.6) is very similar. We compute the quadratic approximation ℓ_Q about the current parameters, and then we use coordinate descent to solve the penalised weighted least squares problem

$$\min_{\boldsymbol{\theta}}\{-\ell_Q(\beta_i, \boldsymbol{\gamma}_i, c_{i,1}, \ldots, c_{i,m}) + \lambda\|\boldsymbol{\gamma}_i\|_1\}.$$

Algorithm 3 Estimation Procedure

```
for $i = 1, \dots, d$ do
    Initialise $\lambda$ and $\tilde{\boldsymbol{\theta}}_0$
    Sample basis functions
    Compute $\ell_Q(\tilde{\beta}_i, \tilde{\boldsymbol{\gamma}}_i, \tilde{c}_{i1}, \dots, \tilde{c}_{im})$
    Use cyclic coordinate descent to solve the penalised weighted least-squares problem
        $\min_{\boldsymbol{\theta}} \{-\ell_Q(\tilde{\beta}_i, \tilde{\boldsymbol{\gamma}}_i, \tilde{c}_{i,1}, \dots, \tilde{c}_{i,m}) + \lambda \|\boldsymbol{\gamma}_i\|_1\}.$
        Calculate the coordindate-wise update step for each element in $\boldsymbol{\gamma}_i^\top$, i.e., for $j = 1, \dots, p$,
            $\tilde{\gamma}_{i,j} \leftarrow \frac{\mathcal{S}\left(\sum_{t=1}^n w_{t,i} y_{j,t-1} \left(z_{t,i} - \tilde{y}_t^{(j)}\right), \lambda\right)}{\sum_{t=1}^n w_{t,i} (y_{j,t-1})^2},$
        where $\tilde{y}_t^{(j)} = \tilde{\beta}_0 + \tilde{\boldsymbol{\gamma}}_i^\top \boldsymbol{y}_{t-1} - \tilde{\gamma}_{i,j} y_{j,t-1} + \sum_{k=1}^m \tilde{c}_{i,k} \tilde{\phi}_k(t/n)$ is the fitted value excluding the contribution from $y_{j,t-1}$ and $\mathcal{S}(a, \kappa)$ is the soft-thresholding operator with value
            $\mathcal{S}(a, \kappa) = \begin{cases} a - \kappa & \text{if } a > 0 \text{ and } \kappa < |a| \\ a + \kappa & \text{if } a < 0 \text{ and } \kappa < |a| \\ 0 & \text{if } \kappa \geq |a|. \end{cases}$
        Update the intercept term
            $\tilde{\beta}_i \leftarrow \frac{\sum_{t=1}^n w_{t,i} (z_{t,i} - \sum_{k=1}^m \tilde{c}_{i,k} \tilde{\phi}_k(t/n) - \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1})}{\sum_{t=1}^n w_{t,i}}$
        Update the spline coefficients, i.e. for $k = 1, \dots, m$
            $\tilde{c}_{i,k} \leftarrow \frac{\sum_{t=1}^n w_{t,i} \phi_k(t/n) (z_{t,i} - \tilde{\beta}_i - \tilde{\gamma}_i^\top \boldsymbol{y}_{t-1} - \sum_{j \neq k} \tilde{c}_{i,j} \tilde{\phi}_j(t/n))}{\sum_{t=1}^n w_{t,i} \tilde{\phi}_k^2(t/n)}$
        Repeat steps (7) – (9) until convergence.
end for
```

Simple calculus yields the coordinate-wise update steps for each of the model parameters outlined in Algorithm 1.

4.3 Inference

In general, it is difficult to determine a 'ground truth' for network interactions amongst a population of real neurons. As noted in Zhao et al. (2012), *"...even if microscopic examination of living tissue were feasible, it would still be unclear what the strength and* *sign of a synaptic interaction between two cells might be"*. It is therefore fundamental to recognise the uncertainty associated with our estimates of the interaction matrix. However, despite the vast progress made in recent years for the estimation of brain connectivity from neural recordings, little work has been done to assess the uncertainty associated with these estimates. In this section, we propose an inference procedure for the construction of confidence intervals for entries in $\mathbf{\Gamma}$, with the overarching aim of quantifying the uncertainty around our estimates of the interactions between neural processes.

4.3.1 Desparsifying the Regularised Estimator

The construction of confidence intervals for the interaction matrix is not a straightforward task. Due to the Lasso regularisation in (4.2.6), entries in $\hat{\boldsymbol{\gamma}}_i \in \mathbb{R}^d$ can be exactly zero, leading to a discontinuous limiting distribution which may have probability one at zero whenever the true parameter value is zero (Waldorp and Haslbeck, 2024). As such, existing techniques for calculating standard errors are not suitable (Tibshirani, 1996) and standard methods for bootstrapping also do not work (Fu and Knight, 2000; Chatterjee and Lahiri, 2011). To circumvent this problem, we will construct a desparsified estimator of the coefficient vector $\boldsymbol{\gamma}_i$, for all $i = 1, \ldots, d$. This technique was pioneered by Van de Geer et al. (2014) in the context of linear and generalised linear models, and laid the ground work for inference procedures in other high-dimensional models (see for example Javanmard and Montanari, 2014; Gueuning and Claeskens, 2016).

Consider the negative log-likelihood for a particular dimension i

$$\ell(\beta_i, \boldsymbol{\gamma}_i, c_{i,1}, \ldots, c_{i,m}) = -\sum_{t=1}^{n} y_{t,i} \log\{p_i(t)\} + (1 - y_{t,i}) \log\{1 - p_i(t)\}, \tag{4.3.1}$$

where $p_i(t) = [1 + \exp\{-\beta_i - \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} - f_i(t/n)\}]^{-1}$. We obtain the following expressions for the first and second partial derivatives of (4.3.1) with respect to $\boldsymbol{\gamma}_i \in \mathbb{R}^d$

$$\dot{\ell}_{\boldsymbol{\gamma}_i} = \frac{\partial}{\partial \boldsymbol{\gamma}_i}\ell(\beta_i, \boldsymbol{\gamma}_i, c_{i,1}, \ldots, c_{i,m}) = -\sum_{t=1}^{n}\{y_{t,i} - p_i(t)\}\boldsymbol{y}_{t-1},$$

$$\ddot{\ell}_{\boldsymbol{\gamma}_i} = \frac{\partial}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^\top}\ell(\beta_i, \boldsymbol{\gamma}_i, c_{i,1}, \ldots, c_{i,m}) = \sum_{t=1}^{n} p_i(t)\{1 - p_i(t)\}\boldsymbol{y}_{t-1}^\top \boldsymbol{y}_{t-1}. \tag{4.3.2}$$

Letting $\hat{\Sigma} := \ddot{\ell}_{\hat{\boldsymbol{\gamma}}_i}$, we can then define the desparsified estimator of the coefficient vector $\boldsymbol{\gamma}_i$, for each $i = 1, \ldots, d$ as

$$\hat{\boldsymbol{\gamma}}_i^{desp} = \hat{\boldsymbol{\gamma}}_i - \hat{\Theta}\dot{\ell}_{\hat{\boldsymbol{\gamma}}_i}, \tag{4.3.3}$$

where $\hat{\Theta} := \hat{\Sigma}^{-1}$.

Often, spike train data are considered to be high dimensional in the sense that the dimension d (or equivalently the number of neurons) is large in comparison to the sample size n (Pinkney et al., 2024). However, under a binned spike count representation of these data, it is likely that $d < n$ meaning that the matrix inverse $\hat{\Sigma}^{-1}$ can be easily obtained. Conversely, in scenarios where $d > n$ or even $d \approx n$ it would be necessary to construct a 'relaxed' inverse $\hat{\Theta}$ due to singularity of $\hat{\Sigma}$. Indeed, a variety of methods have been proposed for this purpose, though the most widely used are that of nodewise regression (Van de Geer et al., 2014) and convex optimisation (Javanmard and Montanari, 2014)

4.3.2 Constructing Confidence Intervals

Our desparsification approach is essentially the method used for generalised linear models, adapted to include an additional smooth component via $f_i(t/n)$. In the classical setting (i.e., without $f_i(t/n)$) the term $\hat{\Theta}\dot{\ell}_{\hat{\boldsymbol{\gamma}}_i}$ removes the shrinkage bias induced by the Lasso penalty, thereby preventing the asymptotic distribution from containing a point-mass at zero. This correction term restores smoothness, yielding an approximately linear estimator. Consequently, a central limit theorem applies, implying that each of the entries in $\hat{\boldsymbol{\gamma}}_i$ are asymptotically normal, thus enabling the construction of confidence intervals (Van de Geer et al., 2014). Here, we assume that a similar argument holds, despite the inclusion of our additional model component, $f_i(t/n)$. More specifically, let $\hat{\boldsymbol{\gamma}}_i^{desp}$ denote the desparsified estimator defined in (4.3.3). Under certain regularity conditions, we expect that

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}_{ij}^{desp} - \boldsymbol{\gamma}_{ij})/\hat{\sigma}_{ij} = V_{ij} + o_p(1),$$

where V_{ij} converges weakly to a $\mathcal{N}(0,1)$-distribution and $\hat{\sigma}_{ij}^2 = \hat{\Theta}_{jj}$. An asymptotic $(1-\alpha)100\%$ confidence interval for Γ_{ij}^* is given by $\hat{\gamma}_{ij}^{desp} \pm \Phi^{-1}(1-\alpha/2)\sigma_{ij}/\sqrt{n}$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. While a full theoretical analysis for this modification to the desparsified estimator is left for future work, our empirical results (Section 4.4) provide preliminary support for the validity of our proposed approach.

4.4 Synthetic Experiments

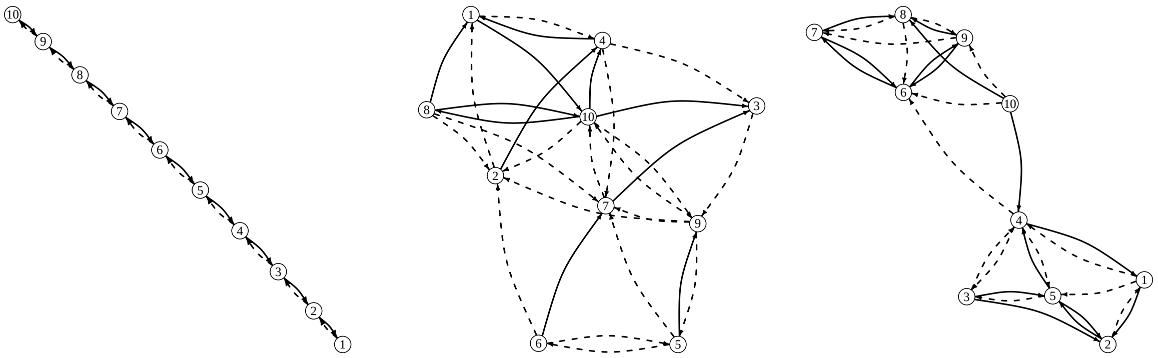
4.4.1 Simulation Setting

We use synthetic data to assess the accuracy of the BAPLA model in recovering estimates of the interaction matrix and non-stationary firing rates in various simulation settings. In the first instance, we simulate binary time series data according to model (4.2.1) for multivariate processes of dimension $d = 10$, $d = 50$ and $d = 100$, to accurately reflect the scale of our real data ($d = 82$). Moreover, we obtain simulated data for $n = 1000, 5000$ and $10,000$ time points, which is equivalent to recording periods (in length seconds (s)) of 1s, 5s and 10s, provided we select $\delta = 1$ms.

We mainly focus on three types of network structure: a simple one in which only neighbouring neurons interact, a random one in which interactions are assigned arbitrarily, and a more complex one with connected sub-networks, in which edges are more common within a sub-network than between sub-networks. Mathematically, these are more commonly referred to as chain graphs (CG), Erdos-Renyi (ER) random graphs, and graphs generated via the stochastic block (SB) model—examples can be seen in Figs. 4.4.1a, 4.4.1b, and 4.4.1c respectively.

In our simulations, we generate realistic neural firing patterns via careful specification and parametrisation of the BAPLA model. Firstly, we allow for both excitatory and inhibitory interactions in the simulated networks, encoded by positive and negative entries in the interaction matrix, respectively. Secondly, to mimic the behaviour of neural spike trains observed in stimulus-response experiments, we set $f_i(t/n)$ in model (4.2.1) as either the scaled probability density function (pdf) for a normal or gamma distribution, and vary this throughout our different simulation scenarios. Thirdly, we consider different specifications for the intercept term β_i, which determines the baseline firing rate of the neurons. In our population of real neurons, we observe firing rates in the range 10Hz - 200Hz (Figure 4.5.1). To approximate this behaviour, we set $\beta_i = -2.6$, realising a firing rate of approximately 70Hz, in which case we observe that approximately 16% of the simulated data points are spiking events, i.e., $y_{t,i} = 1$. We also consider a setting where $\beta_i = 0.1$. In doing so, we are able to assess the impact of different neural firing rates on our estimation procedure. Going forward, we will refer to the settings $\beta_i = -2.6$ and $\beta_i = 0.1$ as low- and high-firing, respectively. Since our focus here is on the estimation of the interaction matrix, we specify networks with a similar number of edges, enabling a fair comparison between the different simulation settings. Moreover, each interaction matrix has both positive and negative entries of the same magnitude, $|\Gamma_{ij}| = 0.3$.

The performance of the BAPLA model is assessed across a variety of scenarios including: varying the complexity of the network (CG, ER or SB), the type of interaction (excitation or inhibition), the dimensionality of the network, the size of the data set



(a) (b) (c)

Figure 4.4.1: Illustrations of the different network structures used in the simulations with $d = 10$ neurons. (a) Chain graph. (b) Erdős-Rényi graph. (c) Stochastic block model. Dashed edges are inhibitory interactions ($\Gamma_{ij} < 0$) and solid edges are excitatory interactions ($\Gamma_{ij} > 0$).

(1s, $2s$, 5s or 10s recording periods), the baseline firing rate and the specification of the non-stationary firing rate (normal or gamma). For each combination of model parameters, we obtain parameter estimates over $N = 100$ Monte Carlo simulations. We compare estimates of the interaction matrix to the ground truth using the relative mean squared error (RMSE) defined as $\text{RMSE}(\hat{\Gamma}) = \|\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_F^2 / \|\mathbf{\Gamma}\|_F^2$. Estimates of the intercept term and non-stationary firing rate are assessed using a mean squared error criterion, which we denote by MSE_β and MSE_f, respectively. We also report results on the recovery of the sparsity pattern of the interaction matrix using the area under receiver operating characteristic curve (AUROC). Numerical results averaged over 100 repetitions are presented in Table 4.4.1.

Confidence intervals for entries in the interaction matrix are constructed using the methods described in Section 4.3. We calculate the average coverage (Avgcov) and average length (Avglen) of the confidence intervals for entries in the interaction matrix, corresponding to entries in the edge set $E(\mathbf{\Gamma})$ (4.2.3) or its compliment $E^c(\mathbf{\Gamma})$, i.e., non-zero and zero coefficients respectively. Specifically, we report empirical versions of

$$\text{Avgcov}_{E(\mathbf{\Gamma})} = \frac{1}{|E(\mathbf{\Gamma})|} \sum_{(i,j)\in E(\mathbf{\Gamma})} \mathbb{P}(\Gamma_{ij} \in \text{CI}_{ij})$$

and

$$\text{Avglen}_{E(\mathbf{\Gamma})} = \frac{1}{|E(\mathbf{\Gamma})|} \sum_{(i,j)\in E(\mathbf{\Gamma})} \text{length}(\text{CI}_{ij}),$$

where CI_{ij} denotes the confidence interval for the $(i, j)^{th}$ element of $\mathbf{\Gamma}$. Analogous definitions are given for the set $E^c(\mathbf{\Gamma})$.

4.4.2 Simulation Results

Examining the results over all simulation settings, it is clear to see that more data, i.e. longer recording periods, yields more favourable results in terms of both RMSE and the AUROC scores for the interaction matrix. We also observe lower RMSE and higher AUROC scores for the high-firing ($\beta_i = 0.1$) setting compared to that of the low-firing. Importantly, we note that the diagonal elements of the Fisher information (4.3.2) are maximised when $p_i(t) \approx \text{logit}^{-1}(0.1) \approx 0.5$. Thus, the high-firing scenario is (by design) a more informative setting and therefore we should expect better estimates of the model parameters in this case. In general, lower-dimensional settings yield more favourable results, as evidenced by the lower RMSEs in cases where $d = 10$ and $d = 50$ compared to $d = 100$.

In terms of complexity of the network structure, our results show that (perhaps as expected) the BAPLA model performs best on the most simple network (CG) achieving an AUROC score of > 0.74 under the high-firing setting regardless of the recording period length and > 0.96 when $n \geq 5000$. Comparatively, the AUROC scores of the ER and SB models are > 0.82 under the high-firing setting for $n \geq 5000$. However,

Table 4.4.1: Simulation results averaged over 100 replications for estimating the parameters of the BAPLA model.

			Estimation				Inference			
d	β_0	n	RMSE_{Γ}	MSE_{β_0}	MSE_f	AUROC	$\text{AvgCov}_{E(\mathbf{\Gamma})}$	$\text{AvgCov}_{E^c(\mathbf{\Gamma})}$	$\text{AvgLen}_{E(\mathbf{\Gamma})}$	$\text{AvgLen}_{E^c(\mathbf{\Gamma})}$
					Chain Graph					
10	0.10	1000	0.21	0.03	0.59	0.82	0.94	0.95	0.57	0.57
		5000	0.05	0.05	0.60	0.96	0.93	0.95	0.26	0.26
		10000	0.03	0.05	0.61	0.98	0.92	0.95	0.18	0.18
	-2.60	1000	0.86	0.51	5.37	0.56	0.95	0.96	1.90	1.90
		5000	0.27	0.04	0.28	0.71	0.95	0.95	0.82	0.82
		10000	0.16	0.04	0.24	0.73	0.94	0.95	0.58	0.58
50	0.10	1000	0.32	0.02	0.56	0.76	0.94	0.95	0.61	0.61
		5000	0.09	0.04	0.59	0.98	0.94	0.95	0.27	0.27
		10000	0.05	0.04	0.60	1.00	0.94	0.95	0.19	0.19
	-2.60	1000	0.94	0.39	4.46	0.53	0.96	0.96	2.36	2.36
		5000	0.40	0.03	0.27	0.69	0.96	0.95	0.89	0.89
		10000	0.27	0.02	0.22	0.72	0.95	0.95	0.61	0.62
100	0.1	1000	0.35	0.02	0.55	0.74	0.95	0.95	0.63	0.63
		5000	0.10	0.03	0.59	0.98	0.95	0.95	0.27	0.27
		10000	0.05	0.04	0.60	1.00	0.94	0.95	0.19	0.19
	-2.60	1000	0.96	0.34	3.54	0.52	0.95	0.96	2.90	2.91
		5000	0.45	0.03	0.27	0.68	0.96	0.96	0.93	0.93
		10000	0.29	0.02	0.22	0.70	0.95	0.95	0.63	0.63
					Erdos-Renyi Random Graph					
10	0.10	1000	0.88	0.02	0.33	0.61	0.94	0.94	0.63	0.63
		5000	0.58	0.03	0.09	0.82	0.87	0.92	0.28	0.28
		10000	0.53	0.04	0.07	0.87	0.84	0.87	0.20	0.20
	-2.60	1000	1.03	0.33	1.30	0.51	0.94	0.94	1.70	1.69
		5000	0.92	0.03	0.11	0.57	0.94	0.94	0.73	0.73
		10000	0.85	0.02	0.08	0.63	0.92	0.93	0.51	0.51
50	0.10	1000	0.95	0.13	0.34	0.56	0.95	0.95	0.65	0.65
		5000	0.60	0.07	0.12	0.86	0.94	0.95	0.28	0.28
		10000	0.45	0.05	0.08	0.92	0.94	0.95	0.20	0.20
	-2.60	1000	1.02	0.14	1.44	0.51	0.95	0.95	2.06	2.07
		5000	0.97	0.02	0.14	0.53	0.95	0.95	0.79	0.79
		10000	0.96	0.02	0.10	0.55	0.95	0.95	0.55	0.55
100	0.10	1000	0.96	0.09	0.29	0.55	0.95	0.95	0.68	0.68
		5000	0.65	0.05	0.10	0.84	0.95	0.95	0.29	0.29
		10000	0.50	0.04	0.07	0.91	0.95	0.95	0.20	0.20
	-2.60	1000	1.03	0.11	1.09	0.50	0.95	0.95	2.45	2.45
		5000	0.98	0.01	0.11	0.52	0.95	0.95	0.82	0.82
		10000	0.96	0.01	0.08	0.55	0.95	0.95	0.56	0.56
					Stochastic Block Model					
10	0.10	1000	0.85	0.10	0.47	0.63	0.94	0.94	0.61	0.63
		5000	0.59	0.06	0.31	0.83	0.90	0.94	0.27	0.28
		10000	0.42	0.05	0.28	0.93	0.85	0.93	0.19	0.20
	-2.60	1000	1.00	0.26	2.01	0.54	0.94	0.96	1.69	2.42
		5000	0.90	0.04	0.20	0.58	0.93	0.95	0.74	1.05
		10000	0.80	0.04	0.17	0.65	0.88	0.94	0.52	0.74
50	0.10	1000	0.93	0.10	0.45	0.58	0.95	0.95	0.66	0.68
		5000	0.65	0.07	0.30	0.83	0.94	0.95	0.29	0.30
		10000	0.52	0.06	0.29	0.90	0.93	0.95	0.20	0.21
	-2.60	1000	1.03	0.27	2.27	0.51	0.95	0.95	2.08	2.56
		5000	0.97	0.03	0.20	0.53	0.95	0.95	0.84	1.06
		10000	0.95	0.03	0.16	0.56	0.94	0.95	0.59	0.74
100	0.10	1000	0.94	0.11	0.47	0.57	0.95	0.95	0.68	0.71
		5000	0.61	0.07	0.31	0.87	0.94	0.95	0.29	0.30
		10000	0.47	0.07	0.30	0.93	0.94	0.95	0.20	0.21
	-2.60	1000	1.04	0.24	1.88	0.51	0.95	0.95	2.25	2.79
		5000	0.98	0.03	0.20	0.52	0.95	0.95	0.82	1.07
		10000	0.95	0.03	0.16	0.55	0.95	0.95	0.57	0.75

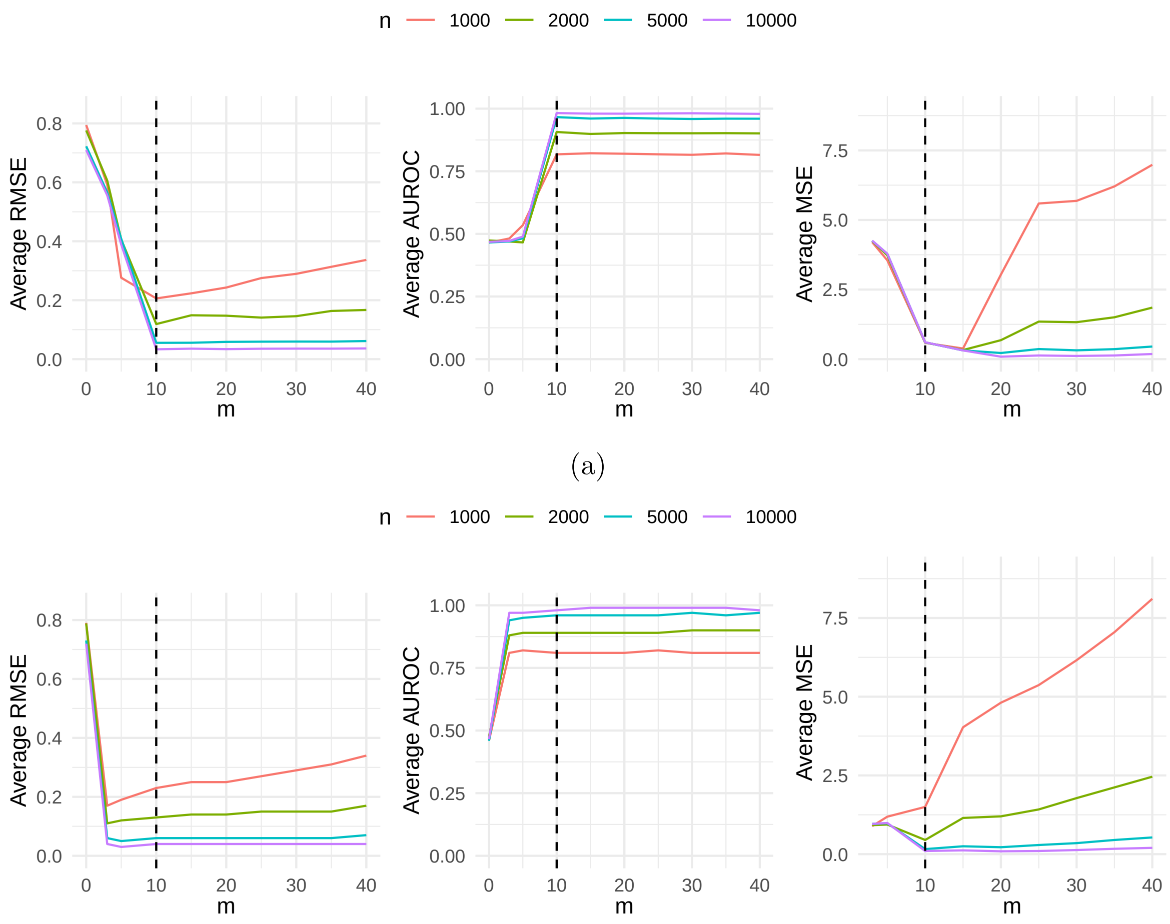
for the low-firing setting, the AUROC scores fall as low as 0.51 even for the longest recording period, highlighting the difficulty observed in detecting neuronal interactions in these settings.

Turing our attention now to the recovery of the intercept term and non-stationary firing rates, we see that in general, the MSE reduces as a function of the sample size n as expected. Importantly, we note that in the CG setting $f_i(t/n)$ is proportional to a normal pdf, whereas in the ER setting it is proportional to a gamma pdf. In the SB setting, specifications of $f_i(t/n)$ are alternated from block to block. Examining the results across all scenarios, we see that ER network yields more favourable results in terms of MSE_f, compared to that of the CG and SB networks, suggesting that the BAPLA model better recovers the shape of the gamma pdf compared to that of the normal pdf. Interestingly, we highlight that the recovery of the non-stationary firing rate appears to be insensitive to the dimensionality of the problem. That is, MSE_f is comparable across the low and high-dimensional settings. This is in contrast to estimates of the interaction matrix, which were largely more favourable in lower dimensional settings compared to higher-dimensional settings.

Results for the inference procedure with the de-sparsified Lasso estimator are also presented in Table 4.4.1. In general, the estimator appears to perform better for the entries in $E^c(\mathbf{\Gamma})$ compared to those in $E(\mathbf{\Gamma})$. However, we would argue that both scenarios yield sufficiently adequate coverage results (nominal coverage is 0.95). As expected, the average length of the confidence intervals reduce as a function of n and are generally shorter for the high-firing setting compared to that of the low-firing.

4.4.3 Sensitivity Analysis

In this section, we investigate the consequences of specifying a fixed number of basis functions m for the estimation of the non-stationary firing rate $f_i(t/n)$ for each $i = 1, \ldots, d$. Keeping in mind that our primary goal is to estimate the effective connec-



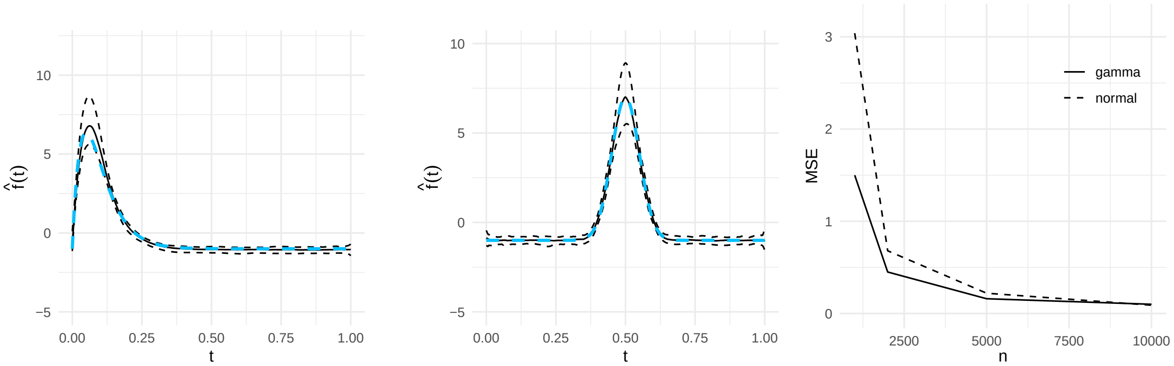
(a)

(b)

Figure 4.4.2: Simulation results averaged over 100 samples for estimating the interaction matrix with RMSE (left) and AUROC (middle) and the non-stationary firing rate (right) with a range of m basis functions for (a) normal $f_i(t/n)$ and (b) gamma $f_i(t/n)$.

tivity amongst a population of neurons, we show here that accurate estimation of the interaction matrix is not overly sensitive to the estimation of $f_i(t/n)$, provided that m is sufficiently large (e.g., $m \geq 10$). As an illustrative example, we mimic the simulation scenario in Section 4.4.1 above, in which the interaction matrix is specified by a chain graph (Fig 4.4.1a) with dimension $d = 10$. Data is simulated according to model (4.2.1) with $\beta_i = 0.1$ and $f_i(t/n)$ is specified as either the scaled pdf of a normal or gamma distribution. As before, we generate data for $n = 1000, 2000, 5000$ and $10,000$ time points.

Our goal is to investigate how sensitive estimation of the interaction matrix is to



(a) (b) (c)

Figure 4.4.3: Estimates of $f_i(t/n)$ for the (a) gamma and (b) normal model with 95% confidence intervals (dashed lines) and compared to the ground truth (dashed blue line). Mean squared error as a function of n is shown in (c).

the number of basis functions used to estimate the non-stationary firing rate $f_i(t/n)$. Coupled with this, we are interested in exploring whether our choice of $m = 10$ basis functions is sufficient for accurate estimation of $f_i(t/n)$. To do so, we conduct a sensitivity analysis in which we fit the model multiple times for various values of m, i.e., $m \in \{3, 5, 10, 15, 20, 25, 30, 35, 40\}$, and evaluate various performance metrics associated with the parameter estimates of the BAPLA model. In doing so, our aim is to show that there exists some $a > 0 \in \mathbb{Z}^+$ such that for $m > a$, both the estimation error and edge structure of the recovered interaction matrix $\mathbf{\Gamma}$ does not considerably improve. Thus, selecting a 'large enough' m is sufficient for the purposes of this work. For completeness, we also investigate the case in which $m = 0$ referring to model (4.2.1) without including the $f_i(t/n)$ term, i.e., a standard logistic model. This scenario is included to motivate the need for an additional model component when there are inherent non-stationarities in the data.

Figure 4.4.2 shows the average error over 100 replications for estimating the parameters of the BAPLA model using a range of m basis functions. Figure 4.4.2a shows the RMSE, AUROC and MSE_f for a BAPLA model with $f_i(t/n)$ specified as a scaled normal pdf, where the dashed line at $m = 10$ is indicative of the scenario conducted in our earlier simulation study. Whilst we can see that both the average RMSE and AUROC of the interaction matrix do not improve beyond $m = 10$, the average MSE_f continues to decrease (for larger n) until $m = 20$ before plateauing. Therefore, in this particular scenario, one might conclude that a choice of $m = 10$ is sufficient for accurate estimation of the interaction matrix. Furthermore, while selecting a larger m could indeed help improve estimates of $f_i(t/n)$, this could come at a cost, as evidenced by the increase in average RMSE (for smaller n). In Figure 4.4.2b, we give results for the BAPLA model when $f_i(t/n)$ is specified as a scaled gamma pdf. In this scenario, it appears that a choice of $m = 5$ would be sufficient for accurate estimation of the interaction matrix as evidenced by the RMSE plot and AUROC plot. Moreover, unlike the previous example, the plots show that selecting $m > 10$ basis functions would not necessarily improve estimates of the non-stationary firing rate. These results align with our earlier findings from the simulation studies, where we observed that the BAPLA model with $m = 10$ basis functions was better able to recover the shape of the gamma pdf compared to that of the normal pdf. Based on these results, we suggest that the original choice of $m = 10$ basis functions is indeed sufficient for accurate estimation of the interaction matrix. Moreover, we select $m^* = 10$ and $m^* = 20$ as the optimal number of basis functions for the gamma and normal models, respectively. Figures 4.4.3a and 4.4.3b show estimates of $f_i(t/n)$ for the gamma and normal models averaged across 100 replications (solid black line) with empirical 95% confidence bands (dashed lines) and the ground truth in blue. In Figure 4.4.3c we plot $\mathrm{MSE}_f(m^*)$ as a function of n, highlighting the somewhat consistent estimation of the non-stationary firing rate as evidenced by the fact that $\mathrm{MSE}_f(m^*) \to 0$ as $n \to \infty$. Further details of the sensitivity analysis can be found in the Supplementary Material, where additional simulation results are also presented.

4.5 Identifying Neural Connectivity

4.5.1 Data Description and Preprocessing

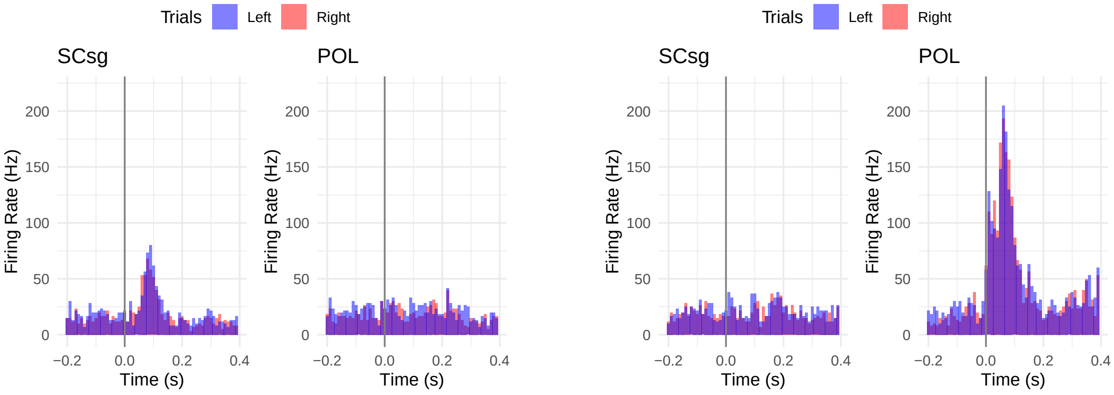
In this section, we use our proposed BAPLA model to analyse a dataset from the DANDI (Distributed Archives for Neurophysiology Data Integration) archive. DANDI is the BRAIN Initiative[1] supported data archive for publishing and sharing neurophysiology data including electrophysiology, optophysiology and behaviour time series, from various neuroscience experiments.

Here, we consider a spike train dataset from Steinmetz et al. (2019) where Neuropixels probes were used to record from a population of neurons in various regions of the mouse brain. In this experiment, neural activity was recorded over a series of experimental trials, while mice performed a visual discrimination task. On each trial, visual stimuli were presented on either the left side, right side, both sides or neither side of a screen at varying contrasts. Mice could earn a water reward by turning a wheel with their forepaws to indicate which side had highest contrast, i.e., the side in which the visual stimulus was most prominent. In the event that neither stimulus was present, mice could earn a reward by keeping the wheel still for 1.5 seconds. Furthermore, if both the left and the right stimulus had equal non-zero contrast, then the mice earned rewards randomly for turning the wheel left or right.

During the experiment, Neuropixel probes were used to record activity from various regions of the mouse brain. At most three probes were inserted at a time, in the left hemisphere, which gave rise to simultaneous recordings from hundreds of neurons in multiple brain regions during each recording session. In total there were 92 probe insertions over 39 sessions in 10 mice. To illustrate the use of the BAPLA model to identify neural interactions, we will consider data from a single mouse during a single recording session, corresponding to the "sub-Hench_ses-20170617T120000.nwb" file on

---

[1] `https://braininitiative.nih.gov/`



(a) (b)

Figure 4.5.1: Firing rate (PSTH plot) aligned to (a) stimulus onset and (b) wheel movement for example neurons in the SCsg and POL regions of the mouse brain.

the DANDI archive.

The experiment in Steinmetz et al. (2019) was designed to assess the distribution of neurons encoding vision, choice, action, and behavioural engagement across the mouse brain. Here, we attempt to build on their analysis, and construct networks of neuronal interactions for data aligned to either stimulus onset (encoding vision) or wheel movement (encoding action). In our analysis, we use spiking data from experimental trials in which visual stimuli were presented with the highest contrast. More specifically, we construct four artificial datasets by aligning the spike times either to stimulus onset or wheel movement, and then further split the data according to 'left trials' (where the stimulus appeared on the left side of the screen) or 'right trials' (where stimulus appeared on the right side). For each dataset, the raw spikes are downsampled with a 1 ms bin size, ensuring that at most one spike can occur per time bin, resulting in a binary time series representation of the neural data. Our goal is to apply the BAPLA model to each dataset individually, in an attempt to visualise neural connectivity in these differing experimental conditions.

Figure 4.5.1 shows firing rates (PSTH plots) aligned to (a) stimulus onset and (b) wheel movement for example neurons in the superior colliculus, superficial grey layer (SCsg) and the posterior limiting nucleus of the thalamus (POL) regions of the mouse brain. Left trials are in blue, while right trials are in red. In these plots, we observe an increase in the firing rate of the SCsg neuron following onset of the visual stimulus, but no response following wheel turns, while the converse is true for the POL neuron. Throughout, we will refer to neurons who do not respond to stimulus onset or initiation of the wheel movement as ‘stationary’, and those whose firing rate increases as ‘non-stationary’.

Initial exploratory analysis revealed that individual neurons do not necessarily fire in every single experimental trial. Therefore, to avoid subsequent issues with the estimation procedure, we choose to remove these trials (in which there are no spiking events) from our analyses. Moreover, to enable a fair comparison between the left and right trials for a specific experimental condition, we analyse neural data from the most active, i.e., highest firing, $l = 60$ trials, yielding an effective sample size of length 36 seconds (60 trials each of length 600ms). Finally, neurons are only selected for analysis if at least 10 spikes are observed per trial on average. As we saw in Figure 4.5.1, individual neurons may or may not respond at different stages in the experimental procedure. However, since neurons are selected for analysis based on their firing rates, this means that some neurons used in the analysis of the wheel movement data are excluded from the analysis of the stimulus onset data, and vice versa. Therefore, to enable a fair comparison between the groups, all neurons are plotted in our estimated networks of neural connectivity (Figure 4.5.2) where white nodes represent neurons excluded from the analysis of a particular dataset. Overall, we consider a subset of 82 neurons from 11 different brain regions that exhibit sufficiently high average firing rates in the period of interest, i.e., $-0.2s$ before and 0.4s after onset of the visual stimulus or initiation of the wheel movement.

Remark 4.5.1. *The discussed preprocessing procedure is necessary since neurons with few spikes may cause computational difficulties for the proposed BAPLA method. Importantly, we recognise that the inferred network structures may therefore primarily* *reflect interactions among high-firing neurons, rather than low firing neurons.*

To estimate the neural connectivity among our population of neurons, we consider the following adaptation to the regularised maximum likelihood estimator (4.2.6) in which we also incorporate the trial-like structure of the data. Letting $L_l(\boldsymbol{\theta}, \phi_i(t/n), \lambda)$ be the trial-specific loss according to (4.2.4), we obtain parameter estimates by maximising the joint likelihood according to

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{l=1}^{\#trials} L_l(\boldsymbol{\theta}, \tilde{\phi}_i(t/n), \lambda), \tag{4.5.1}$$

assuming independence across trials. As before, we solve the above optimisation problem (4.5.1) using a coordinate descent algorithm and select the regularisation parameter using a BIC-type criterion. The update steps, accounting for the trial-like structure of the data, are given in the supplementary material. Moreover, we estimate the non-stationary firing rate using $m = 6$ basis functions (see supplement for sensitivity analysis).

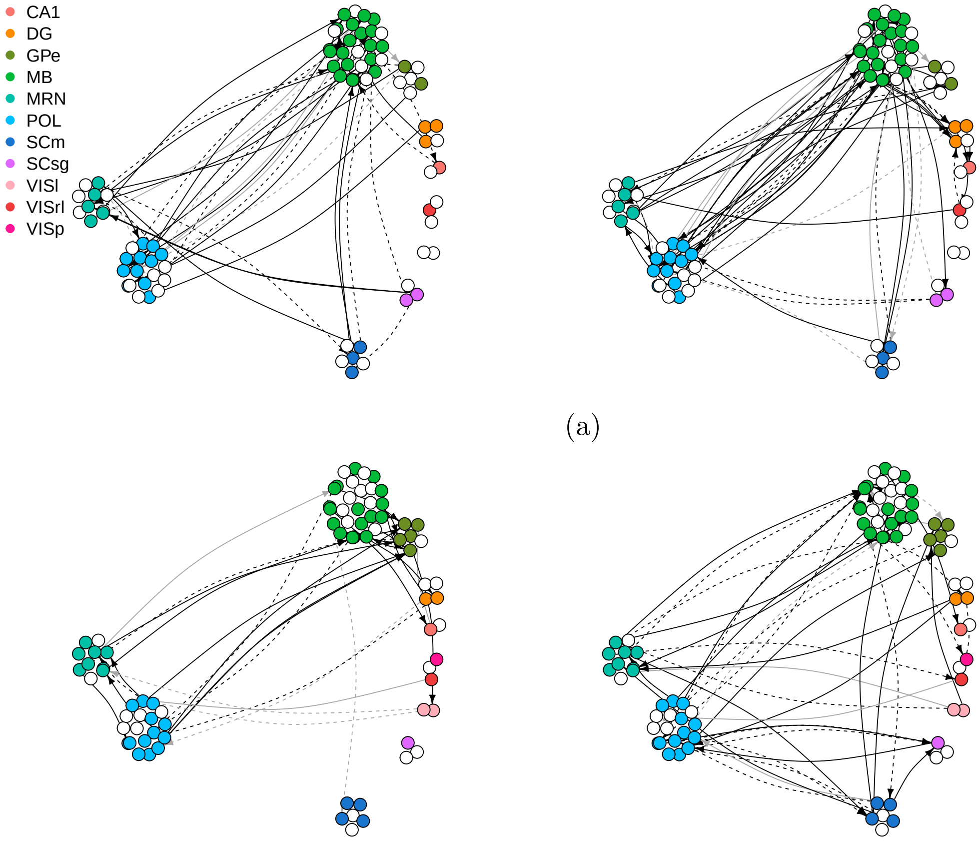
4.5.2 Graphical Representation of Neural Connectivity

Graphical representations of the estimated interaction matrices are shown in Figure 4.5.2. In these graphs, each node represents a neuron (colour coded by brain region[2]) and a directed edge from node i to node j represents neuron i's influence upon neuron j, accounting for the behaviour and activity of all the other neurons. White nodes represent the neurons which are omitted from the analysis due to low firing rates. A solid line represents an excitatory influence (i.e., $\Gamma_{ij} > 0$) and a dashed line indicates an inhibitory influence (i.e., $\Gamma_{ij} < 0$). Absence of an edge between two neurons means there is no interaction detected via the BAPLA model.

Networks of the detected interactions aligned to stimulus onset and wheel movement

---

[2] `https://atlas.brain-map.org/`



(a)

(b)

Figure 4.5.2: Estimated networks for left and right trials aligned to (a) stimulus onset and (b) wheel movement. Solid lines represent excitatory edges and dashed lines indicate inhibitory edges. White nodes represent neurons omitted from the analysis due to insufficient firing rates and grey edges denote entries in the interaction matrix deemed insignificant by our inference procedure.

are presented in Figures 4.5.2a and 4.5.2b, respectively. For the stimulus onset data, more edges are detected for the right trial dataset (68) compared to the left trial dataset (49). This is also true for the wheel movement dataset, where we observed 51 edges for right trials compared to only 38 edges for left trials. In both cases, the BAPLA model detects more excitatory edges than inhibitory ones. However, this is somewhat as expected since it is generally harder to detect inhibitory edges, especially if neurons exhibit low firing rates because further inhibition is limited by a floor at zero (Zhao et al., 2012). In general, our BAPLA model identifies more between region interactions compared to those in the same region (see Table 4.5.1 for further details).

			Interaction Type (%)		Region Type (%)	
Dataset	Trial	No. Edges	Excitatory	Inhibitory	Within	Between
Stimulus Onset	Left	49	61.22	38.78	24.49	75.51
	Right	68	69.12	30.88	14.71	85.29
Wheel Movement	Left	38	65.79	34.21	15.79	84.21
	Right	51	58.82	41.18	19.61	80.39

Table 4.5.1: Summary statistics for estimated interaction matrices.

For the stimulus onset data, we found that 44% of the total number of interactions are between neurons in the midbrain (MB) and the posterior limiting nucleus of the thalamus (POL). Interactions between these brain regions are part of a wider circuit known to facilitate crucial brain functions like sensory processing and motor control (Inagaki et al., 2022). While POL neurons also interact with neurons in other midbrain nuclei such as MRN and SCm, we found that neurons in the thalamus (POL) do not interact much with each other; an observation also highlighted by Jager et al. (2021). In particular, we identified only 2 edges in rightward trials and 3 edges in leftward trials among POL neurons. Interactions among midbrain nuclei (MB, MRN, SCm and SCsg) account for $\approx 36\%$ of the total number of interactions, and these regions also interact with neurons in the basal ganglia (GPe). Overall, our results for the stimulus onset data align with the findings in Steinmetz et al. (2019) where it is highlighted by neuroscientists that neurons encoding vision were found in a pathway comprising primarily of visual areas, such as the thalamus (POL) and superficial superior colliculus (SCs), and other structures such as the basal ganglia (GPe) and several midbrain nuceli (SCm, MRN).

For the wheel movement dataset, we found that 75% of the total number of interactions are between neurons in the midbrain nuclei (MB, MRN, SCm and SCsg), the posterior limiting nucleus of the thalamus (POL) and the basal ganglia (GPe). Interactions among these particular brain regions are central to the function of the so-called cortico-basal ganglia-thalamo-cortical (CBGTC) loop. The CBGTC is one of the fundamental network motifs in the brain, and revealing its structural and functional organisation is critical to understanding cognition, sensorimotor behaviour and the natural history of many neurological and neuropsychiatric disorders (Foster et al., 2021). In general, neural connectivity appears to be more widespread across the different brain regions for the wheel movement dataset compared to the stimulus onset data, particularly for the right trials. Our results therefore corroborate the findings of Steinmetz et al. (2019) who highlight that neurons encoding action were spread throughout all recorded regions, while neurons encoding visual stimuli were more constrained to a particular pathway.

We construct confidence intervals for entries in the interaction matrix using the desparsified estimator outlined in Section 4.3. If a given interval contains zero, the associated entry in the interaction matrix is deemed insignificant and is consequently set to zero, consistent with the sparsity-inducing nature of our regularised estimator. Using this approach, we find that 8 (9) edges in the left trials and 11 (10) edges in the right trials are deemed insignificant at a 5% significance level for the stimulus onset (wheel movement) dataset. These so-called insignificant edges are highlighted in grey in Figure 4.5.2.

4.5.3 Accounting for Non-Stationary Firing Rates

In this section, we assess to what extent accounting for non-stationary firing rates in our BAPLA model impacts the estimation of the interaction matrix. Figure 4.5.3 shows the estimated interaction matrix under the standard logistic model (which assumes stationary firing rates) for left trials in the wheel movement dataset (left). Generally, more edges are detected under the logistic model (80) compared to our proposed BAPLA method (38), perhaps suggesting that neurons with non-stationary firing rates are more likely to interact; either with each other, or with other stationary neurons. To explore this possibility, we sort the spike train data into two groups, and classify each neuron

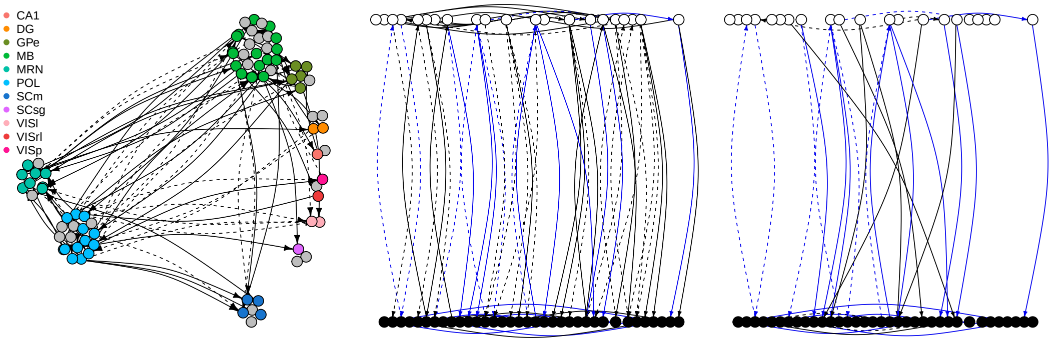


Figure 4.5.3: Estimate of the interaction matrix for left trial data in the wheel movement dataset with standard logistic model for all regions (left) and split into stationary and non-stationary neurons (middle). Non-stationary neurons are depicted by white nodes and stationary neurons are in black. Figure on right shows estimate of the interaction matrix obtained using the BAPLA method. Common edges between the BAPLA model and logistic model are in blue.

as either stationary or non-stationary as evidenced by their firing rate.

Figure 4.5.3 shows estimates of the interaction matrix under the logistic model (middle) compared to that obtained using our BAPLA method (right). In these plots, the black and white nodes represent stationary and non-stationary neurons, respectively, and edges common to both estimates are shown in blue. Displaying the networks in this way, we can see that accounting for non-stationary firing rates reduces the number of edges, both within the group of non-stationary neurons and between other stationary neurons. In particular, 19 edges are detected between non-stationary neurons under the logistic model, compared to only 4 for the BAPLA model. Additionally, the number of edges between the groups reduces from 44 to 21 when we use our proposed BAPLA method. Overall, our results suggest that explicit modelling of the firing rate is essential to reduce the number of spurious connections in the estimated interaction matrix.

4.6 Discussion

To conclude, we have proposed a Bernoulli autoregressive partially linear additive model to identify excitatory and inhibitory interactions among a population of non-stationary neural spike trains. Our results demonstrate that our estimation procedure works well in a variety of scenarios, and that explicit modelling of the non-stationary firing rate is essential for accurate recovery of the interaction matrix. We also propose and empirically validate an inference procedure to construct confidence intervals for our estimates of neural connectivity. Furthermore, we have demonstrated the usefulness of our model in practice using a real dataset of scientific interest, where we study interactions between neural processes using spike train data from Steinmetz et al. (2019).

This paper extends the current literature for identifying neural connectivity in two ways: 1) by accounting for non-stationary firing rates in the estimation procedure, and 2) by providing confidence intervals for entries in the interaction matrix. However, there are many possible and interesting avenues for future work. From an application point of view, extensions of the BAPLA model might include the addition of some categorical covariates or behavioural variables, specific to the neuroscience experiment of interest. One can also consider the choice of a different set of basis functions for modelling the trend, e.g. if we suspected a piecewise constant trend then one could utilise the Haar family of wavelets (Vidakovic, 2009). Methodologically, our method could be extended to include an additional penalty term in the regularised maximum likelihood estimator (4.2.4) for automatic selection of the number of basis functions, or to enforce smoothness conditions on $f_i(t/n)$. For the purposes of this work, we argue that the benefits of including an additional penalty term do not outweigh its complications, nor its increase in computational complexity. However, we believe that in some scenarios and alternative applications this development could be particularly advantageous. Finally, it remains to provide a full theoretical analysis for our adaptation of the desparsified estimator, to construct confidence intervals for entries in the interaction matrix. While our simulation results demonstrate empirical support for the validity of our proposed inference approach, future work could involve tracking the asymptotic properties of the desparsified estimator in the presence of non-linear components. Perhaps the work of Gueuning and Claeskens (2016) or Drikvandi (2025) looking at a combination of ℓ_2 and

ℓ_1 penalisation might provide a reasonable starting point for such analysis.

Chapter 5

Neuronal Network Estimation Methods: Uncovering Cross-Session and Cross-Subject Heterogeneity

5.1 Introduction

As discussed, substantial progress has been made towards understanding the intricate processes that drive brain function. These advances have been fuelled, in part, by the development of new technologies capable of recording the simultaneous activity of hundreds of neurons in the living brain. However, of equal importance, has been the development of new statistical and computational tools that can extract meaningful dependencies from the signalling dynamics of neurons. In this chapter, we focus on the practical application of these methods to real data, assessing the consistency of different neuronal network estimation methods, and revealing important cross-session and cross-subject heterogeneities.

In Chapters 3 and 4, we proposed two new methods for estimating connectivity in the brain network, and showcased their ability to detect neural interactions using data obtained from a single mouse. Naturally, at this stage, two questions arise:

1. How do the proposed estimation methods compare to existing approaches?
2. How do estimates of neural connectivity compare across experimental subjects?

In this chapter, we attempt to answer these questions by providing a more in-depth analysis of the neural data from Steinmetz et al. (2019, 2024) using a variety of computational and statistical tools. These data were previously studied in Chapter 4, where the BAPLA model was used to uncover interactions between neural processes in response to varying visual stimuli. In particular, we analysed brain connectivity using the spike train data from the subject mouse named 'Hench' during a single recording session. Here, we extend our discussion to the multi-session and multi-subject setting, exploring the variability of functional connectivity estimates across recording sessions *and* across experimental subjects. We also compare the methods proposed in Chapter 3 and Chapter 4 to existing approaches for neuronal network estimation, in both the single and multi-subject setting.

Given the variety of neuronal network estimation methods, a number of review papers exist which provide a comprehensive overview of the existing literature. Recent examples include Shahhosseini and Miranda (2022); Chiarion et al. (2023) and El-Yaagoubi et al. (2025) which present methods for brain connectivity analysis using EEG, fMRI and LFP data, respectively. At the mesoscopic level, Abril et al. (2018) reviewed a range of methods for brain connectivity analysis for data derived from multi-electrode recordings and calcium fluorescence imaging. In particular, the authors explored a variety of model-free (descriptive) and model-based (generative) approaches, and highlighted various ongoing challenges associated with the estimation and understanding of individual-neuron level connectivity. Indeed, while single neuron data provide the gold standard for measuring localised activity, they have received comparatively little attention in the statistics literature, relative to macroscopic measurements such as EEG or fMRI. In this chapter, we attempt to narrow this gap by exploring methods for brain connectivity analysis at the individual neuron level, and providing a novel data analysis of the Steinmetz et al. (2019) dataset.

A key part of our analysis involves the comparison of neural connectivity estimates across experimental subjects. Previous studies (albeit for fMRI data) have shown that functional connectivity can be reliably measured within subjects, however variability across subjects is to be expected (Mueller et al., 2013). Moreover, this variability is representative of meaningful differences rather than noise (Finn et al., 2015). Building on this literature, we will examine the variation in cross-subject connectivity patterns using neural spike train data.

As mentioned in the introductory chapter of this thesis, there does not yet exist a set of universally agreed measures to characterise brain activity (Hu et al., 2019). One reason for this, is the absence of a ground-truth connectome, making it difficult to objectively evaluate and compare neuronal network estimation methods. At the individual neuron level, a possible solution lies in using *in-vitro* techniques to map ground-truth connections in controlled experimental conditions. For example, current research involves growing neurons in dissociated cultures and using high-resolution techniques (eg., MEAs, optogenetic stimulation) to directly measure synaptic interactions and reconstruct network connectivity (Poli et al., 2016). However, these datasets are not currently available. In this chapter, we will therefore focus on comparing the consistency of the estimated connectivity patterns, both across methods and experimental subjects, rather than making comparisons to a known ground-truth.

The remainder of this chapter is organised as follows. In Section 5.2, we introduce the datasets used in our analyses. Section 5.3 is dedicated to the discussion of various statistical tools that can be used to infer brain connectivity. In Section 5.4.1, we utilise data from a single subject to make comparisons across the discussed network estimation methods. In Section 5.4.2, we extend our discussion to include comparisons

(a) (b)

Figure 5.2.1: Overview of the experiment. (a) Grey rectangles represent three computer screens surrounding the mouse. Arrows (not visible to the mouse) indicate the rewarded wheel turn direction and the coupled movement of the visual stimulus (X indicates reward for no turn), and the coloured dashed circle (not visible to mouse) indicates the stimulus location at which a reward was delivered. (b) Mice were head fixed with forepaws on the wheel while multiple Neuropixels probes were inserted for each recording. Source: Steinmetz et al. (2019)

across multiple subjects, before concluding with a discussion in Section 5.5.

5.2 Exploratory Data Analysis

In this section, we describe the experimental data from Steinmetz et al. (2019, 2024) that underpin our analyses. While these data have previously been discussed in Chapter 4, we include a more detailed description of the data here, tailored to the analysis conducted in this chapter.

Neural activity was recorded over a series of experimental trials, while mice performed a visual discrimination task. On each trial, visual stimuli were presented on either the left side, right side, both sides or neither side of a screen at varying contrasts. Mice could earn a water reward by turning a wheel with their forepaws to indicate which side had highest contrast, i.e., the side in which the visual stimulus was most prominent. In the event that neither stimulus was present, mice could earn a reward by keeping the wheel still for 1.5 seconds. Furthermore, if both the left and the right stimulus had equal non-zero contrast, then the mice earned rewards randomly for turning the wheel left or right (Figure 5.2.1).

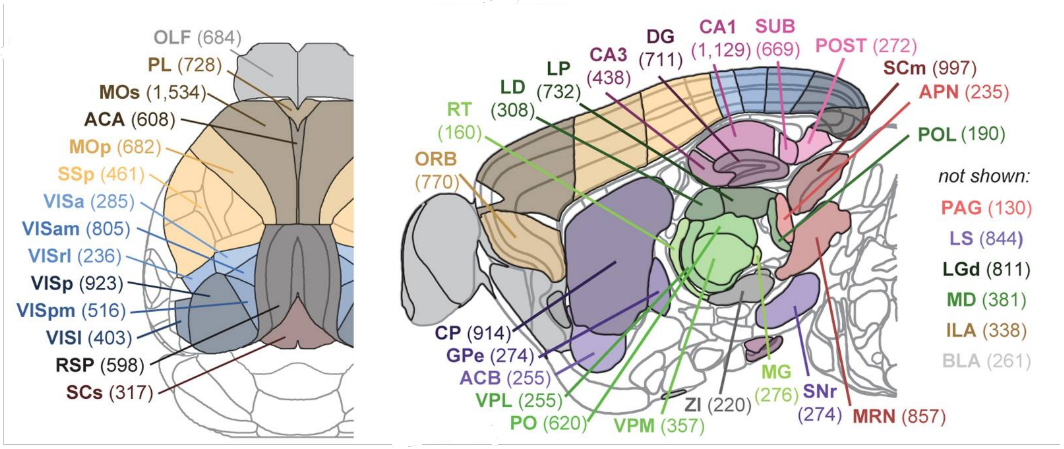


Figure 5.2.2: Summary of recording locations made from each of the 42 brain regions. Top down view of the cortex (left) and sagittal section (right). For each region, the number in parenthesis indicates the total number of recorded neurons. Source Steinmetz et al. (2019).

During the experiment, Neuropixel probes were used to record activity from various regions of the mouse brain. For each mouse, at most three probes were inserted at a time, in the left hemisphere, which gave rise to simultaneous recordings from hundreds of neurons in multiple brain regions during each recording session. In order to sample neurons from a wide range of brain areas, researchers varied the placement of the Neuropixels probes across animals and across recording sessions. This enabled scientists to better understand how neurons encoding action, choice and engagement were distributed across the mouse brain. However, this particular choice of experimental design presents important statistical challenges, especially when attempting to make comparisons between functional connectivity estimates across recording sessions or experimental subjects. In particular, as we will later show in our results section, not all regions are systematically and evenly sampled across the mice. As such, it is difficult to determine to what extent differences in estimated neural connectivity across subjects are representative of meaningful differences, rather than due to simple heterogeneity of the sample.

In total there were 92 probe insertions over 39 sessions in 10 mice, resulting in neural recordings from $30,000$ neurons in 42 brain regions (see Figure 5.2.2 for a visual summary of the recording locations). The data obtained from this experiment are publicly available on the DANDI archive[1]. Table 5.2.1, summarises the data collected from each experimental subject, across the various recording sessions. Upon visual inspection, we can see that the types of neurons recorded in each session is highly variable, both within and between experimental subjects. As discussed, this is likely a direct result of the intentionally varied probe placement, to maximise the number and variety of the recorded neurons (Steinmetz et al., 2019).

However, even if there had been more systematic sampling approach, it is generally not possible to reliably track the same neurons across recording sessions or across experimental subjects. For the latter, a simple explanation lies in the fact that brains are not identical across animals. That is, even if the Neuropixels probes were inserted in the exact same locations for each mouse, we would likely observe differences in the types of recorded neurons. For instance, recent work by Banga et al. (2025) showed that Neuropixels recordings from multiple mice using a repeated site (i.e., targeting the same sterotaxic coordinates) yielded substantial variability in recorded neuronal properties, suggesting that electrophysiology experiments were vulnerable to a lack of reproducibility.

At this point, it is worth highlighting that the experiments discussed in this chapter were a pre-cursor to those carried out by the International Brain Laboratory (IBL) in their quest to understand brain-wide circuits for complex behaviour. The IBL is a collaboration of neuroscience labs across the world, whose key project ‘The Brain-wide Map’ involves collecting and pooling data from multiple mice across multiple labs. When analysing these datasets, there is therefore the added complexity of not only tracking neurons across subjects in one lab, but also across labs. However, we do not consider these additional complexities induced by the multi-lab scenario in this chapter.

---

[1] `https://dandiarchive.org/dandiset/000017`

Table 5.2.1: Location of the recorded neurons for each experimental subject.

Subject	Session	Recorded Brain Regions
Cori	1	{ACA, CA3, DG, LS, MOs, SUB, VISp}
	2	{CA1, POST, VISl, VISpm}
	3	{CA1, DG, LP, MG, MRN, NB, POST, SPF, VISam, VISp}
Forssmann	1	{ACA, CA1, DG, LGd, LSr, MOs, SUB, TH, VISa, VISp, VPL}
	2	{AUD, CA1, SSp, TH}
Hench	1	{CA1, CA3, DG, ILA, LD, LP, LSr, MOs, PL, PO, SUB, TT, VISa, VISp}
	2	{CA1, CA3, LD, LSr, ORBm, PL, TH, TT, VISam, VISl, VPL}
	3	{CA1, DG, GPe, MB, MRN, POL, POST, SCm, SCsg, VISl, VISp, VISrl}
	4	{CP, LSc, LSr, MOp, PT}
Lederberg	1	{ACA, CA1, DG, LGd, LH, MD, MOs, PL, SUB, VISam, VISp}
	2	{ACA, CA1, DG, LGd, MB, MOs, MRN, MR, POL, RN, SCm, SCs, VISam, ZI}
	3	{CA3, LGd, MB, SSp, SSs, TH}
	4	{LD, MEA, RT, VPL, VPM}
Moniz	1	{APN, CA1, DG, LP, POL, SUB, VISam, VISpm}
	2	{CA1, DG, LGd, LP, MG, SPF, SUB, TH, VISa, VISp}
	3	{LD, MOp, SSp, VAL}
Muller	1	{ACA, ACB, IC, MOs, MRN, PL, RSP, SCsg, TT, VISp}
	2	{CA1, DG, LP, PO, VISam}
Radnitz	1	{ACA, DP, ILA, MOs, MRN, PL, RSP, SCig, SCsg, , VISp }
	2	{MOs, MRN, SCi, VISp }
	3	{ACA, ACB, CA1, DG, LD, LP, MOs, PL, PO, VISam }
	4	{APN, CA1, DG, LP, MRN, POL, RSP, SCig, SCsg, VISpm}
	5	{CP, MOp, SSp}
Richards	1	{ACA, CA1, CL, DG, LP, LS, MD, MOs, PO, RSP, SUB, VISa, VPM}
	2	{ CA3, MOs, OLF, ORB, POST, SCm, SNr, TH}
	3	{MOs, MRN, OLF, ORB, PAG, RSP, SCm, SCs }
	4	{BMA, COA, LP, MB, PO, RT}
Tatum	1	{ACA, CA1, DG, LGd, LH, LP, LS, MD, MOs, RSP ,TH, VISam}
	2	{CA, MOs, MRN, PL, SCm, SCsg ,VISam}
	3	{BLA, CA3, GPe, LGd, SNr}
Theiler	1	{ACA, CA1, DG, ILA, LP, MOs, PL, SUB, VISam, VISp}

Generally, it can be difficult to reliably track the same neurons across recording sessions (even for a single mouse) due to electrode drift (Steinmetz et al., 2021). For example, it is common to observe a relative movement of the brain tissue with respect to the probe (Garcia et al., 2024). This causes the electrical signals (e.g. waveforms) to

Region of Interest (ROI)	Brain Regions
Visual Cortex	{VISp, VISl, VISpm, VISam, VISrl, VISa }
Frontal/somatomotor Cortex	{MOp, MOs, PL, ILA, ORB, ACA, SSp, RSP }
Hippocampus	{DG, CA1, CA3, POST, SUB, OLF, BLA }
Basal Ganglia	{CP, GPe ,SNr, ACB, LS}
Thalamus	{LGd, LP, LD, POL, MD, VPL, PO, VPM, RT, MG }
Midbrain	{SCs, SCm, MRN, APN, PAG, ZI }

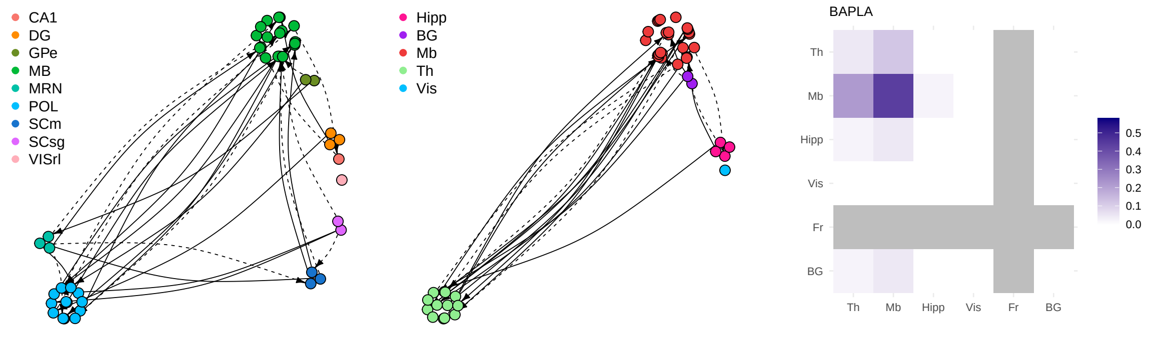
Table 5.2.2: Brain regions categorised into 6 ROIs (c.f. 'areas' in Steinmetz et al., 2019).

become distorted, making it difficult for spike sorters to identify the same neuron over time. For example, even if the electrode shifts only a few microns, the spike waveform can change enough so that the spike sorting algorithm no longer recognises it as the same neuron. Possible solutions to this problem have been discussed in Garcia et al. (2024); van Beest et al. (2025).

Admittedly, we would likely expect to see a larger degree of overlap in the types of neurons recorded if the experimental design involved identical probe placement (Banga et al., 2025). That said, the methods discussed in this chapter, to make comparisons across sessions and across subjects, are useful in either setting.

For the purposes of our discussion, we will therefore attempt to compare neural activity within and between subjects at the so-called region of interest (ROI) level, rather than at the individual neuron level. In the first instance, we will use data obtained from a single mouse (Hench) during a particular recording session (3) to obtain estimates of neural connectivity between individual neurons, using a variety of statistical tools. Then, we will summarise neural connectivity estimates by categorising individual neurons into 6 ROIs according to Table 5.2.2, and examine the connectivity within and between these ROIs. While ROI analysis is ubiquitous in the fMRI literature (Poldrack, 2007) extensions to the single-spiking setting are yet to be explored. This is likely due to the fact that, historically, recording technologies were unable to sample individual neurons from a wide range of brain regions.

Figure 5.2.3 gives an illustration of our analysis process using the BAPLA model applied to the left trial dataset aligned to stimulus onset for a single mouse (during a



(a) (b) (c)

Figure 5.2.3: ROI level summaries for left trial dataset aligned to stimulus onset with (*a*) interaction matrix (*b*) interaction matrix visualised with ROI structure and (*c*) ROI connectivity matrix.

single recording session) previously studied in Chapter 4. In particular, we visualise the estimated interaction matrix in terms of the defined ROIs, before plotting the so-called ROI connectivity matrix (5.4.1) to provide a more comparable estimate of neural interactions. The idea is to repeat this process for a variety of statistical tools, using data obtained from a single mouse, and make comparisons using the ROI connectivity matrix. This will allow us to answer our first question: *how sensitive are estimates of neural connectivity to the choice of statistical methodology?*

Then, to allow for a comparison across experimental subjects, we will obtain estimates of neural connectivity for multiple mice across multiple recording sessions. These estimates will also be analysed at the ROI level using the methods described above (i.e, categorise neurons according to Table 5.2.2 and examine ROI connectivity), enabling us to answer our second question: *how do estimates of neural connectivity compare across experimental subjects?*

5.3 Methods to Infer Brain Connectivity

In this section, we review a range of statistical tools and methods for the estimation of connectivity patterns from neural spike train data. We begin by describing classical methods such as correlation and partial correlation, before discussing spectral and model-based methods.

5.3.1 Classical Methods

Let $Y_{t,i}$ denote the binary time series representation (c.f. Chapter 4) of the spike train data recorded from neuron i at time t, for $i = 1, \ldots, d$ and $t = 1, \ldots, n$. For multiple trials of the same experiment, we use the notation $Y_{t,i}^{(k)}$ to denote the k^{th} trial.

We define the cross-covariance between two spike trains $Y_{t,i}^{(k)}$ and $Y_{t,j}^{(k)}$ as

$$\begin{aligned}\sigma_{ij}^{(k)} &:= \operatorname{cov}\left(Y_{t,i}^{(k)}, Y_{t,j}^{(k)}\right) \\ &= E_t\left[\left\{Y_{t,i}^{(k)} - E\left(Y_{t,i}^{(k)}\right)\right\}\left\{Y_{t,j}^{(k)} - E\left(Y_{t,j}^{(k)}\right)\right\}\right]\end{aligned}$$

where the expectation is taken over time $t = 1, \ldots, n$ within trial k. Given the simultaneous recordings of d neurons in trial k, we can collect all pairwise covariances in the $d \times d$ covariance matrix, denoted $\Sigma^{(k)}$, i.e.,

$$\Sigma^{(k)} = \begin{pmatrix} \sigma_{11}^{(k)} & \sigma_{12}^{(k)} \ldots & \ldots & \sigma_{1d}^{(k)} \\ \sigma_{21}^{(k)} & \sigma_{22}^{(k)} \ldots & \ldots & \sigma_{2d}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1}^{(k)} & \sigma_{d2}^{(k)} \ldots & \ldots & \sigma_{dd}^{(k)} \end{pmatrix}. \tag{5.3.1}$$

The above definition of cross-covariance provides a measure of the instantaneous dependence between pairs of neural spike trains. However, it can often be challenging to interpret as its value depends on the scale (or variability) of the underlying data. Consequently, it is more convenient to estimate the correlation coefficient (5.3.2) which standardises covariance estimates and provides a normalised measure of the linear association between two processes on the interval $[-1, 1]$. This is particularly useful when making comparisons between the strengths of connectivity across different pairs of neural processes.

Correlation

The correlation between two neural spike trains $Y_{t,i}^{(k)}$ and $Y_{t,j}^{(k)}$ is defined as

$$\rho_{ij}^{(k)} := \frac{\operatorname{cov}\left(Y_{t,i}^{(k)}, Y_{t,j}^{(k)}\right)}{\sqrt{\operatorname{var}\left(Y_{t,i}^{(k)}\right)\operatorname{var}\left(Y_{t,j}^{(k)}\right)}}. \tag{5.3.2}$$

While assessing dependence based on correlation is perhaps the most widely accepted and computationally least costly measure of functional connectivity between two neural spike trains, it has a number of limitations. For example, it cannot differentiate between direct and indirect connections. That is, it may include confounding variables that influence the interaction between $Y_{t,i}^{(k)}$ and $Y_{t,j}^{(k)}$, e.g., through another spike train $Y_{t,l}^{(k)}$ in the same network. Moreover, it is unable to detect causal interactions. However, despite its drawbacks, correlation is often used as a first port of call in many neuroscientific studies. Moreover, a winning solution to the First Neural Connectomics Challenge (a competition which provided simulated calcium imaging data and tasked participants with reconstructing the underlying network structure) featured correlation as a key component (Abril et al., 2018).

In the neuroscience literature, a common measure of association between two spike trains is based on the (normalised) cross-correlation histogram (CCH) (Perkel et al., 1967). Mathematically, this amounts to computing correlation estimates for neuron pairs at various time lags, i.e.,

$$\rho_{ij}^{(k)}(\tau) := \frac{\operatorname{cov}\left(Y_{t,i}^{(k)}, Y_{t+\tau,j}^{(k)}\right)}{\sqrt{\operatorname{var}\left(Y_{t,i}^{(k)}\right)\operatorname{var}\left(Y_{t,j}^{(k)}\right)}}.$$

where τ is the lag relative to the reference neuron Y_i. Then, these differences are binned into a histogram over τ. The resulting CCH has the following interpretation: A sharp peak at $\tau = 0$ suggests a direct excitatory interaction or a common input. A peak at $\tau > 0$ indicates that neuron i tends to fire before neuron j, suggesting the directional relationship $i \to j$. Conversely, a peak at $\tau < 0$ suggests that $j \to i$. A flat histogram is indicative of independence and a trough indicates inhibitory interactions. The width of the peak determines the timescale of the interaction, with narrower peaks indicative of fast time-scale interactions. For further details, see Harrison et al. (2013) or Kobayashi and Shinomoto (2025).

Partial Correlation

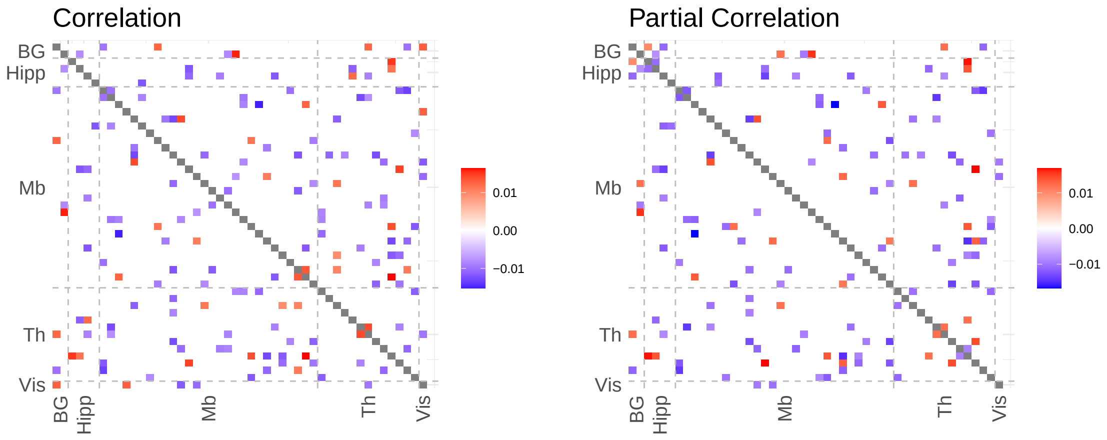
A popular alternative to traditional correlation, is that of partial correlation, which measures the direct dependence between a pair of neural spike trains *after* removing the linear effects from the other measured components in the brain network. Let $\Theta^{(k)} = \{\Sigma^{(k)}\}^{-1}$ denote the inverse covariance matrix,

$$\Theta^{(k)} = \left\{\Sigma^{(k)}\right\}^{-1} = \begin{pmatrix} \Theta^{(k)}_{11} & \Theta^{(k)}_{12} \dots & \dots & \Theta^{(k)}_{1d} \\ \Theta^{(k)}_{21} & \Theta^{(k)}_{22} \dots & \dots & \Theta^{(k)}_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \Theta^{(k)}_{d1} & \Theta^{(k)}_{d2} \dots & \dots & \Theta^{(k)}_{dd} \end{pmatrix}, \tag{5.3.3}$$

for the k^{th} trial. Then, the partial correlation between two processes $Y^{(k)}_{t,i}$ and $Y^{(k)}_{t,j}$, after removing the linear effects of the remaining processes $Y_{\cdot\setminus ij}$, is defined as

$$\rho^{(k)}_{ij|Y_{\cdot\setminus ij}} := -\frac{\Theta^{(k)}_{ij}}{\sqrt{\Theta^{(k)}_{ii}\Theta^{(k)}_{jj}}}, \tag{5.3.4}$$

where $\Theta^{(k)}_{ij}$ denotes the $(i, j)^{th}$ element of the matrix $\Theta^{(k)}$. An estimate of (5.3.3) can be obtained by simple inversion of the estimated covariance matrix (5.3.1) provided that it


Correlation

Partial Correlation

(a)

(b)

Figure 5.3.1: The averaged correlation and partial correlation matrices for the spike train data in Steinmetz et al. (2019) aligned to stimulus onset.

is positive definite. In high-dimensional settings, where $p > n$, or in cases where perfect collinearity between at least one pair of signals exists, the sample covariance matrix is singular, thus preventing the signals from being de-confounded (c.f. Chapters 2, 3).

Correlation Analysis of the Spike Train Dataset

We now implement these classical methods on the spike train data obtained from the subject mouse 'Hench' during the third recording session. In particular, we estimate correlations and partial correlations for neural data aligned to the onset of the visual stimulus which appeared on the left hand side of the screen.

Pairwise covariance estimates are obtained in the usual way, i.e.,

$$\hat{\sigma}_{ij}^{(k)} = \frac{1}{n}\sum_{t=1}^{n}(Y_{t,i}^{(k)} - \mu_i)(Y_{t,j}^{(k)} - \mu_j) \tag{5.3.5}$$

where $\mu_i = \frac{1}{n}\sum_{t=1}^{n} Y_{t,i}^{(k)}$ denotes the sample mean for neuron i, and the resulting correlation estimates are obtained by substitution of (5.3.5) into (5.3.2). Importantly, we note that the measures in (5.3.2) and (5.3.4) were defined for neural spike train data obtained during a single trial. However, in Steinmetz et al. (2019, 2024) data is col-

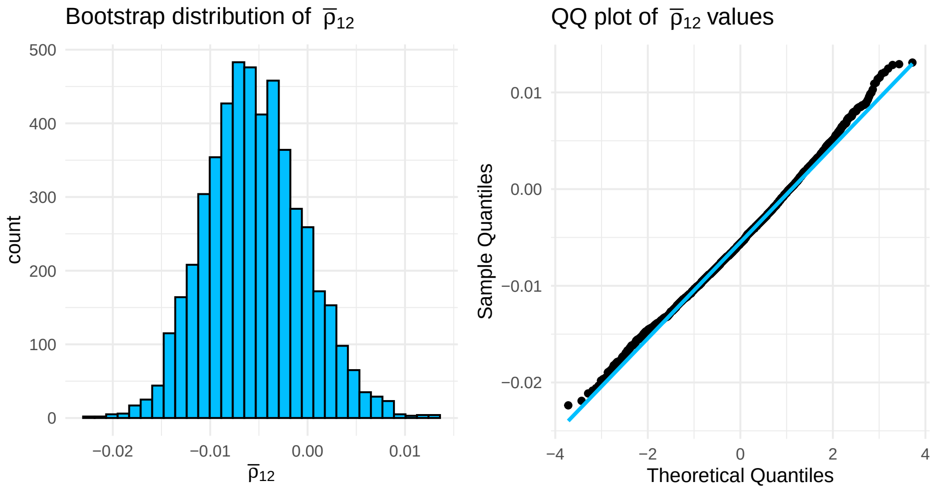


Figure 5.3.2: Bootstrap distribution of the sample mean and QQ-plot of the bootstrap means.

lected over a series of $k = 1, \ldots, m$ experimental trials. As such, we will calculate the average (partial) correlation matrices of the spike train data across trials, and determine statistical significance by testing the hypothesis

$$H_0 : \mathbb{E}_k \left[\rho_{ij}^{(k)}\right] = 0 \quad \text{vs} \quad H_1 : \mathbb{E}_k \left[\rho_{ij}^{(k)}\right] \neq 0,$$

where $\mathbb{E}_k[\cdot]$ denotes the population expectation taken with respect to trials $k = 1, \ldots, m$. Let $\bar{\rho}_{ij} = \frac{1}{m}\sum_{k=1}^{m} \rho_{ij}^{(k)},\ i, j = 1, \ldots, d$, denote the sample average of the (partial) correlations across the $k = 1, \ldots, m$ trials. Using the z-score as our test statistic, i.e.,

$$z = \frac{\bar{\rho}_{ij}}{s/\sqrt{m}},$$

where s is the sample standard deviation and m denotes the number of trials, we reject the null hypothesis if the p-value is less than the significance level α, i.e., $2\,\mathrm{pr}(Z > |z|) < \alpha$ where $Z \sim N(0, 1)$. More specifically, we threshold entries of the average (partial) correlation matrix based on the p-values, and only retain entries which are deemed significantly different from zero (Figure 5.3.1).

Use of the z-test is justified by way of the central limit theorem. That is, while the distribution of correlation values $\rho_{ij}^{(k)}$ across repeated trials ($k = 1, \ldots, m$) may not be normally distributed, the bootstrap distribution of the sample mean (where bootstrap samples are taken across trials for each $i, j = 1, \ldots, d$) is approximately normal (Figure 5.3.2). Of course, a permutation test could serve as an interesting alternative, however, given the observed normality of the bootstrapped sample, one might question whether the additional benefit outweighs the computational burden over a simple z-test.

In general, we observe very small non-zero values of correlation and partial correlation, likely due to the heavily zero-valued data, which are indicative of fairly spurious estimates of neural functional connectivity. Indeed, a limitation of these methods is that we need to decide a-priori on a bin size δ. However, in the current application, this decision is made to be the limit where $\delta = 1$ms corresponds to the refractory period.

5.3.2 Spectral Methods

Correlation and partial correlation summarise the strength of linear dependence between pairs of neural processes in the time domain. The corresponding second order statistics in the frequency domain are that of coherence and partial coherence. These measures provide additional insights into the oscillatory dynamics which drive neural function, revealing not only dependencies between processes, but also the frequencies at which these interactions occur. Indeed, frequency domain methods have been used throughout the literature to estimate brain networks at both the macroscopic (Fiecas and Ombao, 2011; Fiecas and von Sachs, 2014; Euan et al., 2019; Deb et al., 2024) and mesoscopic level (Eichler et al., 2003; Poli et al., 2016; Pinkney et al., 2024).

In this section, we adopt the framework discussed in Chapter 3, whereby the spike train data are modelled in continuous time[2] via the d-dimensional point process $N(t)$. Under the assumption that $N(t)$ is second order stationary (see Chapter 2), its second order properties can be characterised in the frequency domain via the spectral density

---

[2] Spectral methods may also be considered in the discrete time setting, e.g. using binned data.

matrix defined in (2.5.3). In particular, the spectral density matrix

$$S(\omega) = \begin{pmatrix} S_{11}(\omega) & S_{12}(\omega) & \dots & S_{1d}(\omega) \\ S_{21}(\omega) & S_{22}(\omega) & \dots & S_{2d}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ S_{d1}(\omega) & S_{d2}(\omega) & \dots & S_{dd}(\omega) \end{pmatrix},$$

captures the within and between dynamics of the multivariate point process, describing the variance in each process or the covariance between processes that is attributable to oscillations in the data at a particular frequency $\omega \in \mathbb{R}$.

In neuroscience, there are typically 5 frequency bands of interest in the range $(0.5, 50)$Hz. These are: delta $(0.5, 4)$Hz, theta $(4, 8)$Hz, alpha $(8, 12)$Hz, beta $(12, 30)$Hz and gamma $(30, 50)$Hz. By examining neural activity at each of these frequency bands, neuroscientists can gain a better understanding of the underlying processes which drive brain function. For example, in the electroencephalography (EEG) literature, studies have shown that these 5 frequency bands may be used as potential biomarkers for neurological diseases, such as attention deficit-hyperactivity disorder (ADHD) and schizophrenia (Newson and Thiagarajan, 2019).

Admittedly, frequency domain analysis of neural spiking data remains an understudied problem in both the statistics and neuroscience literature, compared to EEG or LFP signals. However, the methodology discussed in Chapter 3 provides a starting point for such analysis, enabling estimation of the dependence structure between neural processes at specified frequencies of interest. Indeed, as neural recording technologies continue to advance, these frequency-domain methods may be able offer new insights beyond those obtained by simple time-domain measures, such as correlation.

Figure 5.3.3 shows the estimated spectra (on the log-scale) for each neuron in the spike train dataset for the subject mouse Hench during the third recording session. Estimates were obtained using the mean-corrected and trial averaged estimator defined

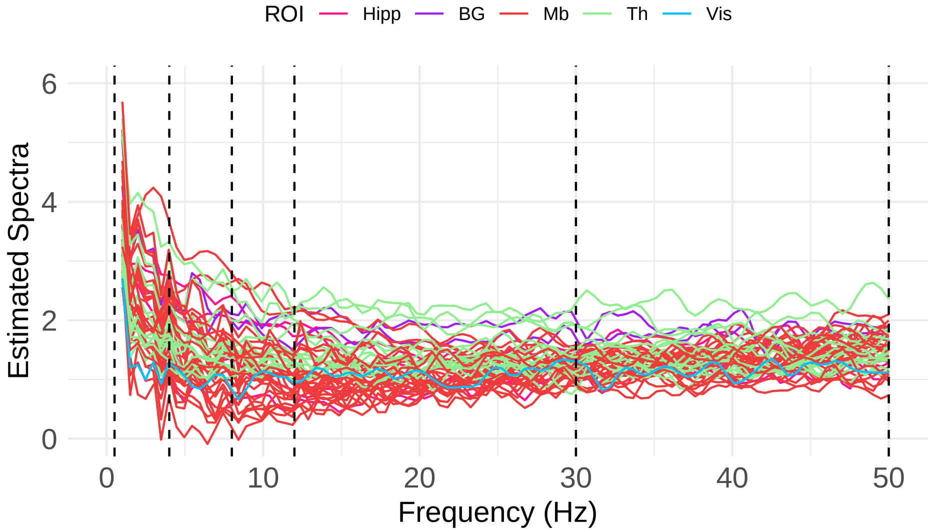


Figure 5.3.3: Estimated spectra (on the log-scale) for each neuron in the spike train dataset for subject mouse Hench during the third recording session. Colours denote the ROI for each spectrum.

in (2.5.6). In the following sections, we discuss methods to estimate neural connectivity within specific frequency bands, the boundaries of which are given by the dashed lines in the plot. However, we must proceed with caution when attempting to interpret the results of our spectral analysis, because it is likely that the fundamental assumption of stationarity does not hold for our spiking dataset. Importantly, if we have a non-stationary mean, then this could show as increased power at lower frequencies in the spectrum. Enhanced power at lower frequencies may also be due to long memory effects, indicative of processes with long-range dependence (Vellmer and Lindner, 2019). Thus, particular care must be taken when attempting to interpret the results in these lower frequency bands.

Coherence

The dependence structure among a population of neural spike train data can be characterised in the frequency domain via the squared spectral coherence (5.3.6). This quantity measures the linear association between a pair of processes at a particular frequency of interest ω, and is defined for two processes $N_i(t)$ and $N_j(t)$ as

$$R_{ij}^2(\omega) := \frac{|S_{ij}(\omega)|^2}{S_{ii}(\omega)S_{jj}(\omega)}, \tag{5.3.6}$$

where $S_{ii}(\omega)$ and $S_{ij}(\omega)$ denote the auto- and cross-spectral densities. If $R_{ij}^2(\omega) = 0$ for all ω, then the two processes are linearly independent. Conversely, a value of 1 indicates a perfect linear relationship. Therefore, the squared spectral coherence (5.3.6) provides a measure of the correlation structure between pairs of process in the frequency domain.

Partial Coherence

Similar to correlation in the time-domain, squared spectral coherence (5.3.6) cannot distinguish a direct effect from an indirect one. This means that the apparent dependence between two neurons may be driven entirely by an unobserved common input. Therefore, an alternative approach is to quantify the direct dependence between a pair of processes *after* taking into account the linear effects of the other processes. This can be achieved using a measure of the squared partial coherence, defined as

$$\rho_{ij}^2(\omega) := \frac{|\Theta_{ij}(\omega)|^2}{\Theta_{ii}(\omega)\Theta_{jj}(\omega)},$$

where $\Theta(\omega) := S(\omega)^{-1}$. This provides a normalised measure on $[0, 1]$ of the partial correlation structure between pairs of processes in the frequency domain.

Coherence Analysis of the Spike Train Dataset

We now implement these frequency-domain methods on the spike train data obtained from the subject mouse ‘Hench’ during the third recording session. In particular, we estimate the coherence and partial coherence for the left-trials stimulus onset dataset, at the five frequency bands of interest. To estimate coherence in a particular frequency band, we first require an estimate of the spectrum.

We define the trial-frequency smoothed estimator as

$$\hat{S}(\omega_*) = \frac{1}{F}\sum_{f=1}^{F}\frac{1}{m}\sum_{k=1}^{m}\bar{d}_{T;k}(\omega_f)\bar{d}_{T;k}^{H}(\omega_f), \tag{5.3.7}$$

where $\bar{d}_{T;k}(\omega)$ denote the mean-corrected coefficients (2.5.5) defined at the frequencies $\omega_f = 2\pi f$ which lie in the specified range of the frequency band, i.e., $f \in (0.5, 4)$Hz for the delta band estimate. Then, pairwise estimates of coherence are obtained by substitution of (5.3.7) into (5.3.6), i.e.,

$$\hat{R}^2_{ij}(\omega_*) = \frac{|\hat{S}_{ij}(\omega_*)|^2}{\hat{S}_{ii}(\omega_*)\hat{S}_{jj}(\omega_*)}.$$

Figure 5.3.4 shows the trial-frequency smoothed estimates of coherence across the delta, theta, alpha, beta and gamma frequency bands. Sparse estimates of coherence are obtained by thresholding the values based on the number of non-zero entries in the correlation matrix, thereby easing comparison between the time- and frequency-domain methods. More specifically, the number of edges is fixed to match those in Figure 5.3.1a, and each frequency band is shown to have 81 edges. Alternatively, we could determine statistical significance of entries in the coherence matrix using a one-sided hypothesis test based on the Goodman distribution (2.5.7) with appropriately modified degrees of freedom.

In general, the strength of the detected interactions decreases as a function of frequency, with the average value of thresholded coherence in the delta band being 0.784 compared to an average value of 0.005 in the gamma band. Additionally, we observe more structure in the lower frequency bands as evidenced by the strong within-region dependence for midbrain neurons in the delta and theta bands. Midbrain neurons also interact with neurons in the thalamus and basal ganglia in these lower frequency bands. The dependence becomes more widespread, both within and between regions, as we move to higher frequency bands (beta, gamma) which is more similar to the results

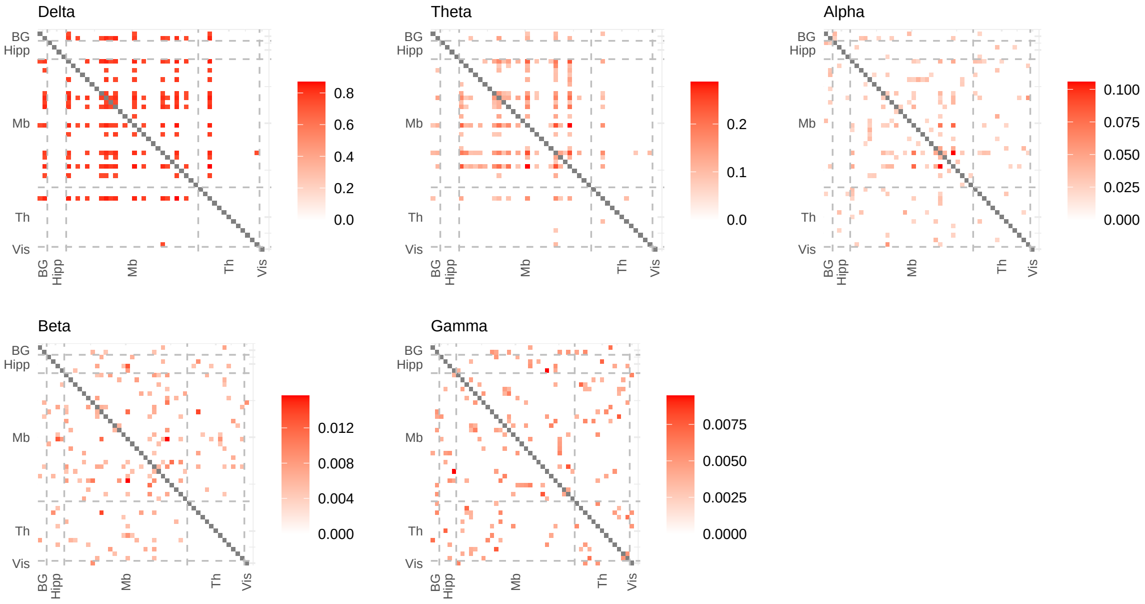


Figure 5.3.4: The trial-frequency smoothed estimates of coherence for the specified frequency bands

observed in the correlation analysis.

Partial coherence estimates are obtained using an adaption of the Lasso-type estimator discussed in Chapter 3, whereby we use (5.3.7) as input to the optimisation problem (3.3.1). This modified estimator is tuned using an eBIC-type criterion, as discussed in Chapter 3. Figure 5.3.5 shows the estimated partial coherence for the delta and theta bands[3]. In general, we observe similarities in structure when comparing to the corresponding coherence estimates, with a large proportion of the detected edges falling within the midbrain region. That said, more edges are detected using partial coherence, namely, 109 in the delta band and 205 in the theta band, compared to the 81 edges identified by our correlation-based estimator. However, these are of course point estimates, and further would involve testing for statistical significance, likely resulting in a reduction in the number of detected edges.

Interestingly, we observe that the partial coherence highlights more clearly, the neurons which drive the within region connectivity. That is, while the coherence estimates

---

[3] The estimated partial coherence at higher frequency bands were extremely sparse, and are therefore omitted here for brevity.

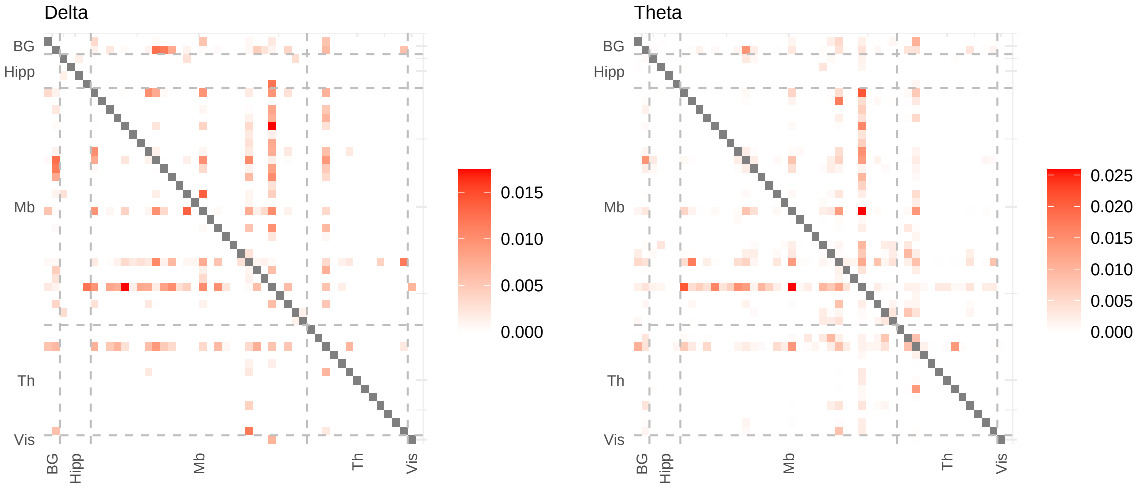


Figure 5.3.5: Estimated pairwise partial coherence for the delta and theta bands.

(Fig 5.3.4) showed that all neurons in the midbrain were connected, partial coherence was better able to discriminate between the level of connectivity for individual neurons.

Overall, we conclude that spectral methods have the potential to offer informative insights into neural functional connectivities. In particular, by summarising frequency-specific interactions, they are able to reveal patterns of neural connectivity which may remain hidden in the equivalent time-domain measures. A caveat, however, is that the spiking data should be stationary. In cases of non-stationary spiking data, the power at lower frequencies in the spectrum may be inflated, resulting in spurious low-frequency coherence.

5.3.3 Model-Based Methods

In this section, we discuss model-based methods which can be used to estimate the latent connectivity structure in the brain. Generally, these methods assume that the spike train data are generated according to some underlying statistical model. Then, estimates of neural connectivity are obtained by determining the model parameters which best fit the observed data. Three examples of model-based methods (Zhao et al., 2012; Song et al., 2013; Hall et al., 2018) were presented in Section 2.6.7 of this thesis. Two further examples of model-based approaches, in the form of the logistic and BAPLA models, were discussed in Chapter 4. In this section, we extend our discussion to include the method of Wang et al. (2025) before comparing the estimated network structures on the spike train data obtained from the subject mouse 'Hench'.

NeuroNetLearn

Recall the framework discussed in Section 2.6.6 for the graphical modelling of multivariate point processes data. In particular, we consider the class of Hawkes process models that allow for both excitatory and inhibitory effects (Costa et al., 2018; Chen et al., 2017a) via careful specification of the intensity functions (2.6.10).

In this model, the so-called connectivity coefficient φ_{ij} determines the underlying structure of the network, where φ_{ij} represents the influence of unit j's past events on unit i's intensity. Therefore, the connectivity coefficient provides an insight into the Granger causal interactions of the multivariate process (Shojaie and Fox, 2022). More specifically, a positive φ_{ij} implies that the past events of unit j excite future events of unit i, whereas a negative φ_{ij} represents an inhibitory influence of one units past events on the other.

As discussed, modern spike train datasets are high-dimensional, consisting of a large number of neurons recorded over a comparatively shorter time period. Therefore, recovering the underlying connectivity structure (via estimation of $\varphi \in \mathbb{R}^{p \times p}$) is not a straightforward task. That said, a number of estimation procedures have been developed for use in this setting. Specifically, ℓ_1-regularised estimation procedures have proved particularly popular (Hansen et al., 2015; Chen et al., 2017a).

Building on these sparsity inducing estimators, Wang et al. (2025) developed a statistical inference procedure, enabling hypothesis testing and the construction of confidence intervals for each entry in the connectivity matrix φ. In short, the authors extend the de-correlated score test framework of to the high-dimensional point process setting. Then, they develop confidence intervals by extending the work of Ning and Liu (2017) for VAR models, to the setting of multivariate Hawkes processes. The inference procedure is implemented in the `NeuroNetLearn` Python package, which is publicly available on PyPI. We refer the reader to Wang et al. (2025) for further details.

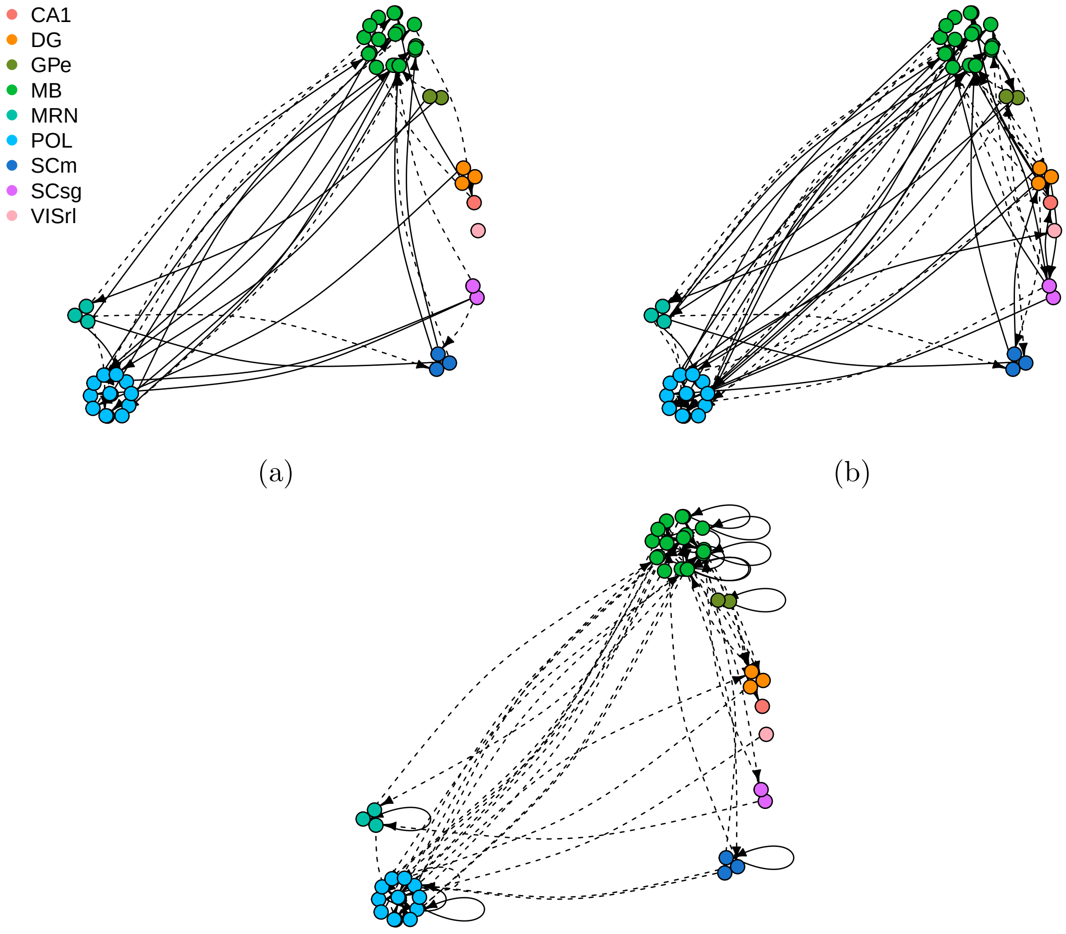
Model-Based Analysis of the Neural Spike Train Dataset

We now implement these model-based approaches on the spike train data obtained from the subject mouse 'Hench' during the third recording session. Estimates of neural connectivity using the method of Wang et al. (2025) are obtained for each trial via the `NeuroNetLearn` package in Python. These estimates are then averaged across trials, and statistical significance is determined via the hypothesis test discussed in Section 5.3.1.

For the logistic and BAPLA models, we obtain parameter estimates by maximising the joint likelihood (4.5.1), assuming independence across trials. The optimisation problem is solved using a coordinate descent algorithm and the strength of the regularisation is determined via a BIC-type criterion. In the case of the BAPLA model, the non-stationary firing rate is estimated using $m = 6$ basis functions. For these latter models, statistically significant edges are determined via the desparsification approach discussed in Chapter 4.

Graphical representations of the estimated interaction matrices are shown in Figure 5.3.6. In these graphs, each node (colour coded by brain region) represents a neuron and a directed edge from node i to node j represents neuron i's influence upon neuron j. Only statistically significant edges are displayed in the networks. A solid line represents an excitatory influence (i.e., $\Gamma_{ij} > 0$) and a dashed line indicates an inhibitory influence (i.e., $\Gamma_{ij} < 0$). Absence of an edge between two nodes means there is no interaction detected by the model.

In general, we found that more edges were detected under the logistic and NNL



(a)

(b)

(c)

Figure 5.3.6: Estimated networks for the left trials aligned to stimulus onset for (a) BAPLA (b) GLM and (c) NeuroNetLearn methods.

models (71 and 65, respectively), compared to a sparser estimate obtained under the BAPLA model (41). However, all three models identified more between region interactions compared to those within the same region (see Table 5.3.1). Furthermore, all models detected a large number of interactions between neurons in the midbrain (MB) and the posterior limiting nucleus of the thalamus (POL). These regional interactions (between MB and POL neurons) accounted for $51\%, 38\%$ and 46% of the total number of interactions in the BAPLA, logistic and NNL models, respectively. A similar result was discussed in Chapter 4, where it was highlighted that interactions between these brain regions are part of a wider circuit known to facilitate crucial brain functions like

		Interaction Type (%)		Region Type (%)	
Model	No. Edges	Excitatory	Inhibitory	Within	Between
BAPLA	41	60.98	39.02	26.83	73.17
Logistic	71	59.15	40.85	21.13	78.87
NNL	65	18.46	81.54	26.15	73.85

Table 5.3.1: Summary statistics for estimated interaction matrices.

sensory processing (Inagaki et al., 2022).

Interestingly, while the models are somewhat similar in terms of structure (i.e., similar patterns of within/between region connections) there are indeed obvious differences in terms of the type of interaction detected. In particular, while the logistic and BAPLA models detect a larger proportion of excitatory interactions, the NNL model detects solely inhibitory interactions between neurons. This is a somewhat unusual finding given that it is well known that the relative percentage of inhibitory neurons is much smaller than excitatory neurons in the mouse brain (Kajiwara et al., 2021). Furthermore, previous studies have shown that inhibitory interactions between neurons are harder to detect than excitatory ones (Zhao et al., 2012).

Admittedly, the theoretical framework of Wang et al. (2025) only guarantees support recovery, not sign recovery. Hence, while the connectivity matrix α *does* have a scientific interpretation (in terms of excitation/inhibition) it is not guaranteed to be reliably estimated. As such, the distinction between excitatory and inhibitory edges is omitted from the simulation studies and data application section in Wang et al. (2025).

The NNL model also detects a number of self-excitatory edges, while the logistic and BAPLA models do not. Upon further investigation (Appendix D) we found that self-excitatory edges were also detected by the logistic model in the other 3 datasets (i.e., wheel-movement datasets for left and right trials and the right trials dataset aligned to stimulus onset). Consequently, we hypothesise that direct modelling of the non-stationary trend possibly masks the effects of auto-correlation, thereby reducing possible self-excitatory edges. This highlights a clear difference of the BAPLA method compared to the logistic and NNL models.

Overall, the different model-based approaches produce varied estimates of neural connectivity when applied to our spike train dataset. While certain features appeared broadly consistent across the methods, the type of neuronal interaction (i.e., excitatory/inhibitory) differed substantially. This highlights a key ongoing challenge in the estimation and understanding of functional brain networks, in that it is difficult to draw reliable conclusions from different network estimation methods. In particular, in the absence of a known ground truth, it is difficult to determine which method, if any, provides the most accurate representation of the underlying network structure. However, these results demonstrate the importance of cautious interpretation when trying to infer network structures from model-based methods, highlighting the need for continued research in this area.

5.4 Comparison of Methods

Given the variety of methods available to estimate network structures from spike train data, it is important to assess how sensitive estimates of neural connectivity are to the choice of statistical methodology. In this section, we attempt to answer this question more broadly, by analysing the network structure at the ROI level, as described in Section 5.2. We begin by analysing data obtained from a single mouse (Hench) during a single recording session (3) in subsection 5.4.1. Then, we move to the multi-session and multi-subject setting in subsection 5.4.2.

5.4.1 Single Subject Analysis

Figure 5.4.1 shows the estimated interaction matrices for the BAPLA, logistic and NNL models, where each node (neuron) is coloured by ROI (see Table 5.2.2). To enable

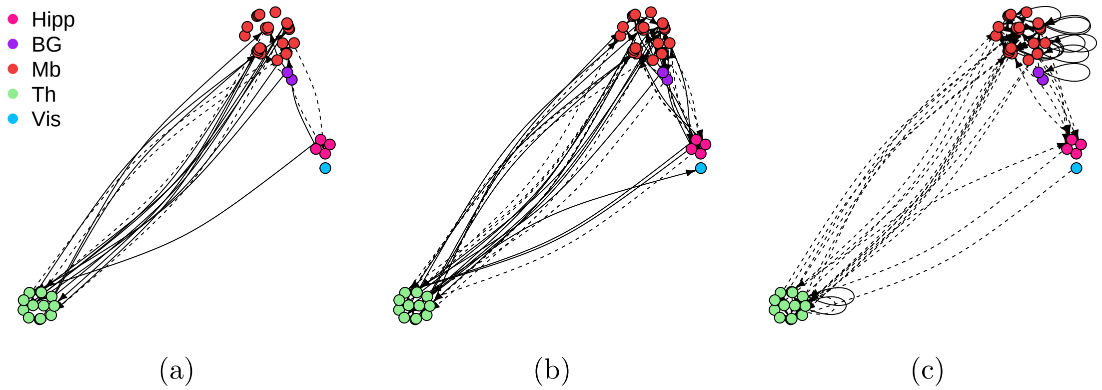


Figure 5.4.1: Interaction networks for (a) BAPLA (b) logistic and (c) NNL models colour coded by cluster.

further comparisons between all of the discussed methods, we appeal to the so-called ROI connectivity matrix, whose $(i, j)^{th}$ entry is defined as

$$\mathrm{E}_{ij} := \frac{\text{Number of edges from ROI } j \text{ to ROI } i}{\text{Total Number of edges}}, \tag{5.4.1}$$

which gives a measure of the empirical probability of an edge falling within a certain region of the adjacency matrix (Newman, 2003), corresponding to the interaction of ROI's i and j.

Figure 5.4.2 shows the ROI connectivity matrices for each of the time domain methods presented in the previous section, alongside the spectral coherence estimate for the alpha band. The grey rows and columns indicate that there were no frontal (Fr) neurons recorded for the subject mouse 'Hench'. The remaining $(i, j)^{th}$ entries represent the empirical probability of an edge from ROI j to i, with a value of 0 shown in white.

In general, the methods appear broadly similar, each suggesting that there are strong within region interactions in the midbrain, and also strong interactions between neurons in the midbrain and the thalamus. Of course, there are some structural differences. For example, the classical methods produce symmetric ROI connectivity matrices, compared to the model-based methods which are directional. Moreover, the model-based methods provide sparser estimates of ROI connectivity, perhaps highlight-

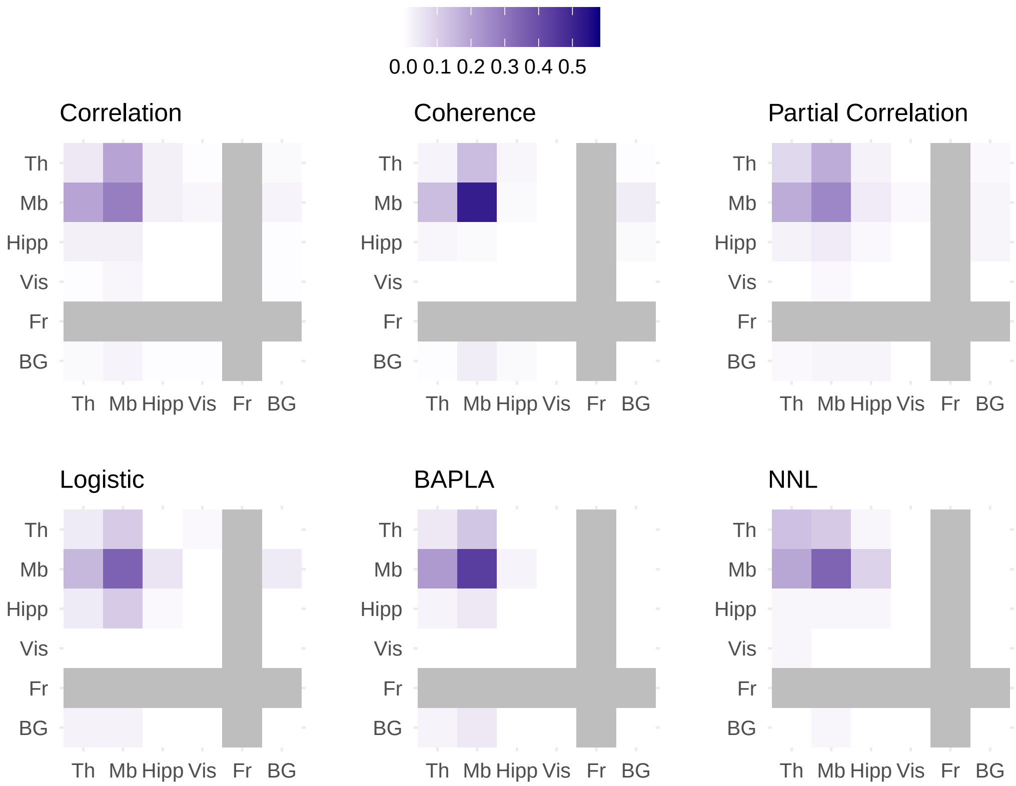


Figure 5.4.2: ROI connectivity matrices for the left-trial dataset aligned to stimulus onset.

ing more clearly, the interactions between the midbrain and the thalamus. As mentioned in Chapter 4, interactions between these brain regions are part of a wider circuit known to facilitate crucial brain functions like sensory processing and motor control (Inagaki et al., 2022). However, as midbrain neurons account for 58% of the total number of recorded neurons, there is a risk of bias due to the unbalanced sample size of the ROIs.

Figure 5.4.3 shows the ROI connectivity matrices for the partial coherence estimates obtained using the regularised estimator proposed in Chapter 3. The ROI connectivity matrices at the lower frequency bands (delta, theta, alpha) somewhat resemble those shown in Figure 5.4.2, with most of the dependence detected between the midbrain and thalamus and midbrain and basal ganglia ROIs. However, the ROI connectivity matrices obtained at higher frequency bands show that the dependence is more widespread, perhaps indicating that high-frequency interactions resemble nothing more

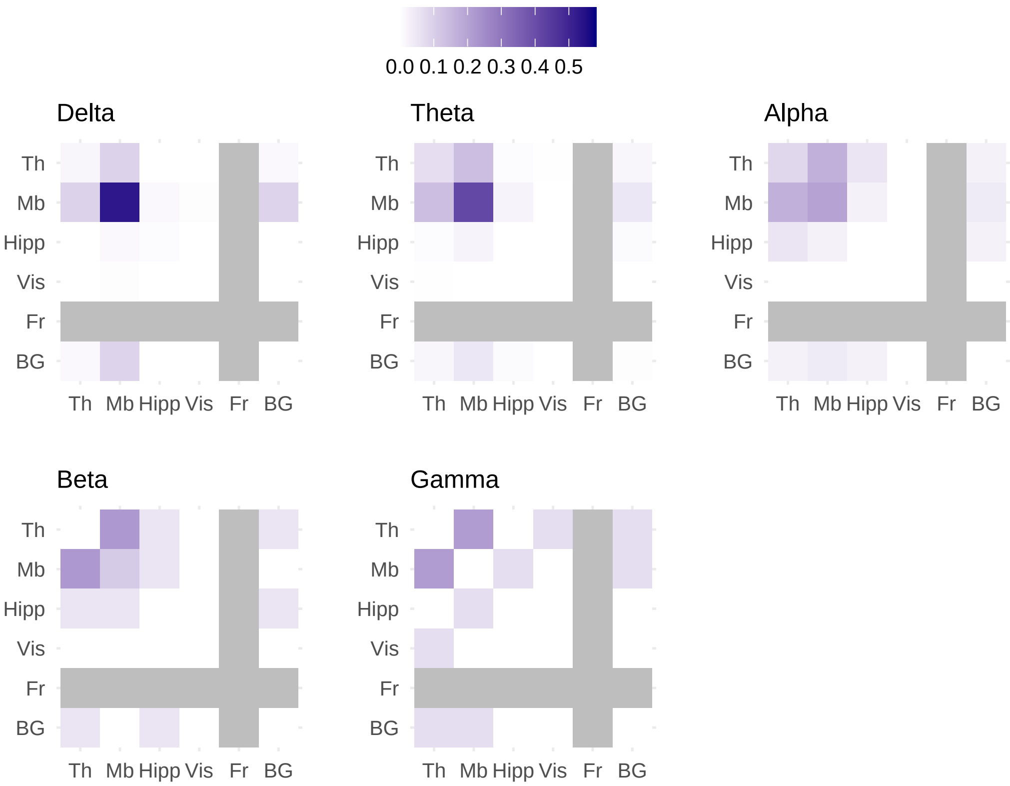


Figure 5.4.3: ROI connectivity matrices for partial coherence estimates at each of the considered frequency bands.

than random associations.

5.4.2 Multi-Subject Analysis

In this section, we switch our focus towards a multi-subject analysis of the neural spiking dataset. In particular, we are interested in comparing estimates of neural connectivity across the experimental subjects to reveal important similarities, or indeed differences, in how certain brain regions interact.

Several preprocessing steps were necessary prior to an application of the discussed network estimation methods to the experimental data. In particular, for each mouse, we use the following preprocessing procedure to obtain neural spike trains from each of the different recording sessions. Initial exploratory analysis revealed that individual neurons do not necessarily fire in every single experimental trial. Therefore, to avoid

Subject	Session	BG	FR	Hipp	Mb	Th	Vis	Total
Hench	1	0	23	8	0	40	2	73
	2	0	8	11	0	44	6	69
	3	4	0	5	42	17	2	70
	4	14	36	0	0	4	0	54
Lederberg	1	0	0	16	0	28	7	51
	2	0	6	2	65	11	0	84
	3	0	1	0	9	5	0	15
	4	0	0	0	0	22	0	22
Tatum	1	12	5	5	0	30	6	58
	2	0	0	4	42	0	2	48
Moniz	1	0	0	10	48	9	3	70
	2	0	0	6	0	27	1	34
Richards	1	19	1	3	0	0	0	23
	2	0	3	1	13	0	0	17
	3	0	0	0	17	10	0	27

Table 5.4.1: Number of neurons in each ROI across sessions for the top 5 mice.

subsequent issues with the network estimation methods, we chose to remove these trials (in which there were no spiking events) from our analyses. Moreover, to enable a fair comparison between the left and right trials for a specific experimental condition, we analyse neural data from the most active, i.e., highest firing, $k = 40$ trials, yielding an effective sample size of length 24 seconds (40 trials each of length 600ms). Neurons are only selected for analysis if they have an average firing rate of at least 10 spikes per trial. Lastly, we only consider the subset of neural data which is aligned to the onset of the visual stimulus. This is in contrast to Chapter 4 where we also investigated neural interactions in response to the onset of the wheel movement.

Table 5.4.1 shows the results of this process for the spike train data aligned to the onset of the visual stimulus, which appears on the left hand side of the screen. For the purposes of our discussion, we will focus on 5 mice with the largest number of recorded neurons - Hench, Lederberg, Tatum, Moniz and Richards. Overall, we consider a subset of 719 neurons from 6 different ROIs across 5 mice that exhibit sufficiently high average firing rates in the period of interest, i.e., $-0.2s$ before and 0.4s after onset of the visual

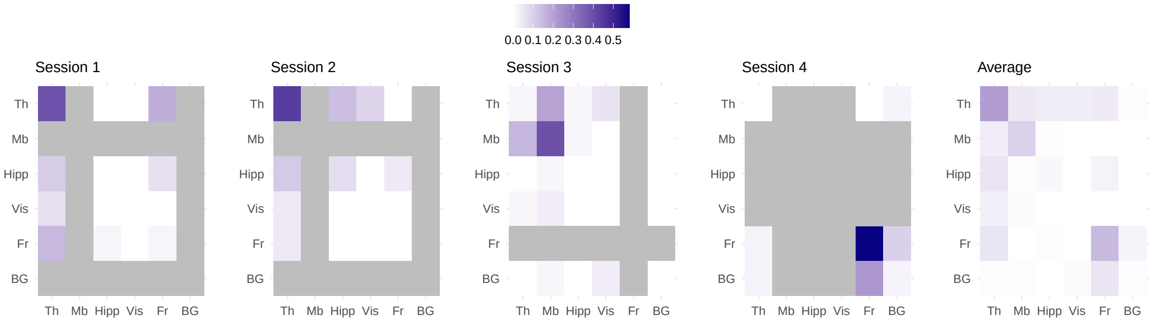


Figure 5.4.4: ROI connectivity matrices obtained using BAPLA method for subject mouse 'Hench' in each recording session. Results are based on left trial dataset aligned to stimulus onset.

stimulus.

Estimates of neural connectivity were obtained for each mouse, during each recording session, using correlation, partial correlation, the logistic model and the BAPLA model, as described in Section 5.3. Then, the ROI connectivity estimates were averaged across the recording sessions to provide an overall measure of the network structure. Figure 5.4.4 demonstrates this process with the subject mouse Hench for estimates of neural connectivity obtained using the BAPLA method.

Figures 5.4.5 and 5.4.6 show the ROI connectivity matrices obtained for the left and right trial datasets, respectively. Estimates are obtained for the top 5 mice using both exploratory (correlation and partial correlation) and model-based (logistic model and BAPLA model) methods. Due to time constraints, we omit the NNL and spectral models from our multi-subject analysis[4].

Generally, we observe more similarities in the ROI connectivity matrices for different methods applied to same mouse, compared to those across different mice using the same method. For example, all of the statistical methods detected strong within region interactions in the Basal Ganglia (BG) for the subject mouse 'Richards'. However, interactions within the BG were not detected in any other mouse. Of course, one can

---

[4] The NNL method had the longest runtime compared to all other approaches, e.g., 3hrs 28mins compared to 5.10 mins for BAPLA method on Ubuntu 22.04.01 with an Intel(R) Xeon(R) CPU.

deduce from Table 5.4.1 that such differences likely arise due to the fact that the samples of neurons recorded from each subject are highly variable. Thus, due to the nature of the dataset, we expect larger variability between mice, than estimation procedures.

5.5 Discussion

In this chapter, we have studied a range of statistical techniques to estimate neural connectivity from spike train data and provided a novel data analysis of the Steinmetz et al. (2019, 2024) dataset. In short, our aim was twofold: 1) to better position the methodologies proposed in Chapter 3 and 4 within the existing literature for the estimation and understanding of functional brain networks and 2) to compare neural connectivity estimates across experimental subjects.

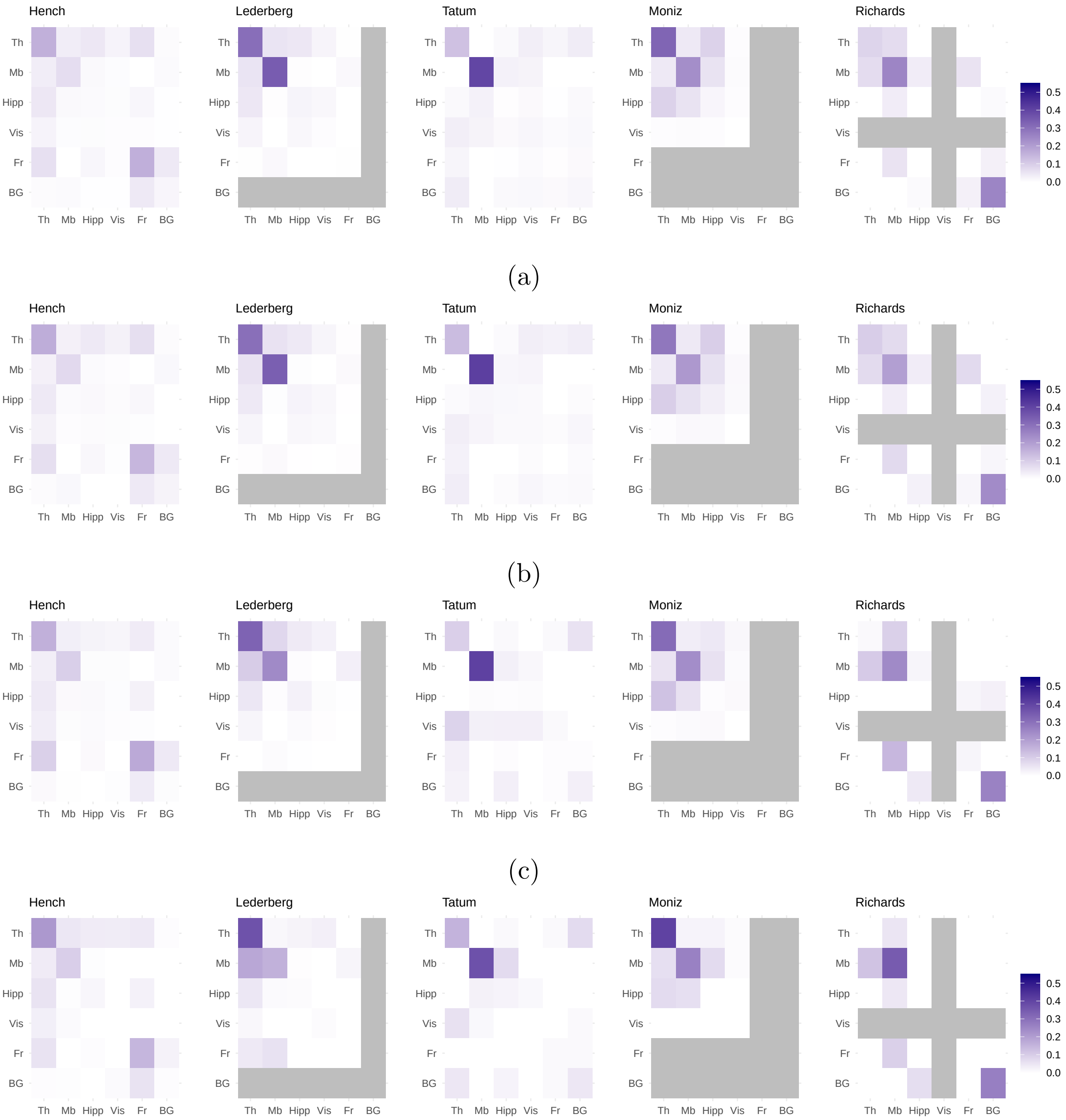
In terms of the single-subject analysis, our results showed that the different statistical methods were largely comparable, as evidenced by the similarities in the estimated ROI connectivity matrices. This overlap (in terms of structure) suggested that certain features of the network were particularly strong, and were insensitive to the chosen methodology.

By construction, the model-based approaches tended to produce sparser and more interpretable results compared to the exploratory methods. Contrastingly, the frequency domain methods revealed varying strengths of dependence at the different frequency bands of interest. In summary, the comparative nature of the single-subject analysis demonstrated important similarities in the functional brain networks obtained from a single mouse, during a single recording session.

Findings from the multi-subject analysis reinforced this notion of consistency among the statistical approaches. In particular, similarities in the ROI connectivity matrices across methods were observed in multiple mice, not solely the subject mouse 'Hench'. In terms of comparisons across mice for a single method, our results showed that there were indeed differences in the estimated network structures. However, these differences were somewhat expected, due to the heterogeneous sample of neurons considered for each subject (Table 5.4.1).

To enable a truly fair comparison across mice, we would require a dataset which had repeated measurements of neural activity from the same anatomical brain regions, for multiple mice. However, it is generally difficult to obtain such a dataset (Banga et al., 2025). More specifically, since neurons differ in their exact positions and morphology, a one-to-one matching across mice almost impossible.

Thus, since ROIs are conserved across animals, while individual neurons are not, ROI-level summaries are essential for meaningful cross-subject comparisons.



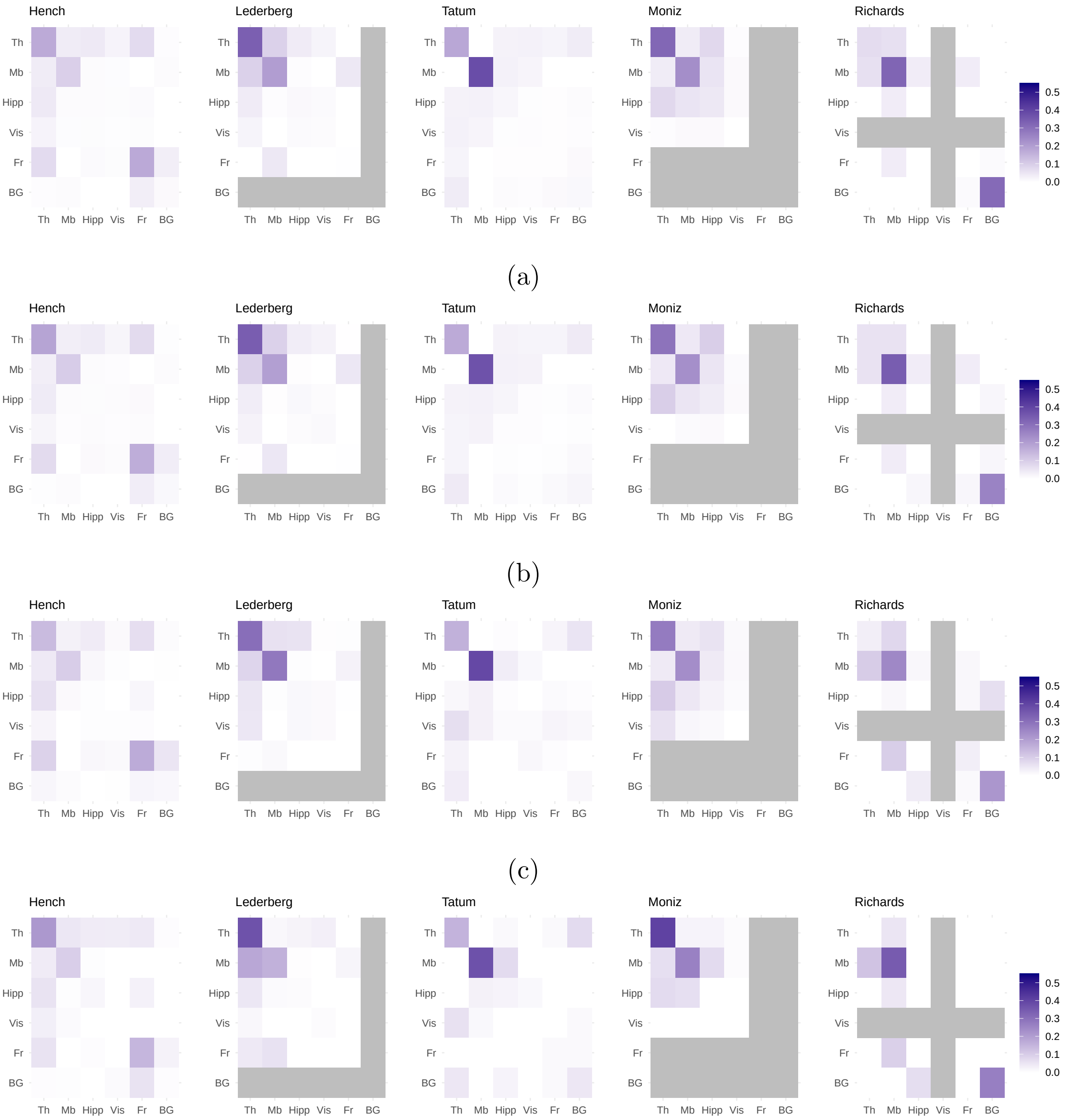
(a)

(b)

(c)

(d)

Figure 5.4.5: ROI connectivity matrices for left trial dataset aligned to stimulus onset. Results obtained for top 5 mice using (a) correlation (b) partial correlation (c) logistic model and (d) BAPLA Model.



(a)

(b)

(c)

(d)

Figure 5.4.6: ROI connectivity matrices for right trial dataset aligned to stimulus onset. Results obtained for top 5 mice using (a) correlation (b) partial correlation (c) logistic model and (d) BAPLA Model.

Chapter 6

Conclusions and Further Work

Estimating brain connectivity from neural spike train data presents exciting statistical challenges. Indeed, as neural recording technologies continue to evolve, it is crucial that the statistics literature progresses in parallel to enable efficient analysis of the resulting complex and high-dimensional datasets. This thesis has introduced novel statistical methodology and theory to aid this effort. To conclude, we present a summary of the key findings from each chapter, alongside related discussion on open challenges which remain as possible avenues for future research.

6.1 Summary of Contributions

Chapter 3 introduced novel methodology for the estimation of the inverse spectral density matrix from high-dimensional point process data. The main advantage of our approach is its ability to estimate the inverse spectrum when the dimension of the multivariate process is large, i.e., $p > m_T$, unlike existing methods which break down in high-dimensional settings. From a theoretical perspective, our main aim was to better understand the behaviour of our multi-trial periodogram estimate of the spectrum, in the context of high-dimensional point processes. Then, we applied regularisation-based methods to enable more efficient estimation of the inverse spectrum in these settings. From a methodological perspective, our framework combined a Whittle pseudo-likelihood with ridge and Lasso type penalities. Particular focus was given to the Lasso-type estimator, which eased interpretation of the spectrum and enabled association of the point process with a form of graphical model.

In Chapter 4, we estimated neural connectivity in a completely different framework, overcoming some of the practical limitations discussed in Chapter 3. In particular, we proposed a Bernoulli autoregressive partial linear additive (BAPLA) model to identify both excitatory and inhibitory interactions among a population of non-stationary neural spike trains. Our work extended the current literature for the identification of neural connectivity in two ways: 1) by accounting for non-stationary firing rates in the estimation procedure, and 2) by providing confidence intervals for entries in the interaction matrix. Synthetic experiments showed that our estimation procedure worked well in a variety of scenarios and that explicit modelling of the non-stationary firing rate was essential for accurate recovery of the interaction matrix. We also demonstrated the usefulness of our model on the spike train data from Steinmetz et al. (2019), where we studied the interactions between neural processes in response to various stimulus-response type neuroscience experiments.

In Chapter 5, we compared the methods presented in Chapters 3 and 4 with existing approaches in the literature. Utilising both single- and multi-subject data from the experiments of Steinmetz et al. (2019, 2024), our aim was to explore the ongoing challenges associated with the estimation and understanding of functional and effective brain networks. Our results showed that, according to our measure of ROI connectivity, there were indeed similarities between the different statistical methods. By construction, the model-based approaches produced sparser estimates of neural connectivity, yielding more interpretable results compared to the classical approaches. However, our comparative analyses demonstrated important similarities in the estimated functional brain networks. This finding was reinforced by the multi-subject analysis which showed that similarities across methods were observed in multiple mice, not only the subject mouse Hench. However, our results also showed important differences across mice for a single method, highlighting the limitations of our analysis due to the heterogeneous sample of neurons considered for each subject.

6.2 Future Work

In this section, we discuss a number of interesting avenues for future work which go beyond the ideas presented in this thesis. In Section 6.2.1, we discuss ideas extensions to the spectral framework presented in Chapter 3. Section 6.2.2 outlines possible adaptations to the methods proposed in Chapter 4. Finally, ongoing challenges for the practical implementation and comparison of methods are discussed in 6.2.3

6.2.1 Advances in Spectral Analysis

A natural next step to advance spectral analysis methods in the point process framework, would be to quantify the uncertainty associated with estimates of the inverse spectrum, and resulting partial coherence graphs. Existing work for the statistical inference of high-dimensional inverse covariance estimation was studied in Janková and van de Geer (2015). More specifically, given a sparse estimate of the precision matrix (e.g. graphical Lasso) the authors 'de-sparsified' it to remove the shrinkage bias introduced by the regularisation. Furthermore, under certain assumptions (e.g., sparsity and sub-Gaussian data) they showed that the individual entries in the resulting non-sparse estimator were asymptotically Gaussian, thereby enabling the construction of confidence intervals.

For our equivalent estimator in the frequency domain, it could therefore be useful to construct a version of Theorem 1 (Janková and van de Geer, 2015) to enable the construction of confidence intervals for entries in $\hat{\Theta}(\omega)$. That is, given a sparse estimate of the inverse spectrum, we would be interested in constructing a similar de-sparsified estimate, namely,

$$\hat{T}(\omega) := 2\hat{\Theta}(\omega) - \hat{\Theta}(\omega)\hat{S}(\omega)\hat{\Theta}(\omega).$$

Then, by the asymptotic complex normality of our Fourier coefficients, we hypothesise that

$$\underbrace{\sqrt{m}(\hat{T}_{ij}(\omega) - \Theta_{ij}(\omega))/\sigma_{ij}(\omega)}_{X_{ij}} \rightarrow^{d} \mathcal{N}^{c}(0,1), \tag{6.2.1}$$

where $\sigma_{ij}^2 = \Theta_{ii}(\omega)\Theta_{jj}(\omega) + \Theta_{ij}^2(\omega)$ is the asymptotic variance, and the Lasso estimate $\hat{\Theta}(\omega)$ might be used as a plug in estimator in place of the unknown $\Theta(\omega)$ for the construction of confidence intervals. Now, supposing that (6.2.1) holds and that $\mathrm{Re}(X_{ij}) \perp \mathrm{Im}(X_{ij})$, we would have, from the properties of the complex normal distribution, that

$$\mathrm{Re}(X_{ij}) \sim N(0, 1/2) \text{ and } \mathrm{Im}(X_{ij}) \sim N(0, 1/2),$$

from which we might be able to construct confidence intervals for entries in $\hat{\Theta}(\omega)$.

A further extension to the methods discussed in Chapter 3, could be to extend our regularisation framework to multi-frequency analysis. For example, to estimate the inverse spectrum at multiple frequencies in a band, we could consider application of a group-Lasso penalty (Danaher et al., 2014) to capture shared zero patterns across frequencies. That is, we might consider an estimator of the form

$$\hat{\Theta}(\cdot) := \mathrm{argmin}_{\Theta(\cdot)\in\mathcal{C}} \left\{ \sum_{\omega\in\mathcal{W}} \Big[-\log\det\{\Theta(\omega)\} + \mathrm{tr}\{\hat{S}(\omega)\Theta(\omega)\} \Big] + \lambda\|\Theta(\cdot)\|_1 \right\},$$

where $\Theta(\cdot) = \{\Theta(\omega)\}_{\omega\in\mathcal{W}}$ defined for frequencies ω in the set $\mathcal{W} = \{\omega_1, \omega_2, \omega_3, \dots\}$ and $\|\Theta(\cdot)\|_1 = \sum_{q\neq r} \sqrt{\sum_{\omega\in\mathcal{W}} |\Theta_{qr}(\omega)|^2}$ is the group-Lasso penalty and $\mathcal{C}$ is a suitably modified constraint set.

Importantly, such an adaptation to the regularisation framework would require further theoretical work, since the analysis provided in Chapter 3 does not go beyond that of single-frequency estimators.

Other obvious avenues include extending the methodology to non-stationary settings. As discussed in Chapters 4 and 5, the assumption of stationarity is often violated in stimulus-response type neuroscience experiments, thereby limiting the usability of the estimator proposed in Chapter 3. A possible answer may lie in extending the estimation methods in Roueff and Sachs (2019) to our particular setting.

Finally, one may consider further extensions beyond the current assumption of non-overlapping trials. Indeed, our regularisation methods, and analysis, could be extended to the more general multi-tapering framework of Walden (2000).

6.2.2 Accounting for Non-Stationary Firing Rates

Building on the methodology proposed in Chapter 4, one possible avenue for future research might involve including an additional penalty term to automatically select the number of basis functions, when estimating the non-stationary firing rate. In the context of our neuronal analysis, we argued that the potential benefits of including such a term would not outweigh its complications, nor its increased computational cost. However, there may indeed be some scenarios or alternative application areas, where this development could be particularly useful.

From a neuroscientific perspective, we might also consider adapting the BAPLA model to include additional categorical covariates or behavioural variables, specific to the neuroscience experiment of interest. For instance, we could include the stimulus type (e.g., visual - grating, chequerboard, blank) or trial type (e.g., success/failure) as categorical variables. It could also be advantageous to consider choosing a different set of basis functions to model the trend. For example, if we suspected a piecewise constant trend, then the Haar family of wavelets could be an interesting alternative (Vidakovic, 2009).

From an analysis point of view, the most obvious extension would involve providing theoretical justification for the adaptation to the desparsified estimator discussed in Section 4.3. Indeed, our synthetic experiments demonstrated empirical support for the validity of our proposed inferential procedure. However, future work might include tracking the asymptotic behaviour of the desparsified estimator in the presence of our additional model components. We hypothesise that the work of Gueuning and Claeskens (2016) or Drikvandi (2025) looking at a combination of ℓ_1 and ℓ_2 penalisation might provide an interesting starting point for such analysis.

Other theoretical improvements could involve extending the mean square error bounds presented in Hall et al. (2018) to include additional model components.

6.2.3 Data-Driven Challenges

The methods presented in Chapter 5, are by no means an exhaustive list of possible approaches for the important task of neuronal network estimation. Indeed, there are a range of approaches, including those discussed in Chapter 2 (e.g., Zhao et al., 2012; Song et al., 2013; Hall et al., 2018) which could serve as interesting alternatives. However, due to time constraints, these were not explored further.

For those that *were* discussed, some important practical challenges remain. For example, when modelling non-stationary neural data, direct application of the proposed spectral methods is somewhat naive. Furthermore, it is unclear to what extent estimating interactions at standard frequency bands makes sense when analysing neural spike train data, compared to standard LFP or EEG analysis.

More generally, further work could include evaluating the discussed methods on a synthetic dataset, where there is a known ground truth. One option might be to simulate data from a range of models (e.g., logistic, BAPLA, Hawkes, etc.) and compare the estimates obtained from the different techniques on each dataset. While we might expect models to perform best on their 'own' dataset, it would be interesting to compare the different results, highlighting which methods are most robust.

Alternatively, synthetic data simulated from a more biologically plausible model (e.g.,integrate-and-fire, Hodgkin–Huxley) could be used as a ground truth dataset. This would allow us to go one step further, and analyse to what extent estimates of functional and effective connectivity align with that of structural (or anatomical) connectivity. For related work in this area, see Poli et al. (2016).

Another area for future research concerns incorporating trial-history effects into existing statistical frameworks, to move beyond the assumption of independence across trials. Indeed, since animals are capable of learning and changing their behaviour over time, this common simplifying assumption is often violated. Extending the methodologies proposed here, to incorporate inter-trial dependencies, awaits future work.

Additionally, there are a number of practical challenges associated with the application of these various statistical methods to real-world spike train datasets. One example relates to the quality of data obtained from the neural recording mechanism. For instance, we saw in Chapter 5 that despite thousands of neurons reported to have been recorded by the Neuropixels probe, less than one hundred (per single mouse session) were selected for analysis based on our preprocessing procedure. Thus, there is potentially a trade-off between the total number of neurons recorded and the quality of the data actually extracted from these tools.[1] Indeed, errors made in the spike sorting process may also contribute to the quality (or lack thereof) of the data.

6.3 Concluding Remarks

To conclude, the increasing abundance of neural spike train data calls for the development of novel statistical tools. Of particular interest, are new methods which can accurately represent interactions between individual neurons in the brain. This is an exciting challenge at the intersection of statistics and neuroscience research, which con-

---

[1] We acknowledge that there are more up-to-date datasets (i.e., recorded using newer technologies) which may or may not suffer from these limitations.

tains a vast number of open problems yet to be solved. The statistical methodologies and theory presented in this thesis aim to make significant contributions to this exciting field. In summary, I hope that this thesis provides a balanced account, for those interested in both statistical theory, and the application of statistics in neuroscience.

Appendix A

Supplementary Material for Chapter 2

In this appendix, we provide further intuition regarding the discrete time approximation of the point process likelihood.

A.1 Discrete Time Approximation

Consider a simple point process with intensity $\Lambda(t)$ and likelihood

$$L = \left\{ \prod_{i=1}^{n} \Lambda(t_i) \right\} \exp\left(- \int_0^T \Lambda(t) dt \right), \tag{A.1.1}$$

where $t_1, \ldots, t_n$ are the event times.

Now, discretising time into bins, $t = 1, \ldots, T/\delta$, of equal width δ we have that $y_t = 1$ if an event occurs in bin t and $y_t = 0$ otherwise. A key assumption in this framework is that the bins are small enough, such that now two events can occur at the same time. As such, the point process representation implies that

$$\text{pr(event in bin } t) = \Lambda(t)\delta + o(\delta),$$

hence we model

$$\mathrm{pr}(y_t = 1) = \Lambda(t)\delta \quad \mathrm{pr}(y_t = 0) = 1 - \Lambda(t)\delta.$$

The discrete time (Bernoulli) likelihood is

$$L = \prod_t \{\Lambda(t)\delta\}^{y_t} \{1 - \Lambda(t)\}^{1-y_t},$$

and the log-likelihood is given by

$$\ell = \underbrace{\sum_{t=1}^{n} y_t \log\{\Lambda(t)\delta\}}_{(I)} + \underbrace{\sum_{t=1}^{n}(1 - y_t)\log\{1 - \Lambda(t)\delta\}}_{(II)}.$$

Let t_k be the bins containing events. We have

$$\begin{aligned}(I) &= \sum_{t:y_t=1} \log\{\Lambda(t)\delta\} \\ &= \sum_{i=1}^{n}\{\log(\Lambda(t_i) + \log(\delta)\} \\ &= \sum_{i=1}^{n} \log\Lambda(t_i) + n\log(\delta),\end{aligned}$$

where the first term matches that given in (A.1.1) and the second term is a constant. Applying the Taylor series expansion of $\log(1 - x)$ around $x = 0$ yields $\log(1 - x) = -x + \mathcal{O}(x^2)$. Therefore, we have $\log\{1 - \Lambda(t)\delta\} = -\Lambda(t)\delta + \mathcal{O}(\delta^2)$ and

$$\sum_{t:y_t=0} \log\{1 - \Lambda(t)\delta\} \approx -\sum_{t:y_t=0} \Lambda(t)\delta.$$

As $\delta \to 0$, we have

$$\sum_t \Lambda(t)\delta \to \int_0^T \Lambda(t)dt.$$

Collecting terms and exponentiating, we have

$$L = \delta^n \left\{ \prod_{i=1}^{n} \Lambda(t_i) \exp\left(-\int_0^T \Lambda(t) dt \right) (1 + o(1)) \right\}.$$

Thus, as the bin size $\delta \to 0$, the Bernoulli likelihood approximates (up to a constant) the point process likelihood.

Appendix B

Supplementary Material for Chapter 3

This Appendix outlines the supplementary material for Chapter 3, subsequently referred to as the main paper. Firstly, in Section B.1, we present the proof of Proposition 3.3.5. Then, in Section B.2, we present additional results for the experimental procedure outlined in the preliminaries section of the main paper. In Section B.3, we give a full description of the ADMM algorithm used for the regularised spectral estimator (RSE) with the lasso penalty. Then in Section B.4, we give additional details of the multivariate Hawkes process used in the synthetic experiments. Finally, in section B.5 we provide further analysis of the Bolding and Franks (2018) dataset and present partial coherence networks obtained for the the theta band.

B.1 Proof of Proposition 3.3.5

Proof. The proof follows a primal dual witness approach, and extends the arguments presented in Ravikumar et al. (2011) to allow for complex positive definite matrices. Similar to the proofs of Propositions 3.3.3 and **??**, we condition on the event $\mathcal{E}$ defined in (3.7.5) in the following analysis. The results are developed for a generic frequency

ω, which we omit below for ease of notation.

As in Ravikumar et al. (2011), the proof is broken down into a series of lemmas, thus easing the comparison between the real and complex analysis. Importantly, we highlight that a similar approach has previously been discussed in Deb et al. (2024).

Lemma B.1.1. *the regularised spectral estimator in* (3.3.1) *with the lasso penalty* (3.3.2) *has a unique solution* $\hat{\Theta}$ *characterised by*

$$\hat{S} - \hat{\Theta}^{-1} + \lambda \hat{Z} = 0, \tag{B.1.1}$$

where $\hat{Z}$ *is an element of the sub-differential* $\partial \|\hat{\Theta}\|_1$ *which has the following entries*

$$Z_{qr} = \mathrm{Re}(Z_{qr}) + i\,\mathrm{Im}(Z_{qr}) = \begin{cases} 1 & \text{if } q = r \\ \left[\mathrm{Re}(\Theta_{qr}) + i\,\mathrm{Im}(\Theta_{qr})\right] / |\Theta_{qr}| & \text{if } q \neq r \text{ and } \Theta_{qr} \neq 0 \\ \in \{\psi \in \mathbb{C} : |\psi| < 1\} & \text{if } q \neq r \text{ and } \Theta_{qr} = 0. \end{cases}$$

Lemma B.1.1 is a modified version of Lemma 3 in Ravikumar et al. (2011) where we have substituted our multi-trial periodogram estimate $\hat{S} \in \mathbb{C}^{p \times p}$ in place of the estimated covariance matrix $\hat{\Sigma} \in \mathbb{R}^{p \times p}$.

Proof. the regularised spectral estimator in (3.3.1) with the lasso penalty is a strict convex program. Therefore, if the minimum is obtained, then it is unique.

Suppose $\lambda > 0$. Then, by Lagrangian duality, we can rewrite our optimisation problem as

$$\min_{\Theta \in \mathcal{C}'} \left\{ -\log\det(\Theta) + \mathrm{tr}\left\{ \hat{S}\Theta \right\} \right\},$$

where $\mathcal{C}' := \{\Theta \in \mathbb{C}^{p \times p} : \Theta \succ 0,\ \Theta^H = \Theta,\ \|\Theta\|_1 \leq c(\lambda)\}$ for some $c(\lambda) < \infty$. Since both the off- and on-diagonal elements are bounded within the ℓ_1-ball, one can see that the minimum is attained, and as argued above, is also unique.

Recall the form of the regularised spectral estimator in (3.3.1) with the lasso penalty. By standard optimality conditions for convex programs, a matrix $\hat{\Theta} \in \mathcal{C}$ is an optimal solution if and only if the zero matrix belongs to the sub-differential of the objective. Equivalently, there must exist a matrix $\hat{Z}$ in the sub-differential of the norm $\|\cdot\|_1$ evaluated at $\hat{\Theta}$ such that

$$\hat{S} - \hat{\Theta}^{-1} + \lambda\hat{Z} = 0.$$

as claimed.

$\square$

Given that Lemma B.1.1 holds, we now outline the primal-dual witness approach which is central to our proof.

Primal-Dual Witness Approach

We define the witness, with oracle knowledge about the edge set $\mathcal{M}$, as

$$\dot{\Theta} = \arg\min_{\Theta(\omega)\in\dot{\mathcal{C}}} \left\{ -\log\det(\Theta) + \operatorname{tr}\left\{\hat{S}\Theta\right\} + \lambda\|\Theta\|_1 \right\}, \tag{B.1.2}$$

where $\dot{\mathcal{C}} := \{\Theta \succ 0 : \Theta^H = \Theta,\ \Theta_{\mathcal{M}^\perp} = 0\}$. Therefore, $\dot{\Theta}$ is a positive definite Hermitian matrix which correctly specifies the true edge set.

Let $\dot{Z}$ be a member of the sub-differential of the regulariser $\|\cdot\|_1$ evaluated at $\dot{\Theta}$. Now, considering each $(q, r) \in \mathcal{M}^\perp$, we write

$$\dot{Z}_{qr} := \frac{1}{\lambda}\left\{-\hat{S}_{qr} + \dot{\Theta}^{-1}_{qr}\right\},$$

which ensures that the primal-dual pair $(\dot{\Theta}, \dot{Z})$ satisfy the optimality condition in (B.1.1) replacing $(\hat{\Theta}, \hat{Z})$.

Finally, it remains to check that

$$|\dot{Z}_{qr}| < 1 \ \forall (q, r) \ \in \mathcal{M}^{\perp}. \tag{B.1.3}$$

This step ensures that $\dot{Z}_{qr}$ satisfies the necessary conditions to belong to the subdifferential. We will subsequently refer to (B.1.3) as the *strict dual feasibility* condition.

In the following discussion, we require some additional notation. Firstly, let $W \in \mathbb{C}^{p \times p}$ denote the discrepancy between the multi-trial periodogram and the true spectral density matrix, i.e.,

$$W = \hat{S} - S. \tag{B.1.4}$$

Secondly, let $\dot{\Delta} = \dot{\Theta} - \Theta_0$ measure the difference between the primal matrix and the ground truth. Finally, let $R(\dot{\Delta})$ denote the remainder of the first-order Taylor expansion of $\partial/\partial\Theta \left(-\log\det(\dot{\Theta})\right)$ around Θ_0. That is,

$$R(\dot{\Delta}) = \dot{\Theta}^{-1} - \Theta_0^{-1} + \Theta_0^{-1}\dot{\Delta}\Theta_0^{-1}. \tag{B.1.5}$$

We now provide sufficient conditions on W and $R(\dot{\Delta})$ for the strict dual feasibility condition in (B.1.3) to hold. The following lemma follows from Lemma 4 in Ravikumar et al. (2011).

Lemma B.1.2. *(Strict Dual Feasibility). Suppose that*

$$\max\left\{\|W\|_{\infty}, \|R(\dot{\Delta})\|_{\infty}\right\} \le \frac{\alpha\lambda}{8}. \tag{B.1.6}$$

Then, the vector $\dot{Z}_{\mathcal{M}^{\perp}}$ *constructed from* (B.1) *satisfies* $\|\dot{Z}_{\mathcal{M}^{\perp}}\|_{\infty} < 1$, *and therefore* $\dot{\Theta} = \hat{\Theta}$.

Proof. Using (B.1.4) and (B.1.5), we can rewrite the optimality condition (B.1.1) as

$$W - R(\dot{\Delta}) + \Theta_0^{-1}\dot{\Delta}\Theta_0^{-1} + \lambda\dot{Z} = 0. \tag{B.1.7}$$

Equation (B.1.7) can be re-written as an ordinary linear equation by 'vectorising' the matrices. We adopt the notation $\mathrm{vec}(M)$ or equivalently $\vec{M}$ for the vector version of matrix M obtained by stacking the rows of M into a single column vector.

Noting that

$$\mathrm{vec}\left(\Theta_0^{-1}\dot{\Delta}\Theta_0^{-1}\right) = \left(\Theta_0^{-1} \otimes \Theta_0^{-1}\right)\vec{\dot{\Delta}} = \Gamma\vec{\dot{\Delta}},$$

we rewrite equation (B.1.7) as two blocks of linear equations in terms of the disjoint decomposition $\mathcal{M}$ and $\mathcal{M}^{\perp}$ as follows:

$$\Gamma_{\mathcal{M}\mathcal{M}}\vec{\dot{\Delta}}_{\mathcal{M}} + \vec{W}_{\mathcal{M}} - \vec{R}_{\mathcal{M}} + \lambda\vec{\dot{Z}}_{\mathcal{M}} = 0 \tag{36a}$$

$$\Gamma_{\mathcal{M}^{\perp}\mathcal{M}}\vec{\dot{\Delta}}_{\mathcal{M}} + \vec{W}_{\mathcal{M}^{\perp}} - \vec{R}_{\mathcal{M}^{\perp}} + \lambda\vec{\dot{Z}}_{\mathcal{M}^{\perp}} = 0. \tag{36b}$$

Here, we used the fact that $\dot{\Delta}_{\mathcal{M}^{\perp}} = 0$ by construction. Further, using the fact that $\Gamma_{\mathcal{M}\mathcal{M}}$ is invertible, we can solve for $\vec{\dot{\Delta}}$ from (36a) and obtain

$$\vec{\dot{\Delta}}_{\mathcal{M}} = \Gamma_{\mathcal{M}\mathcal{M}}^{-1}\left[\vec{R}_{\mathcal{M}} - \vec{W}_{\mathcal{M}} - \lambda\vec{\dot{Z}}_{\mathcal{M}}\right].$$

Substituting this into (36b), we find that

$$\vec{\dot{Z}}_{\mathcal{M}^{\perp}} = \frac{1}{\lambda}\left(\vec{R}_{\mathcal{M}^{\perp}} - \vec{W}_{\mathcal{M}^{\perp}}\right) - \frac{1}{\lambda}\Gamma_{\mathcal{M}^{\perp}\mathcal{M}}\Gamma_{\mathcal{M}\mathcal{M}}^{-1}\left(\vec{R}_{\mathcal{M}} - \vec{W}_{\mathcal{M}}\right) + \Gamma_{\mathcal{M}^{\perp}\mathcal{M}}\Gamma_{\mathcal{M}\mathcal{M}}^{-1}\vec{\dot{Z}}_{\mathcal{M}}.$$

Taking the ℓ_∞ norm of both sides yields

$$
\begin{aligned}
\|\vec{\dot{Z}}_{\mathcal{M}}\|_\infty \le \frac{1}{\lambda}\left(\|\vec{R}_{\mathcal{M}}\|_\infty + \|\vec{W}_{\mathcal{M}}\|_\infty\right) + \frac{1}{\lambda}\vert\!\vert\!\vert\Gamma_{\mathcal{M}^\perp\mathcal{M}}\Gamma_{\mathcal{M}\mathcal{M}}^{-1}\vert\!\vert\!\vert_\infty \left(\|\vec{R}_{\mathcal{M}}\|_\infty + \|\vec{W}_{\mathcal{M}}\|_\infty\right) \\
+ \|\Gamma_{\mathcal{M}^\perp\mathcal{M}}\Gamma_{\mathcal{M}\mathcal{M}}^{-1}\vec{\dot{Z}}_{\mathcal{M}}\|_\infty.
\end{aligned}
$$

Under assumption A3 we know that $\vert\!\vert\!\vert\Gamma_{\mathcal{M}^\perp\mathcal{M}}\Gamma_{\mathcal{M}\mathcal{M}}^{-1}\vert\!\vert\!\vert_\infty \le \|\Gamma_{\mathcal{M}^\perp\mathcal{M}}\Gamma_{\mathcal{M}\mathcal{M}}^{-1}\|_1 \le (1-\alpha)$. Further, we know that $\|\vec{\dot{Z}}_{\mathcal{M}}\|_\infty \le 1$, since $\dot{Z}$ belongs to the sub-differential of the norm $\|\cdot\|_1$ by construction. Hence,

$$
\|\vec{\dot{Z}}_{\mathcal{M}^\perp}\|_\infty \le \frac{2-\alpha}{\lambda}\left(\|\vec{R}_{\mathcal{M}}\|_\infty + \|\vec{W}_{\mathcal{M}}\|_\infty\right) + (1-\alpha).
$$

Applying assumption (B.1.6) we arrive at

$$
\|\vec{\dot{Z}}_{\mathcal{M}^\perp}\|_\infty \le \frac{2-\alpha}{\lambda}\left(\frac{\alpha\lambda}{4}\right) + (1-\alpha) \le \frac{\alpha}{2} + (1-\alpha) < 1,
$$

as claimed. □

Lemma B.1.3. *(Control of remainder). Suppose that* $\|\dot{\Delta}\|_\infty \le \frac{1}{3\kappa_S d}$. *Then*

$$
\|R(\dot{\Delta})\|_\infty \le \frac{3}{2} d \|\dot{\Delta}\|_\infty^2 \kappa_S^3.
$$

Proof. We begin by rewriting the remainder term in (B.1.5) as

$$
R(\dot{\Delta}) = \underbrace{(\Theta_0 + \dot{\Delta})^{-1}}_{A} - \Theta_0^{-1} + \Theta_0^{-1}\dot{\Delta}\Theta_0^{-1}.
$$

Expanding the power series of the complex matrix A yields

$$\begin{aligned}(\Theta_0 + \dot{\Delta})^{-1} &= (\Theta_0(\mathbb{I} + \Theta_0^{-1}\dot{\Delta}))^{-1} \\ &= (\mathbb{I} + \Theta_0^{-1}\dot{\Delta}))^{-1}(\Theta_0)^{-1} \\ &= \sum_{k=0}^{\infty}(-1)^k(\Theta_0^{-1}\dot{\Delta})^k(\Theta_0)^{-1} \\ &= \Theta_0^{-1} - \Theta_0^{-1}\dot{\Delta}\Theta_0^{-1} + \Theta_0^{-1}\dot{\Delta}\Theta_0^{-1}\dot{\Delta}J\Theta_0^{-1},\end{aligned}$$

where $J = \sum_{k=0}^{\infty}(-1)^k(\Theta_0^{-1}\dot{\Delta})^k$. The above expansion can be verified under the assumption that $\|\dot{\Delta}\|_\infty < 1/(3\kappa_S d)$. More specifically, by the sub-multiplicativity of the matrix norm $|\!|\!| \cdot |\!|\!|_\infty$, we have that

$$\begin{aligned}|\!|\!|\Theta_0^{-1}\Delta|\!|\!|_\infty &\leq |\!|\!|\Theta_0^{-1}|\!|\!|_\infty |\!|\!|\dot{\Delta}|\!|\!|_\infty \\ &\leq \kappa_S d\|\dot{\Delta}\|_\infty < 1/3.\end{aligned} \tag{B.1.8}$$

The second inequality follows from the definition of κ_S in (3.3.5) and the fact that $\dot{\Delta}$ has at most d non-zeroes per row/column. We then use our assumption that $\|\dot{\Delta}\|_\infty < 1/(3\kappa_S d)$ to arrive at the final result.

Using the above convergent matrix expansion, we can rewrite the remainder term as

$$R(\dot{\Delta}) = \Theta_0^{-1}\dot{\Delta}\Theta_0^{-1}\dot{\Delta}J\Theta_0^{-1}. \tag{B.1.9}$$

Let e_q denote the unit vector with 1 in position q and zeroes elsewhere. Then, using (B.1.9) and standard vector inequalities we can write

$$\begin{aligned}\|R(\dot{\Delta})\|_\infty &= \max_{qr}\left|e_q'\Theta_0^{-1}\dot{\Delta}\Theta_0^{-1}\dot{\Delta}J\Theta_0^{-1}e_r\right| \\ &\leq \max_q \|e_q'\Theta_0^{-1}\dot{\Delta}\|_\infty \max_r \|\Theta_0^{-1}\dot{\Delta}J\Theta_0^{-1}e_r\|_1 \\ &\leq \max_q \|e_q'\Theta_0^{-1}\|_1\|\dot{\Delta}\|_\infty \max_r \|\Theta_0^{-1}\dot{\Delta}J\Theta_0^{-1}e_r\|_1.\end{aligned}$$

If we define the ℓ_1-operator norm for a generic matrix M as $\vert\vert\vert M\vert\vert\vert_1 := \max_{\|x\|_1=1} \|Mx\|_1$ then we can write

$$\begin{aligned} \|R(\dot{\Delta})\|_\infty &\leq \vert\vert\vert\Theta_0^{-1}\vert\vert\vert_\infty \|\dot{\Delta}\|_\infty \vert\vert\vert\Theta_0^{-1}\dot{\Delta}J\Theta_0^{-1}\vert\vert\vert_1 \\ &\leq \|\dot{\Delta}\|_\infty \vert\vert\vert\Theta_0^{-1}\vert\vert\vert_\infty \vert\vert\vert\Theta_0^{-1}\dot{\Delta}J^H\Theta_0^{-1}\vert\vert\vert_\infty \\ &\leq \|\dot{\Delta}\|_\infty \kappa_S \vert\vert\vert\Theta_0^{-1}\vert\vert\vert_\infty^2 \vert\vert\vert J^H\vert\vert\vert_\infty \vert\vert\vert\Delta\vert\vert\vert_\infty, \end{aligned} \tag{B.1.10}$$

where the second inequality follows from the fact that $\vert\vert\vert M\vert\vert\vert_1 = \vert\vert\vert M^H\vert\vert\vert_\infty$ for any complex matrix M, and the third inequality follows from the definition of κ_S in (3.3.5) and the sub-multiplicativity of the matrix norm $\vert\vert\vert \cdot \vert\vert\vert_\infty$.

To complete the proof, it remains to consider the term $\vert\vert\vert J^H\vert\vert\vert_\infty$ in (B.1.10). By the sub-multiplicative property of matrix norms we find that

$$\vert\vert\vert J^H\vert\vert\vert_\infty \leq \sum_{k=0}^{\infty} \vert\vert\vert\Theta_0^{-1}\dot{\Delta}\vert\vert\vert_\infty^k \leq \frac{1}{1 - \vert\vert\vert\Theta_0^{-1}\vert\vert\vert_\infty \vert\vert\vert\dot{\Delta}\vert\vert\vert_\infty} \leq \frac{3}{2}, \tag{B.1.11}$$

since $\vert\vert\vert\Theta_0^{-1}\Delta\vert\vert\vert_\infty \leq 1/3$ from (B.1.8). Substituting this into (B.1.11) yields the final result.

□

Lemma B.1.4. *(Control of $\dot{\Delta}$). Suppose that*

$$\tilde{r} := 2\kappa_\Gamma \left(\|W\|_\infty + \lambda\right) \leq \frac{1}{3\kappa_S d} \min\left\{1, \frac{1}{\kappa_S^2 \kappa_\Gamma}\right\}. \tag{B.1.12}$$

Then we have the elementwise ℓ_∞ bound

$$\|\dot{\Delta}\|_\infty = \|\dot{\Theta} - \Theta_0\|_\infty \leq \tilde{r}. \tag{B.1.13}$$

Proof. Following the proof of Lemma B.1.1, we know that the restricted problem (B.1.2) has a unique solution $\dot{\Theta}$. If we take partial derivatives of the Lagrangian of the restricted problem (B.1.2) with respect to unrestricted elements $\Theta_{\mathcal{M}}$, then we have that

$$\mathcal{U}_1(\Theta_{\mathcal{M}}) := \hat{S}_{\mathcal{M}} - \Theta_{\mathcal{M}}^{-1} + \lambda \dot{Z}_{\mathcal{M}} = 0. \tag{B.1.14}$$

The above zero gradient condition (B.1.14) is necessary and sufficient for an optimum of the Lagrangian problem, and thus has a unique solution $\dot{\Theta}_{\mathcal{M}}$.

Our aim is to bound the deviation of $\dot{\Theta}_{\mathcal{M}}$ from $\Theta_{0\mathcal{M}}$. This is equivalent to bounding $\dot{\Delta} = \dot{\Theta} - \Theta_0$ since $\Theta_{\mathcal{M}^\perp} = \Theta_{0\mathcal{M}} = 0$. As in Ravikumar et al. (2011), we consider the ℓ_∞-ball defined as

$$\mathcal{B}(\tilde{r}) := \left\{\Theta_{\mathcal{M}} \ \mid \ \|\Theta_{\mathcal{M}}\|_\infty \leq \tilde{r}\right\}, \tag{B.1.15}$$

where $\tilde{r}$ is defined in (B.1.12). The main idea is to show that a solution to the zero gradient condition (B.1.14) is contained within the ball (B.1.15). Consequently, we can conclude that $\dot{\Theta} - \Theta_0$ belongs to this ball.

We define the map $\mathcal{U}_2 : \mathbb{C}^{|\mathcal{M}|} \to \mathbb{C}^{|\mathcal{M}|}$ as

$$\mathcal{U}_2(\vec{\dot{\Delta}}_{\mathcal{M}}) := -(\Gamma_{\mathcal{M}\mathcal{M}})^{-1}(\vec{\mathcal{U}}_1(\Theta_{0\mathcal{M}} + \dot{\Delta}_{\mathcal{M}})) + \vec{\dot{\Delta}}_{\mathcal{M}},$$

where $\vec{\mathcal{U}}_1$ and $\vec{\dot{\Delta}}_{\mathcal{M}}$ denote vectorised versions of $\mathcal{U}_1$ and $\dot{\Delta}_{\mathcal{M}}$ respectively. By construction, we therefore have that $\mathcal{U}_2(\vec{\dot{\Delta}}_{\mathcal{M}}) = 0$ if and only if $\mathcal{U}_1(\Theta_{0\mathcal{M}} + \dot{\Delta}_{\mathcal{M}}) = \mathcal{U}_1(\dot{\Theta}_{\mathcal{M}}) = 0$.

Claim B.1.5. *Suppose that*

$$\mathcal{U}_2(\mathcal{B}(\tilde{r})) \subseteq \mathcal{B}(\tilde{r}).$$

Then, there exists a fixed point $\vec{\dot{\Delta}}_{\mathcal{M}} \in \mathcal{B}(\tilde{r})$. *By the uniqueness of the zero condition we can therefore conclude that* $\|\dot{\Theta}_{\mathcal{M}} - \Theta_{0\mathcal{M}}\|_\infty \leq \tilde{r}$.

Claim 6.2 therefore establishes the bound (B.1.13). A proof of the claim is given in Appendix C of Ravikumar et al. (2011).

□

We now have all the necessary parts to prove Proposition 3.3.5. To proceed, we first need to show that assumption (B.1.6) in Lemma B.1.2 holds. This will allow us to conclude that the restricted estimate is equal to lasso estimate, i.e., $\dot{\Theta} = \hat{\Theta}_G$.

It is easy to see that by choice of our regularisation parameter $\lambda = (8c/\alpha)\sqrt{\log p^\tau / m_T}$, under the event $\mathcal{E}$ (3.7.5), we have $\|W\|_\infty \leq \frac{\alpha}{8}\lambda$. Thus, in order to show that assumption (B.1.6) holds, it remains to demonstrate that $\|R(\dot{\Delta})\|_\infty \leq \frac{\alpha}{8}\lambda$. To do so, we use Lemmas B.1.3 and B.1.4 above.

We start by showing that assumption (B.1.12) holds under for our choice of λ and provided that we have sufficient trials m.

From Lemma 3.7.11 and our choice of regularisation parameter $\lambda = (8c/\alpha)\sqrt{\log p^\tau / m_T}$,

$$2\kappa_\Gamma \left(\|W\|_\infty + \lambda\right) \leq 2c\kappa_\Gamma \left(1 + \frac{8}{\alpha}\right) \sqrt{\frac{\log p^\tau}{m_T}},$$

provided that $c\sqrt{\log p^\tau / m_T} \leq \{\tilde{S}_{qq}(\omega)\tilde{S}_{rr}(\omega)\}^{1/2}/\varphi$.

Moreover, if we have sufficient trials

$$m_T > \left\{6c(1 + 8/\alpha)d\kappa_S^2\kappa_\Gamma \max\{1, \kappa_S\kappa_\Gamma\}\right\}^2 \tau \log p,$$

we can write

$$2c\kappa_\Gamma \left(1 + \frac{8}{\alpha}\right) \sqrt{\frac{\log p^\tau}{m_T}} \leq \frac{1}{3\kappa_S d} \min\left\{1, \frac{1}{\kappa_S^2 \kappa_\Gamma}\right\}. \tag{B.1.16}$$

showing that assumption (B.1.12) is therefore satisfied. We can now apply Lemma B.1.4 and conclude that

$$\|\dot{\Delta}\|_\infty \leq 2\kappa_\Gamma \left(\|W\|_\infty + \lambda\right) \leq 2c\kappa_\Gamma \left(1 + \frac{8}{\alpha}\right) \sqrt{\frac{\log p^\tau}{m_T}}. \tag{B.1.17}$$

Now, we can apply Lemma B.1.3 and equations (B.1.16) and (B.1.17) arriving at

$$\begin{aligned}\|R(\dot{\Delta})\|_\infty &\leq \frac{3}{2} d \|\dot{\Delta}\|_\infty^2 \kappa_S^3 \\ &\leq \left\{ 6\kappa_S^3 d \kappa_\Gamma^2 \left(1 + \frac{8}{\alpha}\right)^2 c \sqrt{\frac{\log p^\tau}{m_T}} \right\} \frac{\lambda\alpha}{8} \\ &\leq \frac{\alpha\lambda}{8}\end{aligned}$$

for sufficient trials

$$m_T > \left\{6c(1 + 8/\alpha)^2 d \kappa_S^3 \kappa_\Gamma^2\right\}^2 \tau \log p.$$

Now that we have shown that assumption (B.1.6) holds, we can conclude that $\dot{\Theta} = \hat{\Theta}_G$. Thus, the estimator $\hat{\Theta}$ satisfies the ℓ_∞-bound (B.1.17) of $\dot{\Theta}$, as claimed in Proposition 3.3.5 (a). Moreover, we have that $\hat{\Theta}_{\mathcal{M}^\perp} = \dot{\Theta}_{\mathcal{M}^\perp} = 0$ and thus $E(\hat{\Theta}) = E(\dot{\Theta})$.

Finally, noting that for all $(q, r) \in \mathcal{M}$ with $|\Theta_{0qr}| > 2c\kappa_\Gamma(1 + 8\alpha^{-1})\sqrt{\log p^\tau / m}$ we have

$$|\hat{\Theta}_{qr}| \geq |\hat{\Theta}_{qr} + \Theta_{0qr} - \Theta_{0qr}| \geq |\Theta_{0qr}| - |\hat{\Theta}_{qr} - \Theta_{0qr}| > 0.$$

Thus, we conclude that $E(\hat{\Theta}_G)$ includes all edges (q, r) with $|\Theta_{0qr}| > 2c\kappa_\Gamma(1 + 8\alpha^{-1})\sqrt{\log p^\tau / m_T}$ as claimed in 3.3.5 (b). Since the above analysis was conditioned on the event $\mathcal{E}$, Proposition 3.3.5 holds with probability $1 - \frac{8}{p^{\tau-2}} \to 1$, for any $\tau > 2$ and sufficient trials as stated.

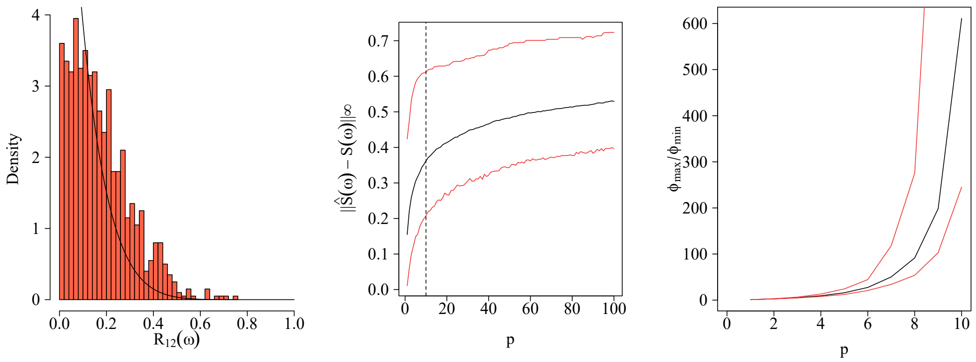
$\square$

B.2 Additional Material for Preliminaries Section

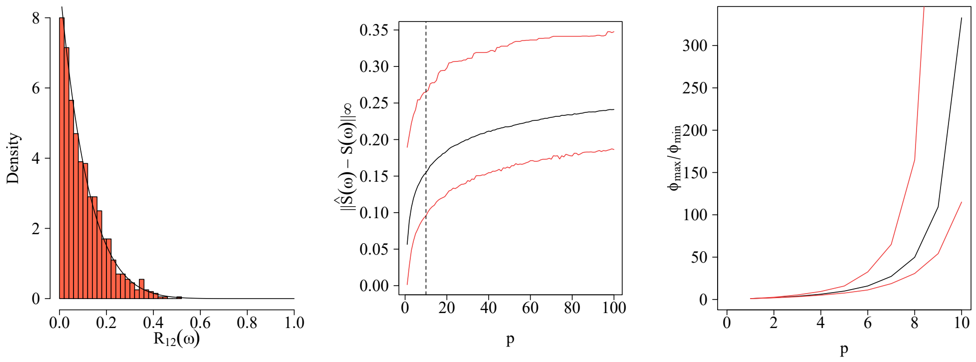
In this section, we present additional results for the experimental procedure outline in the preliminaries section of the main paper. Recall that we we consider a simple homogeneous Poisson point process with rate $\lambda = 1$, and obtain 1000 Monte-Carlo replicates of the multi-trial spectrum with $m = 10$ trials.

The results of the additional experiments are shown in Figure B.2.1, where we investigate the appropriateness of the asymptotic distributional results, for varying values of T. Firstly, we plot the histogram of $\hat{R}_{12}(\omega)$ alongside the theoretical density as defined in (3.2.3), with the substitution $R^2 = 0$. A visual comparison between Fig. B.2.1a and Fig. B.2.1b shows the improvement in fit to the asymptotic distribution when $T = 100$ compared to when $T = 1000$.

We also plot the deviation of the estimator from the true spectrum in terms of the element-wise maximum norm. We note the increased average estimation error for $T = 100$ compared to $T = 1000$. Finally, we highlight the difference in behaviour of the condition number for the estimated spectral matrix when $T = 100$ compared to $T = 1000$.



(a) $T = 100$



(b) $T = 1000$

Figure B.2.1: Results from the experimental procedure in the preliminaries section of the main paper repeated for different values of $T = (100, 1000)$.

B.3 ADMM Implementation

In this section we give details of the ADMM algorithm used to solve the optimisation problem (3.3.1) in the main paper with the lasso penalty. ADMM consists of a decomposition-coordination procedure, whereby solutions to small local sub-problems are coordinated to find a solution to a large global problem (Boyd et al., 2011). Often viewed as a an attempt to merge the advantages of dual decomposition and augmented Lagrangian methods, ADMM has been used throughout the time series literature to solve convex optimisation problems similar to those considered in this work (Jung et al., 2015; Nadkarni et al., 2016; Dallakyan et al., 2022).

We consider first rewriting the the minimisation as

$$\min_{\Theta \in \mathcal{C}} \left\{ -\log\det(\Theta) + \operatorname{tr}\{\hat{S}\Theta\} + \lambda \left\| Z \right\|_1 \right\}$$
$$\text{s.t. } \Theta - Z = 0,$$

where for ease of notation we have dropped the dependence on ω.

Then, following the work of Boyd et al. (2011), we define the augmented Lagrangian of the above problem as

$$L_\tau(\Theta, Z, U) := \operatorname{tr}\{\hat{S}\Theta\} - \log\det(\Theta) + \lambda||Z||_1 + \tau/2||\Theta - Z + U||_F^2,$$

where $|| \cdot ||_F$ denotes the Frobenius norm.

If we begin with arbitrary initialisations for $\Theta^{(0)}, Z^0$ and $U^{(0)}$, then the (scaled) ADMM method has the following update rules at the $(j+1)$ iteration

$$\Theta^{(j+1)} = \arg\min_{\Theta} \left\{ \operatorname{tr}\{\hat{S}\Theta\} - \log\det(\Theta) + \tau/2||\Theta - Z^{(j)} + U^{(j)}||_F^2 \right\} \tag{B.3.1}$$
$$Z^{(j+1)} = \arg\min_{Z} \left\{ \lambda||Z||_1 + \tau/2||\Theta^{(j+1)} - Z + U^{(j)}||_F^2 \right\} \tag{B.3.2}$$
$$U^{(j+1)} = U^{(j)} + \Theta^{(j+1)} - Z^{(j+1)}.$$

One can obtain closed form updates for (B.3.1) and (B.3.2). Consider first equation (B.3.1). If we denote the eigenvalue decomposition of the matrix $\tau(Z^j - U^j) - \hat{S}$ by QCQ^H, where $C = \mathbf{diag}(c_1, \ldots, c_p)$. Then the ADMM update rule (B.3.1) is reduced to the following form

$$\Theta^{(j+1)} = Q\tilde{C}Q^H$$

with diagonal matrix $\tilde{C}$ whose r^{th} diagonal element is $\tilde{c}_r = \frac{c_r + \sqrt{c_r^2 + 4\tau}}{2\tau}$.

The Z-minimisation step can be simplified using element-wise block soft thresholding. The ADMM update rule (B.3.2) becomes

$$Z_{qr}^{(j+1)} := \mathcal{S}_{\lambda/\tau}(\Theta_{qr}^{(j+1)} + U_{qr}^{j}),$$

where $\mathcal{S}$ is the block-thresholding operator defined as

$$S_{\kappa}(W_{qr}) := (1 - \kappa/|W_{qr}|)_{+} W_{qr},$$

where $|W_{qr}| = \sqrt{\mathrm{Re}(W_{qr})^2 + \mathrm{Im}(W_{qr})^2}$, $a_{+} := \max\{a, 0\}$ and $\mathcal{S}_{\kappa}(0) = 0$.

We use standard ADMM stopping criteria based on the primal and dual residuals which converge to zero as the ADMM algorithm proceeds.

B.4 Additional Material for Synthetic Experiments

In this section, we provide additional information regarding the multivariate Hawkes process used in the synthetic experiments described in Section 4 of the main paper.

In all experimental settings, the background intensity ν of the multivariate Hawkes process is set to 0.2 for all dimensions. Furthermore, to allow for an appropriate comparison among the scenarios, each considered setting is parameterised in such a way to ensure that the spectral radius of the Fourier Transform of the excitation function is approximately 0.83. In addition, all entries in the decay matrix $\beta \in \mathbb{R}^{p \times p}$ were set to 0.86 for both settings (a) and (b). The 3×3 excitation matrices for settings (a) and (b) are specified as

$$\alpha^* = \begin{pmatrix} 0 & 0.60 & 0 \\ 0 & 0.40 & 0 \\ 0 & 0 & 0.40 \end{pmatrix} \quad \text{and} \quad \alpha^* = \begin{pmatrix} 0.20 & 0.10 & 0.25 \\ 0.10 & 0.20 & 0.40 \\ 0.25 & 0.40 & 0.20 \end{pmatrix}.$$

This yields the following excitation matrices for models (a) and (b) when $p = 12$

$$\alpha = \begin{pmatrix}
0 & 0.60 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.40 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.40 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.60 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.40 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.40 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.60 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.60 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40
\end{pmatrix}$$

and

$$\alpha = \begin{pmatrix} 0.20 & 0.10 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.10 & 0.20 & 0.40 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0.40 & 0.20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.20 & 0.10 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.10 & 0.20 & 0.40 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0.40 & 0.20 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.20 & 0.10 & 0.25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.10 & 0.20 & 0.40 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.40 & 0.20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.20 & 0.10 & 0.25 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.10 & 0.20 & 0.40 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.40 & 0.20 \end{pmatrix}.$$

For model (c) we consider the following arbitrarily sparse excitation matrix for

$p = 12$

$$\alpha = \begin{pmatrix} 0 & 0 & 0.60 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & 0 & 0 \\ 0.60 & 0 & 0 & 0.80 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.80 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.50 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

For higher dimensions, each of the above α matrices are used to construct p-dimensional processes with block diagonal structure. For example, for dimension $p = 48$, the excitation matrix contains 4 blocks of the above 12×12 matrix where there are no interactions outwith these blocks.

As part of the review process, we considered an additional simulation setting in which the dependence structure of the multivariate Hawkes process was driven by a low-rank excitation matrix $\boldsymbol{\alpha}$, which was also non-sparse. To construct such a matrix, ensuring both stationarity and stability of the multivariate Hawkes process, we generated two random matrices $A \in \mathbb{R}^{p \times \tilde{r}}$ and $B \in \mathbb{R}^{\tilde{r} \times p}$ whose entries were drawn randomly from a uniform distribution, i.e., $a_{qr} \sim U(0, 0.25)$. We then set the excitation matrix $\alpha = AB$ which by definition has rank $\tilde{r}$, and set all entries in the decay matrix $\beta \in \mathbb{R}^{p \times p}$ such that the Fourier Transform of the excitation function is approximately 0.83, as in the other simulation settings. In the $p = 12$ case, this procedure yields the following

Table B.4.1: Simulation results over 100 replications for estimating the inverse spectral density matrix. All results are recorded at a particular frequency $\omega = 0.0628$ and are in the form of mean (standard error). Standard errors of $< 10^{-2}$ are omitted for brevity. Hyphenated entries (-) denote that the multi-trial periodogram matrix could not be inverted.

		Mean Squared Error - ISDM			
p	m	Inverted Periodogram	Ridge	Lasso_1	Lasso_2
12	10	-	0.30	0.33	0.39 (0.01)
	50	1.77 (0.06)	0.27	0.20	0.16
48	10	-	0.03	0.03	0.01
	50	8511.99 (1072.70)	0.02	0.02	0.01
96	10	-	0.01	0.01	0.00
	50	-	0.00	0.01	0.00

excitation matrix (rounded here to 2 decimal places for illustration purposes)

$$\alpha = \begin{pmatrix} 0.04 & 0.03 & 0.03 & 0.03 & 0.04 & 0.05 & 0.02 & 0.07 & 0.03 & 0.06 & 0.08 & 0.07 \\ 0.06 & 0.03 & 0.04 & 0.04 & 0.05 & 0.06 & 0.04 & 0.09 & 0.05 & 0.06 & 0.10 & 0.08 \\ 0.03 & 0.02 & 0.02 & 0.02 & 0.01 & 0.02 & 0.02 & 0.04 & 0.03 & 0.03 & 0.05 & 0.04 \\ 0.07 & 0.04 & 0.05 & 0.04 & 0.06 & 0.07 & 0.05 & 0.11 & 0.06 & 0.08 & 0.12 & 0.09 \\ 0.05 & 0.02 & 0.03 & 0.02 & 0.03 & 0.04 & 0.04 & 0.06 & 0.05 & 0.04 & 0.07 & 0.05 \\ 0.01 & 0.00 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.05 & 0.04 & 0.03 & 0.04 & 0.03 & 0.04 & 0.03 & 0.07 & 0.05 & 0.05 & 0.09 & 0.07 \\ 0.07 & 0.04 & 0.05 & 0.05 & 0.07 & 0.08 & 0.05 & 0.12 & 0.06 & 0.09 & 0.14 & 0.11 \\ 0.05 & 0.03 & 0.04 & 0.04 & 0.06 & 0.06 & 0.04 & 0.09 & 0.04 & 0.07 & 0.11 & 0.08 \\ 0.05 & 0.03 & 0.04 & 0.04 & 0.05 & 0.05 & 0.03 & 0.08 & 0.04 & 0.07 & 0.10 & 0.08 \\ 0.05 & 0.02 & 0.04 & 0.02 & 0.05 & 0.06 & 0.04 & 0.08 & 0.04 & 0.05 & 0.08 & 0.06 \\ 0.04 & 0.03 & 0.04 & 0.03 & 0.06 & 0.07 & 0.03 & 0.09 & 0.04 & 0.07 & 0.10 & 0.08 \end{pmatrix},$$

and consequently, all entries in decay matrix were set to be ≈ 0.73. We report the results (averaged across 100 replications) in Table B.4.1 below. Importantly, we note that the

Ridge and Lasso_1 estimators here are tuned using MSE, whilst the Lasso_2 estimator is tuned using an eBIC-type criterion. In this particular simulation setting, there is very little difference between the performance of the Ridge and Lasso estimators in terms of their MSE. However, as expected, the Lasso estimator shrinks some entries of the inverse spectral density matrix to exactly 0, and is therefore unable to correctly recover the true edge structure of the non-sparse graph. Therefore, in this particular scenario, use of the ridge estimator over the lasso estimator may be more appropriate.

B.5 Additional Material for Data Analysis

In this section, we present additional figures related to the data analysis section of the main paper. Figure B.5.2 shows the estimated partial coherence graphs for the theta band obtained under the 'laser on' and 'laser off' conditions, i.e., at frequencies in the range $(4, 8)$Hz. As was the case for the delta band, in general there are more edges detected for the 'laser on' condition (20 under $10mW/mm^2$ and 13 under $50mW/mm^2$ versus 6 and 11 under $10mW/mm^2$ and 13 under $50mW/mm^2$ for the 'laser off' setting).

The estimated partial coherence graphs were obtained using the Pooled Glasso estimator. Each Pooled estimator was tuned using the eBIC, whereby we search over a grid of λ values and select the value which minimises the eBIC. We record the regularisation parameter selected under each considered scenario in Table B.5.1

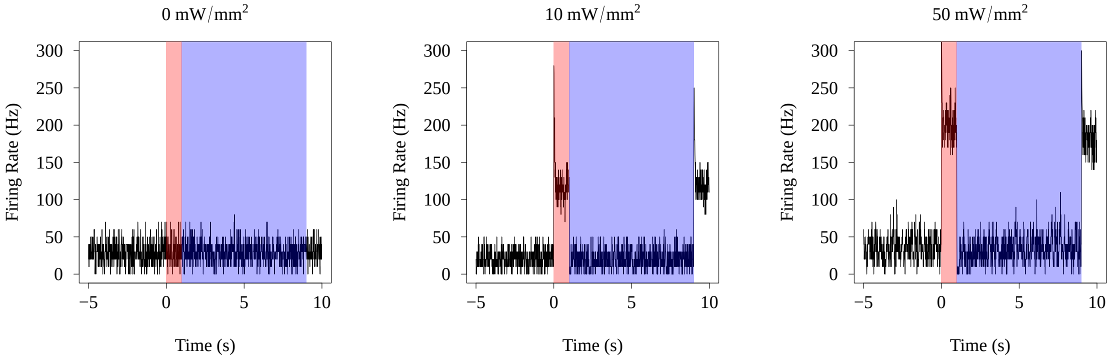
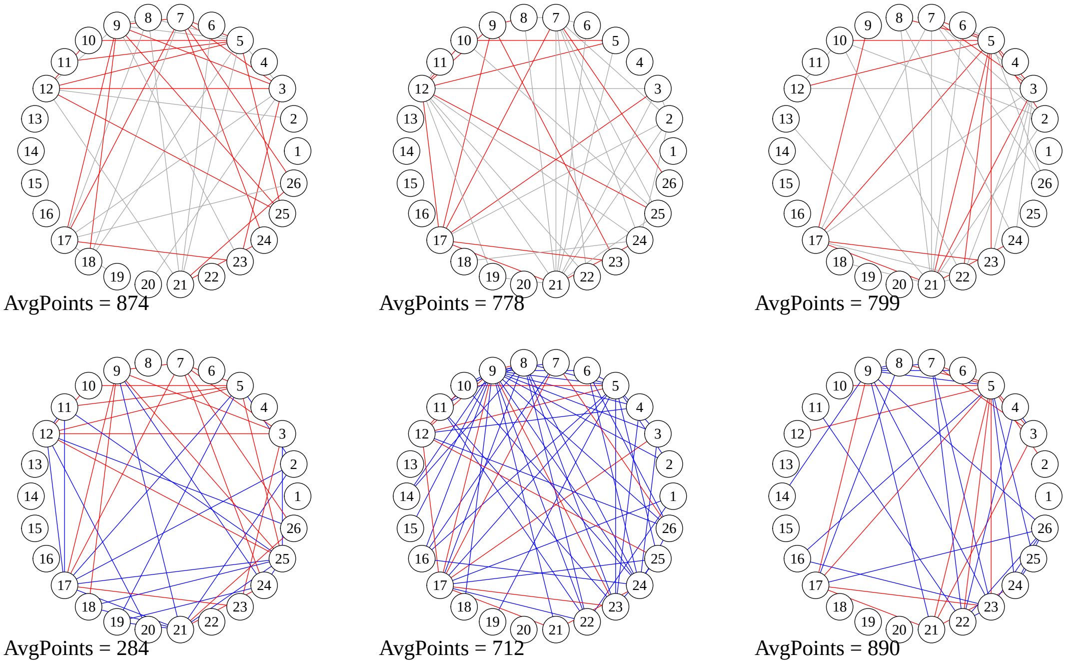


Figure B.5.1: Firing rate plots for the considered experimental settings.



(a) (b) (c)

Figure B.5.2: Estimated networks of neural interactions on the delta band using the spike train data from Bolding and Franks (2018). Each column shows the estimated networks for the laser off (top) and laser on (bottom) condition for laser stimuli (a) $0mw/mm^2$, (b) $10mw/mm^2$ and (c) $50mw/mm^2$. Common edges between the on and off conditions for each level of intensity are shown in red, and edges unique to the laser on condition are shown in blue. All other edges are shown in grey.

Under each scenario, the optimisation problem was solved via ADMM. Figure B.5.3 shows the stopping criteria for the ADMM algorithm for the considered frequency bands under both the 'laser on' and 'laser off' conditions. Since the stopping criteria converges to 0 as the number of iterations increases, we can be confident that the algorithm has converged. For completeness we also show the eBIC curves produced as part of the parameter tuning process.

Condition	Frequency Band	$0mW/mm^2$	$10mW/mm^2$	$50mW/mm^2$
Laser On	Delta	0.351	0.890	0.977
	Theta	0.351	0.464	0.739
laser Off	Delta	0.509	0.509	0.464
	Theta	0.351	0.385	0.351

Table B.5.1: Regularisation parameter values selected via eBIC for each condition at each intensity.

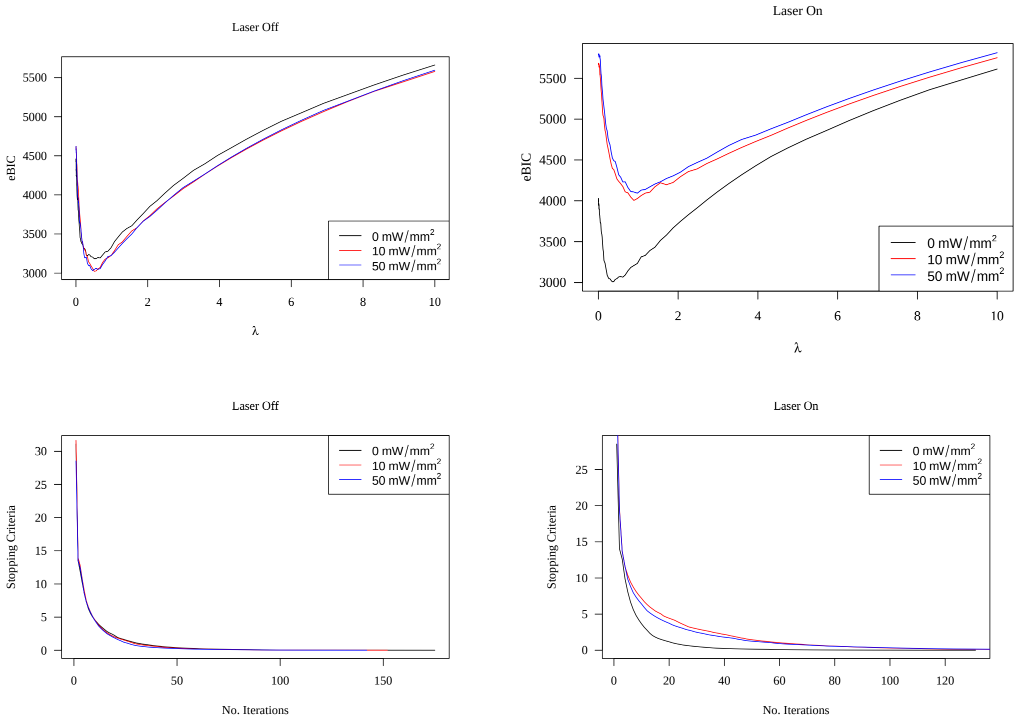


Figure B.5.3: eBIC curves and ADMM stopping criteria for Delta Band.

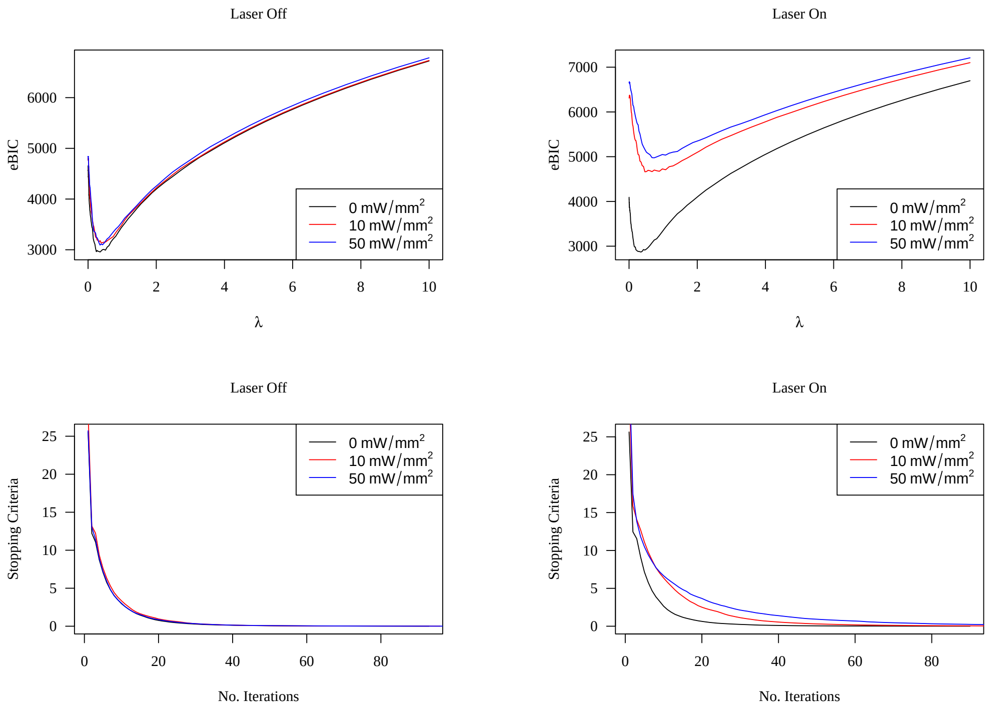


Figure B.5.4: eBIC curves and ADMM stopping criteria for Theta Band.

Appendix C

Supplementary Material for Chapter 4

In this appendix, we present the supplementary material for Chapter 4, subsequently referred to as the main paper. Firstly, we derive the coordinate wise updates outlined in algorithm 1 of the main paper. Then, we give additional results to accompany the simulation study, before presenting supplementary figures related to the data analysis section in the main paper.

C.1 Coordinate-wise updates

Recall the quadratic approximation to the likelihood from Section 4.2 of the main paper

$$\ell_Q(\beta_i, \boldsymbol{\gamma}_i, c_{i,1}, \ldots, c_{i,m}) = -\frac{1}{2}\sum_{t=1}^{n} w_{t,i}\left(z_{t,i} - \beta_i - \sum_{k=1}^{m} c_{i,k}\tilde{\phi}_k(t/n) - \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1}\right)^2 + C\left(\tilde{\beta}_i, \tilde{\boldsymbol{\gamma}}_i, \tilde{c}_{i1}, \ldots, \tilde{c}_{im}\right)^2.$$

Now, consider a coordinate descent step for solving

$$\min_{\boldsymbol{\theta}}\{-\ell_Q(\beta_i, \boldsymbol{\gamma}_i, c_{i,1}, \ldots, c_{i,m}) + \lambda\|\boldsymbol{\gamma}_i\|_1\}.$$

We begin by deriving the coordinate-wise update step for each element in $\boldsymbol{\gamma}_i^\top$, i.e., for each $j = 1, \ldots, d$. Suppose, we have estimates $\tilde{\beta}_i, \tilde{c}_{i,1}, \ldots, \tilde{c}_{i,m}$ and $\tilde{\gamma}_{i,l}$ for $l \neq j$ and we wish to partially optimise with respect to $\gamma_{i,j}$,

$$\begin{aligned}\frac{\partial \ell_Q}{\partial \gamma_{i,j}} &= -\sum_{t=1}^{n} w_{t,i} y_{t-1,j}\left(z_{t,i} - \beta_i - \sum_{k=1}^{m} c_{i,k}\tilde{\phi}_k(t/n) - \sum_{l\neq j}\gamma_{i,l}y_{t-1,l} - \gamma_{i,j}y_{t-1,j}\right)\\ &= \underbrace{-\sum_{t=1}^{n} w_{t,i} y_{t-1,j}\left(z_{t,i} - \beta_i - \sum_{k=1}^{m} c_{i,k}\tilde{\phi}_k(t/n) - \sum_{l\neq j}\gamma_{i,l}y_{t-1,l}\right)}_{\omega_j} + \gamma_{i,j}\underbrace{\sum_{t=1}^{n} w_{t,i}(y_{t-1,j})^2}_{\kappa_j}\\ &= -\omega_j + \gamma_{i,j}\kappa_j\end{aligned}$$

where we define ω_j and κ_j to simplify notation going forward.

For the regularisation term, we note that

$$\lambda\|\boldsymbol{\gamma}_i\|_1 = \lambda\sum_{j=1}^{p}|\gamma_{i,j}| = \lambda|\gamma_{i,j}| + \sum_{l\neq j}|\gamma_{i,l}|,$$

and therefore,

$$\partial_{\gamma_{i,j}}\lambda\|\boldsymbol{\gamma}_i\|_1 = \partial_{\gamma_{i,j}}\lambda|\gamma_{i,j}| = \begin{cases} -\lambda & \text{if } \gamma_{i,j} < 0\\ [-\lambda, \lambda] & \text{if } \gamma_{i,j} = 0\\ \lambda & \text{if } \gamma_{i,j} > 0.\end{cases}$$

Combining these terms, thereby computing the sub-differential of (7) and equating to zero, we arrive at

$$0 = -\omega_j + \gamma_{i,j}\kappa_j + \partial_{\gamma_{i,j}}\lambda\|\boldsymbol{\gamma}_i\|_1$$

$$0 = \begin{cases} -\omega_j + \gamma_{i,j}\kappa_j - \lambda & \text{if } \gamma_{i,j} < 0 \\ [-\omega_j - \lambda, -\omega_j + \lambda] & \text{if } \gamma_{i,j} = 0 \\ -\omega_j + \gamma_{i,j}\kappa_j + \lambda & \text{if } \gamma_{i,j} > 0 \end{cases}$$

For the second case, we must ensure that $\gamma_{i,j} = 0$ is a global minimum, i.e.,

$$\begin{aligned} 0 &\in [-\omega_j - \lambda, -\omega_j + \lambda] \\ -\omega_j - \lambda &\leq 0 \\ -\omega_j + \lambda &\geq 0 \\ -\lambda &\leq \omega_j \leq \lambda. \end{aligned}$$

Finally, we solve the first and third case for $\gamma_{i,j}$ and arrive at

$$\gamma_{i,j} = \begin{cases} \frac{\omega_j + \lambda}{\kappa_j} & \text{for } \omega_j < -\lambda \\ 0 & \text{for } -\lambda \leq \omega_j \leq \lambda \\ \frac{\omega_j - \lambda}{\kappa_j} & \text{for } \omega_j > \lambda. \end{cases}$$

Hence, we can write $\gamma_{i,j} = \frac{\mathcal{S}(\omega_j, \lambda)}{\kappa_j}$, where $\mathcal{S}$ is the soft-thresholding operator. Now, to update the intercept term, consider the partial derivative with respect to β_i

$$\frac{\partial \ell_Q}{\partial \beta_i} = \sum_{t=1}^{n} w_{t,i} \left(z_{t,i} - \beta_i - \sum_{k=1}^{m} c_{i,k}\tilde{\phi}_k(t/n) - \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} \right).$$

Equating to zero yields

$$\beta_i = \frac{\sum_{t=1}^{n} w_{t,i} \left(z_{t,i} - \sum_{k=1}^{m} c_{i,k} \tilde{\phi}_k(t/n) - \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} \right)}{\sum_{t=1}^{n} w_{t,i}}.$$

Finally, for the spline coefficients, we take the partial derivative with respect to $c_{i,k}$ for each $k = 1, \ldots, m$

$$\frac{\partial \ell_Q}{\partial c_{i,k}} = \sum_{t=1}^{n} w_{t,i} \phi_k(t/n) \left(z_{t,i} - \beta_i - \sum_{k=1}^{m} c_{i,k} \phi_k(t/n) - \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} \right).$$

Equating to zero yields,

$$c_{i,k} = \frac{\sum_{t=1}^{n} w_{t,i} \phi_k(t/n) \left(z_{t,i} - \beta_i - \sum_{j \neq k} c_{i,j} \tilde{\phi}_j(t/n) - \boldsymbol{\gamma}_i^\top \boldsymbol{y}_{t-1} \right)}{\sum_{t=1}^{n} w_{t,i} \tilde{\phi}_k^2(t/n)}.$$

C.2 Additional Material for Synthetic Experiments

In this section, additional figures and results are presented for the simulation study in the main paper. Table C.2.1 is the same as Table 1 in the main paper, but with the addition of standard errors. For brevity, we do not report standard errors of $< 10^{-2}$.

Figure C.2.1 shows estimates of $f_i(t/n)$ with $m = (10, 20, 40)$ for the normal and gamma models averaged over 100 repetitions (solid black line) with empirical 95% confidence bands in red and the ground truth in blue. Tables C.2.2 and C.2.3 give additional results for the sensitivity analysis described in the main paper.

Table C.2.1: Simulation results averaged over 100 replications for estimating the parameters of the BAPLA model.All results are in the form of mean (standard error). Standard errors of $< 10^{-2}$ are omitted for brevity.

			Estimation				Inference			
d	β_0	n	RMSE_Γ	MSE_{β_0}	MSE_f	AUC	AvgCov_s	AvgCov_{s^c}	AvgLen_s	AvgLen_{s^c}
Chain Graph										
10	0.10	1000	0.21	0.03	0.59 (0.05)	0.82	0.94	0.95	0.57 (0.01)	0.57
		2000	0.12	0.04	0.60 (0.04)	0.91	0.93 (0.01)	0.95	0.40	0.40
		5000	0.05	0.05	0.60 (0.02)	0.96	0.93 (0.01)	0.95	0.26	0.26
		10000	0.03	0.05	0.61 (0.02)	0.98	0.92	0.95	0.18	0.18
	-2.60	1000	0.86 (0.01)	0.51 (0.21)	5.37 (0.79)	0.56	0.95 (0.01)	0.96	1.90	1.90
		2000	0.61 (0.01)	0.12 (0.06)	1.42 (0.04)	0.64	0.96	0.96	1.31	1.32
		5000	0.27	0.04	0.28 (0.01)	0.71	0.95 (0.01)	0.95	0.82	0.82
		10000	0.16	0.04	0.24	0.73	0.94 (0.01)	0.95	0.58	0.58
50	0.10	1000	0.32	0.02	0.56 (0.05)	0.76	0.94	0.95	0.61	0.61
		2000	0.20	0.02	0.57 (0.03)	0.85	0.95	0.95	0.42	0.42
		5000	0.09	0.04	0.59 (0.02)	0.98	0.94	0.95	0.27	0.27
		10000	0.05	0.04	0.60 (0.02)	1.00	0.94	0.95	0.61	0.61
	-2.60	1000	0.94	0.39 (0.16)	4.46 (0.66)	0.53	0.96	0.96	2.36	2.36
		2000	0.84	0.06	0.76 (0.03)	0.58	0.96	0.96	1.49	1.49
		5000	0.40	0.03	0.27 (0.01)	0.69	0.96	0.95	0.89	0.89
		10000	0.27	0.02	0.22	0.72	0.95	0.95	0.61	0.62
100	0.1	1000	0.35	0.02	0.55 (0.04)	0.74	0.95	0.95	0.63	0.63
		2000	0.22	0.02	0.57 (0.03)	0.83	0.95	0.95	0.43	0.43
		5000	0.10	0.03	0.59 (0.02)	0.98	0.95	0.95	0.27	0.27
		10000	0.05	0.04	0.60 (0.02)	1.00	0.94	0.95	0.19	0.19
	-2.60	1000	0.96	0.34 (0.13)	3.54 (0.50)	0.52	0.95	0.96	2.90	2.91
		2000	0.86	0.07 (0.05)	0.91 (0.03)	0.57	0.96	0.96	1.67	1.68
		5000	0.45	0.03	0.27 (0.01)	0.68	0.96	0.96	0.93	0.93
		10000	0.29	0.02	0.22	0.70	0.95	0.95	0.63	0.63
Erdos-Renyi Random Graph										
10	0.10	1000	0.88	0.02	0.33 (0.03)	0.61	0.94 (0.01)	0.94 (0.01)	0.63 (0.01)	0.63 (0.01)
		2000	0.81	0.02	0.16 (0.01)	0.68	0.92 (0.02)	0.94 (0.01)	0.44	0.44
		5000	0.58	0.03	0.09	0.82	0.87 (0.03)	0.92 (0.02)	0.28	0.28
		10000	0.53	0.04	0.07	0.87	0.84 (0.06)	0.87 (0.04)	0.20	0.20
	-2.60	1000	1.03 (0.01)	0.33 (0.10)	1.30 (0.16)	0.51	0.94	0.94	1.70 (0.01)	1.69 (0.01)
		2000	0.97 (0.01)	0.07 (0.06)	0.52 (0.01)	0.53	0.94	0.94	1.17 (0.01)	1.17 (0.01)
		5000	0.92	0.03	0.11	0.57	0.94 (0.01)	0.94	0.73	0.73
		10000	0.85	0.02	0.08	0.63	0.92 (0.01)	0.93 (0.01)	0.51	0.51
50	0.10	1000	0.95	0.13 (0.02)	0.34 (0.03)	0.56	0.95	0.95	0.65	0.65
		2000	0.87	0.12 (0.01)	0.23 (0.01)	0.65	0.95	0.95	0.45	0.45
		5000	0.60	0.07	0.12	0.86	0.94	0.95	0.28	0.28
		10000	0.45	0.05	0.08	0.92	0.94	0.95	0.20	0.20
	-2.60	1000	1.02	0.14 (0.05)	1.44 (0.18)	0.51	0.95	0.95	2.06 (0.01)	2.07 (0.01)

				Estimation				Inference		
d	β_0	n	RMSE$_\Gamma$	MSE$_{\beta_0}$	MSE$_f$	AUC	AvgCov$_s$	AvgCov$_{s^c}$	AvgLen$_s$	AvgLen$_{s^c}$
		2000	1.00	0.04 (0.08)	0.35 (0.01)	0.51	0.95	0.95	1.32 (0.01)	1.33
		5000	0.97	0.02	0.14	0.53	0.95	0.95	0.79	0.79
		10000	0.96	0.02	0.10	0.55	0.95	0.95	0.55	0.55
100	0.10	1000	0.96	0.09 (0.01)	0.29 (0.02)	0.55	0.95	0.95	0.68	0.68
		2000	0.92	0.07 (0.01)	0.16 (0.01)	0.60	0.95	0.95	0.47	0.46
		5000	0.65	0.05	0.10	0.84	0.95	0.95	0.29	0.29
		10000	0.50	0.04	0.07	0.91	0.95	0.95	0.20	0.20
	-2.60	1000	1.03	0.11 (0.03)	1.09 (0.13)	0.50	0.95	0.95	2.45 (0.01)	2.45 (0.01)
		2000	1.01	0.03 (0.07)	0.32 (0.01)	0.51	0.95	0.95	1.45	1.45
		5000	0.98	0.01	0.11	0.52	0.95	0.95	0.82	0.82
		10000	0.96	0.01	0.08	0.55	0.95	0.95	0.56	0.56
						Stochastic Block Model				
10	0.10	1000	0.85 (0.01)	0.10 (0.01)	0.47 (0.04)	0.63	0.94 (0.01)	0.94	0.61 (0.01)	0.63
		2000	0.77 (0.01)	0.08 (0.01)	0.35 (0.02)	0.72 (0.01)	0.92 (0.01)	0.95	0.43	0.44
		5000	0.59	0.06	0.31 (0.01)	0.83	0.90 (0.02)	0.94 (0.01)	0.27	0.28
		10000	0.42	0.05	0.28 (0.01)	0.93	0.85 (0.04)	0.93 (0.01)	0.19	0.20
	-2.60	1000	1.00 (0.01)	0.26 (0.06)	2.01 (0.27)	0.54	0.94 (0.01)	0.96	1.69 (0.04)	2.42 (0.04)
		2000	0.94 (0.01)	0.09 (0.06)	0.65 (0.02)	0.56	0.93 (0.01)	0.95	1.17 (0.03)	1.67 (0.03)
		5000	0.90	0.04	0.20	0.58	0.93 (0.01)	0.95	0.74 (0.02)	1.05 (0.02)
		10000	0.80 (0.01)	0.04	0.17	0.65	0.88 (0.02)	0.94	0.52 (0.01)	0.74 (0.01)
50	0.10	1000	0.93	0.10 (0.01)	0.45 (0.04)	0.58	0.95	0.95	0.66	0.68
		2000	0.87	0.08 (0.01)	0.36 (0.02)	0.65	0.94	0.95	0.46	0.47
		5000	0.65	0.07	0.30 (0.01)	0.83	0.94	0.95	0.29	0.30
		10000	0.52	0.06	0.29 (0.01)	0.90	0.93	0.95	0.20	0.21
	-2.60	1000	1.03	0.27 (0.08)	2.27 (0.03)	0.51	0.95	0.95	2.08 (0.03)	2.56 (0.01)
		2000	1.00	0.07 (0.05)	0.47 (0.02)	0.51	0.95	0.95	1.38 (0.02)	1.72
		5000	0.97	0.03	0.20	0.53	0.95	0.95	0.84 (0.01)	1.06
		10000	0.95	0.03	0.16	0.56	0.94	0.95	0.59 (0.01)	0.74
100	0.10	1000	0.94	0.11 (0.01)	0.47 (0.04)	0.57	0.95	0.95	0.68	0.71
		2000	0.88	0.09	0.36 (0.02)	0.64	0.95	0.95	0.47	0.48
		5000	0.61	0.07	0.31 (0.01)	0.87	0.94	0.95	0.29	0.30
		10000	0.47	0.07	0.30 (0.01)	0.93	0.94	0.95	0.20	0.21
	-2.60	1000	1.04	0.24 (0.07)	1.88 (0.25)	0.51	0.95	0.95	2.25 (0.01)	2.79 (0.01)
		2000	1.00	0.07 (0.05)	0.44 (0.02)	0.51	0.95	0.95	1.40 (0.01)	1.79
		5000	0.98	0.03	0.20	0.52	0.95	0.95	0.82	1.07
		10000	0.95	0.03	0.16	0.55	0.95	0.95	0.57	0.75

Figure C.2.1: Estimates of $f_i(t/n)$ for the normal model (top) and gamma model (bottom) with 95% confidence intervals in red and compared to the ground truth in blue.

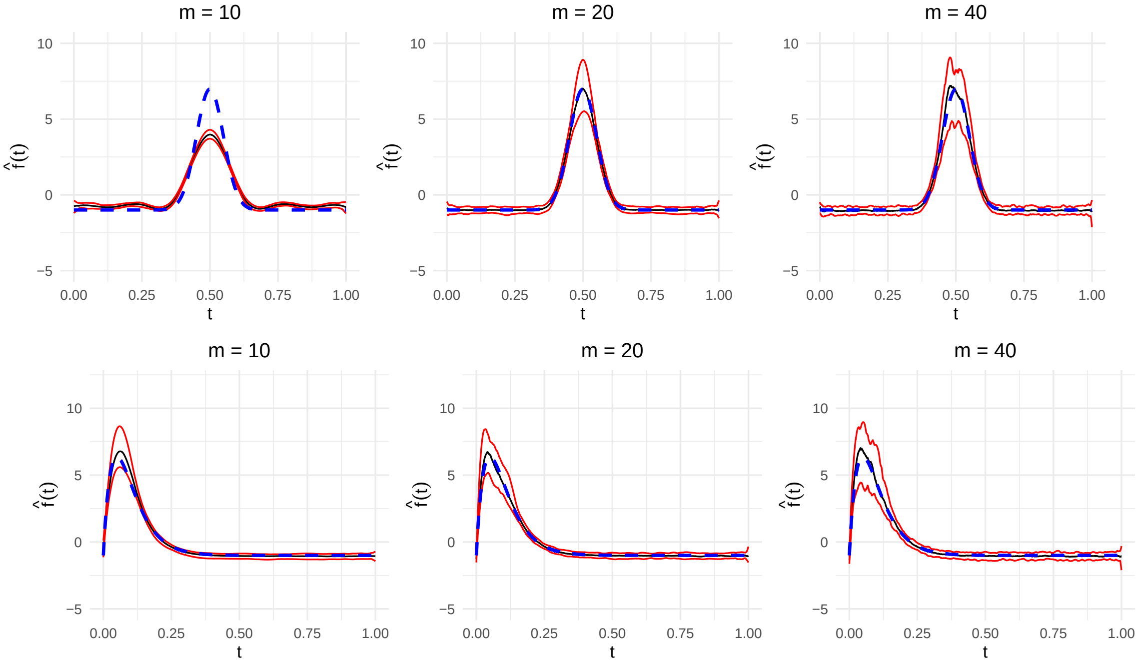


Table C.2.2: Simulation results over 100 replications for estimating the parameters of the BAPLA model with normal $f_i(t/n)$ and varying numbers of basis function.

Type	m	n	RMSE	MSE_{β_0}	MSE_f	AUC	Computation Time (secs)
Normal	5	1000	0.28	1.33	3.55	0.53	6.99
		2000	0.41	2.19	3.73	0.47	15.51
		5000	0.41	2.45	3.78	0.48	24.65
		10000	0.39	2.45	3.78	0.49	67.37
	10	1000	0.21	0.03	0.59	0.82	105.44
		2000	0.12	0.04	0.60	0.91	135.60
		5000	0.06	0.05	0.60	0.97	170.91
		10000	0.03	0.05	0.61	0.98	208.44
	15	1000	0.22	0.04	0.38	0.82	11.00
		2000	0.15	0.02	0.33	0.90	24.29
		5000	0.06	0.02	0.31	0.96	38.14
		10000	0.04	0.02	0.31	0.98	74.30
	20	1000	0.24	0.48	3.04	0.82	118.24
		2000	0.15	0.12	0.68	0.90	140.10
		5000	0.06	0.04	0.22	0.96	204.81
		10000	0.03	0.01	0.09	0.98	220.17
	25	1000	0.28	0.77	5.59	0.82	17.10
		2000	0.14	0.19	1.35	0.90	51.34
		5000	0.06	0.05	0.36	0.96	67.14
		10000	0.04	0.02	0.13	0.98	160.36
	30	1000	0.29	0.70	5.69	0.82	187.52
		2000	0.15	0.17	1.33	0.90	209.93
		5000	0.06	0.04	0.32	0.96	306.69
		10000	0.04	0.01	0.11	0.98	349.81

Type	m	n	RMSE	MSE_{β_0}	MSE_f	AUC	Computation Time (secs)
	35	1000	0.31	0.74	6.21	0.82	32.62
		2000	0.16	0.19	1.50	0.90	115.58
		5000	0.06	0.04	0.36	0.96	144.53
		10000	0.04	0.01	0.13	0.98	254.76
	40	1000	0.34	0.79	6.98	0.82	318.50
		2000	0.17	0.22	1.85	0.90	312.25
		5000	0.06	0.04	0.45	0.96	444.52
		10000	0.04	0.02	0.18	0.98	611.74

Table C.2.3: Simulation results over 100 replications for estimating the parameters of the BAPLA model with gamma $f_i(t/n)$ and varying numbers of basis function.

Type	m	n	RMSE	MSE_{β_0}	MSE_f	AUC
Gamma	5	1000	0.19	0.08	1.19	0.82
		2000	0.12	0.06	0.94	0.89
		5000	0.05	0.09	0.98	0.95
		10000	0.03	0.10	0.99	0.97
	10	1000	0.23	0.32	1.50	0.81
		2000	0.13	0.11	0.45	0.89
		5000	0.06	0.03	0.16	0.96
		10000	0.04	0.02	0.10	0.98
	15	1000	0.25	0.63	4.81	0.81
		2000	0.14	0.19	1.20	0.89
		5000	0.06	0.04	0.22	0.96
		10000	0.04	0.02	0.09	0.99
	20	1000	0.25	0.60	8.11	0.81
		2000	0.14	0.17	2.46	0.89
		5000	0.06	0.03	0.53	0.96
		10000	0.04	0.01	0.20	0.99
	25	1000	0.27	0.65	1.19	0.82
		2000	0.15	0.19	0.94	0.89
		5000	0.06	0.04	0.98	0.96
		10000	0.04	0.02	0.99	0.99
	30	1000	0.29	0.75	1.50	0.81
		2000	0.15	0.22	0.45	0.90
		5000	0.06	0.04	0.16	0.97
		10000	0.04	0.02	0.10	0.99
	35	1000	0.31	0.84	4.81	0.81
		2000	0.15	0.25	1.20	0.90
		5000	0.06	0.05	0.22	0.96
		10000	0.04	0.02	0.09	0.99
	40	1000	0.34	0.95	8.11	0.81
		2000	0.17	0.28	2.46	0.90
		5000	0.07	0.06	0.53	0.97
		10000	0.04	0.02	0.20	0.98

C.3 Additional Material for Data Analysis

C.3.1 Coordinate-wise updates

When we incorporate the trial-like structure of the data, the quadratic approximation to the likelihood becomes

$$\ell_Q(\beta_i, \boldsymbol{\gamma}_i, c_{i,1}, \ldots, c_{i,m}) = -\frac{1}{2} \sum_{l=1}^{\#trials} \sum_{t=1}^{n} w_{t_l,i} \left(z_{t_l,i} - \beta_i - \sum_{k=1}^{m} c_{i,k} \tilde{\phi}_k(t_l) - \boldsymbol{\gamma}_i^\top y_{t_l-1} \right)^2 + C \left(\tilde{\beta}_i, \tilde{\boldsymbol{\gamma}}_i, \tilde{c}_{i1}, \ldots, \tilde{c}_{im} \right)^2.$$

Therefore, the update steps to account for the trial-like structure of data are

$$\tilde{\gamma}_{i,j} \leftarrow \frac{\mathcal{S}\left(\sum_{l=1}^{\#\text{trials}} \sum_{t_l=1}^{n} w_{i,t_l} y_{j,t_l-1} \left(z_{i,t_l} - \tilde{y}_{t_l}^{(j)} \right), \lambda \right)}{\sum_{l=1}^{\#\text{trials}} \sum_{t_l=1}^{n} w_{i,t_l} (y_{j,t_l-1})^2},$$

$$\tilde{\beta}_i \leftarrow \frac{\sum_{l=1}^{\#\text{trials}} \sum_{t_l=1}^{n} w_{i,t_l} (z_{i,t_l} - \sum_{k=1}^{m} \tilde{c}_{i,k} \tilde{\phi}_k(t_l) - \boldsymbol{\gamma}_i^\top y_{t_l-1})}{\sum_{l=1}^{\#\text{trials}} \sum_{t_l=1}^{n} w_{i,t_l}}$$

$$\tilde{c}_{i,k} \leftarrow \frac{\sum_{l=1}^{\#\text{trials}} \sum_{t_l=1}^{n} w_{i,t_l} \phi_k(t_l) (z_{i,t_l} - \tilde{\beta}_i - \tilde{\gamma}_i^\top y_{t_l-1} - \sum_{j \neq k} \tilde{c}_{i,j} \tilde{\phi}_j(t_l))}{\sum_{l=1}^{\#\text{trials}} \sum_{t_l=1}^{n} w_{i,t_l} \tilde{\phi}_k^2(t_l)}.$$

C.3.2 Accounting for Non-stationary Firing Rates

In the main paper, we show that explicit modelling of the firing rate is essential to reduce the number of spurious connections in the estimated interaction matrix. In particular, we give an example using the left trials in the wheel movement dataset, showing that the number of edges reduces, both within the group of non-stationary neurons and between other stationary neurons. Here, we present additional results for the remaining datasets, showing that the conclusions made in the main paper also hold here.

In Figure C.3.2, we sort the spike train data into two groups, and classify each

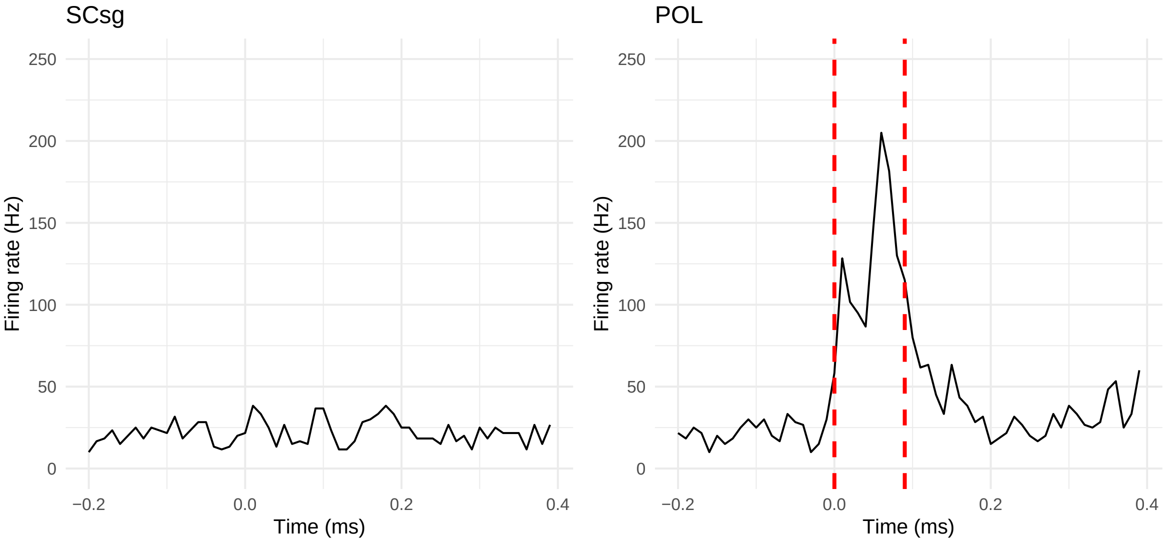
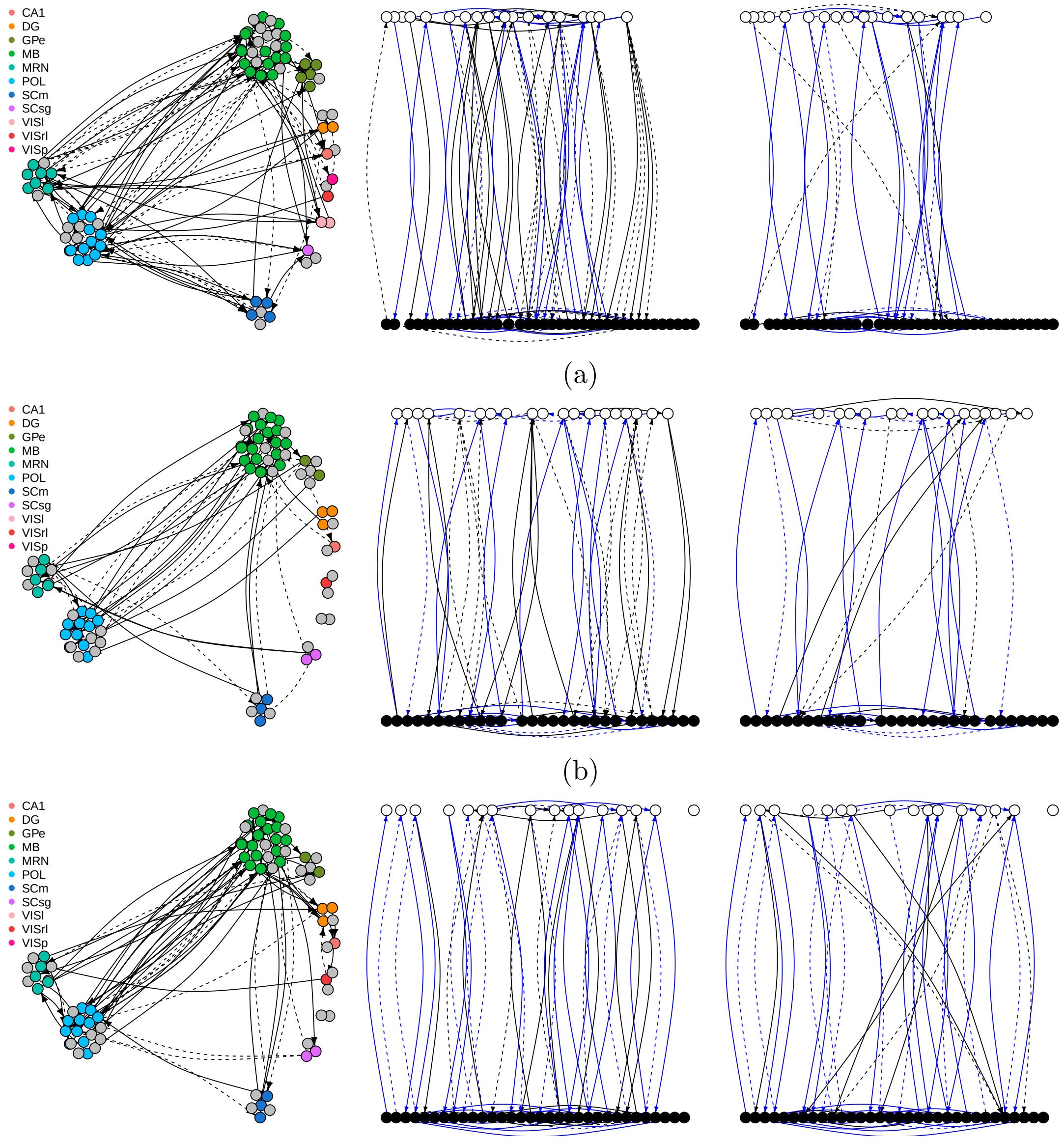


Figure C.3.1: Example neurons classified as stationary (left) and non-stationary (right) via PELT algorithm

neuron as either stationary or non-stationary as evidenced by their firing rate. To do so, we use the the PELT algorithm (Killick et al., 2012), which is readily available in the R package `changepoint`, to detect changes in variance in the firing rate of each neuron. If a change point is detected, the neuron is classified as 'non-stationary', otherwise it is classified as 'stationary'. Figure C.3.1 shows firing rates of two example neurons; one which is classified as stationary and another which is classified as non-stationary. The red dashed line highlights the change-points detected by the PELT algorithm.



(a)

(b)

(c)

Figure C.3.2: Estimate of interaction matrix for (a) right trial data in wheel movement dataset, (b) left trial data in stimulus onset dataset and (c) right trial data with stimulus onset dataset with standard logistic model for all regions (left) and split into stationary and non-stationary neurons (middle). Non-stationary neurons are in white and stationary neurons are in black. Figure on right shows estimate with our BAPLA method. Common edges between BAPLA and logistic model are in blue.

Sensitivity Analysis

In the main paper, we investigate the consequence of specifying a fixed number of basis functions m via a sensitivity analysis using synthetic data simulated from the BAPLA

				Interaction Type (%)		Region Type (%)	
m	Dataset	Trial	No. Edges	Excitatory	Inhibitory	Within	Between
3	Stimulus Onset	Left	67	61.19	38.81	25.37	74.63
		Right	68	66.18	33.82	17.65	82.35
	Wheel Movement	Left	46	60.87	39.13	15.22	84.78
		Right	55	61.82	38.18	9.09	90.91
6	Stimulus Onset	Left	49	61.22	38.78	24.49	75.51
		Right	68	69.12	30.88	14.71	85.29
	Wheel Movement	Left	38	65.79	34.21	15.79	84.21
		Right	51	58.82	41.18	19.61	80.39
10	Stimulus Onset	Left	52	65.38	34.62	28.85	71.15
		Right	59	69.49	30.51	18.64	81.36
	Wheel Movement	Left	35	60.00	40.00	28.57	71.43
		Right	45	62.22	37.78	13.33	86.67

Table C.3.1: Summary statistics for interaction matrices estimated with a varying number of basis functions.

model. In the absence of a ground truth comparison (which is the case in our real data section) it is therefore important to assess the sensitivity of our estimated network structure to the choice of hyperparameter m while keeping preference for the more parsimonious model (i.e., smaller m).

Table C.3.1 shows summary statistics for the estimated interaction matrices obtained with $m = \{3, 6, 10\}$ basis functions. While the number of edges does indeed change depending on the choice of hyperparameter m, our results show that the overall network structures remain fairly consistent across the differing scenarios, as do the scientific conclusions. In particular, we find that in general more edges are detected for the right trial datasets compared to the left, and that the BAPLA model detects more excitatory edges than inhibitory ones. Moreover, the BAPLA model identifies more between region interactions compared to those in the same region. For completeness, we plot estimates of the interaction matrix for each experimental condition, estimated with $m = \{3, 6, 10\}$ basis functions, in Figures 4-7.

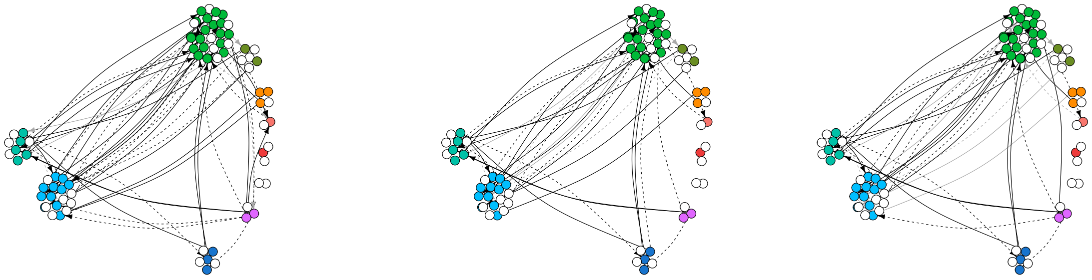


Figure C.3.3: Estimate of the interaction matrix for left trials from the stimulus onset dataset with 3 (left), 6 (middle) and 10 (right) basis functions.

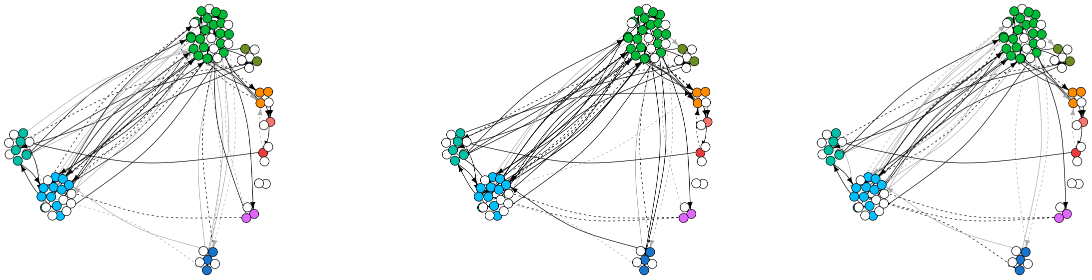


Figure C.3.4: Estimate of the interaction matrix for the right trials from the stimulus onset dataset with 3 (left), 6 (middle) and 10 (right) basis functions.

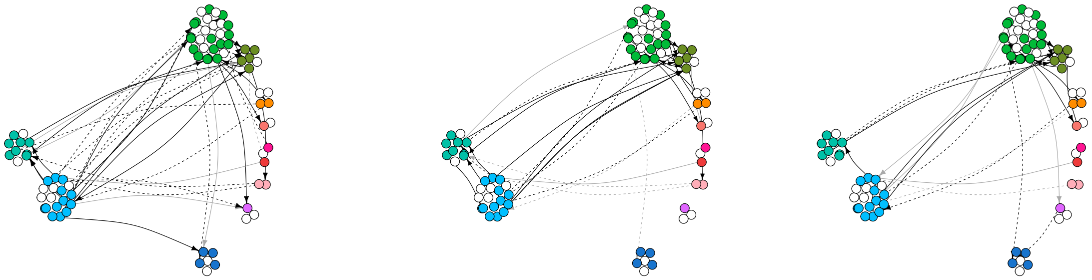


Figure C.3.5: Estimate of the interaction matrix for the left trials from the wheel movement dataset with 3 (left), 6 (middle) and 10 (right) basis functions.

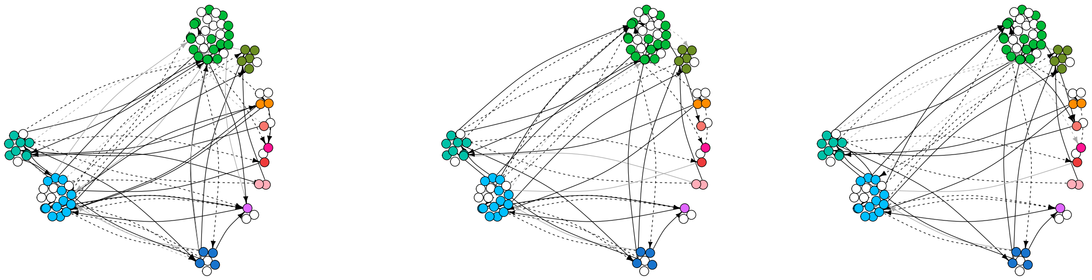


Figure C.3.6: Estimate of the interaction matrix for the right trials from the wheel movement dataset with 3 (left), 6 (middle) and 10 (right) basis functions.

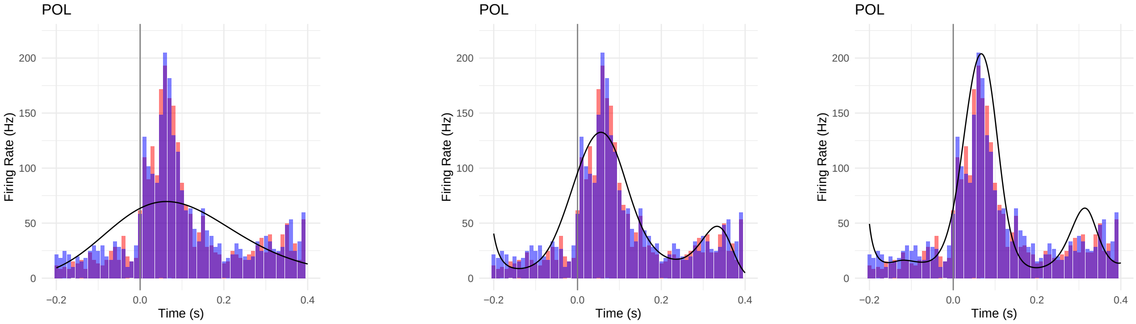


Figure C.3.7: Estimate of the non-stationary firing rate for an example neuron in the wheel movement dataset with (a) 3, (b) 6 and (c) 10 basis functions.

Finally, in Figure C.3.7 we plot estimates of the non-stationary firing rate for an example neuron in the wheel movement dataset with $m = \{3, 6, 10\}$ basis functions. Therefore, we select $m = 6$ as a reasonable number of basis functions for our data analysis section in the main paper.

Data

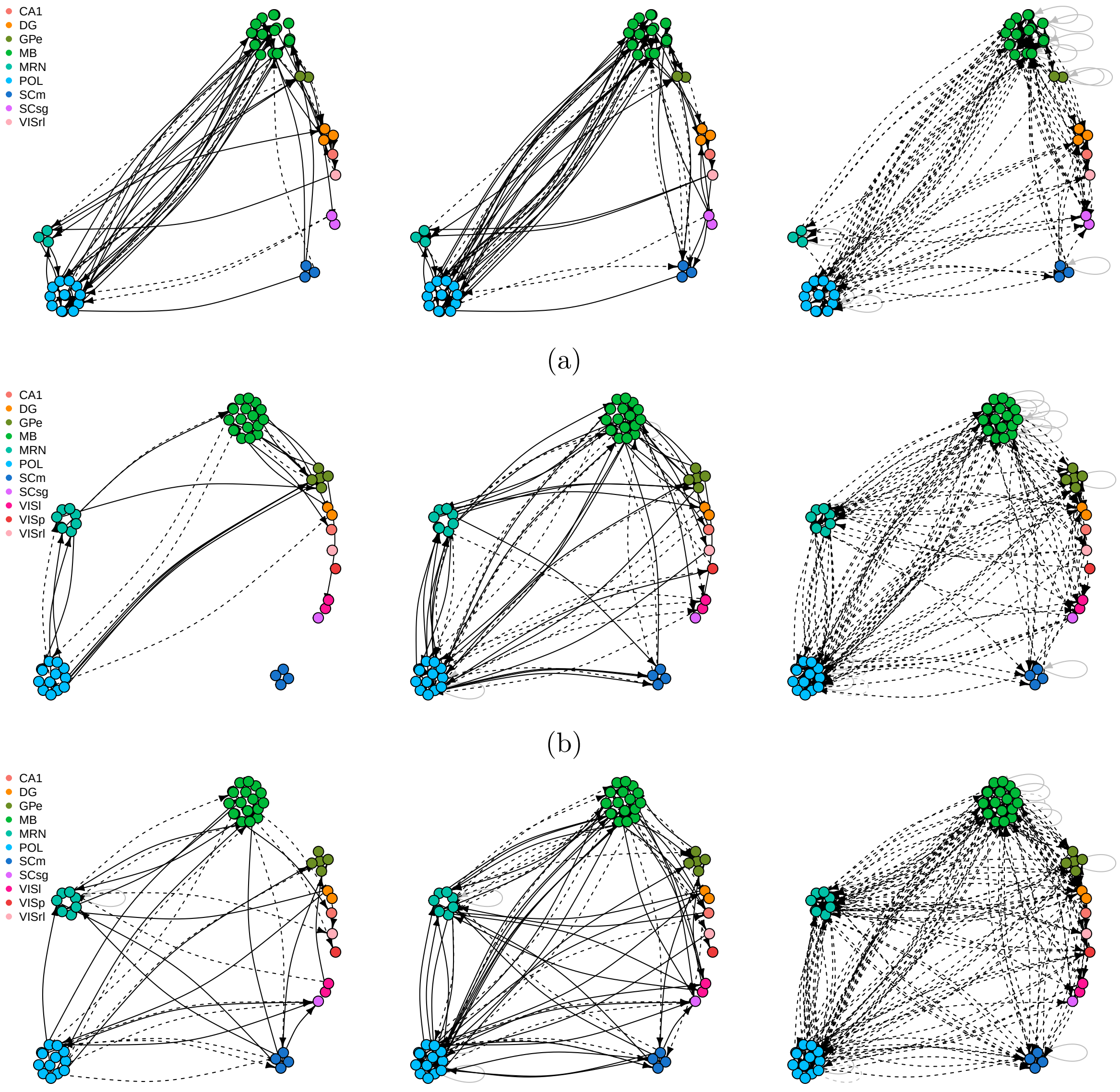
The data that support the findings of Chapter 4 are openly available on the DANDI Archive (RRID:SCR_017571) at `10.48324/dandi.000017/0.240329.1926` (Steinmetz et al., 2024).

Code

The code used to implement the BAPLA model is available at `https://github.com/pinkney19/BAPLA`.

Appendix D

Supplementary Figures for Chapter 5

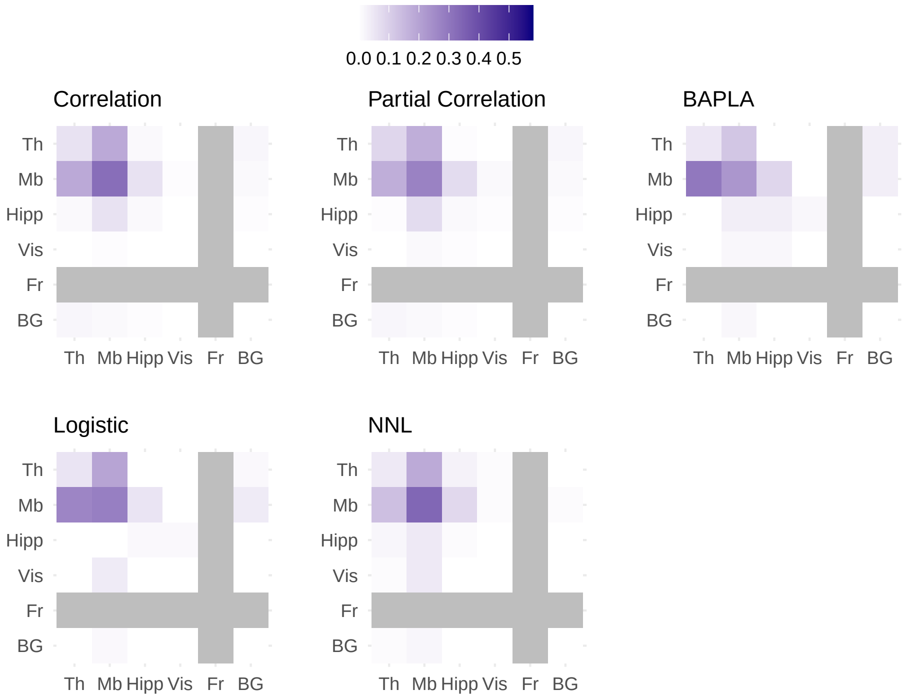


(a)

(b)

(c)

Figure D.0.1: Estimated networks for BAPLA model (left) logistic odel (middle) and NNL model (right) for (a) right trial dataset aligned to stimulus onset (b) left trial dataset aligned to wheel movement and (c) right trial dataset aligned to wheel movement.


Correlation

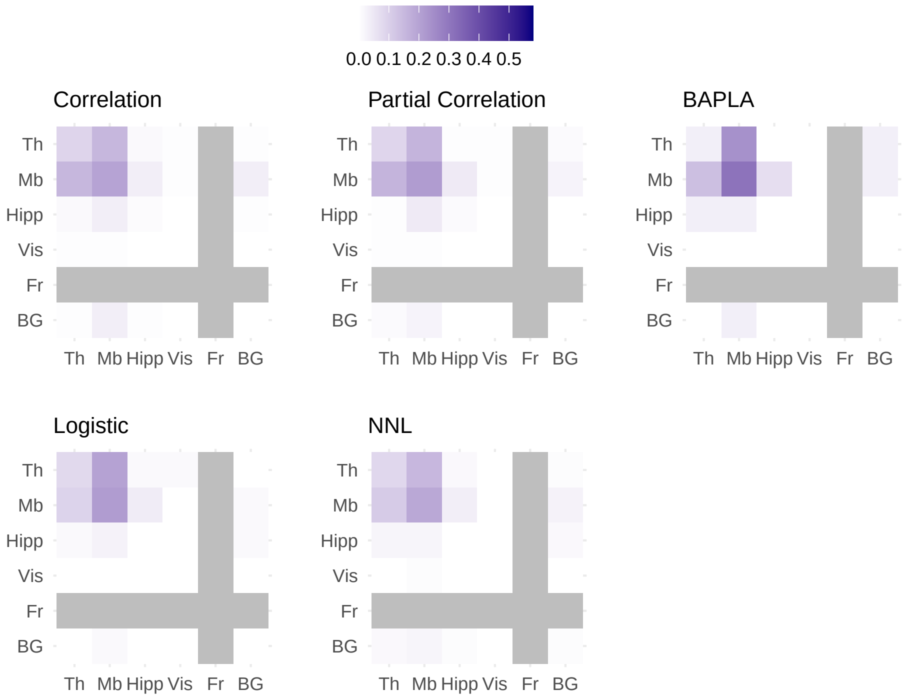
Partial Correlation

BAPLA

Logistic

NNL

Figure D.0.2: ROI connectivity matrices for the right-trial dataset aligned to stimulus onset.


Correlation

Partial Correlation

BAPLA

Logistic

NNL

Figure D.0.3: ROI connectivity matrices for the left-trial dataset aligned to wheel movement.

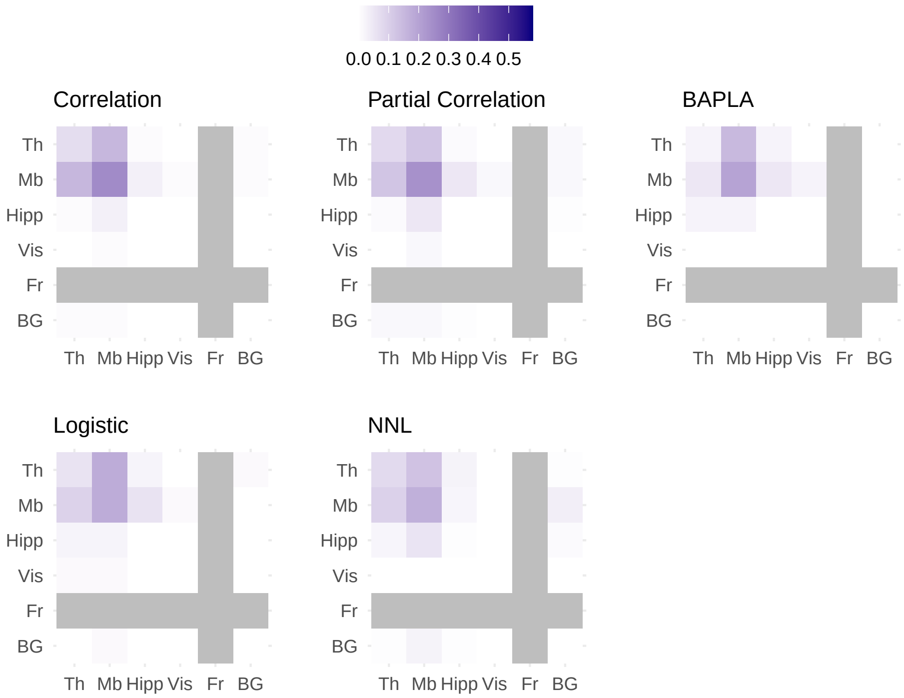


Figure D.0.4: ROI connectivity matrices for the right-trial dataset aligned to wheel movement.

Bibliography

Abril, I. M. d., Yoshimoto, J., and Doya, K. (2018). Connectivity inference from neural recording data: Challenges, mathematical bases and research directions. *Neural Networks*, 102:120–137.

Aertsen, A., Gerstein, G., Habib, M., and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of "effective connectivity". *Journal of neurophysiology*, 61(5):900–917.

Amunts, K., Knoll, A. C., Lippert, T., Pennartz, C. M., Ryvlin, P., Destexhe, A., Jirsa, V. K., D'Angelo, E., and Bjaalie, J. G. (2019). The human brain project—synergy between neuroscience, computing, informatics, and brain-inspired technologies. *PLoS biology*, 17(7):e3000344.

Bair, W., Koch, C., Newsome, W., and Britten, K. (1994). Power spectrum analysis of bursting cells in area mt in the behaving monkey. *Journal of Neuroscience*, 14(5):2870–2892.

Banga, K., Benson, J., Bhagat, J., Biderman, D., Birman, D., Bonacchi, N., Bruijns, S. A., Buchanan, K., Campbell, R. A., Carandini, M., et al. (2025). Reproducibility of in vivo electrophysiological measurements in mice. *Elife*, 13:RP100840.

Bardet, J.-M., Doukhan, P., and León, J. R. (2008). Uniform limit theorems for the integrated periodogram of weakly dependent time series and their applications to Whittle's estimate. *Journal of Time Series Analysis*, 29(5):906–945.

Bartlett, M. S. (1963a). The spectral analysis of point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25(2):264–281.

Bartlett, M. S. (1963b). Statistical estimation of density functions. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 245–254.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227.

Böhm, H. and von Sachs, R. (2009). Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis*, 100(5):913–935.

Bolding, K. A. and Franks, K. M. (2018). Recurrent cortical circuits implement concentration-invariant odor coding. *Science*, 361(6407).

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122.

Brillinger, D. (1972). The spectral analysis of stationary interval functions. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, pages 483–513.

Brillinger, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59(3):189–200.

Brillinger, D. R. (2001). *Time series: data analysis and theory.* SIAM.

Brody, C. D. (1999). Correlations without synchrony. *Neural computation*, 11(7):1537–1551.

Buccino, A. P., Hurwitz, C. L., Garcia, S., Magland, J., Siegle, J. H., Hurwitz, R., and Hennig, M. H. (2020). Spikeinterface, a unified framework for spike sorting. *Elife*, 9:e61834.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications.* Springer Science & Business Media.

Bunea, F. and Xiao, L. (2015). On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. *Bernoulli*, 21:1200–1230.

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature neuroscience*, 7(5):446–451.

Buzsaki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929.

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Carlson, D. and Carin, L. (2019). Continuing progress of spike sorting in the era of big data. *Current opinion in neurobiology*, 55:90–96.

Casile, A., Faghih, R. T., and Brown, E. N. (2021). Robust point-process Granger causality analysis in presence of exogenous temporal modulations and trial-by-trial variability in spike trains. *PLoS computational biology*, 17(1):e1007675.

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, 106(494):608–625.

Chen, S., Shojaie, A., Shea-Brown, E., and Witten, D. (2017a). The multivariate Hawkes process in high dimensions: Beyond mutual excitation. *arXiv preprint arXiv:1707.04928.*

Chen, S., Witten, D., and Shojaie, A. (2017b). Nearly assumptionless screening for the mutually-exciting multivariate Hawkes process. *Electronic journal of statistics*, 11(1):1207.

Chen, Z., Putrino, D. F., Ghosh, S., Barbieri, R., and Brown, E. N. (2010). Statistical inference for assessing functional connectivity of neuronal ensembles with sparse spiking data. *IEEE transactions on neural systems and rehabilitation engineering*, 19(2):121–135.

Chiarion, G., Sparacino, L., Antonacci, Y., Faes, L., and Mesin, L. (2023). Connectivity analysis in EEG data: a tutorial review of the state of the art and emerging trends. *Bioengineering*, 10(3):372.

Cohen, E. A. K. and Gibberd, A. J. (2021). Wavelet spectra for multivariate point processes. *Biometrika*.

Costa, M., Graham, C., Marsalle, L., and Tran, V. (2018). Renewal in Hawkes processes with self-excitation and inhibition. *arXiv preprint arXiv:1801.04645*.

Czanner, G., Grün, S., and Iyengar, S. (2005). Theory of the snowflake plot and its relations to higher-order analysis methods. *Neural Computation*, 17(7):1456–1479.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, 51(2):157–172.

Dahlhaus, R., Eichler, M., and Sandkühler, J. (1997). Identification of synaptic connections in neural ensembles by graphical models. *Journal of Neuroscience Methods*, 77(1):93–107.

Daley, D. J., Vere-Jones, D., et al. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer.

Dallakyan, A., Kim, R., and Pourahmadi, M. (2022). Time series graphical lasso and sparse VAR estimation. *Computational Statistics & Data Analysis*, 176:107557.

Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 76(2):373.

Deb, N., Kuceyeski, A., and Basu, S. (2024). Regularized estimation of sparse spectral precision matrices. *arXiv preprint arXiv:2401.11128*.

Dostrovsky, J., Levy, R., Wu, J., Hutchison, W., Tasker, R., and Lozano, A. (2000). Microstimulation-induced inhibition of neuronal firing in human globus pallidus. *Journal of Neurophysiology*, 84(1):570–574.

Doukhan, P. and Lang, G. (2016). Weak dependence of point processes and application to second-order statistics. *Statistics*, 50(6):1221–1235.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic processes and their applications*, 84(2):313–342.

Doukhan, P. and Neumann, M. H. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Processes and their Applications*, 117(7):878–903.

Drikvandi, R. (2025). High dimensional regression with many nuisance parameters: Both cases of specified and unspecified parameters of interest. *Electronic Journal of Statistics*, 19(1):2923–2957.

Eden, U. T., Gale, J. T., Amirnovin, R., and Eskandar, E. N. (2012). Characterizing the spiking dynamics of subthalamic nucleus neurons in Parkinson's disease using generalized linear models. *Frontiers in integrative neuroscience*, 6:28.

Eichler, M., Dahlhaus, R., and Dueck, J. (2017). Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *Journal of Time Series Analysis*, 38(2):225–242.

Eichler, M., Dahlhaus, R., and Sandkühler, J. (2003). Partial correlation analysis for the identification of synaptic connections. *Biological Cybernetics*, 89(4):289–302.

El-Yaagoubi, A. B., Aslan, S., Gomawi, F., Redondo, P. V., Roy, S., Sultan, M. S., Talento, M. S., Tarrazona, F. T., Wu, H., Cooper, K. W., et al. (2025). Methods for brain connectivity analysis with applications to rat local field potential recordings. *Entropy*, 27(4):328.

Erdős, P. (1960). On the evolution of random graphs. *Publ Math Inst Hungarian Acad Sci*, 5:17.

Euan, C., Sun, Y., and Ombao, H. (2019). Coherence-based time series clustering for statistical inference and visualization of brain connectivity. *The Annals of Applied Statistics*, 13(2):990–1015.

Fiecas, M., Leng, C., Liu, W., and Yu, Y. (2019). Spectral analysis of high-dimensional time series. *Electronic Journal of Statistics*, 13(2):4079–4101.

Fiecas, M. and Ombao, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *Annals of Applied Statistics*, 5(2A):1102–1125.

Fiecas, M. and von Sachs, R. (2014). Data-driven shrinkage of the spectral density matrix of a high-dimensional time series. *Electronic Journal of Statistics*, 8(2):2975–3003.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664–1671.

Foster, N. N., Barry, J., Korobkova, L., Garcia, L., Gao, L., Becerra, M., Sherafat, Y., Peng, B., Li, X., Choi, J.-H., et al. (2021). The mouse cortico–basal ganglia–thalamic network. *Nature*, 598(7879):188–194.

Foygel, R. and Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in neural information processing systems*, 23.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., Qian, J., and Yang, J. (2023). glmnet: Lasso and elastic-net regularized generalized linear models. *Astrophysics Source Code Library*, pages ascl–2308.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175.

Fu, W. and Knight, K. (2000). Asymptotics for lasso-type estimators. *The Annals of statistics*, 28(5):1356–1378.

Gao, Y., Black, M. J., Bienenstock, E., Wu, W., and Donoghue, J. P. (2003). A quantitative comparison of linear and non-linear models of motor cortical activity for the encoding and decoding of arm motions. In *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings.*, pages 189–192. IEEE.

Garcia, S., Windolf, C., Boussard, J., Dichter, B., Buccino, A. P., and Yger, P. (2024). A modular implementation to handle and benchmark drift correction for high-density extracellular recordings. *eneuro*, 11(2).

Gerstein, G. L. and Perkel, D. H. (1972). Mutual temporal relationships among neuronal spike trains: Statistical techniques for display and analysis. *Biophysical journal*, 12(5):453–473.

Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34(1):152–177.

Gueuning, T. and Claeskens, G. (2016). Confidence intervals for high-dimensional partially linear single-index models. *Journal of Multivariate Analysis*, 149:13–29.

Guo, J., Tang, M., Tian, M., and Zhu, K. (2013). Variable selection in high-dimensional partially linear additive models for composite quantile regression. *Computational Statistics & Data Analysis*, 65:56–67.

Hall, E. C., Raskutti, G., and Willett, R. (2016). Inference of high-dimensional autoregressive generalized linear models. *arXiv preprint arXiv:1605.02693*.

Hall, E. C., Raskutti, G., and Willett, R. M. (2018). Learning high-dimensional generalized linear autoregressive models. *IEEE transactions on information theory*, 65(4):2401–2422.

Hansen, N. R., Reynaud-Bouret, P., and Rivoirard, V. (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21:83–143.

Harrison, M. T., Amarasingham, A., and Kass, R. (2013). Statistical identification of synchronous spiking. *Journal of Computatoinal Neuroscience*.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438–443.

Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Hawkes, A. G. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2):193–198.

Hu, L., Fortin, N. J., and Ombao, H. (2019). Modeling high-dimensional multichannel brain signals. *Statistics in Biosciences*, 11(1):91–126.

Huang, J. and Zhang, T. (2010). The benefit of group sparsity. *The Annals of Statistics*, 38(4):1978–2004.

Huang, Y., Brandon, M. P., Griffin, A. L., Hasselmo, M. E., and Eden, U. T. (2009). Decoding movement trajectories through a t-maze using point process filters applied to place field data from rat hippocampal region ca1. *Neural computation*, 21(12):3305–3334.

Inagaki, H. K., Chen, S., Ridder, M. C., Sah, P., Li, N., Yang, Z., Hasanbegovic, H., Gao, Z., Gerfen, C. R., and Svoboda, K. (2022). A midbrain-thalamus-cortex circuit reorganizes cortical dynamics to initiate movement. *Cell*, 185(6):1065–1081.

International Brain Laboratory (2024). Spike sorting pipeline for the international brain laboratory. `https://doi.org/10.6084/m9.figshare.19705522.v4`.

Iremonger, K. J. and Bains, J. S. (2016). Asynchronous presynaptic glutamate release enhances neuronal excitability during the post-spike refractory period. *The Journal of physiology*, 594(4):1005–1015.

Ito, S., Hansen, M. E., Heiland, R., Lumsdaine, A., Litke, A. M., and Beggs, J. M. (2011). Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model. *PloS one*, 6(11):e27431.

Jager, P., Moore, G., Calpin, P., Durmishi, X., Salgarella, I., Menage, L., Kita, Y., Wang, Y., Kim, D. W., Blackshaw, S., et al. (2021). Dual midbrain and forebrain origins of thalamic inhibitory interneurons. *Elife*, 10:e59272.

Janková, J. and van de Geer, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229.

Jarvis, M. and Mitra, P. P. (2001). Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Computation*, 13(4):717–749.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236.

Jung, A., Hannak, G., and Goertz, N. (2015). Graphical lasso based model selection for time series. *IEEE Signal Processing Letters*, 22(10):1781–1785.

Kadirvelu, B., Hayashi, Y., and Nasuto, S. J. (2017). Inferring structural connectivity using ising couplings in models of neuronal networks. *Scientific reports*, 7(1):8156.

Kajiwara, M., Nomura, R., Goetze, F., Kawabata, M., Isomura, Y., Akutsu, T., and Shimono, M. (2021). Inhibitory neurons exhibit high controlling ability in the cortical microconnectome. *PLOS Computational Biology*, 17(4):e1008846.

Kass, R. E., Kelly, R. C., and Loh, W.-L. (2011). Assessment of synchrony in multiple neural spike trains using loglinear point process models. *The annals of applied statistics*, 5(2B):1262.

Keeley, S. L., Zoltowski, D. M., Aoi, M. C., and Pillow, J. W. (2020). Modeling statistical dependencies in multi-region spike train data. *Current opinion in neurobiology*, 65:194–202.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.

Kim, S., Putrino, D., Ghosh, S., and Brown, E. N. (2011). A Granger causality measure for point process models of ensemble neural spiking activity. *PLoS computational biology*, 7(3):e1001110.

Kobayashi, R., Kurita, S., Kurth, A., Kitano, K., Mizuseki, K., Diesmann, M., Richmond, B. J., and Shinomoto, S. (2019). Reconstructing neuronal circuitry from parallel spike trains. *Nature communications*, 10(1):4468.

Kobayashi, R. and Shinomoto, S. (2025). Inference of monosynaptic connections from parallel spike trains: a review. *Neuroscience Research*, 215:37–46.

Kocsis, B., Bragin, A., and Buzsáki, G. (1999). Interdependence of multiple theta generators in the hippocampus: a partial coherence analysis. *Journal of Neuroscience*, 19(14):6200–6212.

Koutitonsky, V., Navarro, N., and Booth, D. (2002). Descriptive physical oceanography of great-entry lagoon, gulf of st. lawrence. *Estuarine, Coastal and Shelf Science*, 54(5):833–847.

Lambert, R. C., Tuleau-Malot, C., Bessaih, T., Rivoirard, V., Bouret, Y., Leresche, N., and Reynaud-Bouret, P. (2018). Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *Journal of Neuroscience Methods*, 297:9–21.

Lang, E. W., Tomé, A. M., Keck, I. R., Górriz-Sáez, J., and Puntonet, C. G. (2012). Brain connectivity analysis: a short survey. *Computational intelligence and neuroscience*, 2012(1):412512.

Laub, P. J., Taimre, T., and Pollett, P. K. (2015). Hawkes processes. *arXiv preprint arXiv:1507.02822*.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Annals of statistics*, pages 1302–1338.

Lauritzen, S. L. (1996). *Graphical Models*, volume 17. Clarendon Press.

Leen, D. A. and Shea-Brown, E. (2015). A simple mechanism for beyond-pairwise correlations in integrate-and-fire neurons. *The Journal of Mathematical Neuroscience (JMN)*, 5(1):17.

Lesage, L., Deaconu, M., Lejay, A., Meira, J. A., Nichil, G., et al. (2022). Hawkes processes framework with a Gamma density as excitation function: application to natural disasters for insurance. *Methodology and Computing in Applied Probability*, pages 1–29.

Lewis, P. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413.

Liew, Y. J., Pala, A., Whitmire, C. J., Stoy, W. A., Forest, C. R., and Stanley, G. B. (2021). Inferring thalamocortical monosynaptic connectivity in vivo. *Journal of Neurophysiology*, 125(6):2408–2431.

Lintas, A., Abe, T., Villa, A. E., and Asai, Y. (2018). Granger causality to reveal functional connectivity in the mouse basal ganglia-thalamocortical circuit. In *International Conference on Artificial Neural Networks*, pages 393–402. Springer.

Magland, J., Jun, J. J., Lovero, E., Morley, A. J., Hurwitz, C. L., Buccino, A. P., Garcia, S., and Barnett, A. H. (2020). Spikeforest, reproducible web-facing ground-truth validation of automated neural spike sorters. *Elife*, 9:e55167.

Marchant, J. K., Ferris, N. G., Grass, D., Allen, M. S., Gopalakrishnan, V., Olchanyi, M., Sehgal, D., Sheft, M., Strom, A., Bilgic, B., et al. (2024). Mesoscale brain mapping: Bridging scales and modalities in neuroimaging–a symposium review. *Neuroinformatics*, 22(4):679–706.

Mazumder, R. and Hastie, T. (2012). The graphical lasso: New insights and alternatives. *Electronic journal of statistics*, 6:2125.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*.

Miller, C. T., Chen, X., Donaldson, Z. R., Marlin, B. J., Tsao, D. Y., Williams, Z. M., Zelikowsky, M., Zeng, H., and Hong, W. (2024). The BRAIN initiative: a pioneering program on the precipice. *nature neuroscience*, 27(12):2264–2266.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J., and Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3):586–595.

Nadkarni, R., Foti, N. J., Lee, A. K., and Fox, E. B. (2016). Sparse plus low-rank graphical models of time series to infer functional connectivity from MEG recordings. *2nd SIGKDD Workshop on Mining and Learning from Time Series*.

Narayanan, N. S. and Laubach, M. (2009). Methods for studying functional interactions among neuronal populations. *Dynamic Brain Imaging: Multi-Modal Methods and In Vivo Applications*, pages 135–165.

Negahban, S., Yu, B., Wainwright, M. J., and Ravikumar, P. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statistical Sciecne*, 27(4):538–557.

Newman, M. E. (2003). Mixing patterns in networks. *Physical review E*, 67(2):026126.

Newson, J. J. and Thiagarajan, T. C. (2019). EEG frequency bands in psychiatric disorders: a review of resting state studies. *Frontiers in human neuroscience*, 12:521.

Ning, Y. and Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, 45:158–195.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31.

Ogata, Y. (1999). Seismicity analysis through point-process modeling: A review. *Seismicity Patterns, their Statistical Significance and Physical Meaning*, pages 471–507.

Okano, H., Sasaki, E., Yamamori, T., Iriki, A., Shimogori, T., Yamaguchi, Y., Kasai, K., and Miyawaki, A. (2016). Brain/MINDS: a japanese national brain project for marmoset neuroscience. *Neuron*, 92(3):582–590.

Ombao, H., Lindquist, M., Thompson, W., and Aston, J. (2016). *Handbook of neuroimaging data analysis*. Chapman and Hall/CRC.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243.

Park, I. M., Meister, M. L., Huk, A. C., and Pillow, J. W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nature neuroscience*, 17(10):1395–1403.

Percival, D. B., Walden, A. T., et al. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967). Neuronal spike trains and stochastic point processes: Ii. simultaneous spike trains. *Biophysical journal*, 7(4):419–440.

Perkel, D. H., Gerstein, G. L., Smith, M. S., and Tatton, W. G. (1975). Nerve-impulse patterns: a quantitative display technique for three neurons. *Brain research*, 100(2):271–296.

Pernice, V., Staude, B., Cardanobile, S., and Rotter, S. (2011). How structure determines correlations in neuronal networks. *PLOS Computational Biology*, 7(5):e1002059.

Pinkney, C., Euan, C., Gibberd, A., and Shojaie, A. (2024). Regularised spectral estimation for high-dimensional point processes. *arXiv preprint arXiv:2403.12908*.

Poldrack, R. A. (2007). Region of interest analysis for fmri. *Social cognitive and affective neuroscience*, 2(1):67–70.

Poli, D., Pastore, V. P., Martinoia, S., and Massobrio, P. (2016). From functional to structural connectivity using partial correlation in neuronal assemblies. *Journal of neural engineering*, 13(2):026023.

Poli, D., Pastore, V. P., and Massobrio, P. (2015). Functional connectivity in in vitro neuronal assemblies. *Frontiers in neural circuits*, 9:57.

Rajala, T. A., Olhede, S. C., Grainger, J. P., and Murrell, D. J. (2023). What is the fourier transform of a spatial point process? *IEEE Transactions on Information Theory*, 69(8):5219–5252.

Rajala, T. A., Olhede, S. C., and Murrell, D. J. (2020). What is the Fourier transform of a spatial point process? *arXiv preprint arxiv:2009.01474*.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980.

Ren, N., Ito, S., Hafizi, H., Beggs, J. M., and Stevenson, I. H. (2020). Model-based detection of putative synaptic connections from spike recordings with latency and type constraints. *Journal of Neurophysiology*, 124(6):1588–1604.

Rey, H. G., Pedreira, C., and Quiroga, R. Q. (2015). Past, present and future of spike sorting techniques. *Brain research bulletin*, 119:106–117.

Rizoiu, M.-A., Lee, Y., Mishra, S., and Xie, L. (2017). A tutorial on Hawkes processes for events in social media. *arXiv preprint arXiv:1708.06401*.

Rodríguez-Álvarez, M. X., Durbán, M., Eilers, P. H., Lee, D.-J., and Gonzalez, F. (2023). Multidimensional adaptive p-splines with application to neurons' activity studies. *Biometrics*, 79(3):1972–1985.

Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515.

Roueff, F. and Sachs, R. V. (2019). Time-frequency analysis of locally stationary hawkes processes. *Bernoulli*, 25:1355–1385.

Saegusa, T. and Shojaie, A. (2016). Joint estimation of precision matrices in heterogeneous populations. *Electronic Journal of Statistics*, 10(1):1341.

Salvador, R., Suckling, J., Schwarzbauer, C., and Bullmore, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):937–946.

Sarmashghi, M., Jadhav, S. P., and Eden, U. (2021). Efficient spline regression for neural spiking data. *Plos one*, 16(10):e0258321.

Shahhosseini, Y. and Miranda, M. F. (2022). Functional connectivity methods and their applications in fMRI data. *Entropy*, 24(3):390.

Shimazaki, H. (2025). Neural coding: Foundational concepts, statistical formulations, and recent advances. *Neuroscience Research*.

Shojaie, A. and Fox, E. B. (2022). Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, 9(1):289–319.

Siegle, J. H., Ledochowitsch, P., Jia, X., Millman, D. J., Ocker, G. K., Caldejon, S., Casal, L., Cho, A., Denman, D. J., Durand, S., et al. (2021). Reconciling functional differences in populations of neurons recorded with two-photon imaging and electrophysiology. *Elife*, 10:e69068.

Song, D., Wang, H., Tu, C. Y., Marmarelis, V. Z., Hampson, R. E., Deadwyler, S. A., and Berger, T. W. (2013). Identification of sparse neural functional connectivity using penalized likelihood estimation and basis functions. *Journal of computational neuroscience*, 35:335–357.

Song, X., Zhang, C., Zhang, J., Zou, X., Mo, Y., and Tian, Y. (2020). Potential linkages of precipitation extremes in Beijing-Tianjin-Hebei region, China, with large-scale climate patterns using wavelet-based approaches. *Theoretical and Applied Climatology*, 141:1251–1269.

Steinmetz, N., Zatka-Haas, P., Carandini, M., Harris, K., and Wang, R. (2024). Distributed coding of choice, action and engagement across the mouse brain. Dataset version 0.240329.1926.

Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., et al. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, 372(6539):eabf4588.

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., and Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 576(7786):266–273.

Sun, Y., Li, Y., Kuceyeski, A., and Basu, S. (2018). Large spectral density matrix estimation by thresholding. *arXiv preprint arXiv:1812.00532*.

Tank, A., Foti, N., and Fox, E. (2015). Bayesian structure learning for stationary time series. *arXiv preprint arXiv:1505.03131*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Tripathy, S. J., Padmanabhan, K., Gerkin, R. C., and Urban, N. N. (2013). Intermediate intrinsic diversity enhances neural population coding. *Proceedings of the National Academy of Sciences*, 110(20):8248–8253.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology*, 93(2):1074–1089.

Tugnait, J. K. (2022). On sparse high-dimensional graphical model learning for dependent time series. *Signal Processing*, 197.

van Beest, E. H., Bimbard, C., Fabre, J. M., Dodgson, S. W., Takács, F., Coen, P., Lebedeva, A., Harris, K. D., and Carandini, M. (2025). Tracking neurons across days with high-density probes. *Nature Methods*, 22(4):778–787.

Van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*.

Vareberg, A. D., Bok, I., Eizadi, J., Ren, X., and Hai, A. (2024). Inference of network connectivity from temporally binned spike trains. *Journal of Neuroscience Methods*, 404:110073.

Vellmer, S. and Lindner, B. (2019). Theory of spike-train power spectra for multidimensional integrate-and-fire neurons. *Physical Review Research*, 1(2):023024.

Vidakovic, B. (2009). *Statistical modeling by wavelets*. John Wiley & Sons.

Walden, A. T. (2000). A unified view of multitaper multivariate spectral estimation. *Biometrika*, 87:767–788.

Waldorp, L. and Haslbeck, J. (2024). Network inference with the lasso. *Multivariate Behavioral Research*, pages 1–20.

Wang, H., Xie, K., Xie, L., Li, X., Li, M., Lyu, C., Chen, H., Chen, Y., Liu, X., Tsien, J., et al. (2019). Functional brain connectivity revealed by sparse coding of large-scale local field potential dynamics. *Brain topography*, 32(2):255–270.

Wang, X., Kolar, M., and Shojaie, A. (2024). Statistical inference for networks of high-dimensional point processes. *Journal of the American Statistical Association*, pages 1–11.

Wang, X., Kolar, M., and Shojaie, A. (2025). Statistical inference for networks of high-dimensional point processes. *Journal of the American Statistical Association*, 120(550):1014–1024.

Wang, X. and Shojaie, A. (2021). Joint estimation and inference for multi-experiment networks of high-dimensional point processes. *arXiv preprint arXiv:2109.11634*.

Warton, D. I. (2008). Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *Journal of the American Statistical Association*, 103(481):340–349.

Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för matematik*, 2(5):423–434.

Xu, H., Farajtabar, M., and Zha, H. (2016). Learning Granger causality for Hawkes processes. In *International conference on machine learning*, pages 1717–1726. PMLR.

Ye, Z., Shelton, A. M., Shaker, J. R., Boussard, J., Colonell, J., Birman, D., Manavi, S., Chen, S., Windolf, C., Hurwitz, C., et al. (2024). Ultra-high density electrodes improve detection, yield, and cell type identification in neuronal recordings. *bioRxiv*, pages 2023–08.

Zhao, M., Batista, A., Cunningham, J. P., Chestek, C., Rivera-Alvidrez, Z., Kalmar, R., Ryu, S., Shenoy, K., and Iyengar, S. (2012). An L_1-regularized logistic model for detecting short-term neuronal interactions. *Journal of computational neuroscience*, 32:479–497.

Zhou, K., Zha, H., and Song, L. (2013). Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes. In *Artificial intelligence and statistics*, pages 641–649. PMLR.

Zhou, P., Burton, S. D., Snyder, A. C., Smith, M. A., Urban, N. N., and Kass, R. E. (2015). Establishing a statistical link between network oscillations and neural synchrony. *PLoS computational biology*, 11(10):e1004549.

Zoltowski, D. and Pillow, J. W. (2018). Scaling the Poisson GLM to massive neural datasets through polynomial approximations. *Advances in neural information processing systems*, 31.