國 立 中 央 大 學

數 學 系
博 士 論 文

Causal Connectivity Analysis for Identifying the Information Flow of Neuronal Networks

研 究 生：邵 培 強

指導教授：單 維 彰 博士

嚴 健 彰 博士

中 華 民 國　　105　　年　　6　　月

Causal Connectivity Analysis for Identifying the Information Flow of Neuronal Networks

Pei-Chiang Shao

A Dissertation
Submitted to Department of Mathematics
College of Science
National Central University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy
in
Mathematics

June 2016

神經元網路訊息的因果分析

摘要

本文主要發展用來判別神經元網路資訊流向的統計分析方法。格蘭傑因果關係（Granger causality, GC）是個用來估計時間序列訊號間因果交互作用相當流行且有效的概念。在過去的二十年來，GC 已成為一個可用來偵測多種不同類型神經活動資料間因果關聯的強大分析方法。然而，在實際運用 GC 到神經科學領域時，有三個主要問題尚須進一步的研究與釐清：第一，原始的 GC 測度被設計成非負實數的形式，因此缺乏用來區分神經元間興奮性與抑制性因果關聯的有效特徵。第二，雖然 GC 已大量被使用，但估算出的因果強度與實際神經元之間突觸權重的關聯仍不清楚。第三，GC 無法直接用來解析大尺度的神經元資料，因為其中的變數數量經常遠遠超過所紀錄到的時間序列樣本數。對於上述前兩個問題，我們利用原始 GC 的架構並根據一個最佳線性預測子（best linear predictor, BLP）的假設，提出了能夠有效解析神經元網路間突觸權重的計算方法。在 BLP 的假設下，GC 可被擴展以同時測量神經元間興奮性與抑制性作用。我們設計了三種不同類型的模擬神經元網路來測試所提的新方法，包含從簡單的線性與近線性網路結構至複雜的非線性網路結構。這些模擬例子驗證了 BLP 假設的正當性以及計算方法的正確性。此方法也被示範性的用來分析大腦前扣帶皮層（ACC）與紋狀體（STR）真實的神經元活動資料。分析結果顯示，在注射 D2 多巴胺受體刺激劑的狀態下，ACC 當中以及由 ACC 投射至 STR 存在顯著的興奮性作用，而 STR 當中存在顯著的抑制作用。

實務上，從腦中擷取大尺度紀錄之神經訊號間的因果交互作用對於腦部特定功能之完整解析與可靠推論來說相當重要。然而，在大尺度訊號中，神經元的數量往往嚴重大過紀錄到的訊號長度；此過大的變數與樣本比值對於大多數現存的統計分析方法來說都會造成相當大的阻礙，甚至使分析方法在數學計算上完全失效。本文接著介紹一個三階段的變數選擇方法，它能夠有效地將一個大尺度變數集合調降成一個只包含有相關性變數的較小變數集，進而讓 GC 能夠應用到此調降的變數集合上。此方法使用（1）正交向前選擇解釋變量，（2）一停止準則以終止前向納入變量，和（3）進一步向後修剪消除不相關變量。對於上段提到的第三個問題，我們以此三階段方法為核心提出一個變量選擇演算法使得 GC 能夠被用來解析大尺度的時間序列資料。我們設計了一個以電位閾值激發的大尺度神經元網路模型來試驗所提方法的一致性。此方法應用在大鼠的行為資料分析中也得到全新的發現。這些框架提供了真實神經元網路分析的新方向。

關鍵詞：格蘭傑因果關係、格蘭傑因果指標、向量自迴歸模型、神經元網路、突觸權重估計、神經突觸指標、變數選擇、大尺度神經元網路、高維度資料分析。

Abstract

This thesis is devoted to developing statistical methods for identifying the information flow of neuronal networks. Granger causality (GC) is a very popular and also useful concept for estimating causal interactions between time-series signals. Over the past two decades, GC has emerged as a powerful analytical method for detecting the causal relationship among various types of neural activity data. However, three problems encountered when GC was used in the field of neuroscience remain not very clear and further researches are needed. Firstly, the GC measure is designed to be nonnegative in its original form, lacking of the trait for differentiating the effects of excitations and inhibitions between neurons. Secondly, how is the estimated causality related to the underlying neuronal synaptic weights? Thirdly, GC can not be applied to large-scale neuronal data, in which the number of variables is far greater than the length of time series. For the first two problems, we propose, under a best linear predictor (BLP) assumption, a computational algorithm for analyzing neuronal networks by estimating the synaptic weights among them. Under the BLP assumption, the GC analysis can be extended to measure both excitatory and inhibitory effects between neurons. The method was examined by three sorts of simulated networks: those with linear, almost linear, and nonlinear network structures. The method was also illustrated to analyze real spike train data from the anterior cingulate cortex (ACC) and the striatum (STR). The results showed, under the quinpirole administration, the significant existence of excitatory effects inside the ACC, excitatory effects from the ACC to the STR, and inhibitory effects inside the STR.

Extracting causal interactions from a large-scale recording of neural ensemble in the brain is very important to a comprehensive understanding or a reliable inference of certain brain functions. However, the oversized ratio between neuron number and signal length causes great difficulties to most of existing statistical methods. This thesis also introduces a three-stage variable selection approach that can be used to effectively reduce the large-scale variable set to a small but relevant one and enables Granger causality to be applied to the reduced set. The method uses (1) an orthogonalized forward selection of input variables, (2) a stopping criterion to terminate the forward inclusion of variables, and (3) a backward elimination to further trim irrelevant variables. For the third problem, we propose a computational algorithm which wraps the above selection approach as its core, enabling GC to work on large-scale time-series data. The method was examined by a large-scale simulated threshold spiking neuron model, and real behavioral data from rats were also analyzed. These frameworks give an insight into the analysis of real neuronal networks.

Keywords: Granger causality; Granger causality index; Vector autoregressive model; Neuronal networks; Synaptic weights estimation; Neuron Synaptic Index; Variable selection; Large-scale neuronal networks; High-dimensional data analysis.

Contents

List of Figures

List of Tables

Chapter 1

Brief Introduction to Neurons

The aim of this chapter is to briefly introduce some elementary notions and terminologies of neuroscience such as, action potentials, chemical synapses, neurotransmitters and postsynaptic potentials. We focus primarily on how neurons generate action potentials and how individual neurons communicate with each other via synapses. Based on these notions, we can then get into some interesting investigations on computational neuroscience such as spike-sorting errors, synaptic weights estimation and neuronal network reconstruction, etc. through the following chapters.

1.1   The neuron

Neurons are specialized cells that can receive and transmit signals. They are the building blocks of the whole nervous system and are responsible for communicating information throughout the body. There are approximately 100 billion neurons in human brain. Most of them consist of three primary components: the cell body (also called soma), the axon, and the dendrites. In general, dendrites are short and tree-like protrusion that extrude from the cell body. The highly branched structures can help increase the surface area of a neuron so that it can extensively receives signals transmitted by other neurons. The received signals will then be passed and joined on the soma.

The axon originates from the end of the soma at a tapered region called the axon hillock, where action potentials are generated and transmitted down the axon. The generation of action potentials will be elaborated in the next section. Unlike dendrites, axons can be very long. Almost all neurons have only one axon but there can be a considerable number of branches on the axon terminal. A neuron with thousands of branches on its axon terminal is not uncommon, and these branches usually spread closely to the dendrites or soma of other neurons. Sometimes, the structural features described above are not sufficient to distinguish dendrites from axons, then functional differences should be considered. Axons conduct action potentials outward from the soma while dendrites convert chemical signals into small electrical potentials and then transmit the impulses inward to the soma. Details will be given in the following two sections.

1.2   Action potentials

Neurons are surrounded by a membrane which possesses lots of ion channels that allow some ions to pass through across the membrane. Some important ions involved in electro-chemical messaging of neuron are: sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), and chloride (Cl^-). The membrane permeability causes different ion concentrations across the membrane, we call this difference in charge between the interior and exterior of the membrane a membrane potential. There are three primary types of ion channels: voltage-gated channels (typically on the axon hillock and the axon), chemically gated channels (typically on the dendrites), and leaky channels (by definition are always opened). The potassium leaky channels make K^+ cross through the membrane more easily than other ions. In the cell, there are many negatively charged macromolecules such as nucleic acid, proteins and ATP, these organic molecules cannot pass through the apolar layer of the membrane. Therefore, positively charged ions outside the cell will be attracted toward a negatively charged intracellular environment. Although the ions on both sides of the membrane try to balance out, it cannot be made due to the following two main reasons: (1) the membrane is selective, it allows only some kinds of ions to pass through. (2) sodium-potassium pumps actively transit three Na^+ out of the cell, and two K^+ into the cell simultaneously. Finally, under no external stimulus, when a steady state is reached on the both sides of the membrane, the voltage difference between them is called the resting membrane potential. The resting potential of a neuron is about -70 mV on average.

Action potentials are induced by different ions crossing the neuron membrane. First, an excitatory stimulus makes little sodium channels to open, increasing the membrane permeability to sodium. If some stimuli integrate and make the membrane potential to reach about -55 mV, then all the sodium channels will open and a large number of Na^+ ions will rush into the cell. We call this critical change level a firing threshold of a neuron and the electrical change a depolarization of the membrane. As Na^+ ions rush in, the membrane potential will move toward 0 mV and then keep going rapidly to about $+40$ mV, but it will not exceed $+60$ mV, the equilibrium potential of Na^+. Then the potassium channels also open and the sodium channels close. More and more K^+ ions move out of the cell so that the membrane potential declines, falling to about -80 mV in less than 2 milliseconds, but it will not exceed -90 mV, the equilibrium potential of K^+. This declining electrical change is called a repolarization of the membrane and is called a hyperpolarization if it goes beyond the resting potential. Finally, the potassium channels gradually close and the membrane potential will return to the resting potential slowly. If the membrane potential does not reach the threshold value, no action potential will occur. On the contrary, when the threshold level is reached, a fixed-size action potential will always generate. This is so-called the "all or none" principle. Here are two special periods: the absolute refractory period and the relative refractory period. The former comes from when the membrane potential hits $+40$ mV, all sodium channels are closed and the potassium channels are opend. During this short period of time, no further action potential can occur. The latter comes from when the membrane potential is more negative than at rest. During this period, it needs more stimuli to create another action potential.

1.3   Synaptic transmissions

Any behavior control in our body involves the neural network with at least two neurons inside. One neuron passes signals to another neuron through a synapse. Synapses are functional conjunctions between neurons, which is usually formed by the axon terminal of the input neuron (called presynaptic neuron) and the dendrite or soma of the target neuron (called postsynaptic neuron). Synaptic transmission can be accomplished either electrically or chemically. Electrical synapses conduct impulses faster and bidirectionally but are less modifiable than chemical synapses, which are the most common type of synapse found in the neural system of higher vertebrates. In this section, we will focus mainly on chemical synapse. A typical neuron possesses thousands of synapses. The synapse functions as a basic structure and place for messaging between neurons, integrating nerve information received from different sources and ensuring unidirectional nerve conduction. All of these are essential for effective and efficient communications between neurons.

The structure of a chemical synapse consists of three parts: the presynaptic axon terminal, the postsynaptic cell and the synaptic cleft between them. The cleft precludes the possibility of direct transmission of action potentials from the presynaptic membrane to the postsynaptic membrane. The main function of the presynaptic part is to release the so-called neurotransmitters, which act as messaging molecules, according to the action potentials arrived. The released neurotransmitters then travel (diffuse) across the synaptic cleft to the postsynaptic cell to alter its membrane potential by binding to some receptor molecules specifically designate for that neurotransmitters. These can produce either excitatory or inhibitory stimulations, depending on the properties of the receptor molecules, i.e., on the types of opened ion channels on the membrane. If the response results in the activation of sodium channels, making more Na^+ ions move into the cell then we call this local depolarization an excitatory postsynaptic potential (EPSP). On the other hand, if the response results in the activation of chloride or potassium channels, making more Cl^- or K^+ ions move into or move out of the cell, respectively, then we call this local hyperpolarization an inhibitory postsynaptic potential (IPSP).

It is worth noting that a single EPSP is not sufficient enough for the postsynaptic membrane to generate an action potential. Also, the density of voltage-gated sodium channels is very low at the dendrites or soma, making it more difficult to do that. In fact, the EPSPs and IPSPs are needed to be passively delivered to the axon hillock, where the density of sodium channels is high, to facilitate the generation of action potentials. However, in the passive delivery process, the amplitude of those PSPs decreases gradually with the distances being passed. Therefore, some forms of summation are required; in other words, those PSPs need to be added together at the axon hillock to surpass the threshold. The firing threshold plays an important role as a filter so that noisy random signals will not be transmitted so easily. Generally speaking, there are two types of mechanisms for summation: the first one is "temporal summation" and the other one is "spatial summation". Temporal summation means that single presynaptic neuron fires frequently such that the released neurotransmitter can successively act on the postsynaptic membrane, causing the superposition of potentials. Spatial summation means that

3

different presynaptic neurons, with their synapses located closely, fire simultaneously or nearly simultaneously so that the PSPs can be superimposed after the passive delivery over an extremely short distance.

In fact, not only the EPSP but the IPSP can take part in the summation process. The IPSPs will cancel out the EPSPs, making the membrane potential farther below the threshold, so that the initiation of an action potential in the postsynaptic neuron can be prevented. The firing bebavior of a neuron usually depends on the stimuli from many different input neurons. Inhibition, in general, plays an important role of regulating the excitatory stimuli for the postsynaptic neuron to fire. Neurons add up these EPSPs and IPSPs constantly in time and over different synapses, the net voltage at the axon hillock determines whether an action potential should be initiated. All of the signals, both excitatory and inhibitory, are transmitted from the input neurons and then integrated at the axon hillock of the target neuron. This process of gathering information and then making decision of neuron is called synaptic integration, which is so crucial for the neurons to perform certain meaningful coordinated neuronal activities in the brain.

Chapter 2

Effects of Spike Sorting Error

2.1 Introduction

In neuroscience research, it is important to identify information flow among multiple neurons in the brain, according to the recorded neural activity data. A powerful method for achieving this is the Granger causality (GC) which arose in economics after being introduced by Wiener and Granger [27, 28, 69]. The GC is a time series inference (TSI) type of method, proposes that if the prediction of one time series can be improved with the knowledge of a second time series, then there is a causal influence from the second time series to the first. This prediction is made by using the vector autoregressive (VAR) model. In this model, if the variance of the prediction error of one time series at the present time can be reduced by including the past values of another series, then the latter is said to Granger-cause the former. This causality can be quantified by the so-called GC Index (GCI) which can be used to determine whether there is any causal interaction between time series. The GC was shown to be effective and has been widely deployed in recent neuroscience research [7, 12–14, 71]. In addition to the time domain GC, other versions of the GC (e.g., frequency, and time-frequency domain) have been developed as well [3, 18]. The time domain formulation of GCI is briefly introduced in the next section, and we refer the reader to an article by [8] for more details about the GC.

Neurons emit action potentials (APs) that are known as *spikes* and play an important role in communicating among cells. The temporal sequence of APs produced by a neuron, which shows its own activity, is also known as a *spike train*. In multi-channel recordings [9], the APs of neurons are detected and differentiated from background electrical noise before single-unit spike trains are used to probe neural behaviors. This technical procedure is called *spike sorting*. However, it is not easy to obtain spike train data that fully agree with the AP because of noise, superimposed APs, and difficulties of differentiating waveforms of APs from different neurons. Spike sorting often introduces unavoidable errors [19, 43]. These errors can roughly be divided into two types, false positives (FPs) and false negatives (FNs). An FP means an error detection of an event that is not a real spike (just an electrical noise) or is a spike from another neuron. Conversely, an FN means that real spikes were not detected or were classified into groups of other neurons. One may be interested in the question: "How FPs and FNs affect the estimation of

functional connectivity among neurons?". This study answered this question analytically and also via numerical simulations. The change in the GCI due to spike sorting errors was derived analytically to form an explicit formula, and a direct discussion of the effects of FPs and FNs is possible. Moreover, numerical simulations were used to verify the analyses. We constructed three types of models for sorting errors: those with uniform, random, and concentrative distributions. That is, errors occur uniformly, randomly, and concentratively in spike trains. Changes in the GCI were computed as these types of spike sorting errors were artificially added to the simulated spike trains, and the effects on the directional interactions were also investigated.

Finally, it is worth noting that spike trains are non-equally spaced data and are regarded as a point process. Interpolation or filtering is usually employed to convert point processes to equally spaced time series. Previous studies on spike trains [37, 75] proposed several methods to convert a time series from being non-equally spaced to equally spaced. This study adopted the procedure of binning to convert spike trains into time series data, which are suitable for GC analyses. Although the GCI between two point processes has been directly defined in [38] recently, we still cannot abandon binning because it reduces the complexity of analysis, and considers also the effect of temporal summation of action potentials in the neuroscience.

The remanider of this chapter is organized as follows. Section 2 presents an analytic formula based on a first order autoregression to show how error processes affect the GCI. Section 3 presents some models for sorting errors and probes the proposed formula further via numerical simulations. Section 4 presents a real data evaluation where the effects of sorting error on the GCI are evaluated using real experimental data. Section 5 provides some suggestions for spike sorting and the discussion.

We remark that the material of this chapter has been published in [61].

2.2   Modeling and analysis

Based on a first order autoregression, we derived an explicit formula for changes in the GCI in terms of four parameters involving the error process. We also investigated the influences of various types of errors on the GCI indicated by the proposed formula.

2.2.1   A short introduction to the GCI

Let x and y be two stationary time series with zero means. The first order linear autoregressive model for x and y is given by

$$\left[\begin{array}{c} x(n) \\ y(n) \end{array} \right] = \mathbf{A} \left[\begin{array}{c} x(n-1) \\ y(n-1) \end{array} \right] + \left[\begin{array}{c} \epsilon(n) \\ \eta(n) \end{array} \right], \tag{2.1}$$

where $\mathbf{A}$ is the model coefficient matrix, and the residuals ϵ and η are zero-mean uncorrelated white noises with covariance matrix $\mathbf{\Sigma}$. Here the variances $Var(\epsilon)$ and $Var(\eta)$ are called prediction errors, which measure the accuracy of the autoregressive prediction.

More specifically, $Var(\eta)$ measures the accuracy of the prediction of $y(n)$ based on the previous values $x(n-1)$ and $y(n-1)$.

Now consider the reduced model that excludes the time series variable x

$$y(n) = B\,y(n-1) + \zeta(n), \tag{2.2}$$

where B is the corresponding model coefficient. The variance $Var(\zeta)$ measures the accuracy of the prediction of $y(n)$ based only on its previous value $y(n-1)$. For η in (2.1) and ζ in (2.2), if $Var(\eta)$ is significantly less than $Var(\zeta)$ in some statistical sense, then we say that x Granger-cause y. This causality can be quantified by the GCI from x to y formulated as:

$$F_{x \to y} = \ln \frac{Var(\zeta)}{Var(\eta)}. \tag{2.3}$$

It is clear that $F_{x \to y} = 0$ when $Var(\eta) = Var(\zeta)$, i.e., x has no causal influence on y, and $F_{x \to y} > 0$ when x Granger-cause y. Notice that $F_{x \to y}$ is nonnegative, i.e., $Var(\eta)$ is bounded above by $Var(\zeta)$, since the full model defined in (2.1) should have a better prediction ability than the reduced model defined in (2.2). Finally, we note that the GCI values should be checked for significance by using hypothesis testing, and more details of the GCI can be found in [20, 27, 28].

2.2.2 An explicit formula

When inaccurate spike sorting occurs, the sorting errors can be regarded as a perturbed error process. For simplicity, we assume that only the source process x has a sorting error and the corresponding error process is denoted by δx. We can assume that δx is zero mean and the model in (2.1) is perturbed as follows when δx is superposed on x:

$$\begin{bmatrix} \{x + \delta x\}(n) \\ y(n) \end{bmatrix} = \tilde{\mathbf{A}} \begin{bmatrix} \{x + \delta x\}(n-1) \\ y(n-1) \end{bmatrix} + \begin{bmatrix} \tilde{\epsilon}(n) \\ \tilde{\eta}(n) \end{bmatrix}, \tag{2.4}$$

where $\tilde{\mathbf{A}}$ is the corresponding model coefficient matrix, and the residuals $\tilde{\epsilon}$ and $\tilde{\eta}$ have the covariance matrix $\tilde{\boldsymbol{\Sigma}}$. Let $S_y := Var(\zeta)$, $S := Var(\eta)$, and $\tilde{S} := Var(\tilde{\eta})$. Since the perturbed quantity δx is superposed only on x, the reduced models for (2.1) and (2.4) are the same as (2.2). Then the original GCI from x to y and the perturbed GCI from $x + \delta x$ to y are

$$F = \ln \frac{S_y}{S} \quad \text{and} \quad \tilde{F} = \ln \frac{S_y}{\tilde{S}}, \tag{2.5}$$

respectively. To investigate the perturbed GCI, we derived an explicit formula for $\tilde{F}$ in terms of four parameters involving δx which are $\xi_1 := E(\delta x_1^2)$, $\xi_2 := E(x_1 \delta x_1)$, $\xi_3 := E(y_2 \delta x_1)$, and $\xi_4 := E(y_1 \delta x_1)$. Further denote $X_0 = E(x_1^2)$, $Y_0 = E(y_1^2)$, $Y_1 = E(y_1 y_2)$, $Z_1 = E(x_1 y_1)$, and $Z_2 = E(x_1 y_2)$. We are now ready to present the formula for $\tilde{F}$.

Proposition 1. In the situation described above, $\tilde{F}$ can be presented explicitly by the following formula (for calculation see Appendix):

$$\tilde{F} = \ln \frac{S_y}{S + \Theta}, \qquad \Theta = (S_y - S)I, \tag{2.6}$$

where

$$I = \cfrac{1}{\begin{aligned} & Y_0(X_0 + \xi_1 + 2\xi_2) - (\xi_4 + Z_1)^2 \\ & \times \Big\{ Y_0(X_0 + \xi_1 + 2\xi_2) - \tfrac{1}{S_y - S}\big[Y_0(\xi_3 + Z_2)^2 \\ & + (Y_0 - S)(\xi_4 + Z_1)^2 - 2Y_1(\xi_3 + Z_2)(\xi_4 + Z_1)\big] \Big\} \end{aligned}} \tag{2.7}$$

Note that since $S + \Theta$ in (2.6) is bounded above by S_y, we have that Θ is upper bounded by $S_y - S$, i.e., I has an upper bound 1.

We end this subsection by the following two remarks.

Remark 1. In the same situation of Proposition 1, the following inequalities hold:

$$Y_0 \geq S_y \geq S \quad \text{and} \quad Y_1 \leq 0. \tag{2.8}$$

According to (2.2), we have $Y_0 = Var(y_1) \geq Var(\zeta) = S_y$. The remainder $S_y \geq S$ just follows by the reason that the prediction error of the reduced model in (2.4) is always less than or equal to that of the full model in (2.1). The latter holds because of the stationary assumption. If $Y_1 = E(y_1 y_2) > 0$, then y will not be stationary. Thus $Y_1 \leq 0$.

Remark 2. The following result can be obtained easily by using (2.5) and (2.6).

$$\begin{aligned} \tilde{F} &> F, & \text{if} \quad I < 0. \\ \tilde{F} &= F, & \text{if} \quad I = 0. \\ \tilde{F} &< F, & \text{if} \quad 0 < I < 1. \\ \tilde{F} &= 0, & \text{if} \quad I = 1. \end{aligned} \tag{2.9}$$

2.2.3 Essential GCI factors

The formula defined by (2.6) and (2.7) is complicated but it reveals the intrinsic property of the perturbed GCI, depending on the four factors ξ_1, ξ_2, ξ_3 and ξ_4 which, by the definition, capture the main properties of error signals, namely, the spike sorting errors. A systematic characterization of these parameters' influences on the GCI would provide a heuristic understanding of the effect of spike sorting error for researchers to make further decisions. Thus, in this subsection we discuss Proposition 1 more by a total of five different situations and relate each of them to a biological meaning.

We now present the following corollaries for investigating the term I defined in (2.7), and this is equivalent to investigating the term $\tilde{F}$ in (2.6).

Corollary 1 ($\xi_2 = \xi_3 = \xi_4 = 0$). We start with the simple case in which the error process δx is uncorrelated with the underlying processes x and y, i.e., $\xi_2 = \xi_3 = \xi_4 = 0$. In this case I can be simplified as follows:

$$I(\xi_1) = \frac{Y_0 \xi_1}{Y_0 \xi_1 + [X_0 Y_0 - Z_1^2]}. \tag{2.10}$$

The above equation shows that I increases as ξ_1 increases, but it is bounded above by the limit $L := \lim_{\xi_1 \to \infty} I(\xi_1) = 1$. Hence, by (2.6), the weakened GCI $\tilde{F}$ is bounded below by $\lim_{\xi_1 \to \infty} \tilde{F}(\xi_1) = \ln \frac{S_y}{S + (S_y - S)L} = \ln \frac{S_y}{S_y} = 0$. In reality, this limit cannot be attained because the variance $\xi_1 = E(\delta x_1^2)$ cannot approach infinity. Thus, the GCI will never vanish if the error process δx is uncorrelated with the underlying process x. We refer to this corollary as FPs being composed of electrical noises or the spikes of other unconnected neurons during spike sorting.

Corollary 2 ($\xi_2 < 0, \xi_3 < 0, \xi_4 < 0$). Suppose the error process δx is negatively correlated with the underlying processes x and y, i.e., ξ_2, ξ_3, and ξ_4 are all negative. Since the perturbed quantity δx is considered to be produced from inaccurate spike sorting, the maximal negative quantity which δx can be is $-x$. Therefore, we have the following constraint:

$$-X_0 \leq \xi_2 < 0, \ -Z_2 \leq \xi_3 < 0, \ -Z_1 \leq \xi_4 < 0, \tag{2.11}$$

and the equalities are attained when $\delta x = -x$.

Now, according to (2.7), (2.8), and (2.11), I is positive, increasing, and bounded above by 1. In other words, $\Theta \to S_y - S$ and $\tilde{F} \to 0$ as $\xi_1 \to \infty$ or $(\xi_2, \xi_3, \xi_4) \to (-X_0, -Z_2, -Z_1)$. This corollary is related to FNs in spike sorting.

Corollary 3 ($\xi_2 < 0, \xi_3 < 0, \xi_4 < 0$ **simplified**). Suppose δx is correlated with the underlying processes as in Corollary 2. If y is further completely induced by x, i.e., y cannot explain itself ($B = 0$ in (2.2)), then we obtain $S_y = Y_0$, $Y_1 = 0$ and (2.7) can be further simplified as:

$$I = \frac{Y_0(X_0 + \xi_1 + 2\xi_2) - (\xi_4 + Z_1)^2 - \frac{Y_0}{S_y - S}(\xi_3 + Z_2)^2}{Y_0(X_0 + \xi_1 + 2\xi_2) - (\xi_4 + Z_1)^2}. \tag{2.12}$$

Equation (2.12) still shows us that $I \to 1$ as $\xi_1 \to \infty$. On the other hand, if ξ_1 is fixed and the negative correlation between x and δx increases (i.e., (ξ_2, ξ_3, ξ_4) decreases simultaneously to $(-X_0, -Z_2, -Z_1)$), then the value I increases to 1 because of the quadratic convergence:

$$(\xi_3 + Z_2)^2 \to 0. \tag{2.13}$$

Note that when (ξ_2, ξ_3, ξ_4) attains the lower bound $(-X_0, -Z_2, -Z_1)$, i.e., $\delta x = -x$, there is nothing left to analyze. Therefore, we do not consider this case.

Corollary 4 $(\xi_2 = 0, \xi_3 > 0, \xi_4 > 0)$. Suppose the error process δx is positively correlated with y, and is uncorrelated with x, i.e., $\xi_2 = 0$, $\xi_3 > 0$, and $\xi_4 > 0$. Since $I = 0$ when $\xi_1 = \xi_2 = \xi_3 = \xi_4 = 0$, it is easy to conclude that, in this case, I is negative and decreasing, i.e., $\tilde{F} > F$ and $\tilde{F}$ increases as ξ_3 or ξ_4 increase. We refer to this corollary as FPs being composed of spikes of some connected neurons which are positively correlated with the target neuron during spike sorting.

Corollary 5 $(\xi_2 > 0, \xi_3 > 0, \xi_4 > 0)$. Suppose the error process δx is positively correlated with both x and y, i.e., $\xi_2 > 0$, $\xi_3 > 0$, and $\xi_4 > 0$. Since I is increasing in ξ_2, we know that $\tilde{F}$ is then decreasing in ξ_2. Therefore, $\tilde{F}$ in this case exhibits the same behavior as that in Corollary 4 if (ξ_3, ξ_4) dominates ξ_2; but $\tilde{F}$ is decreasing if ξ_2 dominates (ξ_3, ξ_4). We refer to this corollary as FPs being composed of spikes of some connected neurons which are positively correlated with both of source and target neurons during spike sorting.

The results of above five corollaries are schematically summarized in Table 2.1.

Table 2.1: A schematic summary of Corollaries 1–5. The factors $\xi_i, i = 1, 2, 3, 4$ and error process δx are described in Section 2.2. $+ (-)$ stands for positive (negative) sign. FP (FN) stands for false positive (negative). $\uparrow (\downarrow)$ stands for increasing (decreasing).

ξ_2	ξ_3	ξ_4	Effects on the GCI	Type	Interpretations on δx
0	0	0	GCI $\downarrow$ as $\xi_1 \uparrow$	FP	Electrical noises or spikes of unconnected neurons.
$-$	$-$	$-$	GCI $\downarrow$ as $\xi_1 \uparrow$	FN	Spike missing.
0	$+$	$+$	GCI $\uparrow$ as $\xi_3 \uparrow$ or $\xi_4 \uparrow$	FP	Connected neurons which are positively correlated with the target neuron.
$+$	$+$	$+$	GCI $\uparrow$ if (ξ_3, ξ_4) dominates ξ_2 GCI $\downarrow$ if ξ_2 dominates (ξ_3, ξ_4)	FP	Connected neurons which are positively correlated with both of source and target neurons.

2.2.4 GCI vs. variance reduction

Because $0 \le S \le S_y$, we set $S = (1-k)S_y$ with $0 \le k \le 1$ and the GCI in (2.5) becomes

$$F = \ln \frac{S_y}{(1-k)S_y} = -\ln(1-k). \tag{2.14}$$

Next, relate F and k through $k = 1 - \exp(-F)$. Since $S_y - S = kS_y$, $k = (S_y - S)/S_y$ represents the percentage of variance reduction. More precisely, it represents the relative decrease in prediction errors from the reduced model (2.2) to the full model (2.1). For example, $k = 0$ means no (0%) variance reduction, and the GCI is equal to zero. On the contrary, $k = 1$ means a total (100%) variance reduction, and the GCI is equal to infinity. Figure 2.1 shows how the GCI relates to k, and it is almost totally reduced when the GCI is equal to 5.

GCI vs. the percentage of variance reduction
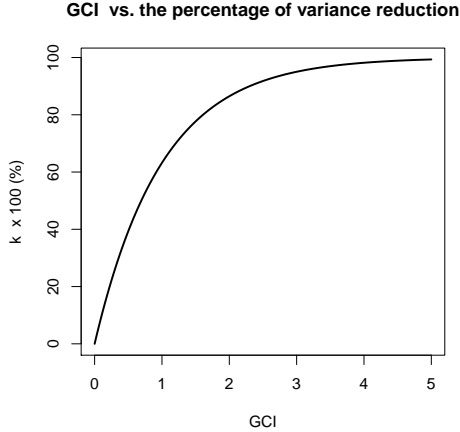
Figure 2.1: The relationship between the Granger Causality Index (GCI) and k. This transforms GCI values to the corresponding percentage of variance reduction in the residual noise process.

2.3 Simulation study

Here, we present some models for sorting errors and further probe formula (2.7) via numerical simulations. For the FP case, we first consider that FPs are made up of electrical noises or spikes of unconnected neurons. Spikes of connected neurons are considered in Section 2.3.4.

2.3.1 Models

We first construct a point process $X = \{p_1, p_2, \ldots, p_N\}$ generated by a Poisson process with rate λ in the time interval $[0, T]$, where N is the total number of points. The second point process $Y = \{q_1, q_2, \ldots, q_N\}$ is then generated by $Y = X + \mathcal{N}(m, \sigma)$, where $\mathcal{N}(m, \sigma)$ is a normal random variable with mean m and standard deviation σ. More precisely, $q_i = p_i + \mathcal{N}(m, \sigma)$, $i = 1, \ldots, N$. The point process Y presents a time lag m with respect to X if $m > 0$ and $\sigma = 0$. This study only considers $m > 0$ and $\sigma > 0$.

Spike sorting errors include two types: fake (FP) and missing (FN) spike events. The fake spike event is an erroneous detection of an event that is not a real spike or is a spike from another neuron. Conversely, a missing spike event means that spikes were not detected or were classified into groups of other neurons. Therefore, the fake spike case can be regarded as adding extra points to a point process. This type of error is denoted by "A-type". The missing spike case can be regarded as removing some points from a point process, and is denoted by "R-type". Forms of the addition or removal of points are considered for uniform, random, and concentrative distributions. The number of fake or missing spike events is rN, where $r(= p\%)$ is the ratio of fake or missing points compared to the original spike train. The following explains the generation of these types in detail.

Uniform-partition addition (PA) model: The extra points $\{\tilde{p}_i : 1 \leq i \leq rN\}$ on the time interval $[0, T]$ form a uniform partition of $[0, T]$. More precisely, $\tilde{p}_i =$

$i\Delta t$, where $\Delta t = \frac{T}{rN-1}$.

Random-uniform addition (UA) model: The extra points $\tilde{p}_i$, $i = 1, \ldots, rN$, on the time interval are generated from a uniform distribution $\mathcal{U}([0,T])$ random variable.

Random-normal addition (NA) model: The extra points $\tilde{p}_i$, $i = 1, \ldots, rN$, on the time interval are generated from a normal distribution $\mathcal{N}(T/2, \sigma_{NA})$ random variable. The standard deviation parameter σ_{NA} represents different degrees of concentration.

Uniform-partition removal (PR) model: A set of reference points $\{\tilde{p}_j : 1 \leq j \leq rN\}$, which is a uniform partition of $[0, T]$ was used to remove spike events from $\{p_i : 1 \leq i \leq N\}$. First, fix $\tilde{p}_j$ and then remove the point that is closest to $\tilde{p}_j$.

Random-uniform removal (UR) model: Points in $\{p_i : 1 \leq i \leq N\}$ are randomly removed using a discrete uniform $\mathcal{U}_d(1, N)$ random variable, i.e., all points have the same probability of being removed.

Middle-succession removal (SR) model: In this model, the removed points are successive and located near the center $T/2$ of the time interval $[0, T]$. The number of spikes is rN.

We note that sorting errors only occur in the source process X, and Y is assumed to be inerrable.

2.3.2 Setup

Set the parameters, $\lambda = 2$, $T = 100$, $m = 0.1$ for the rate of the Poisson process, the total time, and the time lag, respectively. To apply autoregressive modeling, we convert point processes $\{p_i : 1 \leq i \leq N\}$ and $\{q_i : 1 \leq i \leq N\}$ to time series through the procedure of binning with the bin size as the time lag m. Results are obtained from the average of 100 simulations for each random case at fixed error percentage r. To investigate the effects of errors on the GCI, we observe the results from various r, σ, and σ_{NA}.

2.3.3 Simulation results

We note that the simulations and the corresponding results are related to Corollaries 1, 2, and 3 of Section 2.2.3.

A-type vs. R-type

Figure 2.2(a) shows the results for experiment 1 in which $\sigma = 0.02$, $\sigma_{NA} = T/8$, and the error percentage increased by 0.1. In this figure, the GCI of the SR model decreases the fastest among all of the models as r increases. All R-type errors cause information loss and greatly weaken the GCI when the error percentage increases. If $r = 1$, the underlying signal is totally destroyed by the errors, and the causality is undetermined. The PA model produces the largest GCI for a fixed r. Figure 2.2(a) also shows that A-Type GCIs are greater than R-Type GCIs. However, this phenomenon is not always valid. The error

12

in the NA model is uncorrelated with the signal and weakens the GCI more than any of the R-Type models when σ_{NA} is small. Figure 2.2(b) shows that a highly non-stationary process can average out much more underlying causality than the others, for the case of $\sigma_{NA} = T/64$ in the NA model.
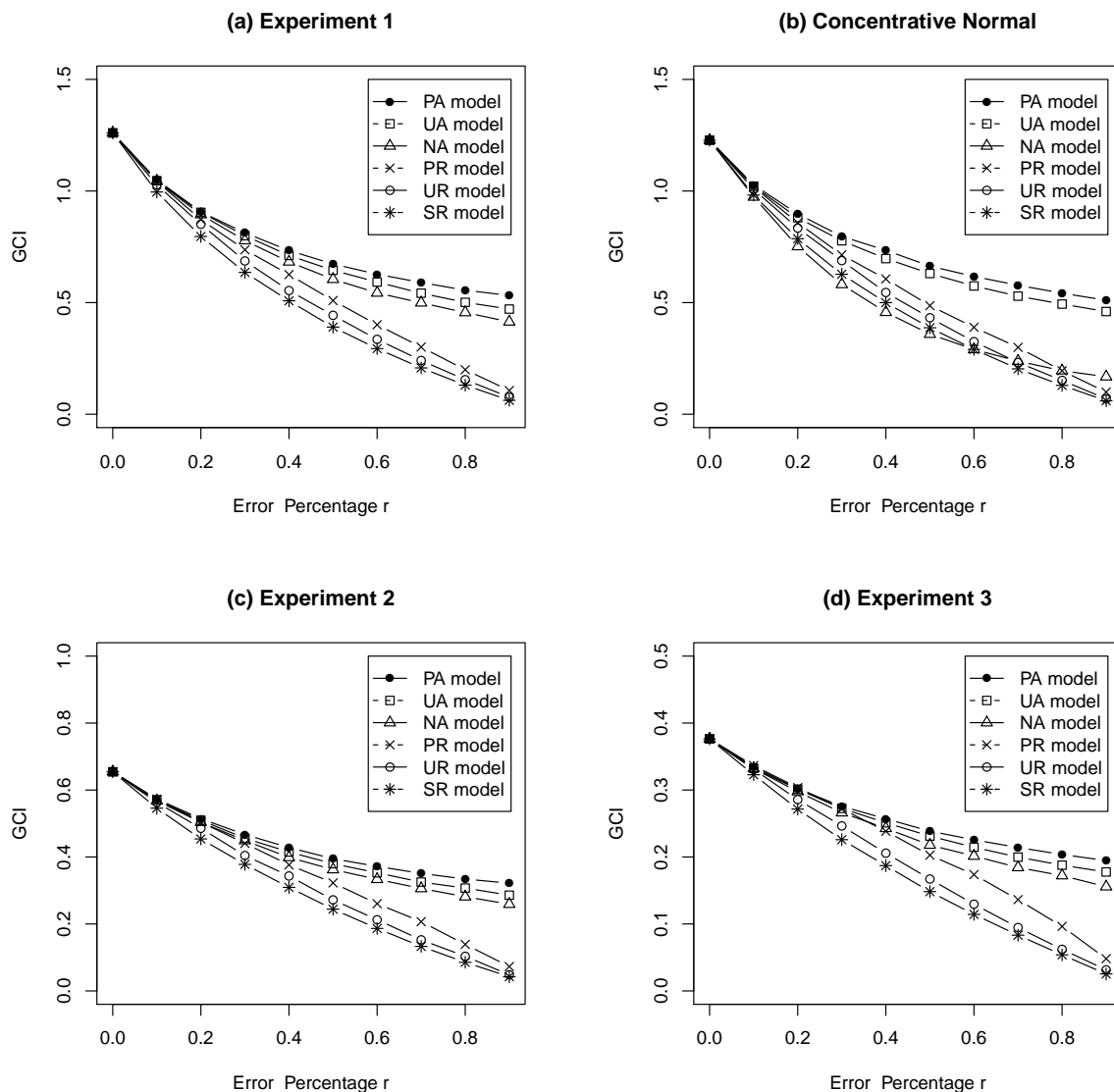
Figure 2.2: (a) Simulation results of Experiment 1, in which the parameter settings were $\lambda = 2$, $T = 100$, $m = 0.1$, $\sigma = 0.02$. (b) Simulation results of a much more concentrative normal, in which the parameter settings were $\lambda = 2$, $T = 100$, $m = 0.1$, $\sigma = 0.02$, and $\sigma_{NA} = T/64$. (c) Simulation results of Experiment 2, in which the parameter settings were $\lambda = 2$, $T = 100$, $m = 0.1$, $\sigma = 0.04$. (d) Simulation results of Experiment 3, in which the parameter settings were $\lambda = 2$, $T = 100$, $m = 0.1$, $\sigma = 0.06$. The error percentages r were all $0 \sim 0.9$, and increased by 0.1.

Standard deviation factor σ

We now investigate the effects of the standard deviation σ on GCIs. Figure 2.2(c) and 2.2(d) show the results for experiments 2 and 3 in which parameter σ is 0.04 and 0.06, respectively. Observations in Figure 2.2(a), 2.2(c), and 2.2(d) show that the behaviors of the profiles with $\sigma = 0.04, 0.06$ closely resemble the behavior of the profile with $\sigma = 0.02$. In the PA and PR models, let F denote the GCI value without a sorting error ($r = 0$), and $\tilde{F}(r)$ denote the GCI value with sorting error of error percentage r. Table 2.2 presents the relative errors of the GCI for r from 0.1 to 0.9 of the PA and PR models in the three experiments, where

$$\text{Relative error}(r) := \frac{\tilde{F}(r) - F}{F} \times 100\%.$$

Table 2.2 shows that Experiment 3 had the flattest PA-curve and PR-curve of the three experiments, and the underlying causality of the PA models were all around 50% off at $r = 0.9$ because the signal to noise ratio $\left(SNR = \frac{Var(x)}{Var(\delta x)}\right)$ was close to 1. Although they were all around 50% off, the corresponding decreases in the percentage of variance reduction are greatly differed, and this can be seen from Figure 2.1. In the PA model, the three F values corresponding to the three experiments were 1.24, 0.64, and 0.38, and the corresponding $\tilde{F}(0.9)$ were 0.71, 0.33, and 0.18, respectively. Equation (2.14) can compute the corresponding relative decreases in the percentage of variance reduction, which were 28.17%, 40.89%, and 46.85%, respectively.

Table 2.2: Relative errors, $\frac{\tilde{F}(r) - F}{F} \times 100\%$, of GCI for $r = 0.1 \sim 0.9$ of the PA and PR models in the three experiments.

Relative errors (%) of the PA model									
Experiment	$r = 0.1$	$r = 0.2$	$r = 0.3$	$r = 0.4$	$r = 0.5$	$r = 0.6$	$r = 0.7$	$r = 0.8$	$r = 0.9$
$1(\sigma = 0.02)$	-16.46	-27.24	-35.29	-41.48	-46.26	-50.26	-53.43	-55.78	-57.60
$2(\sigma = 0.04)$	-12.27	-21.87	-29.41	-34.36	-39.84	-42.63	-46.28	-48.93	-51.20
$3(\sigma = 0.06)$	-11.59	-19.78	-26.99	-32.13	-36.63	-40.89	-43.42	-46.29	-48.41
Relative errors (%) of the PR model									
Experiment	$r = 0.1$	$r = 0.2$	$r = 0.3$	$r = 0.4$	$r = 0.5$	$r = 0.6$	$r = 0.7$	$r = 0.8$	$r = 0.9$
$1(\sigma = 0.02)$	-17.10	-30.16	-41.89	-51.31	-59.78	-68.60	-75.86	-84.02	-92.07
$2(\sigma = 0.04)$	-12.64	-22.71	-32.74	-41.82	-50.63	-59.24	-68.72	-78.04	-88.96
$3(\sigma = 0.06)$	-11.58	-19.21	-27.60	-37.45	-47.44	-54.15	-65.07	-75.34	-87.68

Explanation of the GCI curves

We now discuss the behaviors of the curves in Figure 2.2(a). Because A-type error processes (δx) are uncorrelated with the underlying process (x), curves of the PA, UA, and NA models can be analyzed and directly explained by (2.6) and (2.10). These three curves decrease as the size of the errors (ξ_1) increases. They reach zero only when the error size is infinity, which is actually unfeasible. Therefore, these curves slowly decrease and never reach zero. In other words, the underlying causality remains.

Moreover, the PA-curve is above the UA-curve, and the UA-curve is above the NA-curve. This is because the NA model has the largest error size, followed by the UA model and the PA model at the same error percentage. On the other hand, error processes of the PR, UR, and SR models are correlated with the underlying process, and (2.12) is used instead of (2.10) since ξ_2, ξ_3, and ξ_4 are nonzero.

With approximately the same error size, these curves decrease much more quickly than the uncorrelated case because of the negative correlations. Unlike the former, these curves almost reach zero at the error percentage $r = 0.9$. In addition, the PR-curve is above the UR-curve, and the UR-curve is above the SR-curve. This is because the SR model has the largest error size, followed by the UR and PR models at the same error percentage.

We now discuss the behaviors of the curves in Figure 2.2(b), which shows that A-type errors are not always better than R-type errors. GCI values of the NA model with $\sigma_{NA} = T/64$ are smaller than those of the SR model when the error percentage is between $r = 0.2$ and $r = 0.6$. To see what occurred, we computed ξ_1 to ξ_4 at each error percentage for these two cases, and results are respectively shown in Figure 2.3(a) and 2.3(b). Comparing these two panels indicates that (i) ξ_1 of the NA model is much larger than that of the SR model. According to (2.12), the GCI of the NA model is smaller than that of the SR model. (ii) The SR model has large ξ_2 and ξ_3. According to (2.13), the GCI of the SR model is significantly smaller than that of the NA model when r is quite large, even though the ξ_1 of the SR model is still relatively smaller than that of the NA model. Figure 2.2(b) verifies this analysis showing that the error size of the NA model dominates the negative correlation of the SR model when the error percentage is between $r = 0.2$ and $r = 0.6$. Contrarily, the negative correlation of the SR model dominates the error size of the NA model when $r > 0.6$. Therefore, it is necessary to integrate the error size and the negative correlation to identify the behavior of the GCI, and the reason is that the error size of the NA model is much larger than that of the SR model when $0 \leq r \leq 0.6$.

2.3.4 Supplementary simulations of FPs

We now present supplementary simulations of FPs by considering the case when FPs are composed of connected neurons. There are no supplementary simulations for FNs since they have only one situation as discussed in Corollary 2, and they were simulated in preceding work. The simulations here are devoted to Corollaries 4 and 5 of Section 2.2.3. The following explain the simulations in detail and the model parameters are fixed at $\lambda = 2$, $T = 100$, $m = 0.1$, and $\sigma = 0.02$.

Experiment A: Suppose X and Z are two independent Poisson processes with equal rate λ in $[0, T]$, where N is the total number of points of X. Let Y be another point process generated by $Y = Z + \mathcal{N}(m, \sigma)$. Hence, we have that Y is induced by Z, and X is independent of both Y and Z. Then rN points of Z are randomly added to X, where r denotes the error percentage. After binning with bin size m, the GCI from X to Y as a function of r and the corresponding ξ's are shown in Figure 2.4(a) and 2.4(b). This experiment considers the situation that

(a) The ξ1 to ξ4 of NA model   (b) The ξ1 to ξ4 of SR model
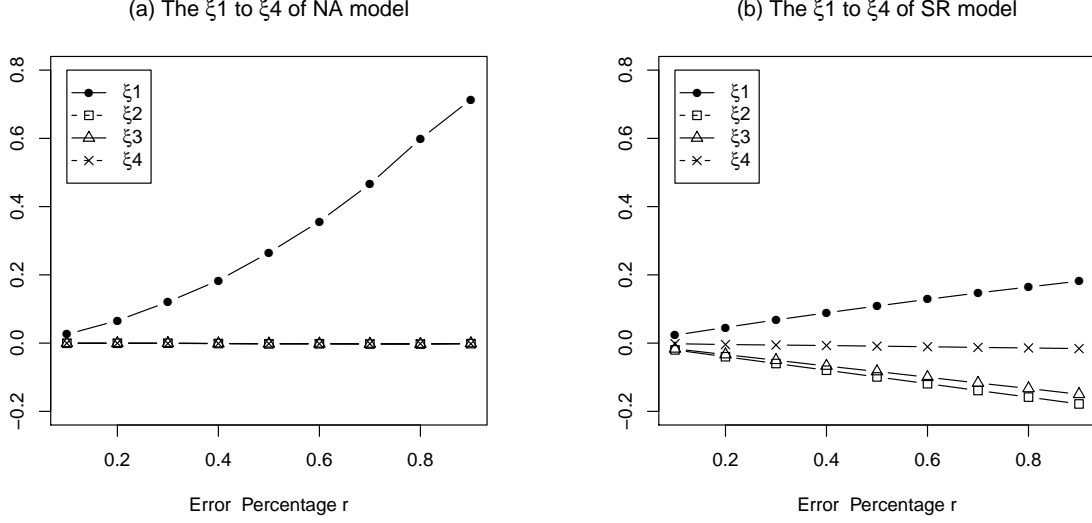
Figure 2.3: (a) ξ_1 to ξ_4 of the NA model, in which the parameter settings were $\lambda = 2$, $T = 100$, $m = 0.1$, $\sigma = 0.02$, $\sigma_{NA} = T/64$. (b) ξ_1 to ξ_4 of the SR model, in which the parameter settings were $\lambda = 2$, $T = 100$, $m = 0.1$, $\sigma = 0.02$. The error percentages r were both $0 \sim 0.9$, and increased by 0.1.

FPs are composed of spikes of connected neurons which are positively correlated with target neurons. Note that the true causality between X and Y is uncorrelated (GCI= 0), which means that the relationship between neurons may be erroneously identified when a FP occurs during spike sorting.

Experiment B: Here we consider another situation that Y is induced by both X and Z, that is, $Y = \{X + \mathcal{N}(m, \sigma)\} \cup \{Z + \mathcal{N}(m, \sigma)\}$, where X and Z are independent Poisson processes with equal rate λ in $[0, T]$. Then rN points of Z are randomly added to X. After binning, the GCI from X to Y and the corresponding ξ's are shown in Figure 2.4(a) and 2.4(c). The result shows that the relationship is correctly identified, but the strength (GCI) is overestimated. Experiments A and B enlighten that the risk of an erroneous causal identification should be estimated by two parts: the relationship (causal or noncausal) and the strength (the magnitude of the GCI).

Experiment C: Suppose Z is a Poisson processes with rate λ in $[0, T]$. Let X and Y be point processes generated by $X = Z + \mathcal{N}(m, \sigma)$ and $Y = X + \mathcal{N}(m, \sigma)$, i.e., X is induced by Z, and Y is induced by X. Then rN points of Z are randomly added to X. After binning, the GCI from X to Y and the corresponding ξ's are shown in Figure 2.4(a) and 2.4(d). The result shows that the effect of the positive-correlation ($\xi_2 > 0$) between Z and X dominates the effect of the positive-correlation ($\xi_3 > 0$, $\xi_4 > 0$) between Z and Y ; thus the GCI decreases. This experiment is devoted to Corollary 5 of Section 2.2.3.
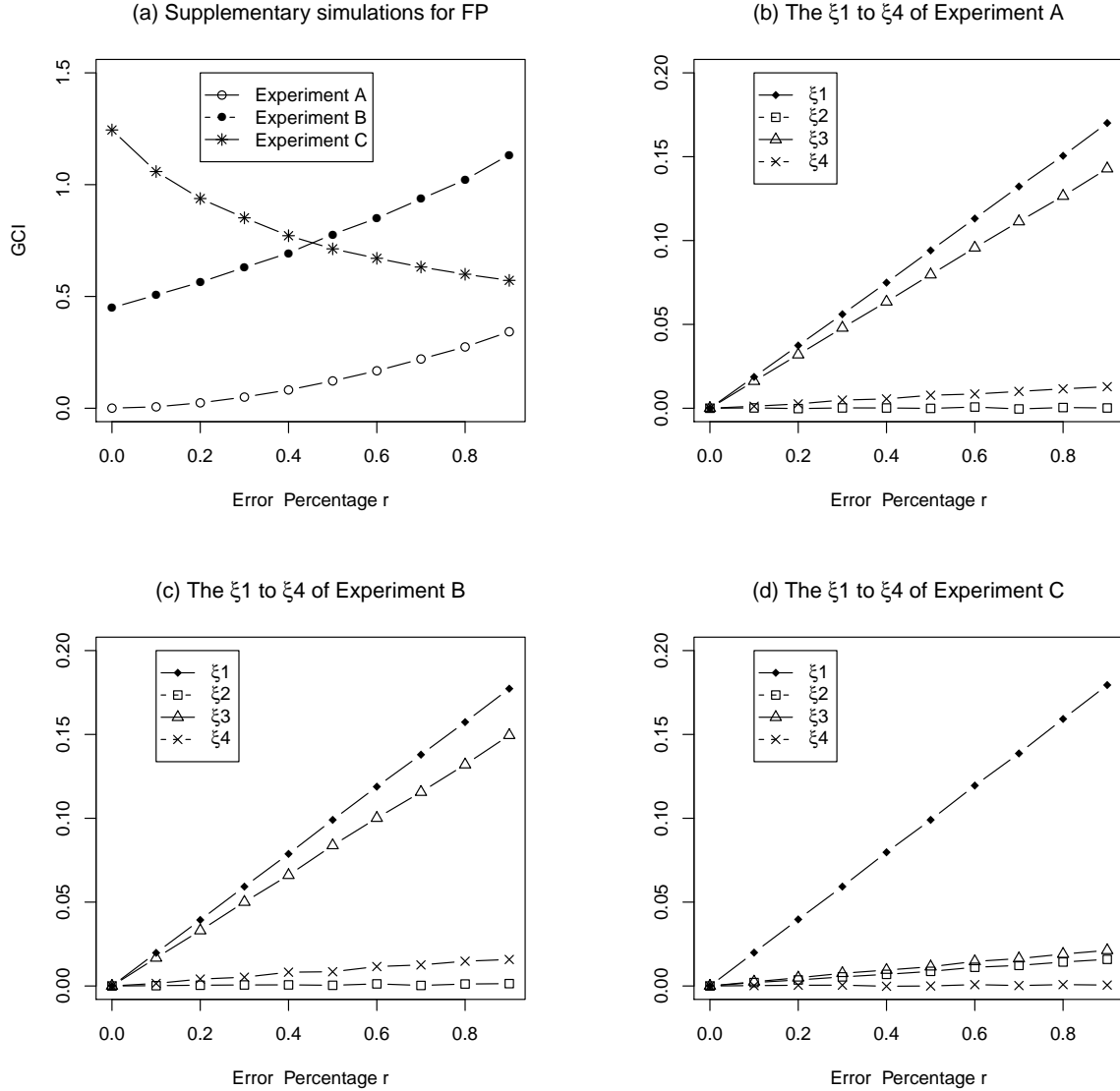
Figure 2.4: (a) Simulation results of Experiment A–C for FP. (b) ξ_1 to ξ_4 of the Experiment A. (c) ξ_1 to ξ_4 of the Experiment B. (d) ξ_1 to ξ_4 of the Experiment C.

2.3.5 Simulation for threshold detection

Spike sorting consists of two parts: AP detection and AP classification, which are based on thresholding and clustering methods, respectively. Here we discuss the relationship between the GCI value and the detecting threshold via simulation. The classification part will be discussed in Section 4 by using real experimental data.

We simulate a sequence of 100 APs, denoted by $\mathcal{A}$, having fixed interspike interval (ISI) length as shown in Figure 2.5(a). Then we add an independent white noise to $\mathcal{A}$ with $SNR = 0.8$ for the background noise. Denoting the standard deviation of the observed noisy data by τ, Figure 2.5(b) and 2.5(c) show the detected spikes with threshold values

being 2.5τ and 3.0τ, respectively. It is easy to see that a lower threshold is tending to result in FPs of spike-detection and a higher one is tending to result in FNs. Now, let pX be the point process obtained from perfect-detection, i.e., pX coincides with $\mathcal{A}$. Let $pY = pX + \mathcal{N}(0.1, 0.02)$ represent the point process obtained from a causal sequence of APs, say $\mathcal{B}$. After binning with bin width 0.1, we can infer the causal relationship between these two sequences of APs, $\mathcal{A}$ and $\mathcal{B}$, by computing the GCI from the binned data of pX to that of pY. To be consistent with our analyses, the errors created by threshold-detection will be put only on the source sequence $\mathcal{A}$. We are now ready to investigate the relationship between GCI value and threshold value by changing the threshold from 1.5τ to 3.5τ, and it is shown in Figure 2.6(a). The result shows that 2.5τ performs the best (i.e., obtaining the largest GCI value), since we know that $\mathcal{B}$ is induced by $\mathcal{A}$. To investigate more deeply into this result, numbers of FPs and FNs are further shown in Figure 2.6(b) and 2.6(c), and these results give us the following findings: (i) Number of FPs increases as threshold value decreases. (ii) Number of FNs increases as threshold value increases. (iii) FPs affect GCI less than FNs since the number of FPs (> 100 at 1.5τ) is much larger than that of FNs (> 60 at 3.5τ) and the decreasing rate (slope) of GCI in $[2.5, 3.5]$ ($= 0.4270$) is larger than the increasing rate (slop) of GCI in $[1.5, 2.5]$ ($= 0.3674$). (iv) As a result of (iii), we can conclude that choosing a threshold lower than the optimal ($= 2.5\tau$) is better and preserves more information than choosing a threshold higher than the optimal.

We now discuss in details the result of this simulation through the following four remarks: (1) For a large threshold, the sorting error is only composed of FNs, without FPs. Therefore the GCI increases as threshold decreases. When the threshold decreases to certain value, the error of FPs occurs. The GCI will reach the maximum and then decreases as the threshold decreases to 0. This result contributes the explanation of the effects of FPs and FNs on the GCI. (2) We can conclude that variations of the GCI are determined by which one of FP and FN to be the dominate sorting error and the total number of FPs and FNs as well for a fixed threshold. (3) Although we cannot choose the optimal threshold in real experiments, (iv) is still useful and it gives a criterion for designing methods of choice of an optimal threshold. (4) Researchers may imitate the procedure of the simulation by using their own $\mathcal{A}$ and background noise to determine the optimal threshold after examining the number of FPs and FNs.

2.4 Real data analysis

Here we design two sorting procedures in real operation and then evaluate the effect of sorting errors on the GCI using real experimental data.

2.4.1 Experimental setup

Neuronal spikes were recorded from the ventroposterior medial (VPM) nucleus of the thalamus and are the same data set used in our previous study [67]. The single-unit recording method is described in Tseng's report (2012). Briefly, spikes were amplified (7000 32,000-fold), filtered (0.25 13 kHz), and digitized at 40 kHz. Recording was performed while a rat was awake. Extracellular single units were recorded in real time

(a) Underlying action potentials

(b) Spike–detection with threshold value 2.5 s.d.

(c) Spike–detection with threshold value 3 s.d.
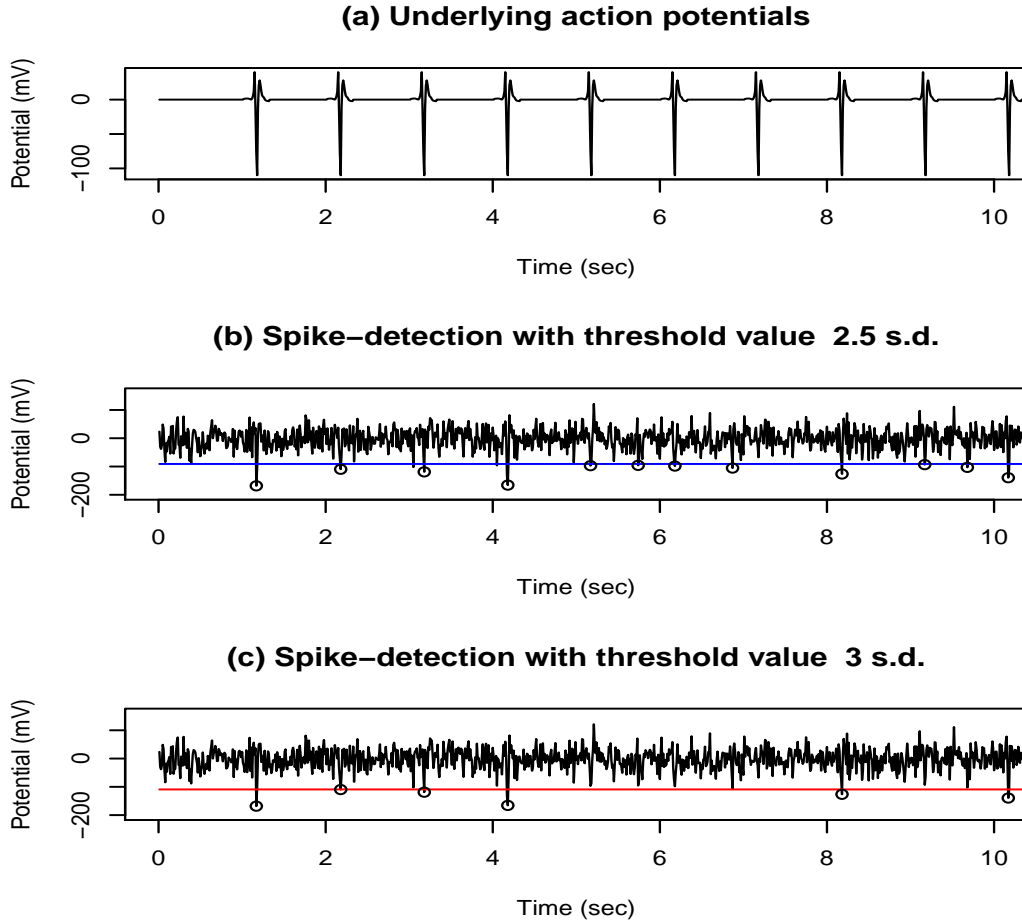
Figure 2.5: (a) The first 10 APs of the underlying AP sequence $\mathcal{A}$. (b) The detected spikes (marked by ○) with threshold value 2.5τ. (c) The detected spikes (marked by ○) with threshold value 3.0τ.

using time-voltage windows and a principle component-based template-matching algorithm (Sort Client, Plexon). Waveforms were saved and re-sorted using Offline Sorter (Plexon), based on principle-component clustering, with a user-defined template. The sample we used here contained 2 or more distinguishable clusters. To evaluate the effect of sorting errors on the GCI, various percentage errors were created from 20%, 40%, 60%, and 80% less or more than the data set of a complete cluster. Shrinkage or expansion of the sample size was calculated based on the difference between a waveform of a neuron and a template, computed by the Offline Sorter (tolerance fit function). Note that cluster expansion included the other cluster of a neuron or noises. We used 6 neurons with an averaged firing frequency of 0.199 Hz, and therefore $\binom{6}{2} \times 2 = 30$ neuron pairs (i.e., GCIs) were derived ((Neuron^i,Neuron^j) $i = 1, \ldots, 6 \quad j \neq i$). Being consistent with our analyses, the errors created by shrinkage or expansion were put only on the source (Neuron^i). In the sequel, we use the FN and FP-procedures to respectively represent shrinkage and expansion operations.

19

(a) GCI result with varying threshold values

(b) Number of False positives

(c) Number of False negatives
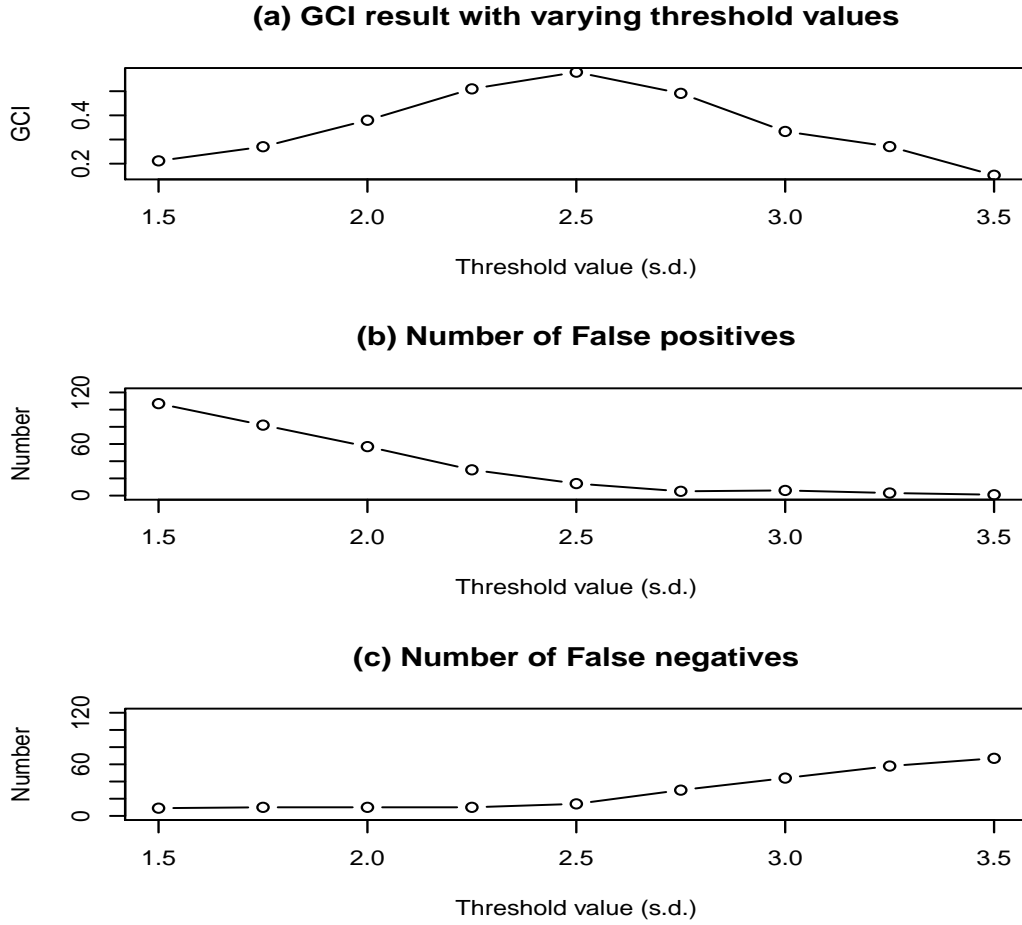
Figure 2.6: (a) The relationship between GCI value and threshold value by changing the threshold from 1.5τ to 3.5τ. (b) Number of FPs increases as threshold value decreases. (c) Number of FNs increases as threshold value increases.

2.4.2 Experimental Results

From these real data we found three GCI patterns which frequently appeared under the above-mentioned sorting procedures, one of them was found under the FN-procedure, and the other two were found under the FP-procedure. These patterns are explained in detail.

FN-decrease: The GCIs of all the neuron pairs (24 of 30 pairs) decreased as the error percentage increased under the FN-procedure (Figure 2.7(a)) except pairs with a zero GCI (6 of 30 pairs). The four corresponding ξ's are shown in Figure 2.7(b), and it shows that the error process induced by the FN-procedure was negatively correlated with processes of both the source and target neurons; thus the GCI decreased (resembling that in Figure 2.3(b)).

FP-decrease: The GCIs of neuron pairs (6 of 30 pairs) decreased as the error percentage increased under the FP-procedure (Figure 2.7(a)) and the four corresponding ξ's of this pattern (Figure 2.7(c)) show that the error process was positively

correlated with process of the source neuron and was uncorrelated with that of the target neuron. The GCI decreased because the effect of ξ_2 dominated the effect of ξ_3 and ξ_4. In other words, the error process was composed of spikes of some connected neurons which were positively correlated with the source neuron.

FP-increase: The GCIs of neuron pairs (16 of 30 pairs) increased as the error percentage increased under the FP-procedure (Figure 2.7(a)) and the four corresponding ξ's of this pattern (Figure 2.7(d)) show that the error process was positively correlated with processes of both the source and target neurons. The GCI increased because the effect of ξ_3 and ξ_4 dominated the effect of ξ_2. In other words, the error process was composed of spikes of some connected neurons which were strongly correlated with the target neuron.

Finally, we note that there are 5 neuron pairs with a zero GCI, and 3 neuron pairs with unchanged GCIs under the FP-procedure.

2.5   Discussion

Because spike sorting errors are almost unavoidable, this study was devoted to investigating how sorting errors affect the identification of information flow among neurons. The analyses of this chapter allowed us to directly discuss the effects of FPs and FNs through the proposed formula, and the results also revealed that they do not have the same effect on spike sorting. In Section 2, we derived an analytic formula (2.7) in terms of factors ξ_1 to ξ_4, and this formula can be used, when incorporating (2.6), to obtain how the GCI changes according to the error signal. Under the FN type of sorting error, we know the error process is negatively correlated with processes of both the source and target neurons; thus the GCI will always be underestimated. On the other hand, under the FP type of sorting error, the GCI may be underestimated or overestimated depending on the error process. If the error process is only composed of electrical noise or spikes of other unconnected neurons, the GCI will be underestimated, but in general, the accuracy is better than cases of the FN type. If the error process is composed of spikes of some positively correlated neurons, then the GCI will be overestimated, and a noncausal neuron pair may be mistaken for a causal pair.

From the perspective of the GC, we provide some suggestions for spike sorting. (i) Missing successive spikes should be avoided as far as possible since this mostly weakens the GCI (Figure 2.2(a), SR model). (ii) A-type errors, which are concentratively added, may weaken the underlying GCI more than R-type errors (Figure 2.2(b), NA model). (iii) During spike detection, choosing a threshold lower than the optimal is better than choosing a threshold higher than the optimal because fake spike events (electrical noise) affect the GCI less than missing spike events (Figure 2.2(a),(c),(d), and Figure 2.6). (iv) During spike classification, cluster shrinkage from the optimal cluster is better than cluster expansion from the optimal because FPs may result in overestimating the GCI and mistaking a noncausal neuron pair for a causal pair (Figure 2.4(a), Experiment A). FNs can only result in an underestimation of the GCI (Figure 2.7(a), FN-decrease), and
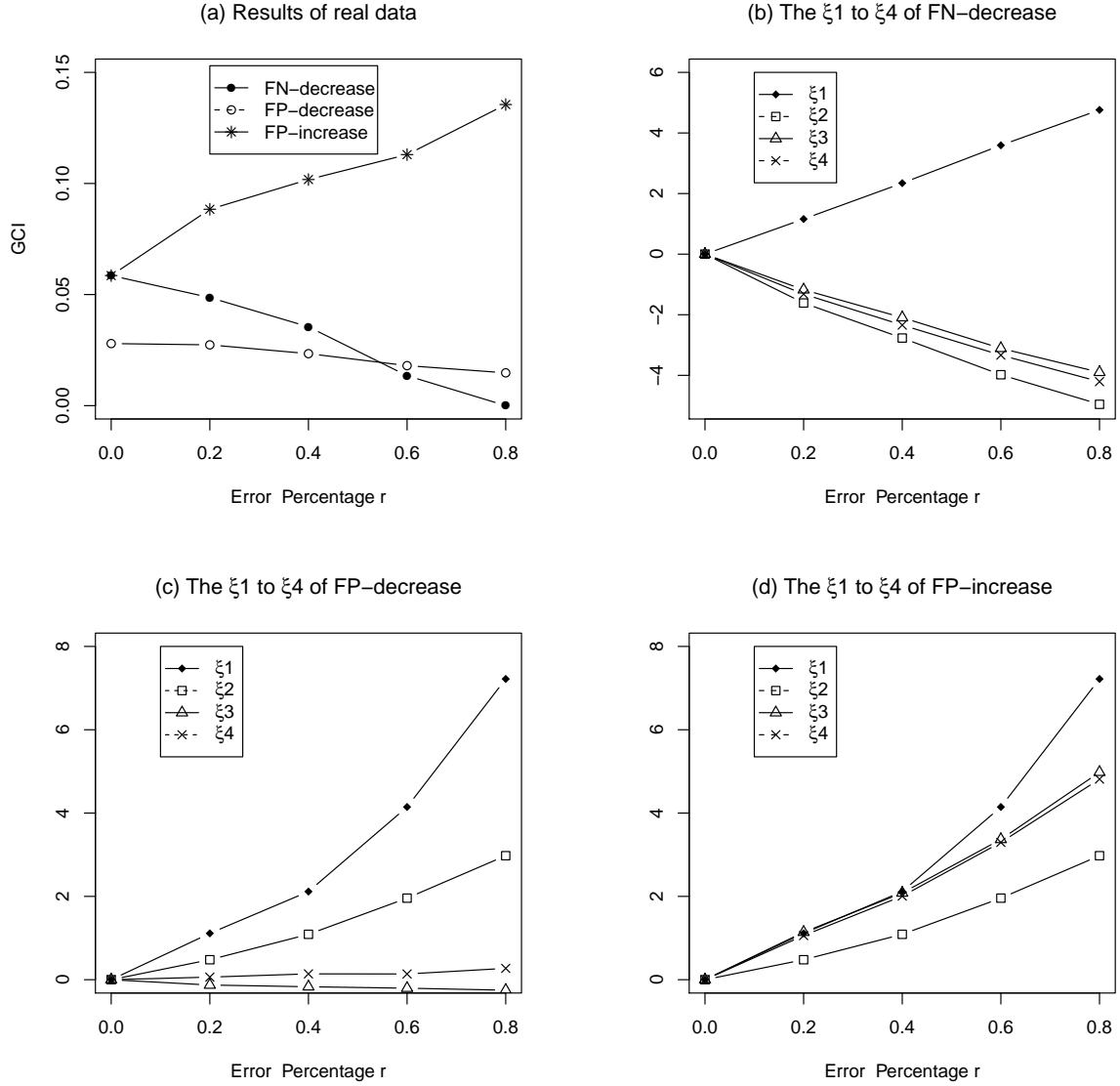
Figure 2.7: (a) Three kinds of Granger Causality Index (GCI) patterns which frequently appeared in real experimental data. (b) ξ_1 to ξ_4 of the false-negative (FN)-decrease pattern. (c) ξ_1 to ξ_4 of the false-positive (FP)-decrease pattern. (d) ξ_1 to ξ_4 of the FP-increase pattern.

this is a relatively conservative and secure strategy for scientific research. For (i) and (ii), in fact, we really cannot avoid missing spikes successively or adding spikes concentratively in the analysis of real data. However, our suggestions still are useful in some cases when recordings are made in the brain regions that neurons are known with complementary intermittence discharge. For example, since the inspiratory-related and expiratory-related neurons coexist in the dorsal and ventral respiratory group [65], and the firing of these two types of neurons are complementary intermittence, so successive missing or concentrative adding of spikes after sorting may occur. We should examine the time series of spike trains

to ensure if the patterns of complementary intermittence are confused after sorting, and to infer which modification would be made when GCIs are calculated. For (iii) and (iv), these two conclusions are just opposite to each other. That is, in spike detection, the FPs are better than the FNs, because the FPs consist of only electrical noises. But in spike classification, the FNs are better than the FPs because the FPs consist of not only electrical noises but also maybe some causal neurons. Finally we note that the way of choosing an optimal threshold or cluster size varies from case to case, since it depends on the sorting method used, and the experimental situation you met. This study is just trying to give a general concept for choosing a better threshold and cluster size.

The results of this study are based on restrictive situations. The analytic formula was obtained from a first order autoregressive model, and the error processes were only superposed on the source process. However, based on these simplifications, the intrinsic properties of the GCI can be seen more clearly than in complete but more complicated situations. Although there are still a lot of concerns on the technical aspect of applying the GCI to determine the relationship among neurons in practice, researchers may be interested in understanding intuitively the effect of spike sorting error before these techniques are really applied, and this is exactly what this chapter wants to provide. Real neuronal networks are much more complex than the simplified assumptions of the analyses and simple models of the simulation. The procedures presented in this study need further development to approach the complex reality. The well-established framework of information theory, for example, might be employed to provide more-credible statistical inferences about true causality in the future.

Chapter 3

Synaptic Weights Estimation

3.1 Introduction

Granger causality (GC) [27, 28] has been shown to be an effective method for analyzing the causal relationship between continuous-valued neural activity data [3,8,13,18] and has been widely deployed in recent neuroscience research. To further understand how neurons cooperate to generate specific brain functions, several extended GC methods were also proposed for identifying directional interactions between neurons through multiple spike trains [38, 41, 51, 57, 74]. Being the fundamental knowledge used in this chapter, the time domain GC analysis will be briefly introduced in the next section and the readers are referred to an article by Barnett and Seth [4] for more details.

The term *synaptic weight* is widely used in neural network research and typically refer to the coupling strength of a connection between two nodes in the network. A large synaptic weight usually means that a large signal (i.e., high-frequency spikes) from the pre-synaptic neuron can result in a large signal of the post-synaptic one. Therefore, in neuroscience and biology, it can also be interpreted as the amount of influence of one neuron has on the firing activity of another.

The spikes of a pre-synaptic neuron are carried by the axon, which will release excitatory or inhibitory neurotransmitter into the synapse. When the post-synaptic neuron receives the neurotransmitter, an excitatory post-synaptic potential (EPSP) or an inhibitory post-synaptic potential (IPSP) is then induced to temporarily depolarize or hyperpolarize the membrane potential. An EPSP makes the neuron more likely to generate an action potential (AP), while an IPSP makes the neuron to do the opposite. However, a single EPSP is not sufficient for the membrane to generate an AP, temporal or spatial summations are required. This means that the firing pattern of the post-synaptic neuron is generally not a direct consequence of the influence of a single pre-synaptic neuron; but a weighted result of the effects of several pre-synaptic neurons with possibly different synaptic weights. Furthermore, IPSPs will diminish EPSPs, playing a much more crucial role of determining whether or not an AP generation will occur at the post-synaptic membrane.

The GC analysis has emerged as a powerful analytical method for estimating the

causal strength of complex networks [58, 59]. However, the effects of excitations and inhibitions could not be differentiated in its original form. Based on the GC, we propose a computational algorithm (presented in Section 3.2.3) under the assumption of best linear predictor (BLP) for analyzing neuronal networks by estimating the synaptic weights among them. The idea of the mathematical assumption BLP is that the weighted voltage-fluctuation of the pre-synaptic neurons should be the best linear explanation for the voltage-fluctuation of the post-synaptic neuron among the network. Using this interpretation, the GC analysis can be extended to measure both excitatory and inhibitory effects between neurons without too much extra computational complexity. The appropriateness of the BLP assumption was examined by three sorts of simulated networks: those with linear, almost linear, and nonlinear network structures. To illustrate the application of the proposed method, real spike trains from the anterior cingulate cortex (ACC) and the striatum (STR) were analyzed.

It is worth noting that spike trains are non-equally spaced data and are regarded as being from a point process. Filtering is usually required for converting them to equally spaced time series for further GC analyses [57]. This study adopted the Gaussian kernel filtering or binning (depending on the situation) to convert spike trains into time series data for the following three main reasons: (1) it reduces the complexity of analysis, and considers also the effect of temporal summation of action potentials, (2) under suitable preprocessing, even short, sparse spike trains can be converted, so that the standard autoregression modeling can be applied [75], (3) most important of all, spike trains can be filtered to form close approximations to the firing rates or the voltage-fluctuations of the underlying neurons [42].

The rest of this chapter is organized as follows. In Section 2, we briefly introduce the so-called Granger causality index, and then extend it to measure both excitatory and inhibitory effects between network nodes by using the BLP assumption. Section 3 presents three network models to ensure the appropriateness of the proposed algorithm. In Section 4, we apply the algorithm to analyze real spike train data. Section 5 provides some discussion about the results obtained from Section 3–4, shortcomings of the method, and related future works.

3.2   Modeling and analysis

Based on the framework of Granger causality analysis and a BLP interpretation of synaptic weights, we propose a procedure for weights estimation and define a synaptic measure between neuronal time series using the estimated weights.

3.2.1 An introduction to the GC

Let $X = (x^1, x^2, \ldots, x^n)$ be a stationary n-dimensional time series process with zero mean. The p-th order linear autoregressive model for X is given by

$$
\begin{cases}
x_t^1 = \sum_{r=1}^{p} a_r^{1,1} x_{t-r}^1 + \cdots + \sum_{r=1}^{p} a_r^{1,n} x_{t-r}^n + \epsilon_t^1 \\
x_t^2 = \sum_{r=1}^{p} a_r^{2,1} x_{t-r}^1 + \cdots + \sum_{r=1}^{p} a_r^{2,n} x_{t-r}^n + \epsilon_t^2 \\
\vdots \\
x_t^n = \sum_{r=1}^{p} a_r^{n,1} x_{t-r}^1 + \cdots + \sum_{r=1}^{p} a_r^{n,n} x_{t-r}^n + \epsilon_t^n,
\end{cases}
\tag{3.1}
$$

where $a_r^{i,j}$ is the projection coefficient from the i-th time series onto the j-th time series at time lag r, representing the coupling strength from node j to node i in the network. The residuals $\epsilon^1, \epsilon^2, \ldots, \epsilon^n$ are zero-mean uncorrelated white noises with covariance matrix Σ. The diagonal entries $\{\Sigma_{ii} = Var(\epsilon^i), i = 1, \ldots, n\}$ measure the accuracy of the autoregressive prediction to each node based on the information from time stamps $t - 1$ to $t - p$.

To see whether the information contained in time series x^j is useful in explaining the state of time series x^i, namely, the importance of node j to node i, we can exclude the time series variable x^j from (4.7) to obtain a reduced [1] $(n-1)$-dimensional autoregressive model with residual series $\eta^{i,j}$ of x^i and corresponding prediction error $\Gamma_{ii}^j = Var(\eta^{i,j})$. Here Γ_{ii}^j measures the accuracy of the prediction of x^i based only on the previous values in time series $\{x^1, \ldots, x^{j-1}, x^{j+1}, \ldots, x^n\}$. If Σ_{ii} in (4.7) is significantly less than Γ_{ii}^j in the reduced model in some suitable statistical sense, then we say that x^j Granger-cause x^i. This causality can be quantified by the GC index from x^j to x^i formulated as:

$$
F_{j \to i} = \ln \frac{\Gamma_{ii}^j}{\Sigma_{ii}}.
\tag{3.2}
$$

It is clear that $F_{j \to i} = 0$ when $\Gamma_{ii}^j = \Sigma_{ii}$, i.e., x^j has no causal influence on x^i, and $F_{j \to i} > 0$ when x^j Granger-cause x^i. Notice that $F_{j \to i}$ is defined only for $j \neq i$ and is always nonnegative, i.e., Σ_{ii} is bounded above by Γ_{ii}^j, since the full model defined in (4.7) should fit the data better than the reduced model. Finally, we note that the GC index (4.8) is significant if the corresponding coefficients $a_r^{i,j}$ are jointly significantly different from zero. This can be assessed via an F-test on the null hypothesis that $a_r^{i,j}$ are zero [29, 60]. The projection coefficients $a_r^{i,j}$ and prediction errors Σ_{ii} can be obtained

---

[1] The $i-$th equation of the reduced model reads $x_t^i = \sum_{r=1}^{p} b_r^{i,1} x_{t-r}^1 + \cdots + \sum_{r=1}^{p} b_r^{i,j-1} x_{t-r}^{j-1} + \sum_{r=1}^{p} b_r^{i,j+1} x_{t-r}^{j+1} +$

$\cdots + \sum_{r=1}^{p} b_r^{i,n} x_{t-r}^n + \eta_{\mathbf{t}}^{\mathbf{i,j}}$, where the b's are the corresponding projection coefficients.

by solving the Yule-Walker equation [39] and an efficient model order can be determined using the Akaike Information Criterion (AIC):

$$AIC(p) = 2\log(\det(\Sigma)) + \frac{2n^2 p}{T}, \tag{3.3}$$

where T is the total length of the time series. More information about the GC can be found in the author's previous work [61] and also [4, 13, 20].

3.2.2 Synaptic weights estimation

Now, consider the multivariate zero-mean time series, $X = (x^1, x^2, \ldots, x^n)$, consisting of the trajectories of membrane voltage of n distinct neurons. Suppose that the i-th neuron is triggered by some other k neurons in the network, say $\{i_1, i_2, \ldots, i_k\}$-th neurons, with synaptic weights $\{\alpha_1^i, \alpha_2^i, \ldots, \alpha_k^i\}$. For convenience, we assume that $1 \le k := k(i) \le n-1$. The case $k = 0$ means that the i-th neuron is not triggered by others, thus is relatively easy to deal with. The weights are assumed to be nonzero, if some α_s^i is zero, then we can just remove the corresponding index i_s from the trigger set. Positive and negative weights represent excitatory and inhibitory influences on the i-th neuron, respectively.

In general, the trigger set $I_i := \{i_1, i_2, \ldots, i_k\}$ and the corresponding weights $\alpha^i := \{\alpha_1^i, \alpha_2^i, \ldots, \alpha_k^i\}$ in the network can not be identified and estimated easily due to the underlying complex dynamics. However, under the assumption of the best linear predictor (BLP) (Definition 1 below), I_i and α^i can be approximated effectively. The results are described in the following proposition.

Definition 1.
Let x and y be two stationary time series with zero-means. Then we say that y forms the best linear predictor (BLP) of x among a variable set $\mathcal{G}$ if $\sigma^2(x|\bar{x}, \bar{y}) < \sigma^2(x|\bar{x}, \bar{z}), \forall z \in \mathcal{G}$, where $\sigma^2(x|\bar{x}, \bar{y}) := min_{p, \{f_r\}, \{d_r\}} E\{x_t - \sum_{r=1}^{p} [f_r y_{t-r} + d_r x_{t-r}]\}^2$.

Proposition 1.
In the situation described above, if further the weighted trajectory $u^i := \alpha_1^i x^{i_1} + \alpha_2^i x^{i_2} + \cdots + \alpha_k^i x^{i_k}$, made by the trigger set and the corresponding weights, forms the BLP to the trajectory of the i-th neuron (namely, x^i) among the whole network; then based on the GC framework, I_i can be completely identified and the estimate of α^i can be obtained as $\hat{\alpha}^i := \{\sum_{r=1}^{p} a_r^{i,i_1}, \sum_{r=1}^{p} a_r^{i,i_2}, \ldots, \sum_{r=1}^{p} a_r^{i,i_k}\}$ up to a scale factor.

Proof:
Let $u^i := \alpha_1^i x^{i_1} + \cdots + \alpha_k^i x^{i_k}$ form the BLP of x^i, then there exist a positive integer p and projection coefficients $\{f_r^i, r = 1, 2, \ldots, p\}$, $\{d_r^i, r = 1, 2, \ldots, p\}$ such that $x_t^i = \sum_{r=1}^{p} [f_r^i u_{t-r}^i + d_r^i x_{t-r}^i] + \epsilon_t^i$, where ϵ^i is a stationary white noise possessing the smallest

27

variance among the whole network. Replacing u^i with the weighted trajectory, we obtain

$$
\begin{aligned}
x_t^i &= \sum_{r=1}^{p} [f_r^i u_{t-r}^i + d_r^i x_{t-r}^i] + \epsilon_t^i \\
&= \sum_{r=1}^{p} [f_r^i(\alpha_1^i x_{t-r}^{i_1} + \cdots + \alpha_k^i x_{t-r}^{i_k}) + d_r^i x_{t-r}^i] + \epsilon_t^i \\
&= \sum_{r=1}^{p} [\alpha_1^i f_r^i x_{t-r}^{i_1} + \cdots + \alpha_k^i f_r^i x_{t-r}^{i_k} + d_r^i x_{t-r}^i] + \epsilon_t^i, \quad (3.4)
\end{aligned}
$$

which represents the underlying but unknown network structure of $\{x^i, x^{i_1}, x^{i_2}, \ldots, x^{i_k}\}$. On the other hand, fitting to data the same equation as the i-th equation in (4.7), we have the following empirical regression (compared to the theoretical regression (3.4))

$$
x_t^i = \sum_{r=1}^{p} [a_r^{i,1} x_{t-r}^1 + \cdots + a_r^{i,n} x_{t-r}^n] + \tilde{\epsilon}_t^i. \quad (3.5)
$$

Let $\bar{I}_i := \{1, 2, \ldots, i-1, i+1, \cdots, n\} - I_i$ be the complement of the trigger set I_i. We note that $\tilde{\epsilon}_t^i \equiv \epsilon_t^i$ if $a_r^{i,s} = 0, \forall r = 1, \ldots, p$ and $s \in \bar{I}_i$; otherwise $\tilde{\epsilon}^i$ and ϵ^i are totally different but with $Var(\tilde{\epsilon}_t^i) = Var(\epsilon_t^i)$ since (3.4) has the smallest residual variance among the whole network and (3.5) has more degree of freedom (coefficients) than (3.4).

If the trajectories of the $\bar{I}_i$-th neurons are stochastically independent of both the i-th and I_i-th neurons, then we have $a_r^{i,s} = 0$, for $r = 1, \ldots, p$ and $s \in \bar{I}_i$. Comparing (3.5) with (3.4), we then have

$$
\sum_{r=1}^{p} a_r^{i,i_1} = \alpha_1^i \sum_{r=1}^{p} f_r^i \quad, \ldots, \quad \sum_{r=1}^{p} a_r^{i,i_k} = \alpha_k^i \sum_{r=1}^{p} f_r^i. \quad (3.6)
$$

Since the projection coefficients in (3.5) can be obtained by solving the Yule-Walker equation or simply by the least-squares method, (3.6) immediately leads to

$$
\frac{\alpha_1^i}{\sum\limits_{r=1}^{p} a_r^{i,i_1}} = \frac{\alpha_2^i}{\sum\limits_{r=1}^{p} a_r^{i,i_2}} = \cdots = \frac{\alpha_k^i}{\sum\limits_{r=1}^{p} a_r^{i,i_k}}, \quad (3.7)
$$

provided $\sum\limits_{r=1}^{p} f_r^i \neq 0$. For α_s^i and $\sum\limits_{r=1}^{p} a_r^{i,i_s}$ to have the same sign for $s = 1, 2, \ldots, k$, we can, without loss of generality, assume that $\sum\limits_{r=1}^{p} f_r^i > 0$. If $\sum\limits_{r=1}^{p} f_r^i < 0$, then $-u^i$ is used to replace u^i.

If some of the trajectories of the $\bar{I}_i$-th neurons are linearly dependent of the i-th or I_i-th neurons, then $a_r^{i,s} \neq 0$, for some $r \in \{1, \cdots, p\}$, $s \in \bar{I}_i$ and the projection

28

coefficients in (3.5) are thus affected, resulting in a biased estimation of (3.7). However, this predicament can be solved by virtue of the assumption of BLP and the concept of GC. Since the ϵ^i in (3.4) has the smallest variance among the whole network, taking out any element of $\{x^s : s \in \bar{I}_i\}$ from the regression (3.5) does not increase the variance of $\tilde{\epsilon}^i$. According to the concept in (4.8), we can correct the model coefficients by ruling out all the useless information of the $\bar{I}_i$-th neurons. $\square$

We end this subsection by the following remarks.

Remark 1.
The idea behind the BLP mathematical assumption is that the weighted voltage-trajectory of the trigger neurons should be the best linear explanation for the voltage-trajectory of the target neuron among the whole network. Based on this interpretation, the GC index can be extended to measure both excitatory and inhibitory effects in virtue of the estimated synaptic weights.

Remark 2.
The synaptic weight α_s^i and the summation of the projection coefficients $\sum_{r=1}^{p} a_r^{i,i_s}$ are forced to have the same sign for $s = 1, 2, \ldots, k$, because positive and negative $\sum_{r=1}^{p} a_r^{i,i_s}$ refer to positive and negative correlations between x^i and x^{i_s}, respectively.

Remark 3.
The case $k = 0$ means that the i-th neuron is not triggered by other neurons in the network, therefore $F_{j \to i} = 0, \forall j$ and there is no synaptic weights to be estimated.

Remark 4.
For readers dealing with sparse networks, L1 regularization (or LASSO) would be an useful technique for fitting to data a sparse regression to avoid overfitting and the problem of multiple testing [1,49]. In this scenario, a computational much more efficient approach would be first running LASSO to learn the network structure and then using GC to get the causal strength.

Remark 5.
Some arguments in the proof such as the stochastic independence and the estimations of projection coefficients assume that the law of large numbers (LLN) holds. In the case of small samples or limited data, estimation errors come into exsitence thus some statistical inferences in the proof may fail. However, the proof holds for most large-sample cases.

3.2.3 The algorithm

Here, we present a step by step algorithm for computing the proposed index (named neuron synaptic index, NSI) from multiple spike train data.

Step 1 : Properly smooth the spike train data by kernel filtering (Gaussian kernels are commonly used) to acquire the approximate membrane voltage trajectories to the underlying voltage evolution of the neurons in the network.

Step 2 : Subtract the mean value from each trajectory to form zero-mean time series and then fit the vector autoregressive model in (4.7). Appropriate model order can be obtained beforehand by using AIC in (4.9).

Step 3 : Compute all the GC indices by (4.8) for all pairs of neurons, i.e., $i, j = 1, 2, \ldots, n$ with $i \neq j$ and also perform F-tests to ensure statistical significance.

Step 4 : For each node $i = 1, 2, \ldots, n$, refine the autoregressive model by ruling out the information about the $\bar{I}_i$-th neurons, i.e., the neurons with insignificant $F_{j \to i}$, to correct the projection coefficients.

Step 5 : For each node, compute the synaptic weights of the trigger neurons by simply summing the projection coefficients up to the model order p as shown in (3.7).

Step 6 : For each node, take the weighted trajectory as a new explanatory time series and then compute the GC index from this weighted time series to that of the node.

Step 7 : Finally, the NSI is then defined to be the l_1-normalized [2] estimated weights obtained in Step 5 multiplying the GC index obtained in Step 6 (see (B.5) in Appendix).

3.3   Simulation study

Here, three sorts of network models are simulated to investigate the proposed algorithm: the linear, almost-linear, and nonlinear networks. The linear network, derived directly from autoregressive framework, examines the validity of the synaptic weights estimation in Proposition 1. Additionally, the almost-linear and nonlinear networks, derived both from integrate-and-fire (IF) neuron models, examine the appropriateness of the BLP assumption in Remark 1 through the reflected subthreshold dynamics of the models. The Matlab code used for this study is available to interested readers upon request.

Significance tests on the GC indices should be corrected for multiple testing and we adopted the approach of False Discovery Rate (FDR) [6] which has greater statistical power than the conservative Bonferroni correction. In this and next sections, significant Granger causality interaction between input neurons and output neurons are calculated using an F-test corrected by FDR for multiple comparison with confidence threshold at $P - value = 0.05$ (i.e., 95% significance level).

---

[2]This means that the estimated weight vector $\hat{\alpha}^i = (\hat{\alpha}_{i_1}, \hat{\alpha}_{i_2}, \ldots, \hat{\alpha}_{i_k})$ is normalized by its l_1 norm $\|\hat{\alpha}^i\|_1 := |\hat{\alpha}_{i_1}| + |\hat{\alpha}_{i_2}| + \cdots + |\hat{\alpha}_{i_k}|$. Then the normalized weight vector $\hat{\alpha}^i / \|\hat{\alpha}^i\|_1$ will have unit l_1 vector norm.

3.3.1 Linear network

A linear network, depicted in Figure 3.1, is presented here. The time series variable w serves as the trajectory of post-synaptic neuron, while x, y, and z serve as the trajectories of pre-synaptic neurons with synaptic weights α, β, and γ, respectively. For readers not familiar with the multivariate settings in Section 3.2, a much clear-cut derivation of the NSI using this simple network is given in Appendix.

Trajectories are generated by the following equations:

$$
\begin{cases}
v_{1,t} = \epsilon_{1,t} \\
x_t = g_1 v_{1,t-1} + g_2 v_{1,t-2} + g_3 v_{1,t-3} + \epsilon_{2,t} \\
v_{2,t} = h_1 x_{t-1} + h_2 x_{t-2} + h_3 x_{t-3} + \epsilon_{3,t} \\
y_t = \epsilon_{4,t} \\
z_t = \epsilon_{5,t} \\
u_t = \alpha x_t + \beta y_t + \gamma z_t \\
w_t = f_1 u_{1,t-1} + f_2 u_{1,t-2} + f_3 u_{1,t-3} + d_1 w_{1,t-1} + d_2 w_{1,t-2} + d_3 w_{1,t-3} + \epsilon_{6,t} \\
v_{3,t} = \epsilon_{7,t}
\end{cases}
\tag{3.8}
$$

where $\epsilon_k \sim^{iid} \mathcal{N}(0,1), k = 1, \ldots, 7$ are zero-mean uncorrelated Guassian white noises.

Settings of the simulation are: $[g_1, g_2, g_3] = [0.4, 0.2, 0.1]$, $[h_1, h_2, h_3] = [0.1, 0.2, 0.4]$, $[f_1, f_2, f_3] = [0.5, 0.3, 0.1]$, and $[d_1, d_2, d_3] = [0.1, 0.3, 0.5]$. Variable x is triggered by v_1 and variable v_2 is triggered by x. Variable v_3 serves as an independent node in the network. The length of each trajectory is 1000. We note that the above coefficients can be replaced by any stable coefficients and the simulation results shown below will remain unaffected. The stability condition [47] ensures (3.8) to generate stationary processes. There have been many convincing examples published [3, 76] showing that the GC framework is a well-established method for identifying the causal relationship among stationary time series. If some unstable coefficients are used instead, the generated time series will be nonstationary, then some signal preprocessing technique (e.g, differencing) will be needed to convert nonstationary processes to stationary processes. More details can be found in [4].

For $[\alpha, \beta, \gamma] = [1.0, 0.5, -0.5]$, averaging from 100 repeated simulations gives: the synaptic weights estimates $[\hat{\alpha}, \hat{\beta}, \hat{\gamma}] = [0.9012, 0.4549, -0.4539]$, the weighted GC index $F_{\hat{\alpha}x + \hat{\beta}y + \hat{\gamma}z \to w} = 0.4515$, the normalized weights $[1.0000, 0.5064, -0.5053]$ (divided by $|\hat{\alpha}|$ instead of $|\hat{\alpha}| + |\hat{\beta}| + |\hat{\gamma}|$ defined in (B.5) for easy comparison with the underlying weights), and the NSIs $[N_{x \to w}, N_{y \to w}, N_{z \to w}] = [0.4515, 0.2286, -0.2281]$ with standard deviations $[0.0359, 0.0396, 0.0347]$. Notice that no knowledge of the parameters for generating the data was used in the estimation procedure, only the generated trajectories were used.

The results show that the normalized weight estimates $[1.0000, 0.5064, -0.5053]$ are consistent with the underlying weights $[1.0, 0.5, -0.5]$, and the collateral variables v_1, v_2, and v_3 do not affect the estimation since $F_{v_1 \to w}$, $F_{v_2 \to w}$, and $F_{v_3 \to w}$ are all zero, thus these information can be ruled out directly by the GC analysis. The NSI (B.5) unites the estimated synaptic weights and the weighted GC index. A positive (negative) value

represents excitatory (inhibitory) synaptic connection and it's magnitude at the same time reflects the degree of synaptic influence.
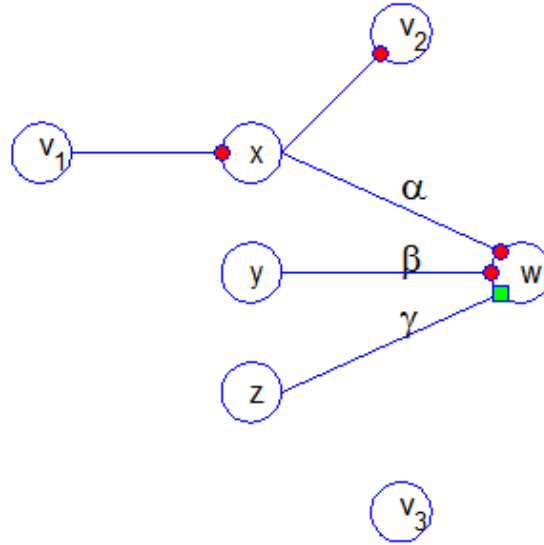
Figure 3.1: A simple linear network. Red circles represent excitation and green squares represent inhibition. Variable w serves as the trajectory of post-synaptic neuron, variables x, y, and z serve as the trajectories of pre-synaptic neurons with synaptic weights $\alpha > 0$, $\beta > 0$, and $\gamma < 0$, respectively. v_1, v_2, and v_3 are collateral variables, consisting of source, target, and independent nodes. A much clear-cut derivation of the NSI using this simple network is given in Appendix.

3.3.2 Almost-linear network

To illustrate the synaptic weights estimation in a neural spiking context, a simple feedforward IF neuron network was simulated (depicted in Figure 3.2). Briefly, Neurons $\#2-5$ were modeled by independent Poisson processes with firing rate λ. Neurons $\#8, 9$ were modeled as single strong inputs by independent Poisson processes with firing rate 1.5λ. Neuron $\#7$ was implemented by a direct discrete time summation of the synaptic inputs α_i (mV), $i = 1, \ldots, 6$ (i.e., the weighted outputs of Neurons $\#1-6$ after some propagation delay), leading to its internal potential that was reset to $V_{reset} = -80$ (mV) and produced a spike when the threshold value $V_{th} = -55$ (mV) was reached. During the refractory period, the potential will linearly recover from V_{reset} to the resting potential $V_{rest} = -70$ (mV). Time resolution was set to be 1 ms and there was a 2 mV decrease/increase of the potential to the V_{rest} every unit time depending on the status of de/hyper -polarizations, respectively to model the diffusion of ions. The internal potential was forced to lie in the range $[E_{K^+}, E_{Na^+}] = [-90, 60]$, the equilibrium potential of K^+ and Na^+, respectively and action potentials were normalized to 30 mV for display. Neurons $\#1, 6$ were implemented in the same way that Neuron $\#7$ was done with synaptic inputs 30 (mV) from Neurons $\#8, 9$ respectively. Neurons $\#10, 11$ were modeled as independent nodes by independent Poisson processes with firing rate λ.
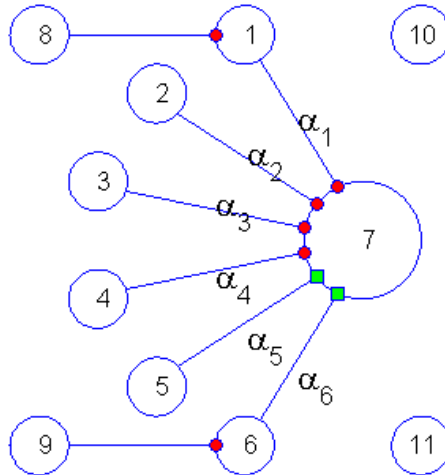
Figure 3.2: A simple feedforward integrate-and-fire neuron network. Red circles represent excitation and green squares represent inhibition. Neurons $\#2 - 5$ are modeled by independent Poisson processes with firing rate λ. Neurons $\#8, 9$ were modeled as single strong inputs by independent Poisson processes with firing rate 1.5λ. Neuron $\#7$ is implemented by a direct discrete time summation of the synaptic inputs $\alpha_1, \ldots, \alpha_6 (mV)$. Neurons $\#1, 6$ are implemented in the same way that Neuron $\#7$ is done with synaptic inputs 30 (mV) from Neurons $\#8, 9$ respectively. Neurons $\#10, 11$ are modeled as independent nodes by independent Poisson processes with firing rate λ. Neurons $\#5, 6$ are inhibitory, i.e., $\alpha_5, \alpha_6 < 0$.

We begin with Simulation 1 in which the synaptic weights were fixed at $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 5$ (mV) and $\alpha_5 = \alpha_6 = -2.5$ (mV), and the propagation delay of each source neuron was set to be 10 ms. 60 sec. voltage-trajectories of Neurons $\#1, 6, 7$ were then simulated according to the way described above. The first 1 sec. of the trajectory of Neuron $\#7$ with input rate $\lambda = 40$ Hz is shown in Figure 3.3 and the corresponding simulated spike train data is shown in Figure 3.4. The subthreshold trajectory of Neuron $\#7$ is not very regular due to the lack of self dynamics, compared to the nonlinear network (3.9a) introduced next. However, in this case it faithfully reflects the effects of the input neurons, that is, the actual degree of effects of the input neurons are to be proportional to the corresponding synaptic weights. To analyze the network directly through the simulated spike train data, a Gaussian kernel filtering with bandwidth 5 ms. was performed to obtain an approximation of the subthreshold dynamics of each neuron in the network, the result is depicted in Figure 3.5. Based on the filtered data, both GC and NS indices were computed for different input rates $\lambda =$40, 60, and 80 Hz. In each case, the indices were both obtained from the average of 100 simulations and the results are summarized in Table 3.1.

We can find from Table 3.1 that although the GC indices correctly identify the direction of information flow among the network, the effects of excitations and inhibitions could not be differentiated directly by the sign of the indices since they are by definition

the voltage-trajectory of Neuron #7
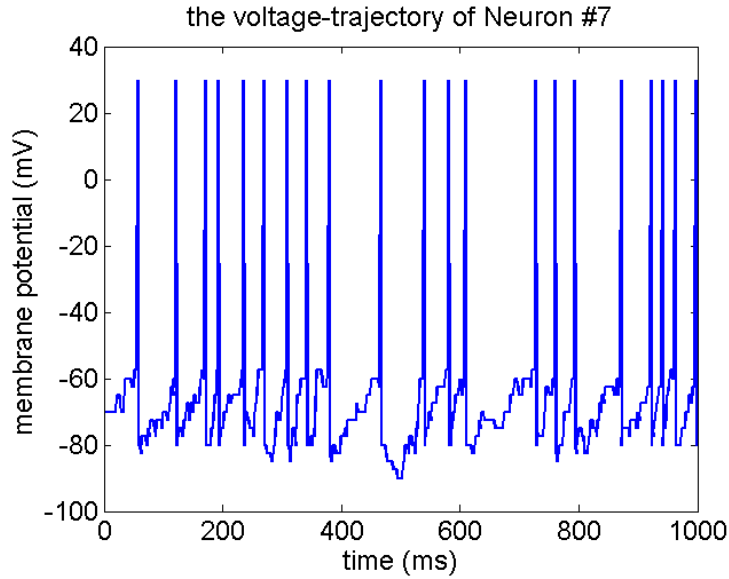
Figure 3.3: The first 1 sec. of a simulated voltage-trajectory of Neuron #7 with input rate $\lambda = 40$ Hz. The simulation was done according to the way described in the context with $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 5$ (mV), $\alpha_5 = \alpha_6 = -2.5$ (mV), and 10 ms. propagation delay. The subthreshold trajectory is not very regular due to the lack of self dynamics, in other words, Neuron #7 is completely triggered by Neurons #1 − 6.

simulated spike train data
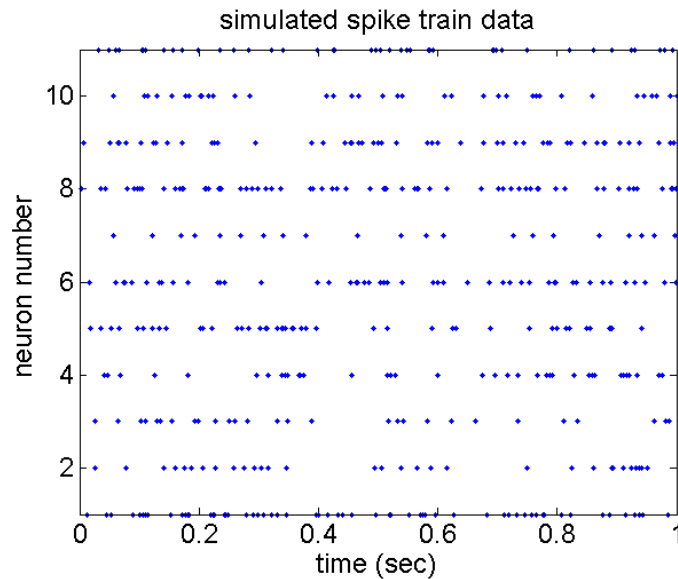
Figure 3.4: The first 1 sec. of a simulated spike train data of the simple feedforward network with $\lambda = 40$ Hz, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 5$ (mV), $\alpha_5 = \alpha_6 = -2.5$ (mV), and 10 ms. propagation delay.

to be nonnegative. From the GC indices, we can only tell that Neurons #2 − 4 have more influences than Neurons #1, 5 have on Neuron #7. Information about the underlying synaptic weights was not provided. As can be found in the lower part of Table 3.1, the synaptic weights were successfully reconstructed from the spike train data by the NS
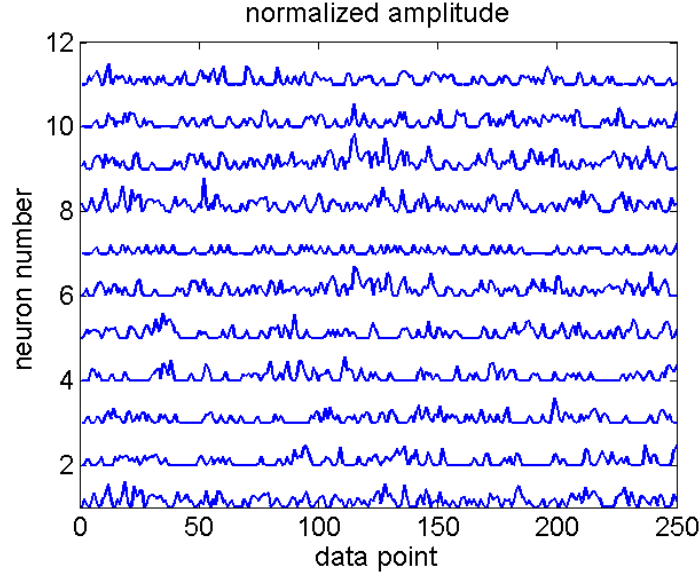
Figure 3.5: A Gaussian kernel filtering with bandwidth 5 ms. was performed on the spike train data depicted in Figure 3.4 to obtain an approximation of the subthreshold dynamics of each neuron in the network. The computations of the GCI and NSI were based on the filtered results and this figure shows the first 250 data points.

Table 3.1: The numerical results of Simulation 1 in Section 3.3.2. The effects of excitations and inhibitions can be differentiated directly by the sign of the NS indices and the ratio of effects between them was close to $5.0 : -2.5 = 1 : -0.5$. Numbers in parentheses are corresponding standard errors.

Input rate	$1 \to 7$	$2 \to 7$	$3 \to 7$	$4 \to 7$	$5 \to 7$	$6 \to 7$	$8 \to 1$	$9 \to 6$
Granger Causality Index (GCI)								
$\lambda = 40$	0.0223	0.1972	0.1906	0.1919	0.0497	0.0066	2.9976	3.0085
	(0.0035)	(0.0125)	(0.0155)	(0.0129)	(0.0059)	(0.0031)	(0.0868)	(0.1016)
$\lambda = 60$	0.0411	0.2770	0.2768	0.2769	0.0791	0.0115	2.5959	2.6063
	(0.0038)	(0.0138)	(0.0141)	(0.0156)	(0.0076)	(0.0020)	(0.0573)	(0.0829)
$\lambda = 80$	0.0591	0.3466	0.3425	0.3444	0.1020	0.0156	2.3026	2.3235
	(0.0080)	(0.0188)	(0.0155)	(0.0168)	(0.0083)	(0.0037)	(0.0531)	(0.0575)
Neuron Synaptic Index (NSI)								
$\lambda = 40$	0.1155	0.1145	0.1140	0.1137	-0.0571	-0.0555	2.9972	3.0083
	(0.0043)	(0.0040)	(0.0052)	(0.0046)	(0.0034)	(0.0069)	(0.0868)	(0.1018)
$\lambda = 60$	0.1530	0.1502	0.1505	0.1503	-0.0756	-0.0754	2.5959	2.6054
	(0.0052)	(0.0048)	(0.0045)	(0.0054)	(0.0038)	(0.0027)	(0.0570)	(0.0819)
$\lambda = 80$	0.1799	0.1795	0.1793	0.1783	-0.0893	-0.0898	2.3030	2.3237
	(0.0044)	(0.0049)	(0.0059)	(0.0053)	(0.0036)	(0.0046)	(0.0532)	(0.0571)

indices in the sense that the ratio between excitatory and inhibitory sources was close to $5.0 : -2.5 = 1 : -0.5$ for all different input rates. We note that the GC and NS indices from Neurons #8 − 12 to Neuron #7 are all zero (i.e., insignificant). As the results show, a large NSI does not necessarily imply a large GCI. That is, a strong synaptic transmission can not always guarantee a strong causal relationship; it depends also on the firing pattern/timing of the source and the coordination with other neurons. So, from

this perspective, NSI can be treated as a better proxy for synaptic weights rather than a new causality measure. GCI provides information on causal structure while NSI provides complementary information on synaptic transmission.

In Simulation 2, the input rate λ was fixed at 60 Hz while the synaptic weights varied. Let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 5$ (mV) and $\alpha_5 = \alpha_6 = -k \times 5$ (mV). Three different weight-ratio $k = 0.5, 1.0, 1.5$ were considered, and the computed NS indices are presented in Table 3.2. We can find that the ratio between excitatory and inhibitory sources was still close to $1 : -k$ as weight-ratio changes.

Table 3.2: The numerical results of Simulation 2 in Section 3.3.2. The input rate λ was fixed at 60 Hz, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 5$ (mV) , and $\alpha_5 = \alpha_6 = -k \times 5$ (mV). The ratio of effects between excitatory and inhibitory sources was still close to $1 : -k$ as weight ratio changes. Numbers in parentheses are corresponding standard errors.

weight ratio	$1 \to 7$	$2 \to 7$	$3 \to 7$	$4 \to 7$	$5 \to 7$	$6 \to 7$	$8 \to 1$	$9 \to 6$
	Neuron Synaptic Index (NSI)							
$k = 0.5$	0.1530	0.1502	0.1505	0.1503	-0.0756	-0.0754	2.5959	2.6054
	(0.0052)	(0.0048)	(0.0045)	(0.0054)	(0.0038)	(0.0027)	(0.0570)	(0.0819)
$k = 1.0$	0.1077	0.1072	0.1063	0.1067	-0.1025	-0.1031	2.5928	2.6151
	(0.0041)	(0.0031)	(0.0044)	(0.0043)	(0.0034)	(0.0037)	(0.0857)	(0.0671)
$k = 1.5$	0.0615	0.0621	0.0612	0.0625	-0.0857	-0.0866	2.5849	2.5869
	(0.0030)	(0.0031)	(0.0036)	(0.0037)	(0.0029)	(0.0039)	(0.0764)	(0.0735)

3.3.3 Nonlinear network

Following the same network topology (Figure 3.2) presented in the previous subsection, here the dynamics of Neuron #7 was modeled instead by the Izhikevich's simple spiking neuron model [35, 50, 51] as it can provide more neural responses compared to classical IF neurons. Briefly, Izhikevich neurons are modeled by the following 2-D differential equation with an after-spiking resetting:

$$v' = 0.04v^2 + 5v + 140 - u + I \tag{3.9a}$$

$$u' = a(bv - u) \tag{3.9b}$$

$$\text{if } (v \geq 30 \text{ mV}) \text{ then } v \leftarrow c \text{ and } u \leftarrow u + d. \tag{3.9c}$$

The variable v represents the membrane potential of the neuron, u represents a membrane recovery variable, and I represents the total input synaptic current. The parameter a describes the time scale of u, characterizing the recovery rate. The parameter b describes the sensitivity of u to the subthreshold fluctuations of v. Parameters c and d are spike reset values of v and u, respectively. Two sets of parameter values were considered in this study: the fast spiking (FS) neurons ($a = 0.1, b = 0.25, c = -65, d = 2$) and the low-threshold spiking (LTS) neurons ($a = 0.02, b = 0.25, c = -65, d = 2$).

Here, the parameter settings are: $\lambda = 60$ Hz, and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 5$ (mV), $\alpha_5 = \alpha_6 = -k \times 5$ (mV), all of which with 10 ms. propagation delay. The numerical results

for the FS and LTS neuron models with weight-ratio $k = 1.0, 1, 5, 2.0$ are summarized in Table 3.3. The NS indices were obtained from the average of 100 simulations, and the first 0.5 sec. of the voltage fluctuations of the Neuron #7 under FS and LTS models are depicted in Figure 3.6 and Figure 3.7, respectively.

Table 3.3: The numerical results of the simulations in Section 3.3.3. The parameter settings are: $\lambda = 60$ Hz, and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 5$ (mV), $\alpha_5 = \alpha_6 = -k \times 5$ (mV), all of which with 10 ms. propagation delay. Numbers in parentheses are corresponding standard errors.

weight ratio	$1 \to 7$	$2 \to 7$	$3 \to 7$	$4 \to 7$	$5 \to 7$	$6 \to 7$	$8 \to 1$	$9 \to 6$
	NSI of the FS neuron model							
$k = 1.0$	0.0634	0.0688	0.0691	0.0695	-0.0444	-0.0425	2.3089	2.3414
	(0.0092)	(0.0058)	(0.0070)	(0.0065)	(0.0049)	(0.0094)	(0.0514)	(0.0531)
$k = 1.5$	0.0514	0.0565	0.0572	0.0548	-0.0522	-0.0516	2.3269	2.3450
	(0.0095)	(0.0061)	(0.0062)	(0.0059)	(0.0074)	(0.0096)	(0.0551)	(0.0513)
$k = 2.0$	0.0428	0.0467	0.0460	0.0483	-0.0502	-0.0572	2.3309	2.3379
	(0.0097)	(0.0081)	(0.0074)	(0.0082)	(0.0066)	(0.0099)	(0.0423)	(0.0513)
	NSI of the LTS neuron model							
$k = 1.0$	0.0512	0.0536	0.0539	0.0538	-0.0256	-0.0238	2.0241	2.0213
	(0.0095)	(0.0092)	(0.0085)	(0.0071)	(0.0051)	(0.0084)	(0.0351)	(0.0399)
$k = 1.5$	0.0404	0.0438	0.0443	0.0436	-0.0373	-0.0387	1.9861	2.0182
	(0.0093)	(0.0071)	(0.0067)	(0.0059)	(0.0051)	(0.0075)	(0.0434)	(0.0569)
$k = 2.0$	0.0352	0.0394	0.0398	0.0390	-0.0401	-0.0443	2.0244	2.0124
	(0.0093)	(0.0061)	(0.0068)	(0.0042)	(0.0068)	(0.0094)	(0.0554)	(0.0530)

Figure 3.6: The first 0.5 sec. of a simulated voltage-trajectory of Neuron #7 with $k = 1$ under the fast spiking (FS) neuron model ($a = 0.1, b = 0.25, c = -65, d = 2$).

The voltage-trajectories are much regular than that of the almost-linear network (Figure 3.3) due to the self-dynamics of the term $0.04v^2 + 5v + 140$ in (3.9a). It can be
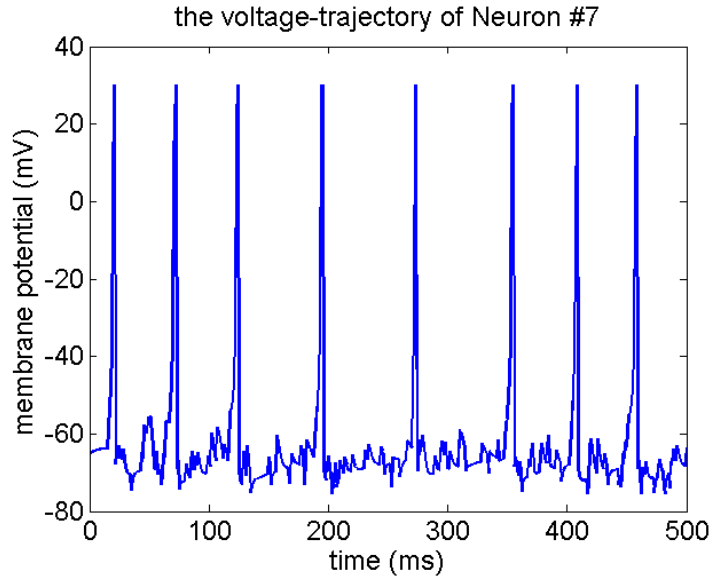
Figure 3.7: The first 0.5 sec. of a simulated voltage-trajectory of Neuron #7 with $k = 1$ under the low-threshold spiking (LTS) neuron model ($a = 0.02, b = 0.25, c = -65, d = 2$).

considered that the nature (or the type) of Neuron #7, to some degree, affects its own behavior; therefore, the effects of the inputs (Neurons $\#1 - 6$) will not be equivalent to the underlying mechanism, meaning that the actual degree of effects of the input neurons will not to be proportional to the corresponding synaptic weights.

Now, it can be found from Table 3.3 that under both the FS and LTS neuron models the negative NSIs from Neurons $\#5, 6$ to Neuron #7 grow in magnitude, relative to the positive ones from Neurons $\#1-4$ to Neuron #7, as weight-ratio k increase. Although the NS indices are not proportional to the underlying weights as what mentioned above, the trend is correctly captured for the increase of the negative NSIs with increasing inhibition strength. Finally, we note that the reason for the absence of the case $k = 0.5$ was that the inhibitory input was so weak to the dynamical system that the resulting NSIs were not significantly different from zero; that is, Neurons $\#5 - 6$ actually did not have any influence on Neuron #7 even though the underlying synaptic weights were not zero.

3.4   Real data analysis

In this section, we illustrate the application of the proposed method to real spike train data and a simulation is also given to examine the validity of the results. Note that significant Granger causality interaction are again calculated using an F-test corrected by FDR for multiple comparison with confidence threshold at $P - value = 0.05$.

3.4.1   Setup and results

Multichannel electrophysiological recording was used for tracking neuronal activity in the anterior cingulate cortex (ACC) and the striatum (STR) and are the same data set

used in our previous study [33]. Briefly, neuronal spikes were recorded from the ACC and the STR of urethane-anesthetized rats after administration of saline or 0.05 or 0.5 mg/kg quinpirole. A multichannel neuronal acquisition processor system (Plexon, Dallas, TX, USA) was used for unit recording, with a filter range of 400 Hz to 8.8 kHz and a sampling rate of 40 kHz. Spikes were further sorted using Offline Sorter (Plexon), based on principle-component clustering with a user-defined template. All animal procedures were approved by the Institutional Animal Care and Use Committee of National Ilan University and adhered to the guidelines established by the Codes for Experimental Use of Animals from the Council of Agriculture, Taiwan.

For this study, data from two independent rats were considered and the numbers of neurons recorded were: 8 (ACC, Rat#1), 9 (STR, Rat#1), 16 (ACC, Rat#2), and 15 (STR, Rat#2). The neurons were randomly put into 2 groups (ACC, Rat#1), 3 groups (STR, Rat#1), 4 groups (ACC, Rat#2), and 3 groups (STR, Rat#2), and each group had 4 (8/2) neurons (ACC, Rat#1), 3 (9/3) neurons (STR, Rat#1), 4 (16/4) neurons (ACC, Rat#2), and 5 (15/3) neurons (STR, Rat#2). After the random grouping described above, the single unit spike trains in each group were pooled as a whole for investigating the brain network. Hence there were 5 (2 + 3) pools in Rat#1 and 7 (4 + 3) pools in Rat#2.

After binning with bin width 2 sec., the GCIs between these random pools can then be computed (Section 3.2.3). Twenty minutes after quinpirole injection, 400 sec. data from both Rat#1 and Rat#2 were used to compute the GCIs. Significant GCIs were found only when certain random group appear, meaning that certain neurons should be pooled together to perform causality. These specific combinations are summarized in Table 3.4, and the corresponding GCIs and NSIs are summarized in Table 3.5. The results show that, under the quinpirole administration, there were excitatory effects inside the ACC (Figure 3.8(a)), excitatory effects from the ACC to the STR (Figure 3.8(b) and 3.8(c)), and inhibitory effects inside the STR (Figure 3.8(d)).

Notice that, for single-input case, e.g., in rat 1 from Pool #2 (ACC) to Pool #3 (STR), both GCI (0.1232) and NSI (0.1116) reflect the degree of causal effect. However, the NSI will be more appropriate than the GCI since the NSI is obtained by fitting a more refined autoregressive model (Step 4 in Section 3.2.3). For multiple-input case, e.g., in rat 2 from Pool #2 (ACC) to Pool #5 (STR) and from Pool #7 (STR) to Pool #5 (STR), the GCI (0.0879 and 0.1598) still reflects the degree of causal effects while the NSI (0.1445 and -0.1058) reflects the degree and the type of synaptic transmission. From this perspective, NSI can be treated as a new complement to provide information on synaptic weights that original GCI does not provide.

Finally, we have to note that, under the saline administration, the same combinations performed no significant GCIs (i.e., GCIs = 0). Furthermore, the GCIs between the ACC of Rat#1 and the STR of Rat#2, and the GCIs between the STR of Rat#1 and the ACC of Rat#2 were all computed, and they were all zero.

Table 3.4: Groups found performing significant NSIs in Section 3.4.1. The numbers of neurons recorded were: 8 (ACC, Rat#1), 9 (STR, Rat#1), 16 (ACC, Rat#2), and 15 (STR, Rat#2). Each group had $8/2 = 4$, $9/3 = 3$, $16/4 = 4$, and $15/3 = 5$ neurons. Hence there were $2 + 3 = 5$ pools in Rat#1 and $4 + 3 = 7$ pools in Rat#2.

Rat no.	Pool no.	location	elements
1	1	ACC	2,4,5,7
1	2	ACC	1,3,6,8
1	3	STR	3,5,7
1	4	STR	1,4,6
1	5	STR	2,8,9
2	1	ACC	4,8,10,16
2	2	ACC	1,5,6,11
2	3	ACC	3,7,9,12
2	4	ACC	2,13,14,15
2	5	STR	6,7,8,9,15
2	6	STR	1,3,10,11,14
2	7	STR	2,4,5,12,13

Table 3.5: The NSIs between the pooled data from the combinations summarized in Table 3.4.

Rat no.	From	To	GCI	NSI
1	Pool #2 (ACC)	Pool #1 (ACC)	0.1966	0.1035
1	Pool #2 (ACC)	Pool #3 (STR)	0.1232	0.1116
1	Pool #2 (ACC)	Pool #4 (STR)	0.1163	0.1995
2	Pool #2 (ACC)	Pool #5 (STR)	0.0879	0.1445
2	Pool #2 (ACC)	Pool #6 (STR)	0.1098	0.1121
2	Pool #7 (STR)	Pool #5 (STR)	0.1598	-0.1058

3.4.2 Implications of the pooled data

A spike train obtained by superimposing individual spike trains and disregarding where each spike came from is called a *pooled spike train* [25]. Adjacent neurons usually work together with each other to generate suitable pooled spike trains to perform specific tasks. An illustration is provided in Figure 3.9. On the cause side (left brown), the pooled train (pool 1) can be considered as a collective input with respect to the effects of temporal or spatial summation of one of the following two types: (i) the additive effect produced by many PSPs that have been generated from several very close synapses on the same post-synaptic neuron at the same time. (ii) the additive effect produced by many PSPs that have been generated from several synapses which have similar effects on the axon hillock of the same post-synaptic neuron. On the effect side (middle blue), the pooled train (pool 2) can be considered as a collective output, which represents the total discharge of a group of cooperative neurons in function. Again, the collective output (pool 2) can then be treated as a collective input to trigger others (right purple).
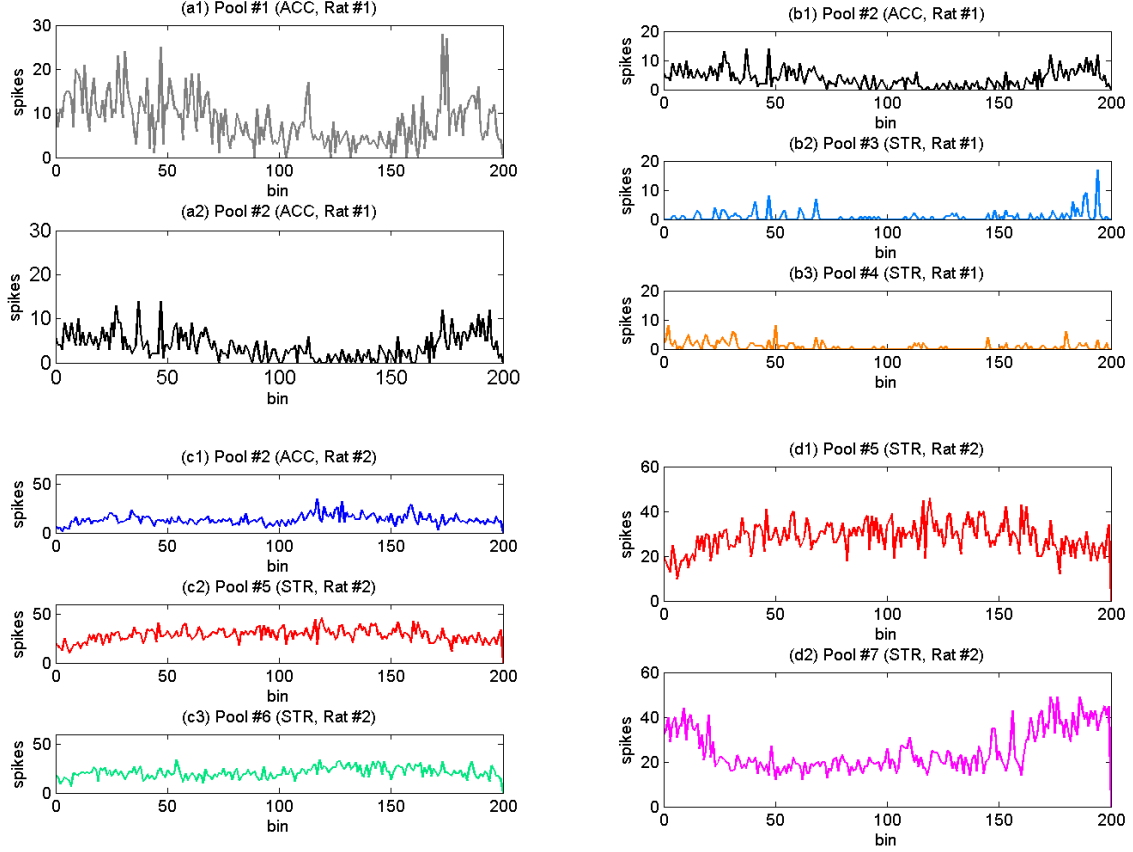
Figure 3.8: The firing trajectories (400 sec. with 2 sec. bin) under the quinpirole administration. (a) excitatory effects inside the ACC (NSI = 0.1035). (b) excitatory effects from the ACC to the STR (NSIs = 0.1116, 0.1995). (c) excitatory effects from the ACC to the STR (NSIs = 0.1445, 0.1121). (d) inhibitory effects inside the STR (NSI = −0.1058). It can be found that positive (negative) NSIs exhibit positive (negative) correlations in the fluctuations of the signals.

Cross correlations can be dramatically amplified by pooling, that is, weak correlations between pairs of neurons in two populations can lead to strong correlations between the summed activity of these two populations [56]. Similar results should hold for the GC analyses. To check this, a simulation is designed as follows: Let $P = \{p_1, p_2, \ldots, p_n\}$ be a Poisson spike train of time length T with firing rate λ. Let Q be the output spike train of the almost linear system in Section 3.3.2 with input P, synaptic weight w, and time delay d. Since P will be treated as a pooled spike train, we uniformly decompose it into k sub-trains $\{P_i, i = 1, \ldots, k\}$, that is, each $p_j \in P$ has the same probability to be distributed into the sub-train P_i, for $j = 1, \ldots, n$ and $i = 1, \ldots, k$. As a result, $\cup_i P_i = P$, $\cap_i P_i = \emptyset$, and the firing rate of each P_i is λ/k. Further, let $\{U_i, i = 1, \ldots, m\}$ be m uncorrelated spike trains with $\{P_i, i = 1, \ldots, k\}$ to serve as the role of environment neurons. The distribution of each U_i is also Poisson with rate λ/k. Now, for $T = 10$ (sec.), $\lambda = 20$ (spikes/sec.), $w = 8$ (mV), $d = 10$ (ms.), $k = 5$, $m = 2$, and bin width 0.1 (sec.), the GCI from P to Q is 0.2655, a quite large value; while the GCIs from P_i to Q is

41
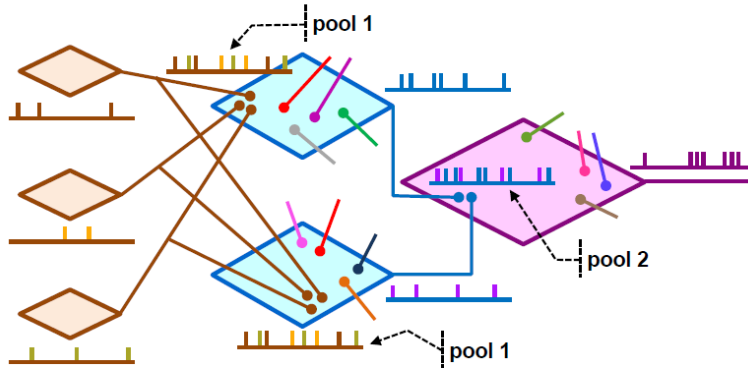
Figure 3.9: An illustration for pooled data. On the cause side (left brown), the pooled train (pool 1) can be considered as a collective input with respect to the effects of temporal or spatial summation of one of the two types ((i) and (ii) in the context.) On the effect side (middle blue), the pooled train (pool 2) can be considered as a collective output, which represents the total discharge of a group of cooperative neurons in function. Again, the collective output (pool 2) can then be treated as a collective input to trigger others (right purple).

about 0.005, a very low causality. The results are obtained from the average of 100 such simulations, and the first 2 sec. of one of the realizations is shown in Figure 3.10; where P is labeled neuron #1, P_i, neuron #2 − 6, U_i, neuron #7 − 8, and Q is labeled neuron #9.

To check the appropriateness of random grouping used in the previous subsection, 5 spike trains are randomly chosen from P_i and U_i and then be pooled together to compute the GCI from such pooled data to the target Q. The averaged results are: when these 5 spike trains are all chosen from P_i then the GCI attends the maximum 0.2655. When 4 is chosen from P_i and 1 from U_i the GCI is destroyed and is 0.0641. Finally, when 3 is chosen from P_i and the other 2 are from U_i, the GCI is 0.0209. The results show that if the pooled data contains the spikes of other irrelevant neurons then the GCIs will be small and destroyed.

For real-world spike train data, individual neurons usually perform weak contributions to each other while groups of neurons perform very significant contributions. In the former case, causal influences are difficult to be detected via most statistical methods, grouping and pooling are usually needed to enhance the causation. Since our data set is small, random grouping approach is both reasonable and sufficient to explore the network structure. Significant NSIs are also found within an acceptable period of time. How to efficiently group neurons is absent in our current analysis, but has been being studied via numerical simulations. Efficient grouping strategy is an interesting research topic and
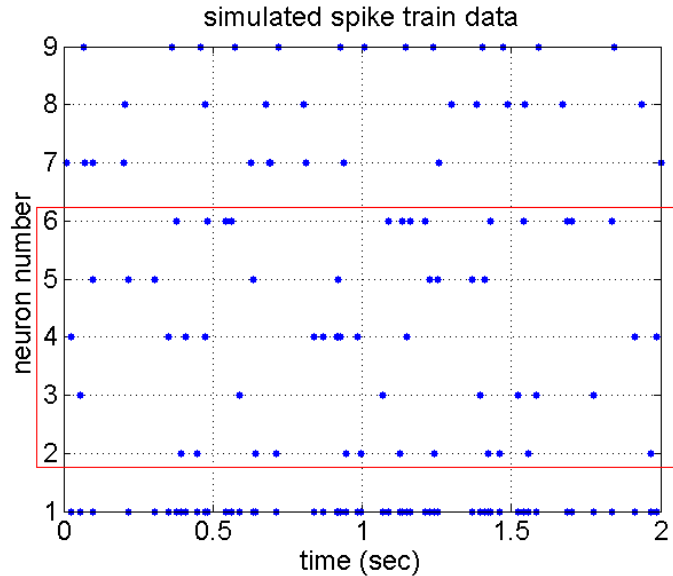
Figure 3.10: The first 2 sec. simulated spike trains of one of the realizations. The parameters are: $T = 10$ (sec.), $\lambda = 20$ (spikes/sec.), $w = 8$ (mV), $d = 10$ (ms), $k = 5$, $m = 2$, and bin width 0.1 (sec.), The source spike train P is labeled neuron #1, the decomposed trains P_i are labeled neuron #2 − 6, the environment trains U_i are labeled neuron #7 − 8, and the target Q is labeled neuron #9.

will be included in a separate article in the future.

3.5 Discussion

The original Granger causality index is by definition nonnegative, thereby lacking of the trait for differentiating the effects of excitations and inhibitions between neurons. Inspired by the concept that the firing pattern of the post-synaptic neuron is generally a weighted result of the effects of several pre-synaptic neurons with possibly different synaptic weights; a computational algorithm was proposed (Section 3.2.3) under a BLP assumption for analyzing neuronal networks by estimating the synaptic weights among them. The extended Granger causality index, the NSI, was shown to be able to measure both excitatory and inhibitory effects between neurons by several numerical simulations. The method was also illustrated to analyze real spike train data from the ACC and the STR. Significant NSIs were found only when certain random groups appear (Table 3.4 and 3.5); and the results showed, under the quinpirole administration, the significant existence of excitatory effects inside the ACC, excitatory effects from the ACC to the STR, and inhibitory effects inside the STR.

Mathematically, the proposed NSI is just a weighted version of the original GCI, that is, the NSI is obtained by multiplying the normalized weights with the GCI from the weighted trajectory to the target trajectory. Physiologically, under the prerequisite conditions that (i) all of the connectivity relations between neurons are correctly identified and (ii) the behavior of influences follow the vector autoregressive model with some

finite order; the original GCI can reveal the degree of causal influences between neurons while the extended NSI can reveal the degree and the type of synaptic transmission. Theoretically, the GCI and NSI can only be approximations or even be spurious if either of these two conditions fail to hold. It seems to be a very strong constraint for researchers to apply these approaches in practice. However, things are not really that bad. The simulated IF spiking networks in Section 3.3.2-3.3.3 are tested to be correctly captured by vector autoregressive model and show a consistency between network structures and the proposed indices. Finally and also importantly, note that significant GCI or NSI do not necessarily signify a anatomical connectivity since they are fundamentally statistical concepts. Treated with care, the GCI and NSI could be useful for researchers to infer possible relationships between network structure or to construct a description of network dynamics in neuroscience.

It should be pointed out that the significance of the NSI can be checked directly via the significance test on the GCI, since the proposed method is GC-based. In addition, if voltage-trajectories are available then the method can be applied directly without any ambiguity. For spike train data, however, the role of voltage-trajectory can be replaced by the trajectory of firing rate estimated using binning or Gaussian kernel filtering. The high firing rate in the simulations are to ensure significant inhibitory effects among the simple networks. In practice, the high firing rate can be considered as a consequence of pooling spike trains. Low firing rate leads to sparse spike trains, for which the Gaussian kernel filtering will introduce highly artificial signals which will hinder the autoregression modeling for computing the GCI. The solution is applying some suitable preprocessing [75] or applying binning [63], which is more stable for converting sparse spike trains.

In closing, it is worth noting some related articles for possible future works. (i) In Section 3.3.3, the NSI was shown to successfully capture the behavior of synaptic transmissions. Although the NSI values are correlated with the underlying synaptic weights, it is not clear what the actual relationship might be. Cadotte et al. have found that under certain settings, the GCI and the synaptic weight has the following nonlinear relationship [12]:

$$F_{Y \to X} = \frac{1}{1 + 384e^{-02124*sw}}, \tag{3.10}$$

where sw denotes the synaptic weight. Similar results for the NSI may be derived in the future using (3.10). (ii) When faced with most scientific computing problem, a linear model is generally a first, basic, and winning strategy to try. Neural spiking networks are highly complicated and nonlinear, a linear model could be considered to be satisfactory if it can well approximates the behavior of a nonlinear dynamics to some degree. In the current study, the behavior of the simulated nonlinear network is well captured by our approach; obtaining the exact ratio of the synaptic weights can be set as an important objective for us to strive. Our proposed method and the BLP assumption could be reformulated and generalized to handle nonlinear dynamics by means of the nonparametric kernel modelling [48] in the future. (iii) This work did not consider any hidden network structure in both theoretical and numerical parts. The partial Granger causality, proposed by Guo et al. [30], may be used to extend the method to deal with the effects

of exogenous inputs or hidden neurons. (iv) Granger causality is originally designed to measure effect, not mechanism [5]. Numerical evidences showed that Granger causality can, to some degree, be used to infer the underlying mechanism. The GCI only use the information of residual noises, there are still some useful information which could be extracted from the regression coefficients [32]. The NSI uses the summation of regression coefficients (3.7), other forms of information may be developed in the future. (v) The Gaussian filtering and binning techniques link the spike train data and the statistical models for continuous signals, leading to both mathematically easy derivations and computationally efficient algorithms. However, distortions may arise after the filtering is applied. A generalized linear model (GLM)-based point process framework was proposed for directly applying the GC on neural spike trains without any filtering [38]. A conceptually similar but more robust measure, called directed information, was also proposed [55]. The modality-independent nature allows the measure to characterize statistically causal relationship between arbitrary stochastic processes. A more sophisticated coupling model was also proposed [53]. It's parameters consist of a bank of stimulus filters, spike-history filters, and couping filters. Splines can also be used to fit nonlinearity in the stimulus filter. The mathematics used in these frameworks are more involved than that used in this chapter, but we surmise that the proposed algorithm and assumption can somehow be translated into them to obtain similar or more powerful results in the future. (vi) Recently, a new framework of spatio-temporal Granger causality has been proposed to reliably estimate the Granger causality from experimental datasets possessing time-varying properties [44]. The NSI may be extended to its dynamic version for automatically analyzing experimental datasets without laborious jobs on windowing.

Chapter 4

Large-Scale Sparse Neuronal Networks

4.1 Introduction

In neuroscience, identifying causal dependencies between multiple time-series signals is a topic of great interest. Regression methods and time-series models have been extensively employed in attempt to uncover the underlying brain connectivity. One of the most popular and successful method is the Granger causality (GC) which proposes that if the prediction of one time series can be improved with the knowledge of a second time series, then there is a causal influence from the second one to the first one. The prediction mentioned above is made by using autoregressive models [27] and a GC index (GCI) is also derived to quantify the strength of caual interactions. The GC is shown to be effective in distinguishing between direct and indirect causal connections and has applications across various domains such as: electroencephalography/magnetoencephalography (EEG/MEG) [2,26,30], local field potential (LFP) [13,24], functional magnetic resonance imaging (fMRI) [40,45,73], calcium imaging (CI) [21], and multiple spike trains [18,38,41,74]. Recently, based on the GC framework, we developed a new index called neuron synaptic index (NSI) to measure the synaptic weights between neurons in a neuronal network [62]. GCI is designed in its original form to measure causal connectivities while NSI is designed to measure structural connectivities, and in general they differ from each other. The GCI and NSI will be shortly introduced in Section 4.2.5 and 4.2.6, respectively. More details about the GC can be found in [4,8,12,20,31].

Nowadays, due to the development of multi-electrode array and optogenetic techniques, more and more large-scale time series data become available. When the number of variables is much larger than the length of time series, traditional implementation of the GC encounters serious problems such as solvability of large-sclae underdetermined system, high computational cost/complexity, and multiple statistical testing. The same problem exists in microarray dataset of large-scale gene regulatory networks [15,49,76]. To cope with this high-dimensional situation, penalized regression methods (e.g., L1-regularization or LASSO, LASSO variants) have been applied. Although these methods lead to a considerable progress for the analysis of high-dimensional sparse data, there still leaves room for improvement. (1) The regularization parameter is not easy to tune, con-

ducting time-consuming cross-validation (CV) is needed. (2) CV usually gives not very sparse solutions which might contains lots of irrelevant variables, leading to overfitting problems. (3) Lasso can fail to distinguish irrelevant variables that are highly correlated with relevant variables [72]. (4) Lasso may lead to severe bias for large regression coefficients [22]. GC has proven to be effective and powerful in the case where the number of variables is much less than the length of time series. For neural data, short time-series length is often used to ensure the stationarity of the data and the time-invariance of the underlying structure since neural data are usually only locally stationary and locally time-invariant. However, the number of variables is capable of being reduced if the underlying connection to each neuron is sparse, meaning that the number of connecting edges is quite small compared to the number of possible edges. Therefore, our direction is to first, for each neuron, reduce the large-scale variable set to a small one which may contains a few irrelevant variables but should contains all the relevant variables; then, for each neuron, apply the GC to the associated reduced set.

The assumption of network sparsity has the following important meanings: the first one is for mathematical reasons that oversized ratio between variable number and series length causes extreme obstacles to any existing statistical method. So, limit the network complexity is currently the best way to cross the obstacle. The second one is for the evidence found that many brain areas have sparse structures. For instance, cortical neurons were found to be quite sparsely connected relative to the population of neurons in a cell's neighborhood [10, 11, 16, 52, 66]. The third one is for practical use that picking out sparse but pivotal variables from a large-scale dataset may be quite satisfactory to most experimental purposes. Based on the sparsity assumption, we propose a computational algorithm (in Section 4.2.4) for automatically and effectively selecting relevant neurons from a large scale neuronal network. The kernel wrapped in the algorithm is a three-stage model selection approach (Section 4.2.1–4.2.3), which consists of (1) an orthogonalized forward selection of input variables that circumvents difficulties with inverting high-dimensional matrices, (2) a stopping criterion to terminate the forward inclusion of variables, and (3) a backward elimination of variables according to the criterion to ensure the consistency of the approach. The performance was checked by a large-scale simulated threshold spiking neuron model, and real data from rats executing empathic prosocial behaviors were also analyzed.

4.2 Methods

4.2.1 Forward stepwise regression via the OGA iteration

Consider the multiple linear regression model

$$y_t = \sum_{j=1}^{p} \beta_j x_{tj} + \epsilon_t, \ t = 1, \dots, n, \tag{4.1}$$

with p predictor variables $x_{t1}, x_{t2}, \cdots, x_{tp}$ that are uncorrelated with the zero mean random noises ϵ_t. Estimating the regression coefficients $\{\beta_j, j = 1, \dots, p\}$ by standard regression methods confronts difficult computational and statistical problems (e.g., covariance

matrix inversion, multiple statistical testing) when p is larger than n. Here we introduce a stepwise regression method, called the orthogonal greedy algorithm (OGA), which uses componentwise linear regression so that the problems can naturally be avoided.

First, initialize with $\hat{y}_k(\cdot) = 0$ and compute at the end of the kth iteration the residual $U_t^{(k)} := y_t - \hat{y}_k(\mathbf{x}_t), 1 \leq t \leq n$, which represents the unexplained portion of y_t at the k-th stage. Second, choose the most correlated variable $x_{\hat{j}_{k+1}}$ with $U_t^{(k)}$, that is, set

$$\hat{j}_{k+1} = \arg \min_{1 \leq j \leq p} \sum_{t=1}^{n} (U_t^{(k)} - \tilde{\beta}_j^{(k)} x_{tj})^2, \tag{4.2}$$

where $\tilde{\beta}_j^{(k)} = \sum_{t=1}^{n} U_t^{(k)} x_{tj} / \sum_{t=1}^{n} x_{tj}^2$ is the projection coefficient of x_{tj} onto $U_t^{(k)}$. Third, orthogonalize the predictor variables sequentially so that componentwise linear regression can be applied at each stage, circumventing difficulties with high-dimensional matrix inversion. Let $\mathbf{X}_j = (x_{1j}, \ldots, x_{nj})^\top$ be the realization of predictor variable x_{tj}. Suppose $\mathbf{X}_{\hat{j}_1}, \mathbf{X}_{\hat{j}_2}^\perp, \ldots, \mathbf{X}_{\hat{j}_k}^\perp$ are the orthogonal vectors already acquired in the previous stages, we can then compute $\mathbf{X}_{\hat{j}_{k+1}}^\perp = \mathbf{X}_{\hat{j}_{k+1}} - \hat{\mathbf{X}}_{\hat{j}_{k+1}}$, where $\hat{\mathbf{X}}_{\hat{j}_{k+1}}$ is the projection of $\mathbf{X}_{\hat{j}_{k+1}}$ onto span$\{\mathbf{X}_{\hat{j}_1}, \mathbf{X}_{\hat{j}_2}^\perp, \ldots, \mathbf{X}_{\hat{j}_k}^\perp\}$. Finally, update the predictor by

$$\hat{y}_{k+1}(\mathbf{x}_t) = \hat{y}_k(\mathbf{x}_t) + \hat{\beta}_{\hat{j}_{k+1}}^{(k)} x_{t,\hat{j}_{k+1}}^\perp, \tag{4.3}$$

where $\hat{\beta}_{\hat{j}_{k+1}}^{(k)} = \sum_{t=1}^{n} U_t^{(k)} x_{t,\hat{j}_{k+1}}^\perp / \sum_{t=1}^{n} (x_{t,\hat{j}_{k+1}}^\perp)^2$.

4.2.2 Stopping rule and consistent model selection

A stopping rule is needed to terminate the forward inclusion of the OGA iteration described above; otherwise all of the p variables will ultimately be included, which is useless and unnecessary. It is shown that, under certain general and sparsity conditions, one can terminate the OGA iteration after $K = O((n/\log p)^{1/2})$ steps, that is, after K variables have been included [34]. Generally, $K \ll p$.

In addition to the stopping rule, we also need a criterion to select the smallest set of all relevant variables along the OGA path (an input variable is called "relevant" if its associated β_j is nonzero, and "irrelevant" otherwise). Here we introduce a variable-selection consistent criterion, called the "high-dimensional information criterion" (HDIC). Specifically, for $\emptyset \neq J \subset \{1, \ldots, p\}$, let $\hat{\sigma}_J^2 = n^{-1} \sum_{t=1}^{n} (y_t - \hat{y}_{t;J})^2$, where $\hat{y}_{t;J}$ denotes the fitted value of y_t when $\mathbf{Y} = (y_1, \ldots, y_n)^\top$ is projected onto span$\{\mathbf{X}_j, j \in J\}$. The criterion is then defined as the following:

$$\text{HDIC}(J) = n \log \hat{\sigma}_J^2 + \sharp(J) w \log p, \tag{4.4}$$

$$\hat{k} = \arg \min_{1 \leq k \leq K} \text{HDIC}(\hat{J}_k), \tag{4.5}$$

where $\hat{J}_k = \{\hat{j}_1, \ldots, \hat{j}_k\}$ and different choices of w correspond to differenct criteria. The case $w = \log n$ corresponds to high-dimensional Bayesian information criterion (HDBIC); without the $\log p$ factor, (4.4) reduces to the usual Bayesian information criterion (BIC). The case $w = c \log \log n$ with $c > 2$ corresponds to Hannan-Quinn criterion (HDHQ).

4.2.3 Further trimming to exclude irrelevant variables

The procedure introduced so far consists of three main parts: (i) carrying out the OGA iteration K times, (ii) computing HDIC($\hat{J}_k$) at the end of the kth iteration, (iii) choosing the k that minimizes HDIC($\hat{J}_k$) over $1 \leq k \leq K$ after the OGA iteration terminates. Although it can be shown to select the smallest set of all relevant variables with probability approaching 1, some irrelevant variables may still be included along the OGA path if one of the required mathematical conditions is violated. To exclude irrelevant variables under these circumstances, we further use the HDIC to define a subset $\hat{N}$ of $\hat{J}_{\hat{k}}$ by

$$\hat{N} = \{\hat{j}_l : \text{HDIC}(\hat{J}_{\hat{k}} - \{\hat{j}_l\}) > \text{HDIC}(\hat{J}_{\hat{k}}), 1 \leq l \leq \hat{k}\} \quad \text{if} \quad \hat{k} > 1, \tag{4.6}$$

and $\hat{N} = \{\hat{j}_1\}$ if $\hat{k} = 1$. Note that (4.6) only requires the computation of $\hat{k} - 1$ additional least squares estimates (the $\hat{j}_{\hat{k}}$ itself must satisfy HDIC($\hat{J}_{\hat{k}} - \{\hat{j}_{\hat{k}}\}$) > HDIC($\hat{J}_{\hat{k}}$), therefore need not to be tested again). The idea is that if there is a redundant variable lies in $\hat{J}_{\hat{k}}$, then the HDIC will be even less when the redundant variable is picked out.

4.2.4 The computational algorithm

Here, we present a step by step algorithm for automatically and effectively selecting relevant neurons from a large scale neuronal network. We use OGA, HDIC and Trim to represent the procedure introduced in Section 4.2.1, 4.2.2, and 4.2.3, respectively.

Step 1 : Properly smooth the spike train data by binning or kernel filtering [42, 63, 75] to form time series matrix $S \in \mathbb{R}^{N \times n}$ where N denotes the total number of neurons and n denotes the data length. (For time-series data, this step can be omitted) Then for each neuron do Step2–Step4

Step 2 : Choose a candidate time lag $m \in \mathbb{N}$ and then convert the autoregressive $AR(m)$ model of S into multiple linear regression format with $X \in \mathbb{R}^{(n-m) \times p}$ and $y \in \mathbb{R}^{(n-m) \times 1}$ where $p = Nm \gg n$.

Step 3 : Run OGA+HDIC+Trim to get the suggested variable set $J(m)$.

Step 4 : Return to Step2 several times and then choose the optimal $J(m)$ which has the smallest $HDIC(J(m)) := n \log \hat{\sigma}^2_{J(m)} + \sharp(J(m)) w \log p(m)$.

Step 5 : Finally, use these feasible small-size variable sets to do further sophisticated modeling.

Once the network scale has been reduced to an acceptable range, some causal inference methods, such as Granger causality index and neuron synaptic index can then be applied. The following subsections give an outline of these methods and more details can be found in [62].

4.2.5 Granger causality index

Let $X = (x_1, x_2, \ldots, x_N)$ be a stationary N-dimensional time series process with zero mean. The m-th order linear autoregressive model for X is given by

$$
\begin{cases}
x_{t,1} = \sum_{r=1}^{m} a_r^{1,1} x_{t-r,1} + \cdots + \sum_{r=1}^{m} a_r^{1,N} x_{t-r,N} + \epsilon_{t,1} \\
x_{t,2} = \sum_{r=1}^{m} a_r^{2,1} x_{t-r,1} + \cdots + \sum_{r=1}^{m} a_r^{2,N} x_{t-r,N} + \epsilon_{t,2} \\
\vdots \\
x_{t,N} = \sum_{r=1}^{m} a_r^{N,1} x_{t-r,1} + \cdots + \sum_{r=1}^{m} a_r^{N,N} x_{t-r,N} + \epsilon_{t,N},
\end{cases}
\tag{4.7}
$$

where regression coefficient $a_r^{i,j}$ indicates the coupling strength from x_j to x_i. The residuals $\epsilon_1, \epsilon_2, \cdots, \epsilon_N$ are zero-mean uncorrelated white noises with covariance matrix Σ. The diagonal entry $\{\Sigma_{ii} = Var(\epsilon_i), i = 1, \ldots, N\}$ measures the prediction error of $\{x_i, i = 1, \ldots, N\}$, respectively, based on the information from time stamps $t - 1$ to $t - m$.

To see the importance of x_j to x_i, we can exclude the variable x_j from (4.7) to obtain a reduced $(N - 1)$-dimensional autoregressive model with residual series $\eta^{i,j}$ of x_i and corresponding prediction error $\Gamma_{ii}^{j} = Var(\eta^{i,j})$. If Σ_{ii} is significantly less than Γ_{ii}^{j} in some suitable statistical sense, then we say that x_j Granger-cause x_i. This causality can be quantified by the GC index from x_j to x_i formulated as:

$$
F_{j \to i} = \ln \frac{\Gamma_{ii}^{j}}{\Sigma_{ii}}.
\tag{4.8}
$$

It is clear that $F_{j \to i} = 0$ when $\Gamma_{ii}^{j} = \Sigma_{ii}$, i.e., x^j has no causal influence on x^i, and $F_{j \to i} > 0$ when x^j Granger-cause x^i. The significance of the GC index (4.8) can be tested via an F-test on the corresponding coefficients $a_r^{i,j}$ [29, 60]. Parameters $a_r^{i,j}$ and Σ_{ii} can be estimated by solving the Yule-Walker equation [39] and an efficient model order can be obtained by minimizing the Akaike Information Criterion:

$$
AIC(m) = 2 \log(\det(\Sigma)) + \frac{2N^2 m}{n},
\tag{4.9}
$$

where n is the total length of the time series.

4.2.6 Neuron synaptic index

Consider the multivariate zero-mean membrane voltage time series, $X = (x_1, x_2, \ldots, x_N)$, of N distinct neurons. Suppose that the i-th neuron is triggered by some other k neurons in the network, say $\{i_1, i_2, \ldots, i_k\}$-th neurons, with synaptic weights $\{\alpha_1^i, \alpha_2^i, \ldots, \alpha_k^i\}$. Positive and negative weights represent excitatory and inhibitory influences on the i-th neuron, respectively.

Under the assumption of the best linear predictor (BLP) (Definition 1 below), $I_i :=$ $\{i_1, i_2, \ldots, i_k\}$ and $\alpha^i := \{\alpha^i_1, \alpha^i_2, \ldots, \alpha^i_k\}$ can be estimated effectively. The results are described in the following proposition.

Definition 1.
Let x and y be two stationary time series with zero-means. Then we say that y forms the best linear predictor (BLP) of x among a variable set $\mathcal{G}$ if $\sigma^2(x|\bar{x}, \bar{y}) < \sigma^2(x|\bar{x}, \bar{z}), \forall z \in \mathcal{G}$, where $\sigma^2(x|\bar{x}, \bar{y}) := min_{p, \{f_r\}, \{d_r\}} E\{x_t - \sum_{r=1}^p [f_r y_{t-r} + d_r x_{t-r}]\}^2$.

Proposition 1.
In the situation described above, if further the weighted trajectory $u^i := \alpha^i_1 x_{i_1} + \alpha^i_2 x_{i_2} + \cdots + \alpha^i_k x_{i_k}$, forms the BLP of x_i among the whole network; then based on the GC framework, I_i can be completely identified and α^i can be estimated as $\hat{\alpha}^i := \{\sum_{r=1}^m a_r^{i,i_1}, \sum_{r=1}^m a_r^{i,i_2}, \ldots, \sum_{r=1}^m a_r^{i,i_k}\}$.

The neuron synaptic index then can be computed through the following steps:

Step 1 : Compute all the GC indices by (4.8) for all pairs of neurons and also perform F-tests to ensure statistical significance.

Step 2 : For each node $i = 1, 2, \ldots, N$, refine the autoregressive model by ruling out the neurons with insignificant $F_{j \to i}$ to correct the regression coefficients.

Step 3 : For each node, compute the synaptic weights of the trigger neurons by simply summing the regression coefficients up to the model order m.

Step 4 : For each node, take the weighted trajectory as a new explanatory time series and then compute the GC index from this weighted time series to that of the node.

Step 5 : Finally, the NSI is defined to be the l_1-normalized estimated weights obtained in Step 3 multiplying the GC index obtained in Step 4.

4.3   Results

4.3.1   Simulation

In this section, we report the simulation results and performance of OGA+HDHQ and OGA+HDBIC through three different types of models: the multiple linear regression, autoregressive time-series and threshold spiking neuron models. The Matlab code used for this study is available to interested readers upon request.

Multiple linear regression model

Consider the regression model

$$y_t = \sum_{j=1}^q \beta_j x_{tj} + \sum_{j=q+1}^p \beta_j x_{tj} + \epsilon_t, \ t = 1, \ldots, n, \tag{4.10}$$

where $\beta_{q+1} = \cdots = \beta_p = 0, p \gg n, \epsilon_t$ are i.i.d. $N(0, \sigma^2)$ and are independent of the x_{tj}. The regressors are designed as

$$x_{tj} = d_{tj} + \eta w_t, \ j = 1, \ldots, p, \tag{4.11}$$

where $\eta \geq 0$ tunes the strength of correlation between input variables, and $(d_{t1}, \ldots, d_{tp}, w_t)^\top$, $1 \leq t \leq n$ are i.i.d. normal with zero mean and identity covariance matrix I. Simple calculation gives $Corr(x_{tj}, x_{tk}) = \frac{\eta^2}{1+\eta^2}$, which increases with $\eta > 0$. The input variables are mutually independent when $\eta = 0$.

Consider (4.10) with $q = 5$, $(\beta_1, \ldots, \beta_5) = (3, -3.5, 4, -3.6, 3.2)$, and $\sigma = 1$. We choose $K = D(n/\log p)^{1/2}$ with $D = 5$ and choose $c = 2.01$ for HDHQ. We have also allowed D to vary between 3 and 10 and c to vary among 2.01, 2.51, 3.01, 3.51 and 4.01, but the results are quite similar for the different choices. The results are shown in Table 4.1. For the sample size $n \geq 100$, OGA includes all of the 5 relevant regressors within K iterations, and for about 99% of the 1000 simulations, the HDHQ and HDBIC successfully identify the smallest correct model, irrespective of whether the candidate regressors x_j are uncorrelated ($\eta = 0$) or highly correlated ($\eta = 2$) with y. The OGA+HDBIC+Trim has the best performance, choosing the smallest correct model in all simulations. For comparison, we have also included the performance of OGA+BIC. Table 4.1 shows that for $n \geq 100$, OGA+BIC always chooses the largest model $\hat{J}_K$ along the OGA path.

Define the mean squared prediction error as follows:

$$\text{MSPE} = \frac{1}{1000} \sum_{l=1}^{1000} (\sum_{j=1}^{p} \beta_j x_{n+1,j}^{(l)} - \hat{y}_{n+1}^{(l)})^2, \tag{4.12}$$

where $x_{n+1,1}^{(l)}, \ldots, x_{n+1,p}^{(l)}$ are the regressors associated with $y_{n+1}^{(l)}$, the new outcome in the lth simulation run, and $\hat{y}_{n+1}^{(l)}$ denotes the predictor of $y_{n+1}^{(l)}$. The MSPEs of OGA+BIC are quite larger than that of other methods when $n \geq 100$ due to serious overfitting. The traditional BIC can include all relevant variables but it also includes all irrelevant variables. The large p leads to many spuriously significant regression coefficients if one does not adjust for multiple testing [6]. The factor used in (4.4) of HDIC can be regarded as such adjustment.

Autoregressive time-series model

Consider the autoregressive model

$$y_t = 0.95 y_{t-r_0} + \sum_{j=1}^{q} \beta_j x_{t-r_j,j} + \sum_{j=q+1}^{N-1} \sum_{r=1}^{m} \beta_{r,j} x_{t-r,j} + \epsilon_t, \ t = 1, \ldots, n, \tag{4.13}$$

where $\beta_{1,q+1} = \cdots = \beta_{m,N} = 0$, ϵ_t are i.i.d. $N(0, \sigma^2)$ and are independent of x_{tj}. Here, for simplicity, we assume that $\{y, x_j, j = 1, \ldots, q\}$ contributes to y only through single lags $\{r_j, j = 0, \ldots, q\}$, respectively. The regressors x_j are designed in the same way as (4.11). After converting to multiple linear regression format, the total number of candidate variables will be $p := Nm \gg n$.

Table 4.1: Frequency, in 1000 simulations, of including all five relevant variables (Correct), of selecting exactly the relevant variables (E), of selecting all relevant variables and i irrelevant variables (E_i), and of selecting the largest model which includes all relevant variables (E^*).

η	n	p	Method	E	E_1	E_2	$E_{3\sim5}$	E^*	Correct	MSPE
0	50	1000	OGA+HDHQ	818	78	23	8	1	928	5.416
			OGA+HDHQ+Trim	873	45	8	2	0	928	5.408
			OGA+HDBIC	864	44	16	0	0	924	5.533
			OGA+HDBIC+Trim	923	1	0	0	0	924	5.526
			OGA+BIC	0	0	0	0	928	928	10.598
	100	2000	OGA+HDHQ	991	9	0	0	0	1000	0.057
			OGA+HDHQ+Trim	991	9	0	0	0	1000	0.057
			OGA+HDBIC	999	1	0	0	0	1000	0.053
			OGA+HDBIC+Trim	1000	0	0	0	0	1000	0.053
			OGA+BIC	0	0	0	0	1000	1000	1.214
	200	4000	OGA+HDHQ	1000	0	0	0	0	1000	0.023
			OGA+HDHQ+Trim	1000	0	0	0	0	1000	0.023
			OGA+HDBIC	1000	0	0	0	0	1000	0.023
			OGA+HDBIC+Trim	1000	0	0	0	0	1000	0.023
			OGA+BIC	0	0	0	0	1000	1000	0.942
2	50	1000	OGA+HDHQ	592	132	55	38	2	819	10.908
			OGA+HDHQ+Trim	780	37	1	0	1	819	10.884
			OGA+HDBIC	614	116	51	20	0	801	12.247
			OGA+HDBIC+Trim	800	1	0	0	0	801	12.230
			OGA+BIC	0	0	0	0	819	819	14.682
	100	2000	OGA+HDHQ	989	11	0	0	0	1000	0.062
			OGA+HDHQ+Trim	995	5	0	0	0	1000	0.062
			OGA+HDBIC	994	6	0	0	0	1000	0.059
			OGA+HDBIC+Trim	1000	0	0	0	0	1000	0.059
			OGA+BIC	0	0	0	0	1000	1000	1.152
	200	4000	OGA+HDHQ	1000	0	0	0	0	1000	0.027
			OGA+HDHQ+Trim	1000	0	0	0	0	1000	0.027
			OGA+HDBIC	1000	0	0	0	0	1000	0.027
			OGA+HDBIC+Trim	1000	0	0	0	0	1000	0.027
			OGA+BIC	0	0	0	0	1000	1000	1.022

Consider (4.13) with $q = 5$, $(\beta_1, \ldots, \beta_5) = (3, -3.5, 4, -3.6, 3.2)$, $m = 3$, $(r_0, r_1, \ldots, r_5) = (3, 1, 2, 3, 2, 1)$ and $\sigma = 1$. We choose $K = D(n/\log p)^{1/2}$ with $D = 5$ and choose $c = 2.01$ for HDHQ. Define the mean squared prediction error

$$\text{MSPE} = \frac{1}{1000} \sum_{l=1}^{1000} (y_{n+1}^{(l)} - \hat{y}_{n+1}^{(l)})^2, \tag{4.14}$$

where $y_{n+1}^{(l)}$ is the new outcome in the lth simulation run, and $\hat{y}_{n+1}^{(l)}$ denotes the predictor

of $y_{n+1}^{(l)}$. This MSPE is also used in the next subsection because in practice we only observed the data y and the true model coefficients are unknown. The simulation results are shown in Table 4.2. The results are quite similar to those presented in Table 4.1 except that more samples are needed for the methods to perform well when $\eta > 0$. This is mainly due to the structural discrepancy between autoregressive and linear regression models for the former has temporal correlations between lag variables while the latter has ideal i.i.d. samples.

Table 4.2: Simulation results of autoregressive time-series model (see notation in Table 4.1).

η	n	p	Method	E	E_1	E_2	$E_{3\sim5}$	E^*	Correct	MSPE
0	50	750	OGA+HDHQ	793	94	22	9	1	919	5.491
			OGA+HDHQ+Trim	859	44	11	4	1	919	5.490
			OGA+HDBIC	847	54	13	0	0	914	6.047
			OGA+HDBIC+Trim	912	2	0	0	0	914	6.046
			OGA+BIC	0	0	0	0	919	919	10.964
	100	1500	OGA+HDHQ	995	5	0	0	0	1000	1.017
			OGA+HDHQ+Trim	995	5	0	0	0	1000	1.017
			OGA+HDBIC	1000	0	0	0	0	1000	1.019
			OGA+HDBIC+Trim	1000	0	0	0	0	1000	1.019
			OGA+BIC	0	0	0	0	1000	1000	2.278
1	100	750	OGA+HDHQ	776	178	33	13	0	1000	1.142
			OGA+HDHQ+Trim	983	16	1	0	0	1000	1.120
			OGA+HDBIC	790	167	30	13	0	1000	1.139
			OGA+HDBIC+Trim	999	1	0	0	0	1000	1.118
			OGA+BIC	0	0	0	0	1000	1000	2.300
	200	1500	OGA+HDHQ	991	9	0	0	0	1000	1.120
			OGA+HDHQ+Trim	995	5	0	0	0	1000	1.120
			OGA+HDBIC	996	4	0	0	0	1000	1.119
			OGA+HDBIC+Trim	1000	0	0	0	0	1000	1.119
			OGA+BIC	0	0	0	0	1000	1000	2.077

Threshold spiking neuron model

To illustrate the variable selection method in a neural spiking context, a large-scale sparse feedforward I&F neuron network is simulated (depicted in Figure 4.1). Briefly, excluding the target neuron, there are $N-1$ neurons in total, but only $q = 5$ neurons have inputs to the target. Three of the five have excitatory effects on the target and the other two hava inhibitory effects. The target neuron is implemented by a direct discrete time summation of the synaptic inputs $\{\beta_1, \ldots, \beta_5\}$ with propagation delay $\{r_1, \ldots, r_5\}$ respectively. Besides, the background noisy input is set to be i.i.d. $N(\mu, \sigma)$ at every time step with time resolution 1 ms. The internal potential is reset to $V_{reset} = -80$ (mV) and produces a spike when the threshold value $V_{th} = -55$ (mV) is reached. During the refractory period, the potential will linearly recover from V_{reset} to the resting potential

$V_{rest} = -70$ (mV). To model the diffusion of ions, there is a 2 mV decrease/increase of the potential to the V_{rest} every unit time depending on the status of de/hyper -polarizations, respectively. The internal potential is forced to lie in the range $[E_{K^+}, E_{Na^+}] = [-90, 60]$, which are the equilibrium potentials of K^+ and Na^+.
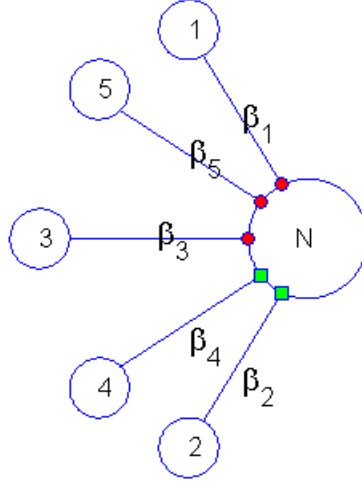
Figure 4.1: The simulated feedforward neuronal network. Red circles represent excitation and green squares represent inhibition. Source neurons $1 - 5$ have synaptic inputs $\beta_1, \ldots, \beta_5$ with propagation delay $r_1, \ldots, r_5$, respectively. The target neuron N is implemented by a direct discrete time summation of the synaptic inputs. The background noisy input is set to be i.i.d. $N(\mu, \sigma)$ at every time step with time resolution 1 ms.

The rest $N - 1 - q$ neurons are independent of the target neuron, but all the $N - 1$ neurons are modeled as Poisson point processes (X_{tj}) with same firing rate:

$$X_{tj} = D_{tj} + W_t, \ j = 1, \ldots, N - 1, \tag{4.15}$$

where $(D_{t1}, \ldots, D_{t,N-1}, W_t)$ are independent Poisson processes with firing rate $(\lambda, \ldots, \lambda, \eta)$ Hz. So, $\{X_{tj}, j = 1, \ldots, N-1\}$ have the same firing rate $\lambda + \eta$ and are correlated with each other if $\eta \neq 0$. The synaptic weights are fixed at $(\beta_1, \ldots, \beta_5) = (3, -3.5, 4, -3.6, 3.2) \times$ 3 mV, and the associated propagation delay are fixed at $(r_1, \ldots, r_5) = (1, 2, 3, 2, 1) \times 0.5$ sec. Let n donote the data length after binning with bin size 0.5 sec and $p = N * m$ denote the total number of candidate variables, where m is the chosen AR order (will be 3 in this case).

We begin with Simulation 1 in which the associated parameter settings are: $\lambda = 40$ Hz, $\eta = 0$ Hz, $\mu = \sigma = 0$; namely, the candidate input processes are uncorrelated with each other with common firing rate 40 Hz and there are no background noisy inputs to the target neuron. The simulated firing trajectories and the simulation result are shown in Figure 4.2 and Table 4.3, respectively. For about 99% of the 1000 simulations, both HDHQ and HDBIC successfully identify the smallest correct model as sample size n increases. But it is found that, for small sample size, HDBIC is too conservative and performs not very well; while HDHQ+Trim identifies the smallest correct input neurons

in 948 simulations and includes $1 - 2$ extra irrelevant neurons in only 16 simulations. The traditional BIC always chooses the largest set of neurons along the OGA path and causes serious overfitting (largest MSPE).
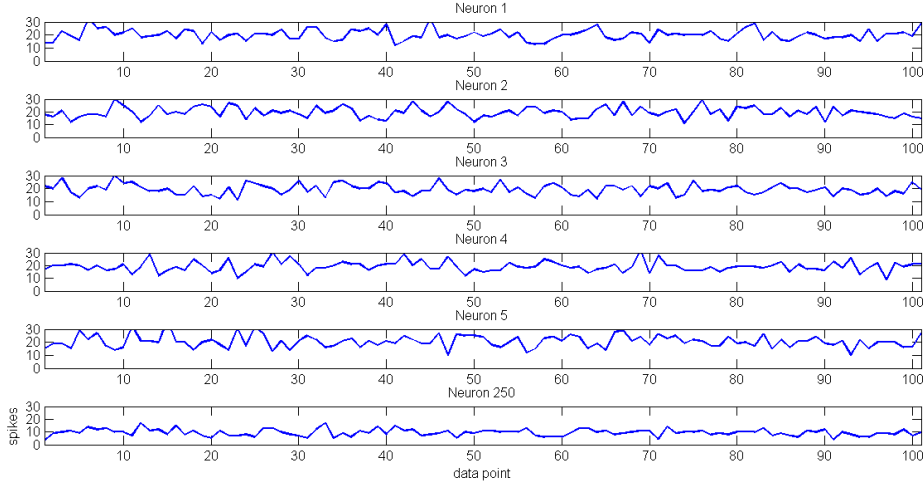
Figure 4.2: The firing behaviors of the source and target neurons in Simulation 1. The associated parameter settings are: $(\beta_1, \ldots, \beta_5) = (3, -3.5, 4, -3.6, 3.2) \times 3$ mV, $(r_1, \ldots, r_5) = (1, 2, 3, 2, 1) \times 0.5$ sec., $N = 250$, $n = 100$, $m = 3$, $p = Nm = 750$, $\lambda = 40$, $\eta = \mu = \sigma = 0$.

Table 4.3: Results of Simulation 1 in threshold spiking neuron model (see notation in Table 4.1). The associated parameter settings are: $\lambda = 40$ Hz, $\eta = 0$ Hz, $\mu = 0$, $\sigma = 0$.

n	p	Method	E	E_1	E_2	$E_{3 \sim 5}$	E^*	Correct	MSPE
100	750	OGA+HDHQ	900	57	6	1	0	964	1.221
		OGA+HDHQ+Trim	948	15	1	0	0	964	1.222
		OGA+HDBIC	760	4	0	0	0	764	1.680
		OGA+HDBIC+Trim	761	0	0	0	0	761	1.680
		OGA+BIC	0	0	0	0	993	993	2.455
200	1500	OGA+HDHQ	995	5	0	0	0	1000	1.188
		OGA+HDHQ+Trim	995	5	0	0	0	1000	1.188
		OGA+HDBIC	999	1	0	0	0	1000	1.185
		OGA+HDBIC+Trim	999	1	0	0	0	1000	1.185
		OGA+BIC	0	0	0	0	1000	1000	2.041
400	2250	OGA+HDHQ	997	3	0	0	0	1000	1.215
		OGA+HDHQ+Trim	997	3	0	0	0	1000	1.215
		OGA+HDBIC	1000	0	0	0	0	1000	1.217
		OGA+HDBIC+Trim	1000	0	0	0	0	1000	1.217
		OGA+BIC	0	0	0	0	1000	1000	1.920

In Simulation 2, the background noise presents with mean value and standard deviation $(\mu, \sigma) = (0.15, 0.1)$. Results are shown in Table 4.4. HDBIC still converges to the smallest set of input neurons as n becomes large enough while performs not well when n is small. HDHQ+Trim performs relatively well when n is small, selecting the correct set of neurons in 969 simulations. However, it tends to include $1-2$ irrelevant neurons at all n since the target neuron in this simulation is not completely triggered by the five input neurons but also by the unmeasurable background noise. Nevertheless, compared to BIC, including $1-2$ irrelevant neurons among a large set of neurons is still acceptable and good enough in practical use.

Table 4.4: Results of Simulation 2 in threshold spiking neuron model (see notation in Table 4.1). The associated parameter settings are: $\lambda = 40$ Hz, $\eta = 0$ Hz, $\mu = 0.15$, $\sigma = 0.1$.

n	p	Method	E	E_1	E_2	$E_{3\sim5}$	E^*	Correct	MSPE
100	750	OGA+HDHQ	770	176	22	1	0	969	1.285
		OGA+HDHQ+Trim	811	153	5	0	0	969	1.279
		OGA+HDBIC	649	14	0	0	0	663	1.937
		OGA+HDBIC+Trim	656	3	0	0	0	663	1.934
		OGA+BIC	0	0	0	0	997	997	2.469
200	1500	OGA+HDHQ	728	263	9	0	0	1000	1.241
		OGA+HDHQ+Trim	728	264	8	0	0	1000	1.241
		OGA+HDBIC	997	3	0	0	0	1000	1.227
		OGA+HDBIC+Trim	997	3	0	0	0	1000	1.227
		OGA+BIC	0	0	0	0	1000	1000	2.128
400	2250	OGA+HDHQ	409	567	24	0	0	1000	1.331
		OGA+HDHQ+Trim	409	567	18	0	0	1000	1.330
		OGA+HDBIC	994	6	0	0	0	1000	1.299
		OGA+HDBIC+Trim	994	6	0	0	0	1000	1.299
		OGA+BIC	0	0	0	0	1000	1000	2.023

In Simulation 3, we set further $(\lambda, \eta) = (30, 10)$ Hz, making the firing behaviors of the candidate input neurons similar; in other words, we make the candidate input variables correlated to each other. Here, the background noise with $(\mu, \sigma) = (0.15, 0.1)$ still keeps. Table 4.5 shows that, compared to Simulation 2, more sample data are needed to ensure good performance. Because the input variables behave similarly, more information (i.e., large sample size) are needed for the method to distinguish them using the associated behavior of the target neuron. For practical use, if neurons from some brain area behave very similarly, then longer recordings will be needed to correctly identify relevant input neurons among them.

Finally, in simulation 4, we consider two extra indirect inputs in the original network (Figure 4.3). Neurons 6 and 7 are simulated as independent Poisson processes with firing rate 40 Hz. The point processes of Neurons 1 and 2 are designed as time shifted version of Neurons 6 and 7 respectively with i.i.d. $N(0.5, 0.1)$ propagation delay on each spike

Table 4.5: Results of Simulation 3 in threshold spiking neuron model (see notation in Table 4.1). The associated parameter settings are: $\lambda = 30$ Hz, $\eta = 10$ Hz, $\mu = 0.15$, $\sigma = 0.1$.

n	p	Method	E	E_1	E_2	$E_{3\sim5}$	E^*	Correct	MSPE
200	750	OGA+HDHQ	768	225	4	0	0	997	1.347
		OGA+HDHQ+Trim	769	226	2	0	0	997	1.347
		OGA+HDBIC	918	3	0	0	0	921	1.359
		OGA+HDBIC+Trim	915	2	0	0	0	917	1.358
		OGA+BIC	0	0	0	0	1000	1000	2.188
400	1500	OGA+HDHQ	473	508	19	0	0	1000	1.376
		OGA+HDHQ+Trim	473	512	15	0	0	1000	1.376
		OGA+HDBIC	989	11	0	0	0	1000	1.363
		OGA+HDBIC+Trim	989	11	0	0	0	1000	1.363
		OGA+BIC	0	0	0	0	1000	1000	2.080

event [61]. The two indirect inputs are considered as very strong and excitatory so that Neurons 1 and 2 have very similar spiking behavior to that of Neurons 6 and 7. The simulated firing trajectories of Neurons 1 and 6 is shown in Figure 4.4 and the simulation result is shown in Table 4.6. Comparing Table 4.6 with Table 4.4, we can find that, for small sample size ($n = 100$), there are much more mistakes made in Simulation 4 since the method has not enough information to distinguish Neuron 1 from Neuron 6 and distinguish Neuron 2 from Neuron 7. However, increasing sample size n rapidly alleviates the mistakes and converges nearly to the true set of input neurons.
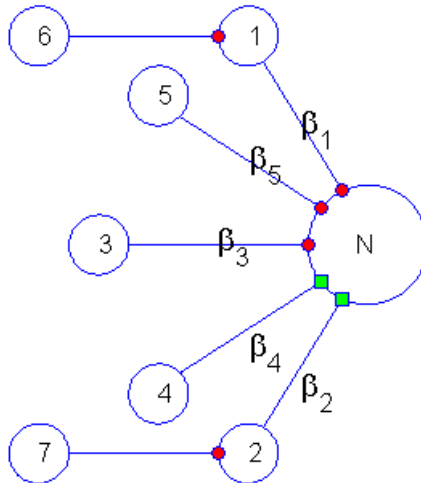
Figure 4.3: Two extra indirect inputs added in the original network. Neurons 6 and 7 are simulated as independent Poisson processes with firing rate 40 Hz. The point processes of Neurons 1 and 2 are designed as time shifted version of Neurons 6 and 7 respectively with i.i.d. $N(0.5, 0.1)$ propagation delay on each spike event.
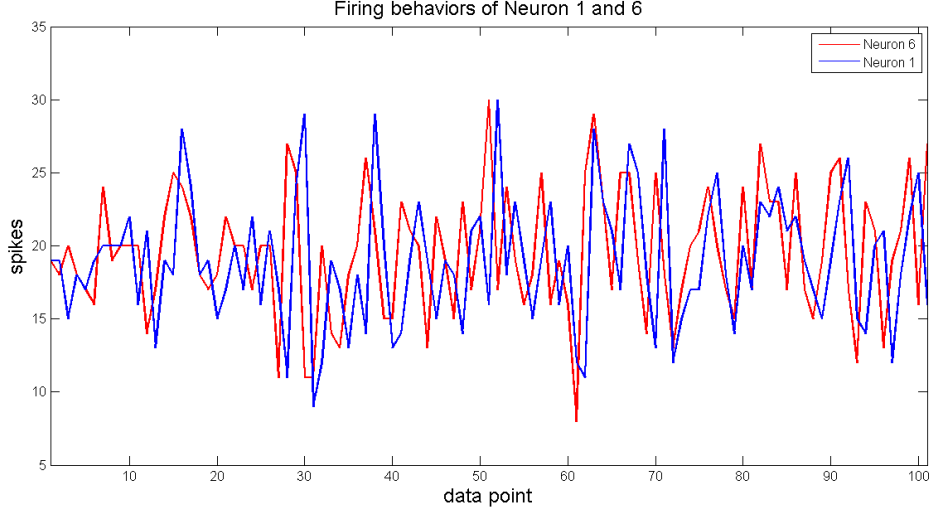
Figure 4.4: The firing behaviors of Neurons 1 and 6. The indirect input (Neuron 6) is considered as very strong and excitatory so that Neuron 1 has a very similar spiking behavior to that of Neurons 6.

Table 4.6: Results of Simulation 4 in threshold spiking neuron model with two extra indirect inputs (see notation in Table 4.1). The associated parameter settings are: $\lambda = 40$ Hz, $\eta = 0$ Hz, $\mu = 0.15$, $\sigma = 0.1$.

n	p	Method	E	E_1	E_2	$E_{3 \sim 5}$	E^*	Correct	MSPE
100	750	OGA+HDHQ	661	144	11	1	0	817	1.427
		OGA+HDHQ+Trim	684	130	3	0	0	817	1.434
		OGA+HDBIC	555	10	0	0	0	565	1.955
		OGA+HDBIC+Trim	558	0	0	0	0	558	1.980
		OGA+BIC	0	0	0	0	875	854	2.616
200	1500	OGA+HDHQ	736	210	3	1	0	950	1.301
		OGA+HDHQ+Trim	737	210	3	0	0	950	1.301
		OGA+HDBIC	946	1	0	0	0	947	1.290
		OGA+HDBIC+Trim	946	1	0	0	0	947	1.290
		OGA+BIC	0	0	0	0	957	957	2.191
400	2250	OGA+HDHQ	451	527	20	0	0	998	1.233
		OGA+HDHQ+Trim	452	531	15	0	0	998	1.234
		OGA+HDBIC	993	4	0	0	0	997	1.241
		OGA+HDBIC+Trim	993	4	0	0	0	997	1.241
		OGA+BIC	0	0	0	0	998	998	1.950

4.3.2 Real data analysis

In this section, we illustrate the application of the proposed method to real spike train data.

59

Experimental setup

• Subjects

All animal procedures were approved by the Institutional Animal Care and Use Committee of National Taiwan University and adhered to the guidelines established by the Codes for Experimental Use of Animals from the Council of Agriculture, Taiwan.

Experiments were conducted on adult female Long-Evans rats weighing 200300 g. At the commence of this study, all subjects were trained to open the gate of a restraining trap. Briefly, rats were acclimated to an arena (50 50 43 cm) for 1 hour each in the first two days. Afterwards rats were fasted overnight and on the next day placed into the arena containing an opened restrainer with rat chow inside. Rats were shaped to attain the skill of gate-opening on the following two days. Rats capable of opening the gate without any help from the experimenter more than ten times in two successive days were defined as an opener. A total of 13 rats completed the experimental task.

• Surgery and recording method

The openers were implanted with stainless steel microwire electrodes into their anterior cingulate cortex (ACC), insular cortex (InC) or primary motor cortex (MI). Two 8-channel microwire array electrodes were implanted in each rat. Briefly, 8 stainless-steel wires individually insulated with Teflon (50 μm OD) were line up linearly with equal inter-electrode distance and a total width of 2.5 mm. Small longitudinal holes were opened in the fronto-parietal bone for implantation into the ACC or InC. The coordinates of the ACC were $1.5-3.5$ mm anterior to and $0.6-0.8$ mm lateral to the bregma, and $1.6-2.0$ mm deep in the cortex. The coordinates of the InC were $1.5-3.5$ mm anterior to and $3.0-5.0$ mm lateral to the bregma, and $4.5-5.0$ mm deep from the surface of the cortex. The coordinates of the MI were $1-3$ mm anterior to and 3 mm lateral to the bregma, and $1.6-2.0$ mm deep in the cortex. Once the electrodes of MI were in the target site, electrical stimulation was employed to ascertain their motor fields. Electrical pulses were delivered from a constant current stimulator (AM system, model 2100) consisting of a train of 7 square-wave pulses, each 0.2 ms in duration, 300 Hz in 100 ms train duration. Intensities of the test electrical stimulation ranged from $30-300$ μA. This stimulation evoked movements of muscle in whisker (22.2 %), neck (33.3 %), or upper limb (33.3 %). No overt body movement could be discerned in the remaining 11% sites of stimulation.

Six rats had implants in the ACC and InC, four rats had implants in ACC and MI, and three rats had implants in InC and MI. A pair of stainless-steel screws (1 mm OD) was placed in the skull bilaterally, 2 mm posterior and 2 mm lateral to bregma for EEG recording. The ground electrode was a stainless-steel screw located over the top of the cerebellum (mid-occipital bone). In addition, several stainless-steel screws were placed in the frontal and parietal bones for anchoring. A pair of seven-stranded stainless-steel wires (793200, A-M systems) was inserted into the neck muscles for EMG recording. After implantation, the holes in the skull and the implanted electrodes were sealed and secured with dental cement.

• Data acquisition

Neuronal activity was recorded using a Multi-channel Neuronal Acquisition Processor

system (MNAP, Plexon, Dallas, TX). The electrical signals were passed from the headset to an amplifier and band-pass filtered (spike signals: $154-13$ kHz, gain: 10,000- to 32,000-fold; EEG and EMG filters: $0.7-170$ Hz, gain: 5000- to 10,000-fold) displayed on an oscilloscope and an audio monitor (Grass AM8). Real-time spike sorting was controlled by SortClient (Plexon), and the sampling rate of individual channels was 40 kHz. Synchronized video signals were acquired through CinePlex (Plexon).

● **Experimental protocol**

After recovery from the surgery for 5 days, the implanted rat was adapted with the telemetry sensor (TBSI, W016020H07K1A) and habituated in the testing arena for the rescuing task. On the testing day, we recorded behavior by videotape and neuronal activity by telemetry of the freely moving rat in the arena. In an experimental session, the implanted rat was acclimated to the test arena for 30 minutes and then a restrainer with a trapped rat was placed into the center of the arena. Control sessions included testing an implanted rat with a restrainer either being empty or containing a toy rat. Each type of session contained ten trials and the three types of trials (30 in total) were randomized in order.

● **Pro-social behaviors in the rescuing task**

Throughout the training and testing sessions, 13 female LE rats completed the 2 experiment tasks after implantation of cortical recording electrode arrays. On the testing day of the rescuing task, each rat was presented with 30 trials of the rescuing test including 10 trials respectively for each of the following three conditions, namely, a restrainer containing another rat, a toy rat, or no rat inside.

Experimental results

A total of 288 single units were recorded from the 13 rats. Among them, 107 units were in the ACC from 10 rats, 82 units in the InC from 9 rats and 82 units in the MI from 7 rats. For each unit, the 5-second spike trains prior to the onset of gate-opening of each trial were aligned and superimposed to obtain a single enhanced 5-second spike train, which is then transformed into a 50-sample time series (through binning with bin size 100 ms.) for causal analysis. For each rat, OGA+HDHQ was used to analyze the causal relationship among all its recorded uints through the time-series signals. The time-lag parameter m was fixed at 5 for all rats and the NSIs were computed once the relevant input and output neurons were identified. Through the method, we have figured out, for each rat and condition, some special groups of neurons (from ACC, InC, or MI) conveyed their neuronal information to some other groups of neurons (from ACC, InC, or MI) before the rat perform the gate-opening act.

In order to investigate the functional significance of the flow information for the selected neurons, we examined the correlation between neuronal activity and gate-opening behavior using the latter as a trigger event. The averaging binned values of selected units with congruent flow direction were transformed into Z scores according to the mean and SD values of the baseline period from 5 s to 10 s prior to the onset of gate-opening. We found enhancement of activities in projecting neurons of InC prior to the gate opening

for a rat (Figure 4.5E) but not a restrainer either being empty or containing a toy rat (Figure 4.6E). Such enhancements also found in neurons of MI projected to ACC and InC (Figure 4.5G and 4.5H). Coherently, these changes were absent when the rat was opening the restrainer without a real rat (Figure 4.6G and 4.6H). Hence, we suggest that the opening acts executed by the rat for another rat may conducted by InC and contained the affective component.

From the view of the recipient units, we also found that the activities of the recipient units in the MI were increased after the opening acts and sustained for more than 5 seconds (Figure 4.7G to 4.7I). These enhancements were correlated with the sniffing and social approaching behaviors of the recorded rat toward another relief rat. These results also suggested that the social behaviors of the subject were highly related to the strengthen limbic-motor connections.
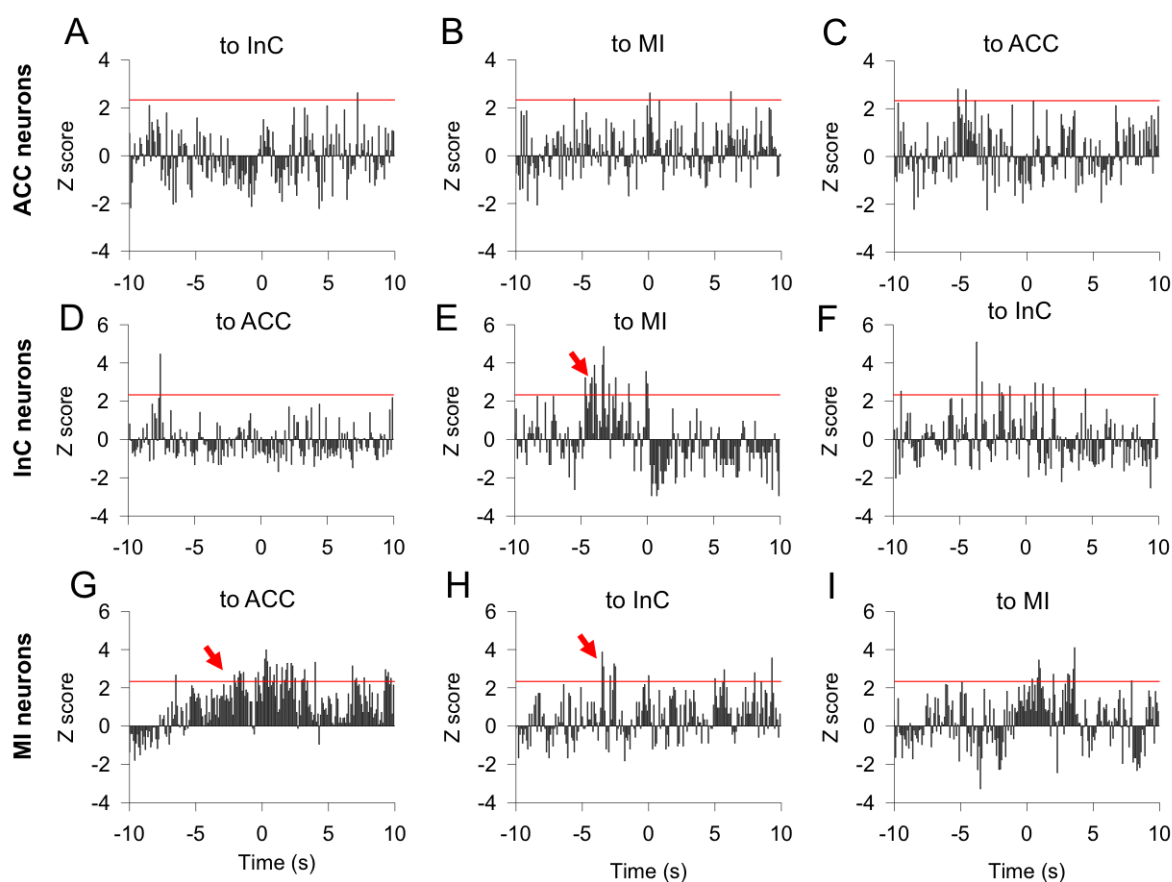
Figure 4.5: Average all the projected ACC, InC and MI unit activity changes when the rat was opening the gate for a conspecific. Values in the Y-axis are normalized Z-score calculated from ACC neurons which projected to InC (A, $n = 8$), MI (B, $n = 5$), and ACC (C, $n = 17$) units; InC neurons which projected to ACC (D, $n = 5$), MI (E, $n = 4$), and InC (F, $n = 15$) units; MI neurons which projected to ACC (G, $n = 4$), InC (H, $n = 5$), and MI (I, $n = 15$) units. The red line represents the 99 % confidence interval. The red arrows above the histograms point to the periods of the activity exceeded mean $+2.33$ SD with 2 consecutive bins. Bin size = 100 ms.
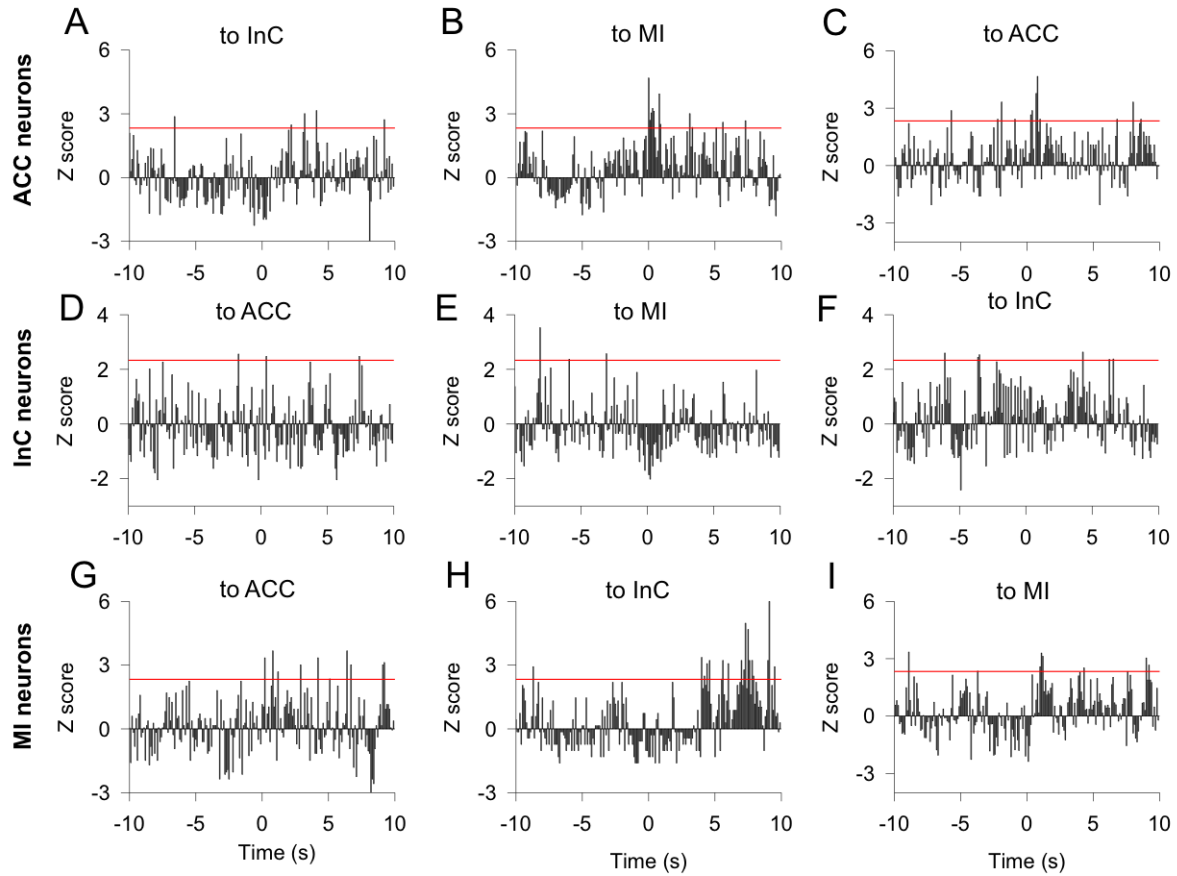
Figure 4.6: Average all the projected ACC, InC and MI unit activity changes when the rat was opening the gate for a toy rat or nothing inside the box. Values in the Y-axis are normalized Z-score calculated from ACC neurons which projected to InC (A, $n = 7$), MI (B, $n = 10$), and ACC (C, $n = 4$) units; InC neurons which projected to ACC (D, $n = 7$), MI (E, $n = 5$), and InC (F, $n = 11$) units; MI neurons which projected to ACC (G, $n = 6$), InC (H, $n = 2$), and MI (I, $n = 12$) units. The red line represents the 99 % confidence interval. Bin size = 100 ms.
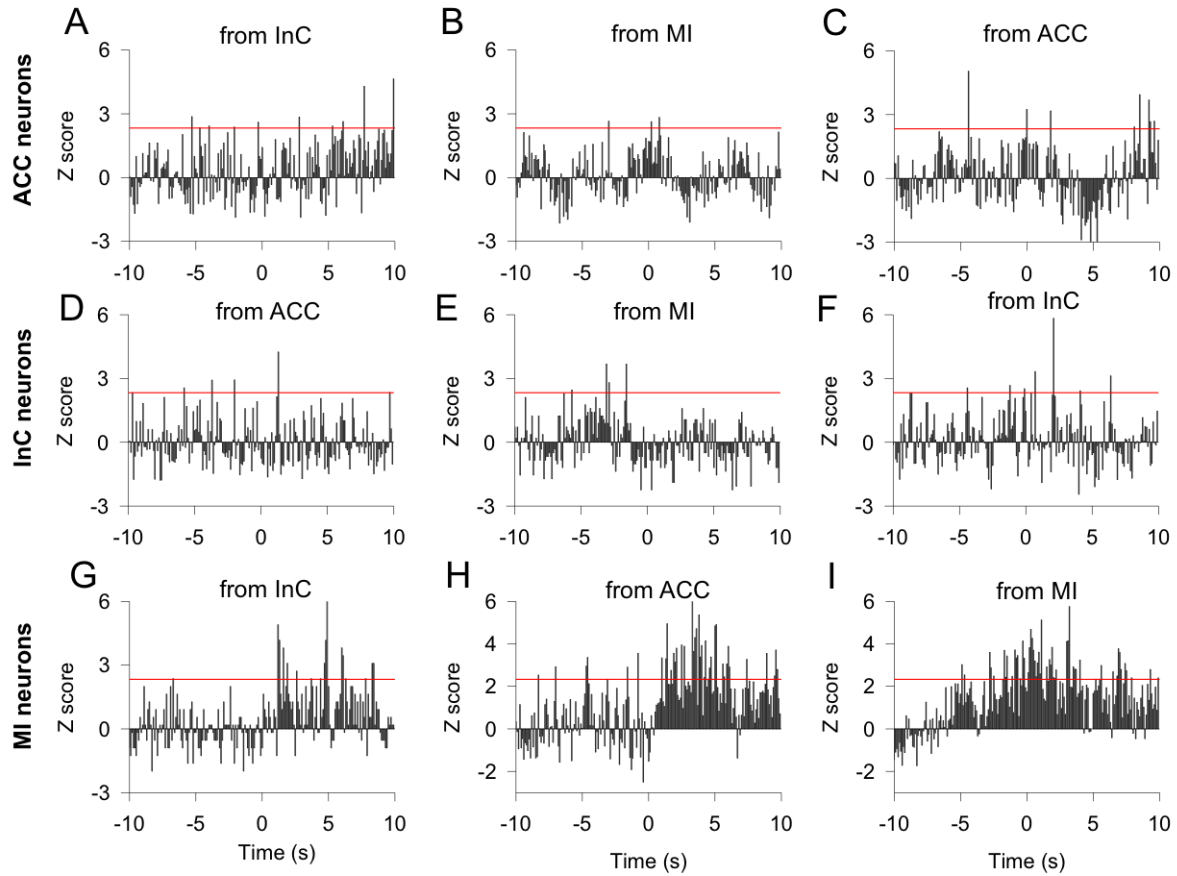
Figure 4.7: Average all the recipient unit activity changes in the ACC, InC and MI when the rat was opening the gate for a conspecific. Values in the Y-axis are normalized Z-score calculated from ACC neurons which received projection from InC (A, $n = 5$), MI (B, $n = 5$), and ACC (C, $n = 19$) units; InC neurons which received projection from ACC (D, $n = 9$), MI (E, $n = 7$), and InC (F, $n = 18$) units; MI neurons which received projection from InC (G, $n = 3$), ACC (H, $n = 6$), and MI (I, $n = 18$) units. The red line represents the 99 % confidence interval. Bin size = 100 ms.
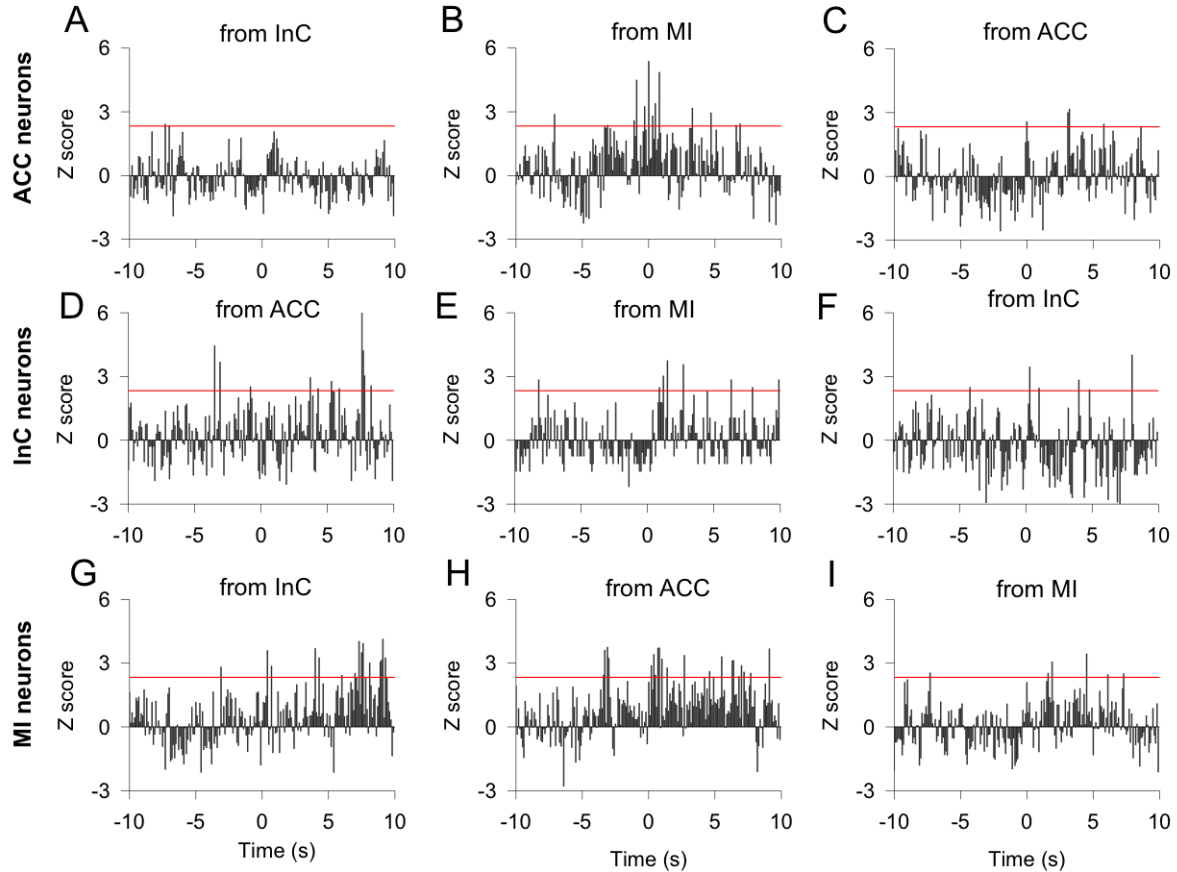
Figure 4.8: Average all the recipient unit activity changes in the ACC, InC and MI when the rat was opening the gate for a toy rat or nothing inside the box. Values in the Y-axis are normalized Z-score calculated from ACC neurons which received projection from InC (A, $n = 8$), MI (B, $n = 6$), and ACC (C, $n = 8$) units; InC neurons which received projection from ACC (D, $n = 7$), MI (E, $n = 2$), and InC (F, $n = 14$) units; MI neurons which received projection from InC (G, $n = 4$), ACC (H, $n = 9$), and MI (I, $n = 12$) units. The red line represents the 99 % confidence interval. Bin size = 100 ms.

4.4   Discussion

At the commence of the experiment, all rats were trained to open the gate of a restraining trap. Once the rat learned the opening skill, it was then implanted with microwire electrodes into the ACC, InC or MI and recovered for another week. A total of 13 rats completed the experimental task. Six rats had implants in the ACC and InC, four rats had implants in ACC and MI, and three rats had implants in InC and MI. On the testing day, a telemetry recording headpiece was connected to the microwire electrode setup and we proceeded to record the behavior and the cortical neuronal activities throughout the experiment. Each rat was presented with 30 trials of the rescuing test including 10 trials respectively for each of the following three conditions, namely, a restrainer containing another rat, a toy rat, or no rat inside.

Evidence from human studies implicates that the ACC and InC are involved in empathy [23, 36, 46, 64, 70]. According to our previous findings, neurons in the ACC and InC showed activity specifically increased prior to rescuing acts and relate to empathic driven prosocial behaviors (Wu et al., 2016). However, the neuronal flow direction and causal information remain unclear. Therefore, we utilize the proposed method to assess this issue. By means of the method, we discover causal relationship among ACC, InC and MI under the empathy-like behavior task. The results show that the activity of the projecting neurons in InC connected with MI was increased significantly and specifically prior to the execution of prosocial-rescuing behaviors. These findings provide an important evidence to support that the empathy-related neurons in InC would convey information to MI to trigger the prosocial behaviors.

To realize the dynamics of neural circuitry under prosocial acts is crucial for elucidating underlying empathic mechanisms. Our data demonstrated that the means by which observing another in trap engages empathically motivated helping behaviors may relate to the increased activity of the neurons in the InC which projected to the MI. In view of the idea that the InC integrates the endogenous and exogenous information of self [17], our results would imply a greater role of the InC in perception of emotions of the others without confusion between self and others, which is an important characteristic of the empathy. In light of our findings, the InC may serve the affective-motivational component, i.e., the perspective and evaluation of subjective discomfort and response to trigger the specific prosocial acts through the neurons in the MI.

Now, back to the mathematical part of the proposed method that should be noticed. For greedy-like algorithms, the most crucial component lies in when should we stop the iterative procedure. As Table 4.3 to Table 4.6 in Section 4.3.1 show, HDHQ performs quite satisfactorily even when the sample size is small and only includes $1 - 2$ irrelevant neurons when the behavior of target neurons can not be fully explained due to the noise. On the other hand, the HDBIC performs well only when the sample size is large enough and is too conservative when the sample size is samll. For practical use, we suggest that HDHQ can be the first stopping criterion to be used. One can also resort to the HDBIC if the information in hand is quite rich (e.g., the sample size $n > 200$). However, the HDHQ can play a more important role than the HDBIC since the sample size is often limited in reality.

Appendix A

Derivation of the Explicit Formula

We first denote $x_k = x(n-k)$ and $y_k = y(n-k)$ for convenience. Then for the model in (2.1), we compute matrices $\mathbf{A}$ and $\boldsymbol{\Sigma}$ by the method of Yule-Walker [54]. Since x and y are stationary, multiply (2.1) from the right by the vector $\begin{bmatrix} x(n-1) & y(n-1) \end{bmatrix}$ and then take the expectation E, we have $\mathbf{R}(-1) = \mathbf{A}\mathbf{R}(0)$, where

$$\mathbf{R}(0) = \begin{bmatrix} E(x_1^2) & E(x_1 y_1) \\ E(x_1 y_1) & E(y_1^2) \end{bmatrix} \text{ and } \mathbf{R}(-1) = \begin{bmatrix} E(x_1 x_2) & E(x_2 y_1) \\ E(x_1 y_2) & E(y_1 y_2) \end{bmatrix}.$$

Thus, we get $\mathbf{A} = \mathbf{R}(-1)\mathbf{R}^{-1}(0)$. Alternatively, $\boldsymbol{\Sigma}$ can be obtained by $\boldsymbol{\Sigma} = \mathbf{R}(0) - \mathbf{A}\mathbf{R}^\top(-1)$ [68]. Substituting $\mathbf{A}$ into $\boldsymbol{\Sigma}$ gives

$$\boldsymbol{\Sigma} = \mathbf{R}(0) - \mathbf{R}(-1)\mathbf{R}^{-1}(0)\mathbf{R}^\top(-1). \tag{A.1}$$

Using the same computation, we have $\tilde{\mathbf{A}} = \tilde{\mathbf{R}}(-1)\tilde{\mathbf{R}}^{-1}(0)$ and $\tilde{\boldsymbol{\Sigma}} = \tilde{\mathbf{R}}(0) - \tilde{\mathbf{A}}\tilde{\mathbf{R}}^\top(-1)$ for the perturbed model in (2.4), where

$$\tilde{\mathbf{R}}(0) = \begin{bmatrix} E(x_1^2 + \delta x_1^2 + 2x_1 \delta x_1) & E(x_1 y_1 + y_1 \delta x_1) \\ E(x_1 y_1 + y_1 \delta x_1) & E(y_1^2) \end{bmatrix}$$

and

$$\tilde{\mathbf{R}}(-1) = \begin{bmatrix} E((x_1 + \delta x_1)(x_2 + \delta x_2)) & E(x_2 y_1 + y_1 \delta x_2) \\ E(x_1 y_2 + y_2 \delta x_1) & E(y_1 y_2) \end{bmatrix}.$$

Substituting $\tilde{\mathbf{A}}$ into $\tilde{\boldsymbol{\Sigma}}$ also gives

$$\tilde{\boldsymbol{\Sigma}} = \tilde{\mathbf{R}}(0) - \tilde{\mathbf{R}}(-1)\tilde{\mathbf{R}}^{-1}(0)\tilde{\mathbf{R}}^\top(-1). \tag{A.2}$$

Using (A.1) and (A.2), and denoting $\delta\mathbf{R}(0) := \tilde{\mathbf{R}}(0) - \mathbf{R}(0)$ and $\delta\mathbf{R}(-1) := \tilde{\mathbf{R}}(-1) - \mathbf{R}(-1)$, it follows that

$$\begin{aligned} \boldsymbol{\Delta} &:= \tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \\ &= \delta\mathbf{R}(0) - \delta\mathbf{R}(-1)\mathbf{R}^{-1}(0)\mathbf{R}(1) \\ &\quad - \tilde{\mathbf{R}}(-1)\mathbf{R}^{-1}(0)\delta\mathbf{R}(1) \\ &\quad + \tilde{\mathbf{R}}(-1)\mathbf{R}^{-1}(0)\delta\mathbf{R}(0)\tilde{\mathbf{R}}^{-1}(0)\tilde{\mathbf{R}}(1). \end{aligned} \tag{A.3}$$

By the definition of S and $\tilde{S}$ defined in (2.5) we know that $\tilde{S} - S = \boldsymbol{\Delta}_{2,2}$, the (2,2)-element of matrix $\boldsymbol{\Delta}$. Hence, we can decompose $\tilde{S}$ into $S + \boldsymbol{\Delta}_{2,2}$. Annoying algebraic computation from (A.3) gives $\boldsymbol{\Delta}_{2,2} = (S_y - S)I$, where I is defined in (2.7), and the formula in (2.6) is obtained by denoting $\Theta = (S_y - S)I$.

Appendix B

Derivation of the NSI using Simple Network

Here, we re-formulate the NSI using the simple network (Figure 3.1). Let $u = \alpha x + \beta y + \gamma z$ form the BLP of w, then there exist p, $\{f_r, r = 1, 2, \ldots, p\}$, and $\{d_r, r = 1, 2, \ldots, p\}$ such that $w_t = \sum_{r=1}^{p}[f_r u_{t-r} + d_r w_{t-r}] + \epsilon_t$, where ϵ is a stationary white noise possessing the smallest variance among $\mathcal{G} = span(\{x, y, z, v_1, v_2, v_3\})$. Replacing u with the weighted trajectory, we obtain

$$
\begin{aligned}
w_t &= \sum_{r=1}^{p}[f_r u_{t-r} + d_r w_{t-r}] + \epsilon_t \\
&= \sum_{r=1}^{p}[f_r(\alpha x_{t-r} + \beta y_{t-r} + \gamma z_{t-r}) + d_r w_{t-r}] + \epsilon_t \qquad \text{(B.1)} \\
&= \sum_{r=1}^{p}[\alpha f_r x_{t-r} + \beta f_r y_{t-r} + \gamma f_r z_{t-r} + d_r w_{t-r}] + \epsilon_t.
\end{aligned}
$$

On the other hand, fitting to data the following empirical regression

$$
w_t = \sum_{r=1}^{p}[a_r x_{t-r} + b_r y_{t-r} + c_r z_{t-r} + g_r v_{t-r} + d_r w_{t-r}] + \tilde{\epsilon}_t, \qquad \text{(B.2)}
$$

where $g_r v_{t-r} := \sum_{k=1}^{3} g_{k,r} v_{k,t-r}$ for convenience.

- If v is stochastically independent of x, y, z, w, then we have $g_r \equiv 0$. Since $\{a_r\}, \{b_r\}, \{c_r\}$ can be obtained through Least-Squares method, comparing (B.2) with (B.1), we have

$$
\sum_{r=1}^{p} a_r = \alpha \sum_{r=1}^{p} f_r, \ \sum_{r=1}^{p} b_r = \beta \sum_{r=1}^{p} f_r, \ \sum_{r=1}^{p} c_r = \gamma \sum_{r=1}^{p} f_r, \qquad \text{(B.3)}
$$

and get

$$
\alpha : \beta : \gamma = \sum_{r=1}^{p} a_r : \sum_{r=1}^{p} b_r : \sum_{r=1}^{p} c_r, \text{ provided } \sum_{r=1}^{p} f_r > 0, \qquad \text{(B.4)}
$$

where $sgn(\alpha) = sgn(\sum_{r=1}^{p} a_r)$, $sgn(\beta) = sgn(\sum_{r=1}^{p} b_r)$, and $sgn(\gamma) = sgn(\sum_{r=1}^{p} c_r)$.

- If v is linear dependent of x, y, z, w, then $g_r \gg 0$ and $\{a_r\}, \{b_r\}, \{c_r\}$ will be affected. However, since ϵ in (B.1) possessing the smallest variance among $\mathcal{G}$, taking out v does not increase the variance of $\tilde{\epsilon}$, therefore we still can correct the model coefficients by ruling out the useless information v.

Finally, the neuron synaptic index from x, y, z to w are defined respectively as

$$
\begin{aligned}
N_{x \to w} &:= \frac{\alpha}{|\alpha|+|\beta|+|\gamma|} F_{u \to w}, \\
N_{y \to w} &:= \frac{\beta}{|\alpha|+|\beta|+|\gamma|} F_{u \to w}, \\
N_{z \to w} &:= \frac{\gamma}{|\alpha|+|\beta|+|\gamma|} F_{u \to w},
\end{aligned}
\tag{B.5}
$$

where $|N_{x \to w}| + |N_{y \to w}| + |N_{z \to w}| = F_{u \to w}$ is the GC index from the weighted trajectory $u = \alpha x + \beta y + \gamma z$ to the target trajectory w.

Bibliography

[1] Arnold, A., Liu, Y., and Abe, N. (2007). Temporal causal modeling with graphical Granger methods. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 7, 66–75.

[2] Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. A., de Vico, Fallani F., Salinari, S., Ursino, M., Zavaglia, M., Ding, L., Edgar, J. C., Miller, G. A., He, B., and Babiloni, F. (2007). Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *Human Brain Mapping*, 28, 143–157.

[3] Baccala, L. A. and Sameshima, K. (2001). Partial directed coherence : a new concept in neural stucture determination. *Biological Cybernetics*, 84, 463–474.

[4] Barnett, L. and Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, 223, 50–68.

[5] Barrett, A. B. and Barnett, L. (2013). Granger causality is designed to measure effect, not mechanism. *Frontiers in Neuroinformatics*, 7, 1–2.

[6] Benjamini, Y. and Hochberg Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57, 289–300.

[7] Bressler, S. L., Richter, C. G., Chen, Y., and Ding, M. (2007). Cortical functional network organization from autoregressive modeling of local filed potential oscillations. *Statistic in Medicine*, 26, 3875–3885.

[8] Bressler, S. L. and Seth, A. K. (2011). Wiener-Granger Causality: A well established methodology. *NeuroImage*, 58, 323–329.

[9] Brown, E. N., Kass, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience*, 7, 456–461.

[10] Brunel, N. (2000). Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons *Journal of Computational Neuroscience*, 8, 183–208.

[11] Bush, P. and Sejnowski, T. (1996). Inhibition synchronizes sparsely connected cortical neurons within and between columns in realistic network models. *Journal of Computational Neuroscience*, 3, 91–110.

[12] Cadotte, A. J., DeMarse, T. B., He, P., and Ding, M. (2008). Causal measures of structure and plasticity in simulated and living neural networks. *PLoS Computational Biology*, 3, 1–14.

[13] Cadotte, A. J., DeMarse, T. B., Mareci, T. H., Parekh, M. B., Talathi, S. S., Hwang, D. U., Ditto, W. L., Ding, M., and Carney, P. R. (2010). Granger causality relationships between local field potentials in an animal model of temporal lobe epilepsy. *Journal of Neuroscience methods*, 189, 121–129.

[14] Cao, Y., Maran, S. K., Dhamala, M., Jaeger, D., and Heck, D. H. (2012). Behavior-related pauses in simple-spike activity of mouse purkinje cells are linked to spike rate modulation. *Journal of Neuroscience*, 25, 8678–8685.

[15] Cheng, D., Bahadori, M., T., and Liu, Y. (2014). FBLG: A simple and effective approach for temporal dependence discovery from time series data. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 14, 382–391.

[16] Churchland, P. S. and Sejnowski, T. J. (1992). *The computational brain.* The MIT press.

[17] Craig, A. D. (2009). How do you feel–now? The anterior insula and human awareness. *Nat Rev Neurosci*, 10(1), 59–70.

[18] Dhamala, M., Rangarajan, G., and Ding, M. (2008). Analyzing information flow in brain networks with nonparametric Granger causality. *NeuroImage*, 41, 354–362.

[19] Deborah, S., Won, D. Y., and Patrick, D. W. (2003). Effects of spike sorting error on information content in multi-neuron recordings. *Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering*, 3, 618–621.

[20] Ding, M., Chen, Y., and Bressler, S. L. (2006). Granger Causality: Basic Theory and Application to Neuroscience. *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications.* Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. 437–460.

[21] Fallani, F. De Vico, Corazzol, M., Sternberg, J., Wyart, C., and Chavez, M. (2015). Hierarchy of Neural Organization in the Embryonic Spinal Cord: Granger-Causality Graph Analysis of Calcium Imaging Data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23, 333–341.

[22] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties.*Journal of American Statistical Association*, 96, 1348–1360.

[23] Fan, Y., Duncan, N. W., de Greck, M., and Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neurosci Biobehav Rev*, 35(3), 903–911.

[24] Ge, T., Kendrick, K.M., and Feng, J. (2009). A Novel Extended Granger Causal Model Approach Demonstrates Brain Hemispheric Differences during Face Recognition Learning. *PLoS Computational Biology*, 5, 11.

[25] Gomez, L., Budelli, R., Saa, R., Stiber, M., and Segundo, J. P. (2005). Pooled spike trains of correlated presynaptic inputs as realizations of cluster point processes. *Biological Cybernetics*, 92, 110–127.

[26] Gow, D. W., Segawa, J. A., Ahlfors, S. P., Lin, and F. H. (2008). Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. *Neuroimage*, 43, 614–623.

[27] Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–438.

[28] Granger, C. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2, 329–352.

[29] Greene, W. H. (2002). *Econometric Analysis, fifth ed.* Prentice-Hall, Upper Saddle River, NJ.

[30] Guo, S., Seth, A. K., Kendrick, K. M., Zhou, C., and Feng, J. (2008). Partial Granger causality – Eliminating exogenous inputs and latent variables. *Journal of Neuroscience Methods*, 172, 79–93.

[31] Guo, S., Ladroue, C., and Feng, J. (2010). Granger Causality: Theory and Applications. *Frontiers in Computational and Systems Biology.* Springer, London 83–111.

[32] Hu, S., Dai, G., Worrel, G. A., Dai, Q., and Liang, H. (2011). Causality analysis of neural connectivity: critical examnination of existing methods and advances of new methods. *IEEE Transactions on Neural Networks*, 22, 829–844.

[33] Huang, J. J., Yen, C. T., Liu, T. L., Tsao, H. W., Hsu, J. W., and Tsai, M. L. (2013). Effects of dopamine D2 agonist quinpirole on neuronal activity of anterior cingulate cortex and striatum in rats. *Psychopharmacology*, 227, 459–466.

[34] Ing, C. K. and Lai, T. L. (2011). A stepwise regression method and consistent model selection for high-dimensional sparse linear models. *Statistica Sinica*, 21, 1473–1513.

[35] Izhikevich, E. M. (2003). Simple models of spiking neurons. *IEEE Transactions on Neural Networks*, 14, 1569–1572.

[36] Jackson, P. L., Rainville, P., and Decety, J. (2006). To what extent do we share the pain of others? Insight from the neural bases of pain empathy. *Pain*, 125(1-2), 5–9.

[37] Kaminski, M., Ding, M., Truccolo, W. A., and Bressler, S. L. (2001). Evaluating causal relations in neural systems : Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics*, 85, 145–157.

[38] 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.

[39] Kitagawa, G. (2010). *Introduction to Time Series Modeling.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

[40] Krueger, F., Landgraf, S., van der Meer, E., Deshpande, G., and Hu, X. (2011). Effective connectivity of the multiplication network: a functional MRI and multivariate Granger Causality Mapping study. *Human Brain Mapping*, 32, 1419–1431.

[41] Krumin, M. and Shoham, S. (2010). Multivariate autoregressive modeling and Granger causality analysis of multiple spike trains. *Computational Intelligence and Neuroscience* 752428.

[42] Lehky, S. R. (2010). Decoding poisson spike trains by gaussian filtering. *Neural Computation*, 22, 1245–1271.

[43] Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9, R53–R78.

[44] Luo, Q., Lu, W., Cheng, W., Valdes-Sosa, P. A., Wen, X., Ding, M., and Feng, J. (2013). Spatio-temporal Granger causality: A new framework. *NeuroImage*, 79, 241–263.

[45] Luo, Q., Ge, T., Grabenhorst, F., Feng, J., and Rolls, E. T. (2013). Attention-Dependent Modulation of Cortical Taste Circuits Revealed by Granger Causality with Signal-Dependent Noise. *PLoS Computational Biology*, 9, 10.

[46] Morrison, I., Lloyd, D., di Pellegrino, G., and Roberts, N. (2004). Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? *Cogn Affect Behav Neurosci*, 4(2), 270–278.

[47] Lutkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. *Springer*.

[48] Marinazzo, D., Liao, W., Chen, H., and Stramaglia, S. (2011). Nonlinear connectivity by Granger causality. *NeuroImage*, 58, 330-338.

[49] Michailidis, G. and d'Alche-Buc, F. (2013). Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues. *Mathematical Biosciences*, 246, 326–334.

[50] Nageswaran, J. M., Dutt, N., Krichmar, J. L., Nicolau, A., and Veidenbaum, A. V. (2009). A configurable simulation environment for the efficient simulation of large-scale spiking neural networks on graphics processors. *Neural Networks*, 22, 791–800.

[51] Nedungadi, A. G., Rangarajan, G., Jain, N., and Ding M. (2009). Analyzing multiple spike trains with nonparametric granger causality. *Journal of Computational Neuroscience*, 27, 55–64.

[52] Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature Neuroscience*, 17, 594–600.

[53] Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M, Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454, 995–999.

[54] Priestley, M.B. (1994). *Spectral Analysis and Time Series.* London: Academic Press.

[55] Quinn, C. J., Coleman, T. P., Kiyavash, N., and Hatsopoulos, N. G. (2011). Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *Journal of Computational Neuroscience*, 30, 17–44.

[56] Rosenbaum, R., Trousdale, J., and Josic, K. (2011). The effects of pooling on spike train correlations. *Frontiers in Neuroscience*, 5, 1–10.

[57] Sameshima, K. and Baccala, L. A. (1999). Using partial directed coherence to describe neuronal ensemble interactions. *Journal of Neuroscience Methods*, 94, 93–103.

[58] Seth, A. K. (2005). Causal connectivity of evolved neural networks during behavior. *Network*, 16, 35–54.

[59] Seth, A. K. and Edelman, G. M. (2007). Distinguishing causal interactions in neural populations. *Neural Computation*, 19, 910–933.

[60] Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*, 186, 262–273.

[61] Shao, P. C., Tseng, W. T., Kuo, C. C., Shann, W. C., Tsai, M. L., and Yen, C. C. (2013). Effects of spike sorting error on the Granger causality index. *Neural Networks*, 46, 249–259.

[62] Shao, P. C., Huang, J. J., Shann, W. C., Yen, C. T., Tsai, M. L., and Yen, C. C. (2015). Granger causality-based synaptic weights estimation for analyzing neuronal networks. *Journal of Computational Neuroscience*, 38, 483–497.

[63] Shimazaki, H. and Shinomoto, S. (2007). A method for selecting the bin size of a time histogram. *Neural Computation*, 19, 1503–1527.

[64] Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), 1157–1162.

[65] Takeda, M. and Matsumoto S. (1997). Discharge patterns of dorsal and ventral respiratory group neurons during spontaneous augmented breaths observed in pentobarbital anesthetized rats. *Brain Res.*, 749(1), 95–100.

[66] Tripp, B. P. and Orchard, J. (2012). Population coding in sparsely connected networks of noisy neurons. *Frontiers in Computational Neuroscience*, 6, 23.

[67] Tseng, W. T., Tsai, M. L., Iwata, K., and Yen, C. T. (2012). Long-Term Changes in Trigeminal Ganglionic and Thalamic Neuronal Activities following Inferior Alveolar Nerve Transection in Behaving Rats. *Journal of Neuroscience*, 32, 16051–16063.

[68] Wei, William W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods, second edition.* California: Addison-Wesley. 391–392.

[69] Wiener, N. (1956). The theory of prediction. *In Modern Mathematics for Engineers, vol. 1.* New York: McGraw-Hill.

[70] Zaki, J., Ochsner, K. N., Hanelin, J., Wager, T. D., and Mackey, S. C. (2007). Different circuits for different pain: patterns of functional connectivity reveal distinct networks for processing pain in self and others. *Soc Neurosci*, 2(3-4), 276–291.

[71] Zhang, L., Chen, G., Niu, R., Wei, W., Ma, X., Xu, J., Wang, J., Wang, Z., and Lin, L. (2012). Hippocampal theta-driving cells revealed by Granger Causality. *Hippocampus*, 8, 1781–1793.

[72] Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7, 2541–2563.

[73] Zhou, Z., Ding, M., Chen, Y., Wright, P., Lu, Z., and Liu, Y. (2009). Detecting directional influence in fMRI connectivity analysis using PCA based Granger causality. *Brain Research*, 1289, 22–29.

[74] Zhou, D., Xiao, Y., Zhang, Y., Xu, Z., and Cai, D. (2014). Granger causality network reconstruction of conductance-based integrate-and-fire neuronal systems. *PLoS ONE*, 9, 2.

[75] Zhu, L., Lai, Y. C., Hoppensteadt, F. C., and He. J. (2003). Probing changes in neural interaction during adaptation. *Neural Computation*, 15, 2359–2377.

[76] Zou, C., Ladroue, C., Guo, S., and Feng, J. (2010). Identifying interactions in the time and frequency domains in local and global networks - A Granger causality approach. *BMC Bioinformatics*, 11, 337.