Detecting causality from observations of dynamic variables

A thesis submitted by

Daniel Rosebrock

in partial fulfillment of the requirements for the degree of

Master of Science in Mathematics

Tufts University

Graduation month and year: May 2015

Adviser: Professor Christoph Börgers

Committee: Professor Bruce Boghosian and Professor Zbigniew Nitecki

Abstract

"Granger causality" is a methodology, first proposed in 1969 by Clive Granger, for detecting causal relationships between two stochastically fluctuating quantities. I will present several equivalent ways of describing this methodology. In Granger's original work, it was assumed that the stochastic quantities are discretized Ornstein-Uhlenbeck processes, that is, AR(1) processes. Thus the deterministic piece of the dynamics is simply relaxation into a stable steady state. If the deterministic dynamics are more complicated, Granger causality is not applicable, but the idea of one of the equivalent versions that I present has a relatively straightforward generalization. I present numerical experiments testing the application of this idea to noisy oscillators that are coupled in one direction or the other, with the goal of deducing the direction of the coupling from observed solutions.

Contents

1 Introduction

Causality can be understood as a relation between one event (a cause) and another event (an effect). The cause must precede the effect, and must have some explanatory power in determining why and how the effect occurred. There are many variations on how to precisely define and determine causality, and there are in fact many different types of causality. For example, if an event A causes an event B in the deterministic sense, then event A is always followed by event B. Consider the following: When you press a key on a piano, a hammer will strike some strings in the body of the instrument, producing a sound which we interpret as a note on the piano. The physical event of pressing the key (cause) deterministically produced the occurrence of the sound (effect) which you hear as a result of the cause. Another type of causality is probabilistic causality. In the probabilistic sense of causality, if an event A probabilistically causes an event B, then the occurrence of event A will increase the probability of the occurrence of event B. For example, smoking is a probabilistic cause for cancer. Not every person who smokes will get lung cancer, but smoking increases the probability that a person will get lung cancer later in life. In most physical situations, probabilistic causality reflects improper knowledge of a deterministic system.

Identifying causal relationships between dynamical systems is a fundamental problem in the natural sciences. In fact, many physical laws are mathematical relationships describing causal interactions between different variables. It would seem that a natural question to ask is, can one determine a causal relationship between two dynamic variables (i.e., two time-dependent variables) from observations? This is precisely the question which the notion of Granger causality is designed to answer in a very simple case.

The idea of prescribing a statistical test to determine causality between two dynamical systems was introduced by Wiener in 1956 [1]. He labeled this type of causality "computationally measurable" and described it as follows: "For two simultaneously measured

signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one." This concept was made more rigorous by the British economist Clive W. Granger in the 1960s [2], in which Granger developed a statistical test for testing for causality between two time series. This method is called "Granger causality", as it only provides evidence for causality, but does not prove the existence of a causal relationship. Granger proposed the following test for causality.

Definition 1.1. Let X and Y be two discrete-time stochastic processes. Suppose we try to predict Y_{k+1} using only the past of Y, and then we try to predict Y_{k+1} using both the past of X and Y. If the second prediction is significantly more successful than the first, than we say that X *Granger causes* Y. "Significantly more successful", of course, needs an explanation, to be given later.

Typically, we are given a set of observations from two discrete-time stochastic processes, X and Y, and we want to determine if there exists a causal relationship between these two processes. We have the 4 following possible scenarios: X G-causes Y, Y G-causes X, there is no G-causation in either direction, or there is G-causation in both directions. The last scenario, where there is mutual causation is known as a feed-back stochastic system. In Granger causality analysis, we have to test for all of these scenarios. Typically, one can set up two hypothesis tests to test for bivariate Granger causality:

$$H_0 : X \text{ does not G-cause } Y, \qquad H_A : X \text{ G-causes } Y.$$

$$H_0 : Y \text{ does not G-cause } X, \qquad H_A : Y \text{ G-causes } X.$$

What Granger causality does not account for are confounding variables. For example, suppose there is a third variable, Z, which influences X and Y. If we do not have any

information about Z, then we are unable to measure the effects of Z on X and Y. One pitfall of Granger causality analysis is that we may falsely conclude that X G-causes Y, when in reality, Z influences X, and shortly thereafter influences Y.

Granger originally proposed his concept of Granger causality in the context of discrete-time stochastic processes. We begin with the concept of causality in a deterministic continuous time dynamical system. Then, we discuss the discrete-time stochastic processes in which Granger first proposed Granger analysis. Finally, we add noise to continuous time dynamical systems, specifically phase-locking oscillators, and look at how the idea of Granger causality can be applied in this more general setting.

2 Detecting Causality in Deterministic Systems

2.1 First Order Linear System

Let $A \in \mathbb{R}^{2\times 2}$ and consider the dynamical system, $\mathbf{x}(t) \in \mathbb{R}^2$ with components x and y, defined by the following differential equation

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{1}$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}.$$

Solutions to this equation will be of the form

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2, \tag{2}$$

where c_1 and c_2 are constants determined by the initial conditions $\mathbf{x}(0) = \mathbf{x}_0$, λ_1 and λ_2 are the eigenvalues of A associated with eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Note that λ_1, λ_2 may be complex, as well as their associated eigenvectors. To see why this is a solution to the above differential equation, simply differentiate both sides of the equation to get

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= c_1 e^{\lambda_1 t}\lambda_1\mathbf{v}_1 + c_2 e^{\lambda_2 t}\lambda_2\mathbf{v}_2 \\ &= c_1 e^{\lambda_1 t}A\mathbf{v}_1 + c_2 e^{\lambda_2 t}A\mathbf{v}_2 \\ &= A\left(c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2\right) \\ &= A\mathbf{x}. \end{aligned}$$

It is clear from Equation (2) that the solution $\mathbf{x}(t)$ will grow exponentially fast if $\mathrm{Re}(\lambda_1) > 0$ or $\mathrm{Re}(\lambda_2) > 0$. The eigenvalues of A, λ_1 and λ_2, will satisfy the equation $\det(A - \lambda I) = 0$.

$$\det(A-\lambda I) = 0 \implies \lambda^2 - \tau\lambda + d = 0 \implies \lambda = \frac{\tau \pm \sqrt{\tau^2 - 4d}}{2},$$

where $\tau = \mathrm{trace}(A)$, $d = \det(A)$. From this, we see that solutions will grow exponentially fast, or will be unstable, if $\tau \pm \sqrt{\tau^2 - 4d} > 0$. Since we are only interested in stable solutions which don't grow exponentially fast, we restrict ourselves to matrices A, for which $\mathrm{trace}(A) < 0$ and $\det(A) > 0$. Here is a sample solution where the fixed point $(0,0)$ is a stable node.

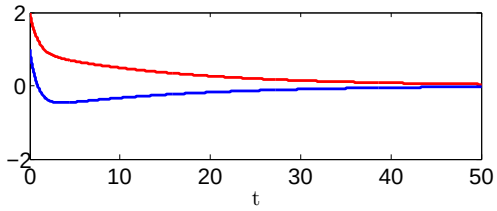

Figure 1: Two components of solution $\mathbf{x}(t)$ satisfying (1)

Suppose we have a solution to this dynamical system, $\{\mathbf{x}_k\}_{k=0}^N$, where $\mathbf{x}_k = \mathbf{x}(\Delta t k)$, and we want to estimate the matrix A from this solution. Pick two points in time $t_1 = \Delta t\, j$ and $t_2 = \Delta t l$, sucht that $t_1 \neq t_2$, and $j, l \in [0, N]$. Next, we can approximate the derivative of $\mathbf{x}$ at times t_1 and t_2 by the following approximations:

$$\mathbf{x}'(t_1) \approx \frac{\mathbf{x}(t_1 + \Delta t) - \mathbf{x}(t_1 - \Delta t)}{2\Delta t} \approx A\mathbf{x}(t_1), \tag{3}$$

$$\mathbf{x}'(t_2) \approx \frac{\mathbf{x}(t_2 + \Delta t) - \mathbf{x}(t_2 - \Delta t)}{2\Delta t} \approx A\mathbf{x}(t_2). \tag{4}$$

Re-writing $\mathbf{x}$ in terms of x and y, we define the 2×2 matrices, B and C to be the following:

$$B = \frac{1}{2\Delta t}\begin{bmatrix} x(t_1+\Delta t)-x(t_1-\Delta t) & x(t_2+\Delta t)-x(t_2-\Delta t) \\ y(t_1+\Delta t)-y(t_1-\Delta t) & y(t_2+\Delta t)-y(t_2-\Delta t) \end{bmatrix}, \tag{5}$$

$$C = \begin{bmatrix} x(t_1) & x(t_2) \\ y(t_1) & y(t_2) \end{bmatrix}. \tag{6}$$

From this, it follows that $B \approx AC$. So,

$$A \approx BC^{-1}. \tag{7}$$

In order to estimate A in this way, we need to ensure that the matrix C has a small condition number. For example, in the previous example, if we pick the times $t_1 = 1$ and $t_2 = 2$, the two vectors $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ are far from dependent. We can plot them in the (x,y)-plane, along with the solution $\mathbf{x}(t)$ in the (x,y)-plane to get the following figure.

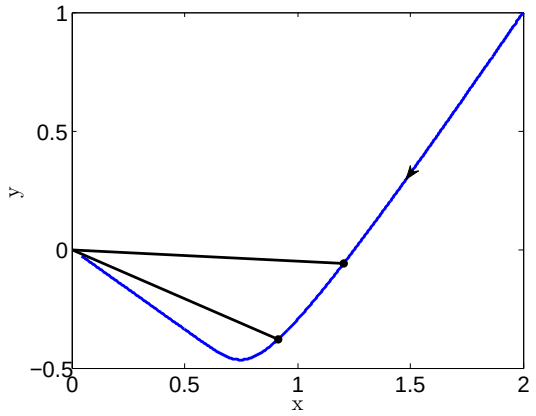

Figure 2: Vectors $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ plotted in the (x,y)-plane

The coefficient matrix for this example was

$$A = \begin{bmatrix} -0.4 & -0.5 \\ -0.5 & -0.8 \end{bmatrix}.$$

Each entry in our estimate for the coefficient matrix using the method described above,

and choosing t_1 and t_2 in a smart way, has an error $O(\Delta t^2)$, since the estimate for $\frac{d\mathbf{x}}{dt}$ has an error of $O(\Delta t^2)$. In fact, using $t_1 = 1$ and $t_2 = 2$, and a step-size of $\Delta t = 0.01$, our estimate for A is accurate to 7 digits in each entry. However, if we use the times $t_1 = 40$ and $t_2 = 41$ to estimate the matrix A, we get the following estimate,

$$A_{est} = \begin{bmatrix} -7.25 & -10.64 \\ 6.25 & 9.17 \end{bmatrix}.$$

This estimate is terrible! The vectors $\mathbf{x}(40)$ and $\mathbf{x}(41)$ are almost linearly dependent, so the condition number for the matrix C is huge, so we cannot invert the matrix accurately, and hence cannot get a good estimate for A using the times $t_1 = 40$ and $t_2 = 41$. This example highlights an interesting point, namely, if we only observe the first order linear system when it is close to a stable node, then we are unable to estimate the matrix A accurately. It is only possible to detect causal interactions between the elements of $\mathbf{x}$ when the system is far enough away from the stable node. Note the issue of ill-conditioning of the matrix C goes away when the fixed point is a spiral.

2.2 Coupled Phase Oscillators

A phase oscillator is defined by the following autonomous ODE

$$\frac{d\theta}{dt} = \omega, \quad \omega > 0. \tag{8}$$

In this equation, θ represents the phase of the oscillator, and ω the phase speed. So, what really oscillates in this model is $\cos(\theta)$ or $\sin(\theta)$. The period of the limit cycle oscillation, the time it takes for the phase oscillator to complete one loop around the unit circle, S^1, is $T = \frac{2\pi}{\omega}$. And so, the frequency of the phase oscillator is $f = \frac{1}{T} = \frac{\omega}{2\pi}$.

We can couple two phase oscillators, θ_1 and θ_2, by adding a coupling term to each phase oscillator which is dependent on the phase difference between the two oscillators.

We define the coupled phase oscillator model as follows:

$$\frac{d\theta_1}{dt} = \omega_1 + c_{12}\sin(\theta_2 - \theta_1), \tag{9}$$
$$\frac{d\theta_2}{dt} = \omega_2 + c_{21}\sin(\theta_1 - \theta_2), \tag{10}$$

where $\omega_1 > 0$, $\omega_2 > 0$, $c_{12} \geq 0$, $c_{21} \geq 0$. Define the *phase difference* between the two phase oscillators to be $\eta := \theta_1 - \theta_2$. So, if the two phase oscillators are phase-locked, then their phase difference is a constant, and $\frac{d\eta}{dt} = 0$. Define $F(\eta) = \frac{d\eta}{dt}$. Then,

$$F(\eta) = \frac{d(\theta_1 - \theta_2)}{dt} = \omega_1 - \omega_2 - (c_{12} + c_{21})\sin\eta.$$

Setting $F(\eta) = 0$, we can solve for the parameters ω_1, ω_2, c_{12}, c_{21}, for which phase-locking is possible.

$$F(\eta) = 0 \iff \omega_1 - \omega_2 - (c_{12} + c_{21})\sin\eta = 0 \iff \sin\eta = \frac{\omega_1 - \omega_2}{c_{12} + c_{21}}.$$

We know $|\sin\eta| \leq 1$. Thus, phase-locking is only possible if

$$\left|\frac{\omega_1 - \omega_2}{c_{12} + c_{21}}\right| \leq 1 \implies \boxed{|\omega_1 - \omega_2| \leq c_{12} + c_{21}}$$

We can determine the stability of the synchronous state by plotting η vs. $F(\eta)$, and observing where $F(\eta)$ crosses the line $F(\eta) = 0$. The synchronous state, or *phase-locked state*, will be stable at values of η where $F(\eta)$ crosses the line $F(\eta) = 0$ with a negative slope.

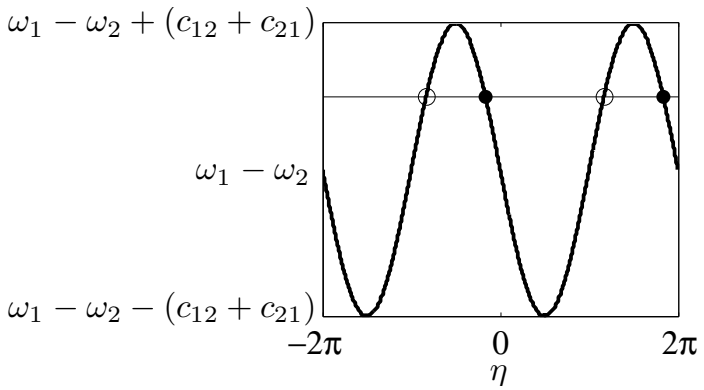

Figure 3: $F(\eta)$ vs. η with stable and unstable fixed points

If we assume $|\omega_1 - \omega_2| \leq c_{12} + c_{21}$ in the coupled phase oscillator model, then we can calculate the exact phase difference of the stable phase-locked state. The phase-locked state occurs at the value of η for which $F(\eta) = 0$ and $F'(\eta) < 0$. We know that $F(\eta) = 0$ iff $\sin\eta = \frac{\omega_1 - \omega_2}{c_{12} + c_{21}}$. Thus, there exists a fixed point at $\eta = \arcsin\left(\frac{\omega_1 - \omega_2}{c_{12} + c_{21}}\right)$. Since we are assuming $|\omega_1 - \omega_2| \leq c_{12} + c_{21}$, we know $-1 \leq \frac{\omega_1 - \omega_2}{c_{12} + c_{21}} \leq 1$, and thus this fixed point lies in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

$$F'(\eta) = \frac{d}{d\eta}\left(\omega_1 - \omega_2 - (c_{12} + c_{21})\sin\eta\right) = -(c_{12} + c_{21})\cos\eta.$$

Thus, $F'(\eta) < 0$ iff $\cos\eta > 0$. Since the fixed point $\arcsin\left(\frac{\omega_1 - \omega_2}{c_{12} + c_{21}}\right) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, this is precisely the stable fixed point of the system, and so the phase locked state will have a phase difference of $\theta_1 - \theta_2 = \arcsin\left(\frac{\omega_1 - \omega_2}{c_{12} + c_{21}}\right)$.

In the phase-locked state, the phase speed of θ_1 and θ_2 are equal, and they are constant. Thus, to calculate the phase speed of the phase-locked state, we just need to calculate the phase speed of either θ_1 or θ_2 when $\theta_1 - \theta_2 = \arcsin\left(\frac{\omega_1 - \omega_2}{c_{12} + c_{21}}\right)$. Calculating $\frac{d\theta_2}{dt}$ at the stable fixed point, we have

$$\begin{aligned}\frac{d\theta_2}{dt} &= \omega_2 + c_{21}\sin\left(\arcsin\left(\frac{\omega_1-\omega_2}{c_{12}+c_{21}}\right)\right)\\ &= \omega_2 + c_{21}\left(\frac{\omega_1-\omega_2}{c_{12}+c_{21}}\right)\\ &= \frac{c_{21}}{c_{12}+c_{21}}\omega_1 + \frac{c_{12}}{c_{12}+c_{21}}\omega_2.\end{aligned}$$

Therefore, the frequency of the phase-locked state will be

$$\frac{1}{2\pi}\left(\frac{c_{21}}{c_{12}+c_{21}}\omega_1 + \frac{c_{12}}{c_{12}+c_{21}}\omega_2\right). \tag{11}$$

We are interested in detecting causality between two phase oscillators from a given set of observations. The motivation behind this question lies at the heart of neuroscience. Neural oscillations are often observed in EEG signals, and it is of great interest to determine if neuronal activity in one brain area influences the activity in another area.

Suppose we are given a set of observations from two phase oscillators, $\{\cos(\theta_k)\}_{k=0}^{N}$, where $\theta_k = [\theta_{1k};\theta_{2k}]'$, and $\theta_k = \theta(\Delta t k)$. Can we determine if θ_1 influences θ_2, or θ_2 influences θ_1? We know that if $c_{12} = 0$, then $\frac{d\theta_1}{dt}$ will be constant for all time t. Similarly, if $c_{21} = 0$, then $\frac{d\theta_2}{dt}$ will be constant for all time t. Suppose $c_{12} = 0$. Then, $\frac{d\theta_1}{dt} = \omega_1$. Thus,

$$\theta_1(t) = \theta_1(0) + \omega_1 t \implies \cos(\theta_1(t)) = \cos(\theta_1(0) + \omega_1 t).$$

Using the sum formula for cosine, we can write this as

$$\cos(\theta_1(t)) = \cos(\theta_1(0))\cos(\omega_1 t) - \sin(\theta_1(0))\sin(\omega_1 t).$$

Differentiating both sides with respect to t, we find:

$$\begin{aligned}\frac{d\cos(\theta_1(t))}{dt} &= -\omega_1\cos(\theta_1(0))\sin(\omega_1 t) - \omega_1\sin(\theta_1(0))\cos(\omega_1 t)\\ &= -\omega_1\left(\sin(\theta_1(0))\cos(\omega_1 t) + \cos(\theta_1(0))\sin(\omega_1 t)\right)\\ &= -\omega_1\left(\sin(\theta_1(0)+\omega_1 t)\right)\\ &= -\omega_1\sin(\theta_1(t)).\end{aligned}$$

Thus, if $c_{12}=0$, then $\left(\cos(\theta_1(t)), \frac{d\cos(\theta_1(t))}{dt}\right)$ will rotate clockwise around the ellipse with vertices $(1,0),(-1,0),(0,\omega_1),(0,-\omega_1)$. Similarly, if $c_{21}=0$, then $\left(\cos(\theta_2), \frac{d\cos(\theta_2)}{dt}\right)$ will rotate clockwise around the ellipse with vertices $(1,0),(-1,0),(0,\omega_2),(0,-\omega_2)$. So, to determine causality from θ_2 to θ_1 in this dynamical system, we can first estimate the derivative of $\cos(\theta_1)$ at each point in time, and plot $\frac{d\cos(\theta_1(t))}{dt}$ vs. $\cos(\theta_1)$. If the trajectory does not strictly follow the path of an ellipse in the plane, then $\cos(\theta_2)$ must influence $\cos(\theta_1)$, assuming $\cos(\theta_1)$ and $\cos(\theta_2)$ are the only variables of interest in this system. We can also do the same type of analysis to check if $\cos(\theta_1)$ influences $\cos(\theta_2)$. Here are some examples.

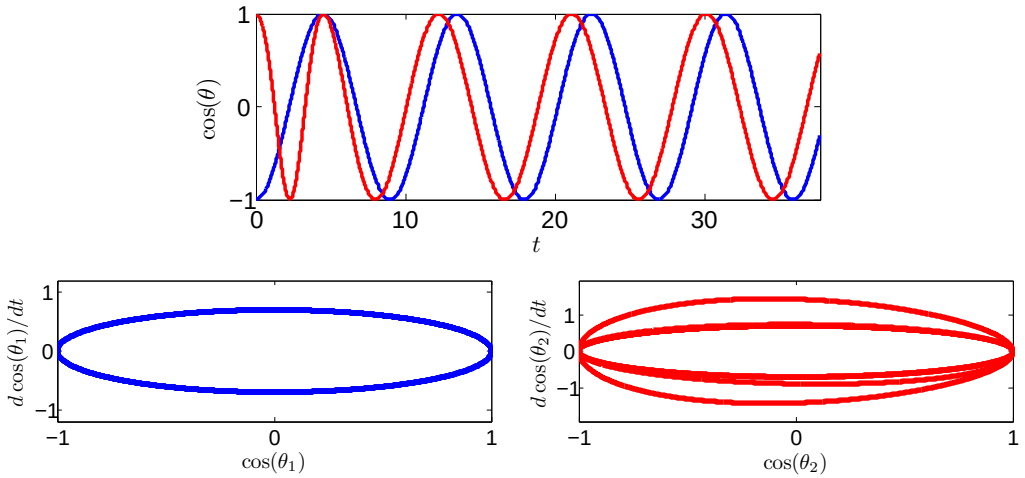

Figure 4: One Way Causality in Coupled Phase Oscillators

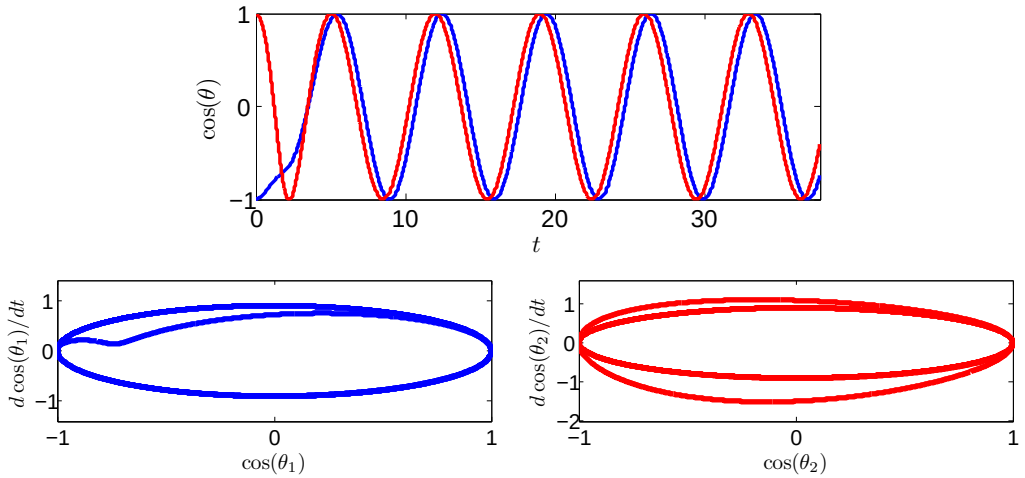

Figure 5: Mutual Causality in Coupled Phase Oscillators

If the two phase oscillators start in a phase-locked state, then we cannot determine the direction of causality without knowing ω_1 and ω_2. Similar to the first-order linear system, in which the transition to the stable fixed point contained information about causality, for the phase oscillators, the transition to the phase-locked state contains information about causality, not the synchronous state itself. However, if the phase oscillators do not phase lock, then we will always be able to determine the direction of causality, regardless of initial conditions.

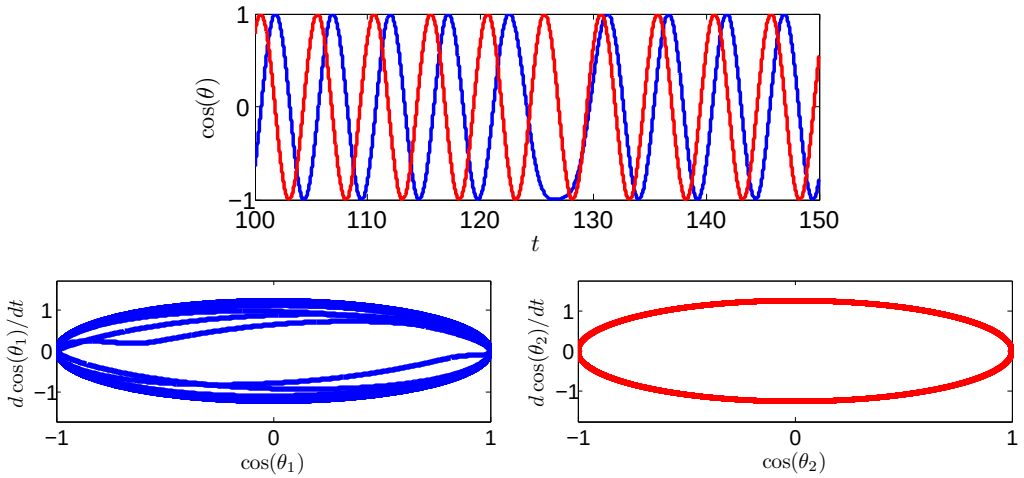

Figure 6: Non Phase-Locking Phase Oscillators

In this example, we see that the phase oscillators are very close to a synchronous state, however, due to the violation of the phase-locking condition, they will never phase lock. A little after time $t = 120$, the phase difference changes dramatically, which can be seen in the first image. During this time interval, the underlying dynamics become clear, which we can see in the bottom two pictures, and we conclude that $\cos(\theta_2)$ influences $\cos(\theta_1)$, but $\cos(\theta_1)$ has no influence over $\cos(\theta_2)$.

2.3 Wilson-Cowan Oscillator

The simplified Wilson-Cowan Oscillator models the network dynamics of an oscillator with two variables, one excitatory (E) and one inhibitory (I), which can be thought of as firing rates of excitatory and inhibitory cells in a neuronal network. In the original 1972 paper [3], Wilson and Cowan defined $E(t)$ and $I(t)$ to be the proportion of E-cells and I-cells firing per unit time at the instant t. The oscillatory nature of the model arises due to the negative feedback the inhibitory cells give to the excitatory cells, after receiving a positive input from the excitatory cells. Here, we present a dramatically simplified version of the model proposed by Wilson himself in 1999 [4]. The model is defined by a system of 2 non-linear ODEs.

$$\frac{dE}{dt} = \frac{E_\infty - E}{5}, \tag{12}$$

$$\frac{dI}{dt} = \frac{I_\infty - I}{10}. \tag{13}$$

When E-cells fire, they give an excitatory input to neighboring cells through synapses. Similarly, when I-cells fire, they give an inhibitory input to neighboring cells. We assume that for a given E and I, the value of E is driven towards a value E_∞ which is an increasing function of E and a decreasing function of I. In the simplified Wilson-Cowan model, E_∞ is a sigmoidal function with the following form:

$$E_\infty = \begin{cases} \dfrac{(c_1E - c_2I + K)^2}{9 + 0.01(c_1E - c_2I + K)^2} & \text{if} \quad c_2I \leq c_1E + K \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

The dynamics of I are modeled in the same way, however, we assume they are not externally driven, and strictly receive an excitatory input from the variable E, that is, they do not self-inhibit.

$$I_\infty = \frac{(d_1E)^2}{9 + 0.01(d_1E)^2}. \tag{15}$$

with $c_1, c_2, d_1 \geq 0$. In the simplified model, Wilson chose the parameters $c_1 = 1.6$, $c_2 = 1$, and $d_1 = 1.5$. These parameters can be thought of as the connection strengths, and represent the average number of excitatory and inhibitory synapses per cell. K is a constant that determines external drive to the E-cells in the network. Here is a network diagram of the Wilson-Cowan oscillator, with a filled in circle representing an inhibitory input, and an open circle representing an excitatory input.

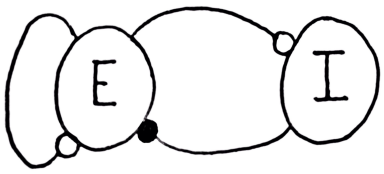

Figure 7: Wilson-Cowan Oscillator Network Diagram

In the Wilson-Cowan oscillator, $E(t)$ and $I(t)$ should be thought of as the proportion of E-cells and I-cells firing per unit time at instant t. So, we assume that $0 \leq E(t) \leq 100$ and $0 \leq I(t) \leq 100$ for all time $t \geq 0$. We can understand the dynamics of the Wilson-Cowan oscillator by looking at the nullclines of both E and I. The nullcline of E occurs along the curve at which $\frac{dE}{dt} = 0$, which occurs when $E = E_\infty$:

$$E = \begin{cases} \frac{(1.6E-I+K)^2}{9+0.01(1.6E-I+K)^2} & \text{if} \quad I \leq 1.6E+K \\ 0 & \text{otherwise} \end{cases}$$

First, we solve for the values of I which satisfy the first equation in the this array, and then restrict ourselves to those values of I for which $I \leq K + 1.6E$. The first equation in this array is a quadratic equation for I, and solving for I as a function of E, we find

$$9E = (1-0.01E)(1.6E-I+K)^2 \implies I = 1.6E+K \pm \sqrt{\frac{9E}{1-0.01E}}$$

Enforcing the restriction that $I \leq 1.6E + K$, we get

$$\boxed{I = 1.6E+K-\sqrt{\frac{9E}{1-0.01E}}}$$

Similarly, we can solve for the nullcline of I, which occurs along the curve $\frac{dI}{dt} = 0$, at which

$$\boxed{I = \frac{(1.5E)^2}{9 + 0.01(1.5E)^2}}$$

We note that the I-nullcline is independent of K. Thus, modifying K only affects the E-nullcline. We can plot these two nullclines in the E, I-phase plane to get the following image for various values of K. The points at which the two nullclines cross are the fixed points of the dynamical system.

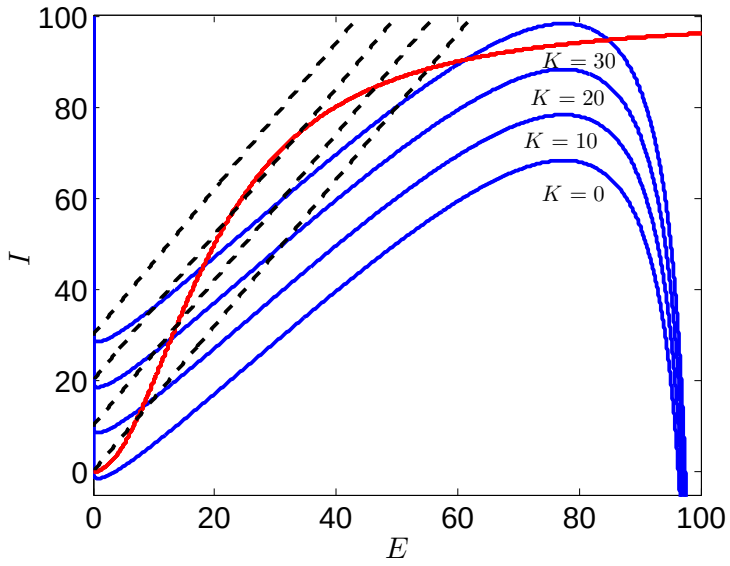

Figure 8: E-nullclines are in blue, I-nullcline is in red, dashed black lines are corresponding lines $I = 1.6E + K$ for each K.

For $K = 0$, the nullclines cross at the point $(0,0)$. As K increases, the point of intersection moves along the I-nullcline, until K reaches a value for which there are 2 crossings of the nullclines. As K continues to increases, a third crossing comes into existence. As K continues to increase further, an annihilation occurs between the two left crossings, and only one exists in the $[0,100] \times [0,100]$ plane.

A transition from a stable fixed point to a periodic solution occurs via a supercritical Hopf bifurcation, in which a stable spiral becomes an unstable spiral and a *periodic solution* (or *stable limit cycle*) appears. Eventually, the stable limit cycle disappears, as K reaches a threshold value, after which the system has a globally attracting node at the right-most

crossing of the two nullclines. To understand this visually, we look at the phase-space portraits of the solution to the Wilson-Cowan Oscillator with various values of K.

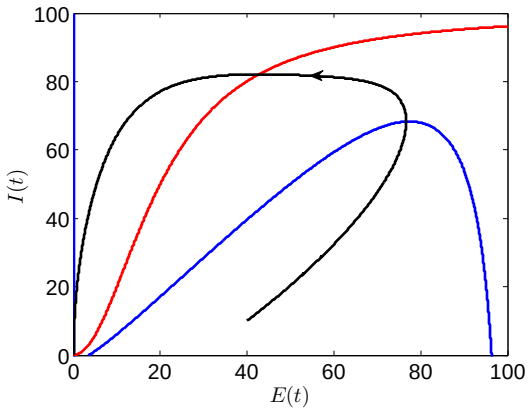

Figure 9: Solution for $K = 0$

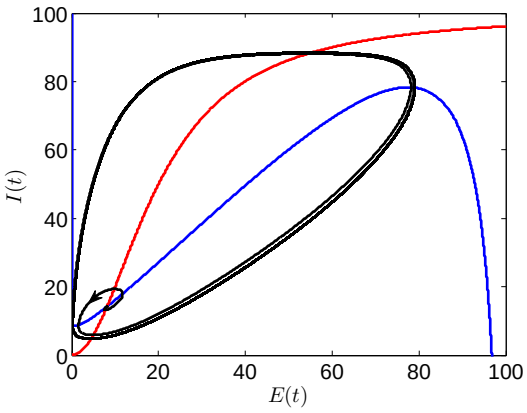

Figure 10: Solution for $K = 10$

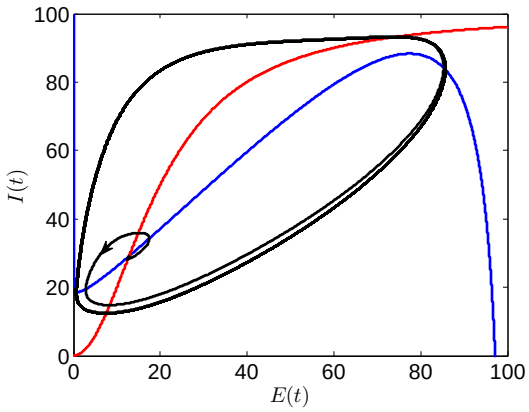

Figure 11: Solution for $K = 20$

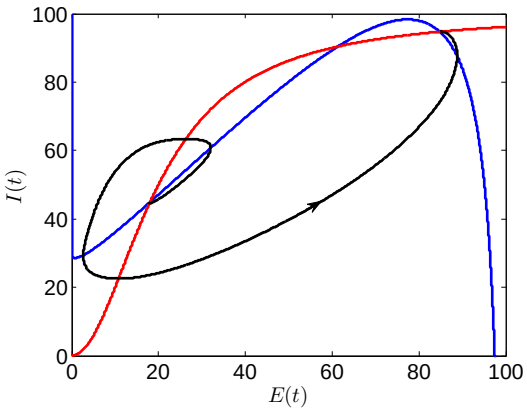

Figure 12: Solution for $K = 30$

For $K = 0$, the stable fixed point occurs at $E^* = 0$ and $I^* = 0$, which also has the physical interpretation that without any external drive, the system approaches a state of zero activity. With enough drive, there is oscillatory behavior, and a periodic solution arises. However, when K gets too large, the external drive overpowers the inhibitory input given by the I-cells, and a large portion of both E-cells and I-cells remain activated. For the E-cells and the I-cells, a periodic solution for $K = 20$ looks like the following.

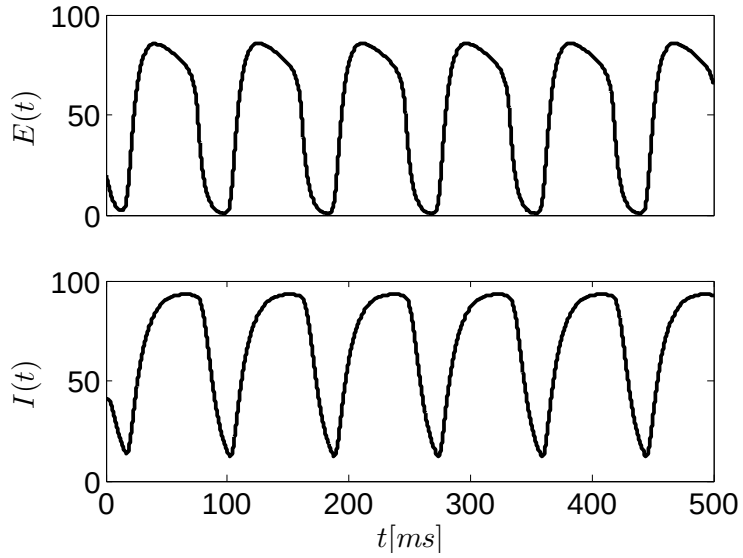

Figure 13: $E(t)$ and $I(t)$ for $K = 20$

Note that time is measured in milliseconds. There are about 6 oscillations over 500 milliseconds, which implies that the periodic solution for $K = 20$ has a frequency of about 12 Hz, called an alpha wave.

2.4 Coupled Wilson-Cowan Oscillators

We are interested in coupling two Wilson-Cowan oscillators, and from a given solution to the two oscillators, determine the direction of causality between them. The dynamics of coupled Wilson-Cowan oscillators have been studied extensively, especially the synchronization of a network of Wilson-Cowan neural oscillators [5] [6]. It has been shown that for certain parameter choices, the coupled Wilson-Cowan oscillator network will produce phase-locking. Here is an example of a coupled phase-locking Wilson-Cowan oscillator network:

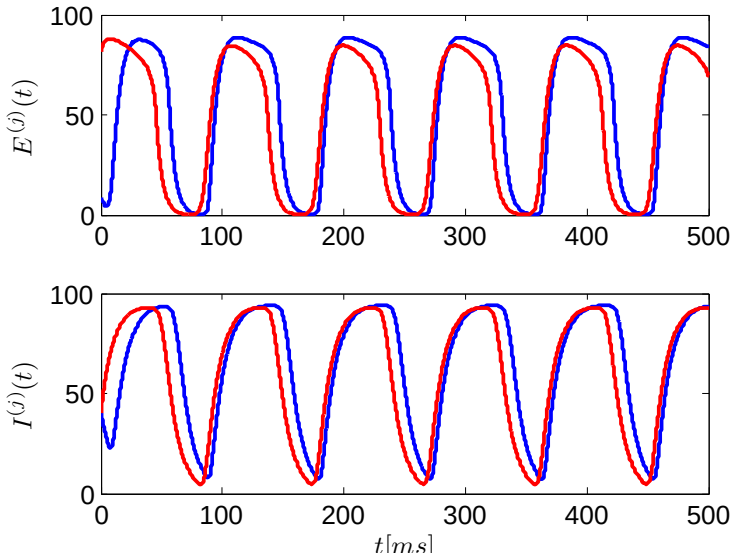

Figure 14: Oscillations in a coupled Wilson-Cowan oscillator

The model we will focus on is of the following form:

$$\frac{dE^{(1)}}{dt} = \frac{E_\infty^{(1)} - E^{(1)}}{5}, \tag{16}$$

$$\frac{dI^{(1)}}{dt} = \frac{I_\infty^{(1)} - I^{(1)}}{10}, \tag{17}$$

$$\frac{dE^{(2)}}{dt} = \frac{E_\infty^{(2)} - E^{(2)}}{5}, \tag{18}$$

$$\frac{dI^{(2)}}{dt} = \frac{I_\infty^{(2)} - I^{(2)}}{10}. \tag{19}$$

Here, $E^{(1)}$ and $I^{(1)}$ are one Wilson-Cowan oscillator, and $E^{(2)}$ and $I^{(2)}$ are a second Wilson-Cowan oscillator. We only couple the E-cells of each Wilson-Cowan oscillator, which translates to the following equations for the $E_\infty^{(i)}$ and $I_\infty^{(i)}$ for $i = 1, 2$.

$$E_\infty^{(i)} = \begin{cases} \dfrac{(\varphi_j E^{(j)} + c_1 E^{(i)} - c_2 I^{(i)} + K)^2}{9 + 0.01(\varphi_j E^{(j)} + c_1 E^{(i)} - c_2 I^{(i)} + K)^2} & \text{if} \quad c_2 I^{(i)} \leq \varphi_j E^{(j)} + c_1 E^{(i)} + K \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where j corresponds to the second Wilson-Cowan oscillator. That is, $j = 1$ if $i = 2$, and $j = 2$ if $i = 1$. As before, set $c_1 = 1.6$, $c_2 = 1$, and $d_1 = 1.5$. The coupling terms between the two Wilson-Cowan oscillators are purely determined by the variables $\varphi_1 \geq 0$, and $\varphi_2 \geq 0$, where φ_1 is the coupling strength from $E^{(2)}$ to $E^{(1)}$ and φ_2 is the coupling strength from $E^{(1)}$ to $E^{(2)}$. The dynamics of $I_\infty^{(i)}$ are modeled as before, for $i = 1, 2$,

$$I_\infty^{(i)} = \frac{(d_1 E^{(i)})^2}{9 + 0.01(d_1 E^{(i)})^2}. \tag{21}$$

The network diagram of the coupled Wilson-Cowan oscillator is as follows

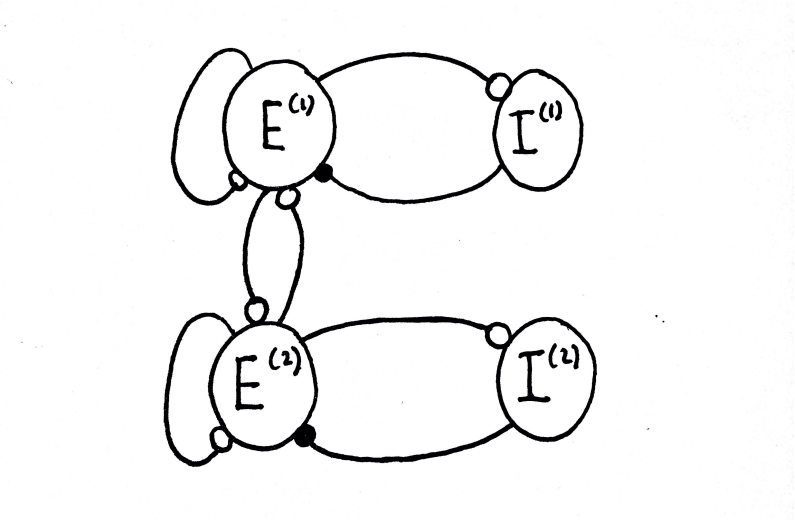

Figure 15: Coupled Wilson-Cowan Oscillator Network Diagram

Coupling the E-cells only is motivated by findings that long-range intracortical connections (within a specific cortex in the brain) arise almost exclusively from excitatory neurons [8] [9]. Furthermore, the target cells of these synaptic connections are other excitatory cells about 80% of the time [10] [11].

We are interested in the following question. Suppose we are given a set of observations from a coupled Wilson-Cowan oscillator, $\{E_k^{(i)}, I_k^{(i)}\}_{k=0}^N$, $i = 1, 2$, where $E_k^{(i)} = E^{(i)}(\Delta t k)$, $I_k^{(i)} = I^{(i)}(\Delta t k)$. From this set of observations, we want to determine if $E^{(1)}$ influences $E^{(2)}$, or vice-versa, or both. That is, we want to determine if $\varphi_1 \neq 0$, or $\varphi_2 \neq 0$, or both. We

restrict ourselves to parameters which produce a phase-locking solution.

In the case of a single Wilson-Cowan oscillator, the existence and uniqueness theorem for first-order ODEs implies that different trajectories in the phase-space plane never intersect [7]. The single Wilson-Cowan oscillator has a two-dimensional phase space, the (E,I)-plane. Thus, if we look at trajectories of a single Wilson-Cowan oscillator in the (E,I)-plane, those trajectories will never intersect. In the case of two coupled Wilson-Cowan oscillators, $(E^{(1)},I^{(1)})$ and $(E^{(2)},I^{(2)})$, if the coupling terms φ_1 and φ_2 are nonzero, then the overall phase space for both Wilson-Cowan oscillators will be four-dimensional. Therefore, when plotting the trajectory of the second Wilson-Cowan oscillator in the two-dimensional $(E^{(2)},I^{(2)})$-subspace of the four-dimensional phase space, the trajectories may intersect. If there is an intersection of the trajectory in the $(E^{(2)},I^{(2)})$-plane, then the first Wilson Cowan oscillator must have a causal influence on the dynamics of the second Wilson-Cowan oscillator. Using this observation, we can determine if $\varphi_1 \neq 0$ and $\varphi_2 \neq 0$, if you start far enough away from the phase-locked state.

Here is an example of two Wilson-Cowan oscillators for which we are able to detect causality from $E^{(1)}$ to $E^{(2)}$ due to the cross-over of trajectories in the $(E^{(2)},I^{(2)})$-plane.

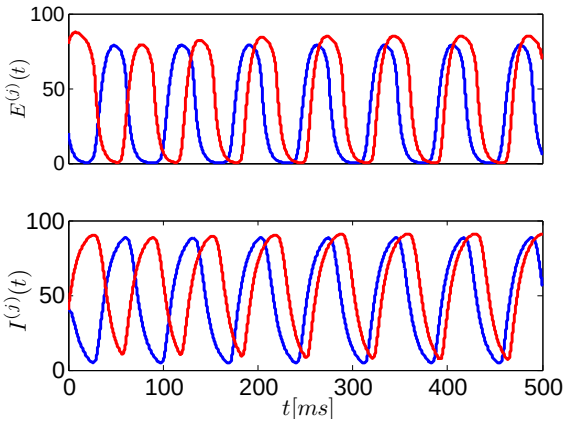

Figure 16: coupled Wilson-Cowan oscillators

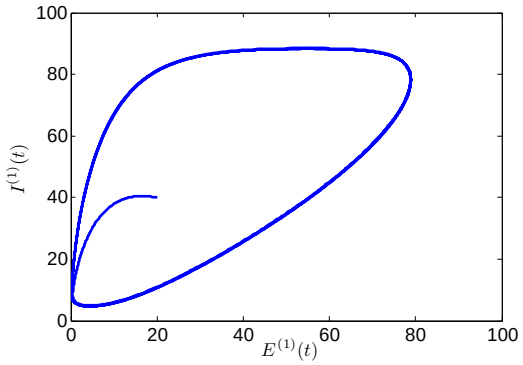

Figure 17: No Crossover in $(E^{(1)}, I^{(1)})$-plane

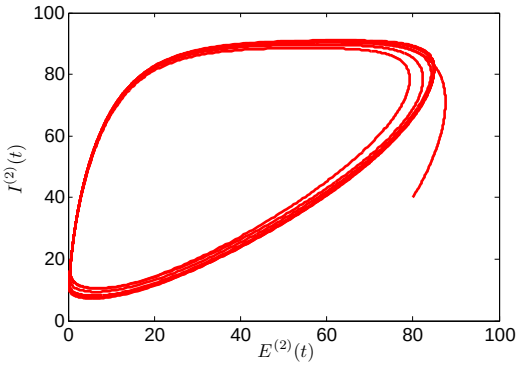

Figure 18: Crossover in $(E^{(2)}, I^{(2)})$-plane

In order to deduce causality from looking at cross-overs in the $(E^{(1)}, I^{(1)})$-plane and $(E^{(2)}, I^{(2)})$-plane, the system has to start far enough away from the stable limit cycle. Furthermore, we have to view all four dynamic variables, $E^{(1)}, I^{(1)}, E^{(2)}$, and $I^{(2)}$. What if we only viewed the excitatory part of each network, $E^{(1)}$ and $E^{(2)}$? Then, this method does not work anymore. Another way we can look for causality in the coupled Wilson-Cowan oscillator model is to estimate the time derivatives of $E^{(1)}(t)$ and $E^{(2)}(t)$, and plot these as functions of $E^{(1)}(t)$ and $E^{(2)}(t)$. We would assume that if $E^{(1)}$ influences $E^{(2)}$, then the time derivative of $E^{(2)}$ will fluctuate more over time than if there was no influence. Keep in mind, this will only happen if the system starts far enough away from the phase-locked state. Using the previous simulation, we plot the estimated time derivatives of $E^{(1)}$ and $E^{(2)}$.

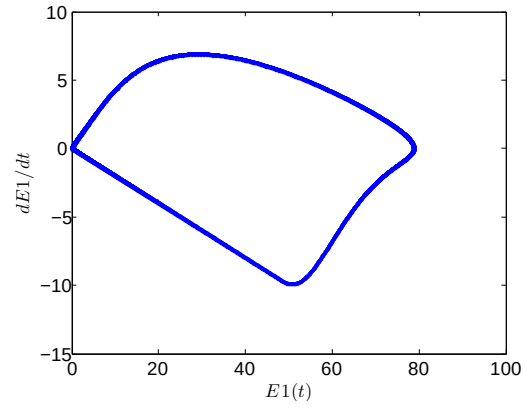

Figure 19: $dE^{(1)}/dt$ vs. $E^{(1)}(t)$.

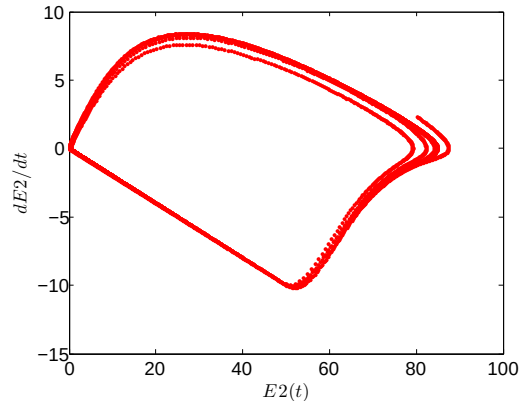

Figure 20: $dE^{(2)}/dt$ vs. $E^{(2)}(t)$.

The plot of time derivatives of $E^{(2)}$ certainly varies more than the plot of time derivatives of $E^{(1)}$. Thus, it also possible to see the influence of $E^{(1)}$ onto $E^{(2)}$ without using any information about the inhibitory component of the networks. The main assumption that we are making with this analysis is the following. If the motion starts far enough away from the stable limit cycle, then if one of the dynamic variables is influencing another, the path or trajectory of the variable that is being influenced will be more volatile, since we are trying to project a high-dimensional phase-space into a low-dimensional space. However, these methods will not work in every case, and are contingent upon the system starting far away from the stable limit cycle. Due to this fact, it would seem that adding a noisy driving force to a dynamical system in a phase-locked state may be causality revealing. Although the noise may drown out some of the dynamics of the system, it will also continuously knock the dynamical system out of a stable steady state.

In the following sections, we will present how to test for causality in a linear discrete-time stochastic process using Granger causality techniques. We will also look at how the presence of a noisy driving force affects our ability to detect causality in phase-locking oscillators in the final section.

3 Granger Causality - An Overview

3.1 Multivariate Gaussian Distribution

In order to understand Granger causality and how we can use it in a discrete time Gaussian process, we will have to review the multivariate Gaussian distribution. A *non-degenerate multivariate Gaussian density on* $\mathbb{R}^n$ is defined by a probability density function of the form

$$f(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|} \exp\left(-\frac{1}{2}(x-\mu)^T(\Sigma\Sigma^T)^{-1}(x-\mu)\right), \tag{22}$$

where μ is a constant vector in $\mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n\times n}$ is invertible, and $|\Sigma|$ denotes the determinant of the matrix Σ. If Σ has rank less than n, then the corresponding multivariate Gaussian does not have a probability density, since Σ is not invertible. For this case, the multivariate Gaussian random variable is said to be *degenerate*.

The non-degenerate multivariate Gaussian random variable in $\mathbb{R}^n$ is defined by two variables, $\mu \in \mathbb{R}^n$ and $\Sigma\Sigma^T \in \mathbb{R}^{n\times n}$, which are the mean, and the variance-covariance matrix of the random variable. Thus, the multivariate Gaussian is defined by its first two moments. We can verify that μ and $\Sigma\Sigma^T$ are the mean and variance-covariance matrix of the multivariate Gaussian, but first let's show that (22) is a probability density function, ie. that it integrates to 1.

$$\begin{aligned}\int_{\mathbb{R}^n} f(x)dx &= \frac{1}{(2\pi)^{n/2}|\Sigma|}\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}(x-\mu)^T(\Sigma\Sigma^T)^{-1}(x-\mu)\right)dx \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma|}\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}(\Sigma^{-1}(x-\mu))^T(\Sigma^{-1}(x-\mu))\right)dx\end{aligned}$$

Let $y = \Sigma^{-1}(x-\mu)$. Then, $dy = |\Sigma^{-1}|dx = dx/|\Sigma|$, and the integral becomes

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2} y^T y\right) dy = \frac{(2\pi)^{n/2}}{(2\pi)^{n/2}} = 1.$$

Next, we show that the mean of a random variable X with density (22) is μ. That is, we want to show

$$\int_{\mathbb{R}^n} (x-\mu) f(x) dx = 0.$$

Let $z = x - \mu$. Then, this integral becomes

$$\frac{1}{(2\pi)^{n/2}|\Sigma|} \int_{\mathbb{R}^n} z \exp\left(-\frac{1}{2}(\Sigma^{-1}z)^T(\Sigma^{-1}z)\right) dz = \frac{1}{(2\pi)^{n/2}|\Sigma|} \int_{\mathbb{R}^n} z \exp\left(-\frac{1}{2}\|\Sigma^{-1}z\|_2^2\right) dz$$

The function $\exp\left(-\frac{1}{2}\|\Sigma^{-1}z\|_2^2\right)$ is even, which means the integrand is odd. Since we are integrating over the entire space $\mathbb{R}^n$, this integral is 0. Finally, we want to show that the variance-covariance matrix of X is equal to $\Sigma\Sigma^T$. In order to do this, we have to calculate the following integral, and show that it equals $\Sigma\Sigma^T$.

$$\int_{\mathbb{R}^n} (x-\mu)(x-\mu)^T f(x) dx =$$

$$\frac{1}{(2\pi)^{n/2}|\Sigma|} \int_{\mathbb{R}^n} (x-\mu)(x-\mu)^T \exp\left(-\frac{1}{2}(x-\mu)^T(\Sigma\Sigma^T)^{-1}(x-\mu)\right) dx.$$

Again, use the change of variables $y = \Sigma^{-1}(x-\mu)$ to get

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \Sigma y y^T \Sigma^T \exp\left(-\frac{1}{2} y^T y\right) dy =$$

$$\Sigma \left[\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} yy^T \exp\left(-\frac{1}{2}y^T y\right) dy \right] \Sigma^T.$$

The expression in the brackets is equal to the identity matrix, so this integral is equal to $\Sigma\Sigma^T$. So, the random variable with density function (22) has mean μ and variance-covariance matrix $\Sigma\Sigma^T$. We can normalize a multivariate Gaussian by doing the following transformation. If X is an n-dimensional multivariate Gaussian random variable with mean μ and variance-covariance matrix $\Sigma\Sigma^T$, then the following linear transformation of X produces a *standard Gaussian vector* with mean 0 and variance-covariance matrix I.

$$Z = \Sigma^{-1}(X - \mu). \tag{23}$$

To see what the density of $\Sigma^{-1}(X-\mu)$ is, we calculate the probability that $\Sigma^{-1}(X-\mu)$ lies in a specific region in $\mathbb{R}^n$. Let $R \subset \mathbb{R}^n$. Then,

$$\begin{aligned} P\left((\Sigma^{-1}(X-\mu) \in R\right) &= P\left(X \in \Sigma R + \mu\right) \\ &= \int_{\Sigma R+\mu} f(x)dx \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma|} \int_{\Sigma R+\mu} \exp\left(-\frac{1}{2}(\Sigma^{-1}(x-\mu))^T(\Sigma^{-1}(x-\mu))\right) dx. \end{aligned}$$

Substituting $y = \Sigma^{-1}(x-\mu)$, this integral becomes

$$\frac{1}{(2\pi)^{n/2}} \int_R \exp\left(-\frac{1}{2}y^T y\right) dy.$$

This integral describes the probability that an $n-$dimensional multivariate random variable with mean zero and variance-covariance matrix I lies in the region $R \subset \mathbb{R}^n$. Thus, the

random variable Z defined by the linear transformation $Z = \Sigma^{-1}(X - \mu)$, has a multivariate Gaussian distribution with mean zero and variance-covariance matrix I, which implies that Z is vector of n uncorrelated Gaussians with mean zero and variance 1. This linear transformation also works in the other direction. That is, if Z is a standard Gaussian vector, then $X = \mu + \Sigma Z$ is multivariate Gaussian with mean μ and variance-covariance matrix $\Sigma\Sigma^T$.

There are many interesting properties of the multivariate Gaussian random variable. Here, we will only present how to solve the conditional distribution of a non-degenerate multivariate Gaussian. In order to do so, we first need to show if two partitions of a multivariate Gaussian random variable are uncorrelated, then they are also independent. Let $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ be uncorrelated and multivariate Gaussian, and suppose the joint random variable, $Z = [X,Y]^T$ is also multivariate Gaussian. Suppose X has mean μ_x and Y has mean μ_y, and $\mu_z = [\mu_x, \mu_y]^T$. If X and Y are uncorrelated, then $\text{cov}(X,Y) = 0$ and $\text{cov}(Y,X) = 0$. So, the variance-covariance matrix of Z will be block diagonal, and will have the form

$$\Sigma\Sigma^T = \begin{bmatrix} \Sigma_x\Sigma_x^T & 0 \\ 0 & \Sigma_y\Sigma_y^T \end{bmatrix},$$

where $\Sigma_x\Sigma_x^T \in \mathbb{R}^{m\times m}$ is the variance-covariance matrix of X, and $\Sigma_y\Sigma_y^T \in \mathbb{R}^{n\times n}$ is the variance-covariance matrix of Y. Thus,

$$|\Sigma\Sigma^T| = |\Sigma_x\Sigma_x^T||\Sigma_y\Sigma_y^T| \implies |\Sigma|^2 = |\Sigma_x|^2|\Sigma_y|^2 \implies |\Sigma| = |\Sigma_x||\Sigma_y|.$$

Also, the inverse of $\Sigma\Sigma^T$ is

$$\left(\Sigma\Sigma^T\right)^{-1} = \begin{bmatrix} (\Sigma_x\Sigma_x^T)^{-1} & 0 \\ 0 & (\Sigma_y\Sigma_y^T)^{-1} \end{bmatrix}.$$

By definition, since Z is non-degenerate multivariate Gaussian, it will have the following probability density function, which we simplify as follows

$$\begin{aligned}
f_Z(z) &= \frac{1}{(2\pi)^{(m+n)/2}|\Sigma|}\exp\left(-\frac{1}{2}(z-\mu_z)^T(\Sigma\Sigma^T)^{-1}(z-\mu_z)\right) \\
&= \frac{1}{(2\pi)^{(m+n)/2}|\Sigma_x||\Sigma_y|}\exp\left(-\frac{1}{2}(x-\mu_x)^T\left(\Sigma_x\Sigma_x^T\right)^{-1}(x-\mu_x)\times\right. \\
&\quad \left.(y-\mu_y)^T\left(\Sigma_y\Sigma_y^T\right)^{-1}(y-\mu_y)\right) \\
&= \frac{1}{(2\pi)^{m/2}|\Sigma_x|}\exp\left(-\frac{1}{2}(x-\mu_x)^T\left(\Sigma_x\Sigma_x^T\right)^{-1}(x-\mu_x)\right)\times \\
&\quad \frac{1}{(2\pi)^{n/2}|\Sigma_y|}\exp\left(-\frac{1}{2}(y-\mu_y)^T\left(\Sigma_y\Sigma_y^T\right)^{-1}(y-\mu_y)\right) \\
&= f_X(x)f_Y(y).
\end{aligned}$$

Thus, the multivariate Gaussian random variables X and Y are independent if they are uncorrelated and components of a multivariate random Gaussian. Now, we want to solve for the conditional distribution of a multivariate Gaussian. Let $Z = [X_1, X_2]^T$ be a multivariate random Gaussian with $X_1 \in \mathbb{R}^m$ and $X_2 \in \mathbb{R}^n$, where both X_1 and X_2 are multivariate Gaussians with means μ_1 and μ_2. We want to solve for the distribution of the conditional random variable $(X_2|X_1 = x_1)$. Let $\Sigma\Sigma^T$ be the variance-covariance matrix of Z, and let

$$\Sigma\Sigma^T = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where $A_{11} \in \mathbb{R}^{m\times m}$ is the variance-covariance matrix of X_1, $A_{22} \in \mathbb{R}^{n\times n}$ is the variance-covariance matrix of X_2, $\text{cov}(X_1, X_2) = A_{12} \in \mathbb{R}^{m\times n}$, and $\text{cov}(X_2, X_1) = A_{21} \in \mathbb{R}^{n\times m}$, such that $A_{12} = A_{21}^T$.

We want to find a linear transformation of Z, which will produce uncorrelated parti-

ations of Z. Define the following linear transformation of the multivariate Gaussian, Z, by

$$L = \begin{bmatrix} I & 0 \\ -A_{12}A_{11}^{-1} & I \end{bmatrix}.$$

Clearly, L is invertible because it is lower triangular and all diagonal entries are non-zero. Thus, LZ will be a multivariate Gaussian. When we multiply Z by L, we get

$$LZ = \begin{bmatrix} I & 0 \\ -A_{12}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ -A_{12}A_{11}^{-1}X_1 + X_2 \end{bmatrix}.$$

Define $Y = X_2 - A_{21}A_{11}^{-1}X_1$, the second partition of LZ. Since Y is a sub-vector of a multivariate Gaussian, it is also Gaussian. Solving for the covariance between Y and X_1, we find:

$$\begin{aligned} \operatorname{cov}(Y, X_1) &= \operatorname{cov}(X_2 - A_{21}A_{11}^{-1}X_1, X_1) \\ &= \operatorname{cov}(X_2, X_1) - \operatorname{cov}(A_{21}A_{11}^{-1}X_1, X_1) \\ &= A_{21} - A_{21}A_{11}^{-1}\operatorname{cov}(X_1, X_1) \\ &= A_{21} - A_{21}A_{11}^{-1}A_{11} \\ &= A_{21} - A_{21} \\ &= 0. \end{aligned}$$

Thus, Y and X_1 are uncorrelated components of a multivariate Gaussian, and hence independent! Now, assume X_1 is given, that is $X_1 = x_1$, and we want to solve for the distribution of the conditional random variable $(X_2 | X_1 = x_1)$. First, we solve for the expectation, and using the fact that Y is independent of X_1, we get:

$$
\begin{aligned}
\mathrm{E}(X_2|X_1=x_1) &= \mathrm{E}\left(Y + A_{21}A_{11}^{-1}X_1|X_1\right) \\
&= \mathrm{E}(Y|X_1) + \mathrm{E}\left(A_{21}A_{11}^{-1}X_1|X_1\right) \\
&= \mathrm{E}(Y) + A_{21}A_{11}^{-1}x_1 \\
&= \mathrm{E}\left(X_2 - A_{21}A_{11}^{-1}X_1\right) + A_{21}A_{11}^{-1}x_1 \\
&= \mathrm{E}(X_2) - A_{21}A_{11}^{-1}\mathrm{E}(X_1) + A_{21}A_{11}^{-1}x_1 \\
&= \mu_2 - A_{21}A_{11}^{-1}\mu_1 + A_{21}A_{11}^{-1}x_1 \\
&= \mu_2 + A_{21}A_{11}^{-1}(x_1 - \mu_1).
\end{aligned}
$$

Solving for the variance of $(X_2|X_1=x_1)$, we get:

$$
\begin{aligned}
\mathrm{var}(X_2|X_1=x_1) &= \mathrm{var}\left(Y + A_{21}A_{11}^{-1}X_1|X_1\right) \\
&= \mathrm{var}(Y|X_1) + \mathrm{var}(A_{21}A_{11}^{-1}X_1|X_1) + \mathrm{cov}(Y|X_1, A_{21}A_{11}^{-1}X_1|X_1) \\
&\quad + \mathrm{cov}(A_{21}A_{11}^{-1}X_1|X_1, Y|X_1) \\
&= \mathrm{var}(Y) + \cancel{\mathrm{var}(A_{21}A_{11}^{-1}x_1)} + \cancel{\mathrm{cov}(Y, A_{21}A_{11}^{-1}x_1)} + \cancel{\mathrm{cov}(A_{21}A_{11}^{-1}x_1, Y)} \\
&= \mathrm{var}(X_2 - A_{21}A_{11}^{-1}X_1) \\
&= \mathrm{var}(X_2) + \mathrm{var}(A_{21}A_{11}^{-1}X_1) - \mathrm{cov}(X_2, A_{21}A_{11}^{-1}X_1) - \mathrm{cov}(A_{21}A_{11}^{-1}X_1, X_2) \\
&= A_{22} + A_{21}A_{11}^{-1}\mathrm{var}(X_1)\left(A_{21}A_{11}^{-1}\right)^T - \mathrm{cov}(X_2, X_1)\left(A_{21}A_{11}^{-1}\right)^T \\
&\quad - \left(A_{21}A_{11}^{-1}\right)\mathrm{cov}(X_1, X_2) \\
&= A_{22} + A_{21}A_{11}^{-1}A_{11}A_{11}^{-1}A_{12} - A_{21}A_{11}^{-1}A_{12} - A_{21}A_{11}^{-1}A_{12} \\
&= A_{22} + A_{21}A_{11}^{-1}A_{12} - 2A_{21}A_{11}^{-1}A_{12} \\
&= A_{22} - A_{21}A_{11}^{-1}A_{12}.
\end{aligned}
$$

Note, in this derivation we have used the fact that the inverse of a symmetric positive definite matrix is symmetric positive definite, that is, since A_{11} is symmetric positive definite, so is A_{11}^{-1}. So, now we have the conditional mean and variance covariance matrix of the random variable $(X_2|X_1 = x_1)$. Furthermore, $(X_2|X_1 = x_1)$ will be multivariate Gaussian. To show this, we would need to calculate the conditional density of $(X_2|X_1 = x_1)$ which is defined as

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_Z(z)}{f_{X_1}(x_1)},$$

since $f_Z(z)$ is the joint probability density of function of $[X_1, X_2]^T$. It is possible to do this, and the pdf will precisely follow a multivariate Gaussian with mean and variance-covariance matrix we solved for above [12].

Consider the case of a multivariate Gaussian random variable where $n = 2$. This case is known as a *bivariate Gaussian*. For this case, we let

$$X = \begin{bmatrix} X \\ Y \end{bmatrix}.$$

The pdf of the bivariate Gaussian, $f(x,y)$, is defined as

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right) \quad (24)$$

with $\sigma_x, \sigma_y > 0$. The mean, μ, and variance-covariance matrix, $\Sigma\Sigma^T$, are defined as

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \Sigma\Sigma^T = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}. \quad (25)$$

The correlation between X and Y is the quantity ρ.

$$\rho = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}. \tag{26}$$

Note that $-1 \leq \rho \leq 1$. Here are some examples of the bivariate pdf and contour plot with various values of ρ.

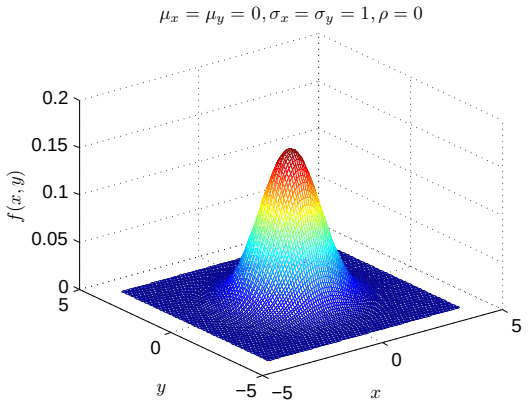

Figure 21: Bivariate Gaussian, $\rho = 0$.

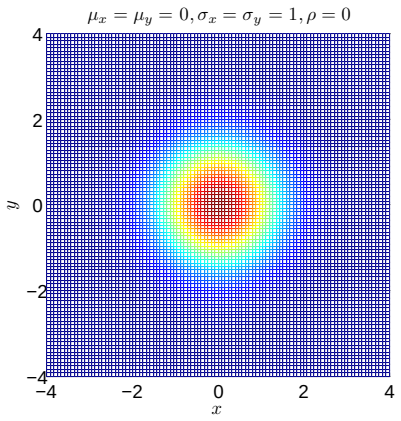

Figure 22: Bivariate Gaussian, $\rho = 0$.

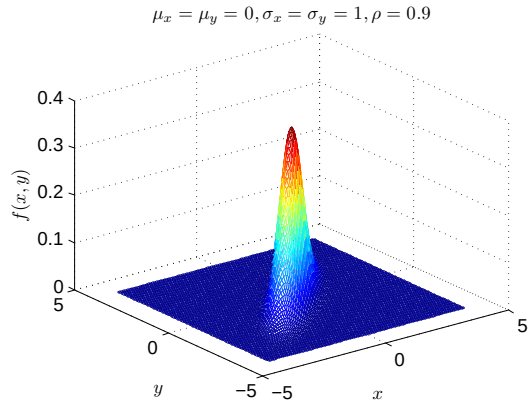

Figure 23: Bivariate Gaussian, $\rho = 0.9$.

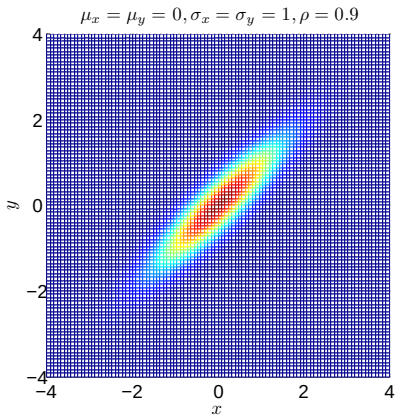

Figure 24: Bivariate Gaussian, $\rho = 0.9$.

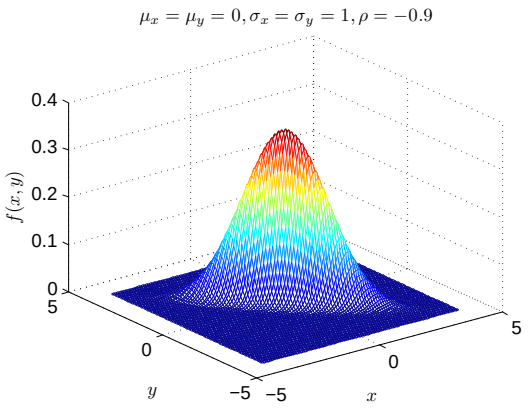

Figure 25: Bivariate Gaussian, $\rho = -0.9$.

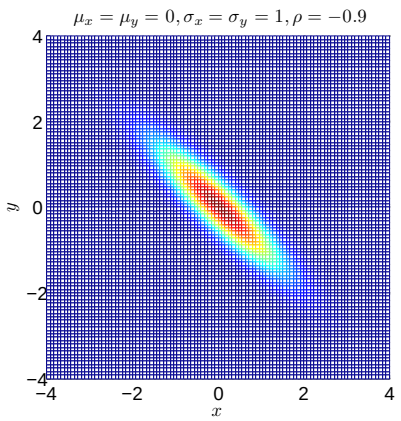

Figure 26: Bivariate Gaussian, $\rho = -0.9$.

Notice that the pdf gets stretched out as ρ deviates from zero. This is because as ρ increases in magnitude towards 1, the variance-covariance matrix gets closer to being singular. When $|\rho| = 1$, the variance-covariance matrix is singular, and the corresponding bivariate Gaussian is degenerate.

3.2 Granger Causality – Two Equivalent Statistical Tests

Let $C \in \mathbb{R}^{2\times 2}$ with spectral radius $\rho(C) < 1$, and $\sigma > 0$. Consider a random sequence, $\{\mathbf{x}_k\}_{k=0,1,2,\ldots}$ of vectors in $\mathbb{R}^2$ satisfying the recursion

$$\mathbf{x}_{k+1} = C\mathbf{x}_k + \sigma \begin{bmatrix} G_{k+1} \\ H_{k+1} \end{bmatrix}, k = 0,1,2,\ldots \tag{27}$$

where G_{k+1} and H_{k+1} are standard Gaussians, independent of each other, and independent of G_j and H_j for $j < k+1$. We also write

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Thus, $x \in \mathbb{R}$ (not bold) is the first component of $\mathbf{x} \in \mathbb{R}^2$ (bold). The condition $\rho(C) < 1$ ensures that $\lim_{k\to\infty} \mathbf{x}_k$ would be $\mathbf{0}$ if σ were zero. In other words, what we are considering is the simplest kind of deterministic dynamics, with convergence to a stable fixed point, perturbed by a white noise process. This process is known as an autoregressive process of order 1, or AR(1) process. Furthermore, the condition that $\rho(C) < 1$ ensures that the process will have an invariant distribution, which we can solve for as follows. Taking expectations of both sides of equation (27).

$$\mathrm{E}\left(\mathbf{x}_{k+1}\right) = \mathrm{E}\left(C\mathbf{x}_k + \sigma \begin{bmatrix} G_{k+1} \\ H_{k+1} \end{bmatrix}\right) = C\mathrm{E}(\mathbf{x}_k).$$

Setting $\mathrm{E}(\mathbf{x}_{k+1}) = \mathrm{E}(\mathbf{x}_k) = \mu$, we get

$$\mu = C\mu,$$

for which the only solution is the zero vector. Next, we can take variance of both sides of equation (27), and setting $\mathrm{Var}(\mathbf{x}_k) = \mathrm{Var}(\mathbf{x}_{k+1}) = VV^T$, we get the following relation for VV^T,

$$VV^T = CVV^TC^T + \sigma^2 I.$$

This type of equation is known as the *discrete Lyapunov equation*, and will have a unique symmetric positive definite solution, VV^T, iff $\rho(C) < 1$ [13]. Since we are assuming $\rho(C) < 1$, there exists a unique symmetric positive definite matrix V, which satisfies this equation. Furthermore, the invariant distribution will be multivariate Gaussian.

The following image shows an example of the process defined in (27).

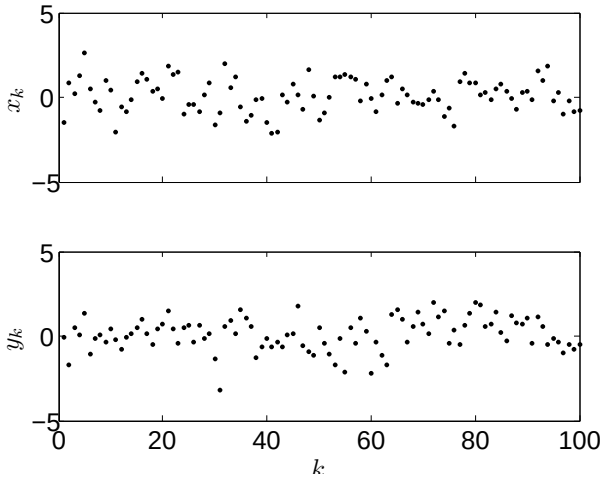

Figure 27: The two components, $\{x_k\}$ and $\{y_k\}$, of $\{\mathbf{x}_k\}_{k=1}^N$ satisfying (27)

This set of observations was produced by the model with coefficient matrix

$$C = \begin{bmatrix} 0.30 & -0.10 \\ 0.20 & 0.40 \end{bmatrix}.$$

Note: This matrix has spectral radius < 1 by Gerschgorin's theorem. Our question now is if we are only given the set of observations in Figure 27, can we tell whether the right upper entry in C is zero or non-zero. Writing

$$C = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

our question is whether $b = 0$ or $b \neq 0$. In other words, our question is whether in the *deterministic* dynamics, y influences (is causal for) x. Similarly, we want to ask if x is causal for y by ascertaining if $c = 0$ or $c \neq 0$. That is we set up the following hypothesis test to test if y "Granger" causes x,

$$H_0 : b = 0, \quad H_A : b \neq 0.$$

To test if x Granger causes y, we set up the following hypothesis test,

$$H_0 : c = 0, \quad H_A : c \neq 0.$$

We can test these two hypotheses using methods from multiple linear regression analysis. In order to do this, we will need to present some of the concepts used in regression analysis from probability and statistics, however, we will mainly present an overview of the theory without going into the minute details.

There are two equivalent statistical tests for testing if $b = 0$. Both of the tests involve comparing a "base" model to a "full" model. For this example, we assume that x is dependent on its own immediate past, and we want to test if x is also dependent on the immediate past of y. Under the null hypothesis, $b = 0$. Therefore, working under the null hypothesis,

$$x_{k+1} = ax_k + \sigma G_{k+1}. \tag{28}$$

This will be our "base" model. Here, we are assuming that x is only dependent on its own past. Furthermore, under the assumptions of the null hypothesis, the noise terms of the base model will be independent, identically distributed Gaussian random variables with mean zero and variance σ^2. Also, they will be independent of x_j for $j < k$. It will be convenient to form the following vectors,

$$Y = \begin{bmatrix} x_N \\ x_{N-1} \\ \vdots \\ x_2 \end{bmatrix}, \quad X = \begin{bmatrix} x_{N-1} \\ x_{N-2} \\ \vdots \\ x_1 \end{bmatrix}, \quad Z = \begin{bmatrix} y_{N-1} \\ y_{N-2} \\ \vdots \\ y_1 \end{bmatrix}.$$

We can use the set of observations $\{\mathbf{x}_k\}_{k=1}^N$ to estimate a using the Least Squares method, from which we get the following estimator for a:

$$\hat{a} = \arg\min_a \|Y - Xa\|_2^2 = (X^T X)^{-1} X^T Y = \frac{\sum_{k=1}^{N-1} x_{k+1} x_k}{\sum_{k=1}^{N-1} x_k^2}. \tag{29}$$

Now, we form a vector of residuals or errors, which we call $\mathbf{e}_b$, the subscript b identifying that these errors are formed from the base model, which is $\mathbf{e}_b = Y - X\hat{a}$. Under the null hypothesis, this estimator for a will be unbiased. Furthermore, under the null hypothesis, the residuals will be independent, identically distributed normal random variables. Just to test this, we plot the residuals of the base model, $e_k = x_k - \hat{a}x_{k-1}$ vs. k to get the following residual plot and histogram plot of the residuals. There are statistical tests to determine if the residuals are independent, identically distributed Gaussians, but we will only use heuristics here.

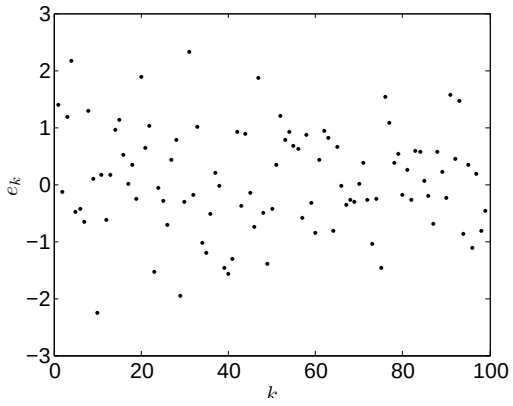

Figure 28: Plot of Residuals

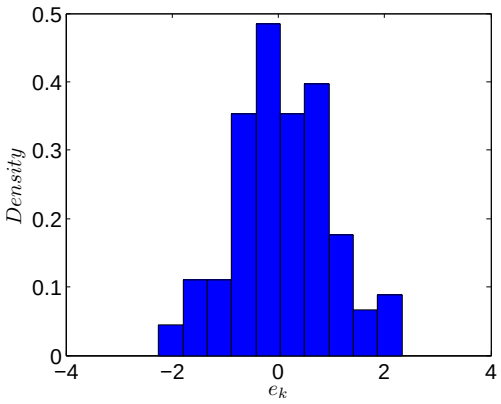

Figure 29: Histogram of Residuals

Under the null hypothesis, the noise terms, σG_k's, are assumed to be independent and identically distributed Gaussians. Therefore, the sum of squared residuals in the base model, $\|\mathbf{e}_b\|_2^2$, will be the sum of $N-1$ squared independent and identically distributed Gaussians. This type of random variable is a χ^2 random variable and follows a χ^2 distribution with $N-2$ degrees of freedom. The $N-2$ degrees of freedom comes from the fact that we are using a total of $N-1$ quantities to estimate the one parameter a. Furthermore, we note that since we are using the Least Squares method to estimate a, it is guaranteed that

the vector of residuals from the base model, $\mathbf{e}_b$, is orthogonal to the vector Y. It is fairly straightforward to show this:

$$\begin{aligned}
\langle Y, \mathbf{e}_b \rangle &= \langle Y, Y - X\hat{a} \rangle \\
&= \langle Y, Y - X((X^T X)^{-1} X^T Y) \rangle \\
&= Y^T Y - Y^T X (X^T X)^{-1} X^T Y \\
&= Y^T Y - (X^T X)^{-1} X^T Y X^T Y \\
&= Y^T Y - (X^T X)^{-1} X^T X Y^T Y \\
&= Y^T Y - Y^T Y \\
&= 0.
\end{aligned}$$

Note, we were allowed to do the above calculation in this way because X and Y are in $\mathbb{R}^{N-1\times 1}$. The next step in this hypothesis test is to form the "full" model, defined by the following equation

$$x_{k+1} = ax_k + by_k + \sigma G_{k+1}.$$

Note that the noise terms in the full model are not necessarily the the same as the error terms in the reduced model. Now, we fix a to be $\hat{a}$, since we assumed it was part of the base model, and we estimate b using the Least Squares method to get the following estimator for b:

$$\hat{b} = \arg\min_b \|Y - X\hat{a} - Zb\|_2^2 = (Z^T Z)^{-1}(Z^T Y - Z^T X\hat{a})) = \frac{\sum_{k=1}^{N-1} (x_{k+1} y_k - \hat{a}_1 x_k y_k)}{\sum_{k=1}^{N-1} y_k^2}. \quad (30)$$

Another requirement to continue the hypothesis test is that the errors in the full regression model are independent, identically distributed Gaussians. Again, we use the heuristic of plotting the residuals and the histogram plot of the residuals of the full model to get the following plots.

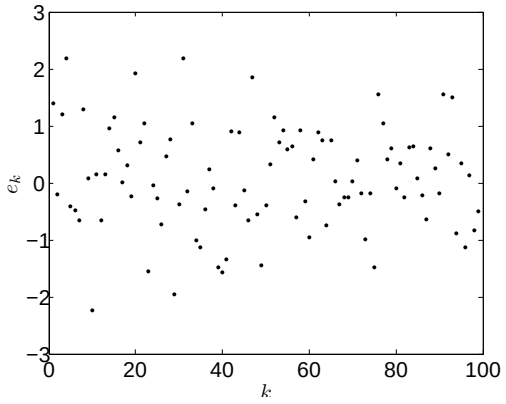

Figure 30: Plot of Residuals

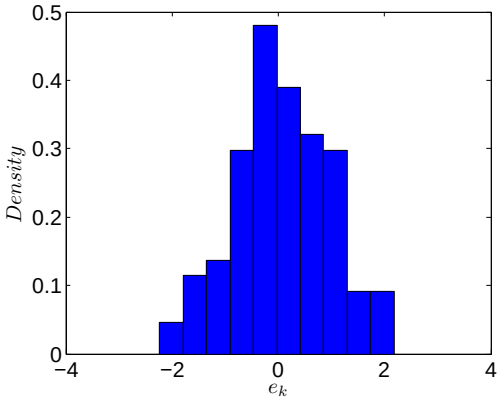

Figure 31: Histogram of Residuals

Now, we can define the vector of residuals from the full model by $\mathbf{e}_f = Y - \hat{a}X - \hat{b}Z$. Note $\mathbf{e}_f = \mathbf{e}_b - \hat{b}Z$. Again, under the null hypothesis that $b = 0$, the square of the 2-norm of this vector, $\|\mathbf{e}_f\|_2^2$, will then be the sum of $N-1$ squared identically distributed Gaussians, so it will be a χ^2 random variable, this time with $N-3$ degrees of freedom. Furthermore, since we are using the Least Squares Estimator to estimate b, the vector of residuals, $\mathbf{e}_f$ will be orthogonal to the vector Z.

We are almost ready to form the test statistic, which involves an F-ratio. A random variable has an F-distribution with parameters d_1 and d_2 if one can write it as the ratio of two appropriately scaled independent χ^2 random variables. That is, if U_1 and U_2 have χ^2 distributions with d_1 and d_2 degrees of freedom respectively, then the random variable $F = \frac{U_1/d_1}{U_2/d_2}$ has an F-distribution.

Under the null hypothesis, $\mathbf{e}_b$ and $\mathbf{e}_f$ are random vectors, which contain independent and identically distributed Gaussian random variables (we verified this heuristically by observing the residual plots and confirming they were normal). Also, the square of the

2-norm of both $\mathbf{e}_b$ and $\mathbf{e}_f$ have χ^2 distributions. Thus, their ratio will have an F-distribution under the null hypothesis, since they are independent random variables.

The F-statistic uses the following test statistic, which compares the variance of the error terms in the base model to the variance of the error terms in the full model in the following way,

$$F = \frac{\|\mathbf{e}_b - \mathbf{e}_f\|_2^2/1}{\|\mathbf{e}_f\|_2^2/(N-3)} \sim F_{1,N-3}. \tag{31}$$

This F-test looks at the marginal effect that Z has on the regressive model, or in other words the marginal effect of including the y_k's in the regressive model for x_{k+1}. One way to understand this quantity is to break it down using what we know about the error terms $\mathbf{e}_b$ and $\mathbf{e}_f$. One way we can do this is to simplify $\|\mathbf{e}_b - \mathbf{e}_f\|_2^2$ in the following way, by using the orthogonality relations from the Least Squares estimate.

$$\begin{aligned}
\|\mathbf{e}_b - \mathbf{e}_f\|_2^2 &= \langle \mathbf{e}_b - \mathbf{e}_f, \mathbf{e}_b - \mathbf{e}_f \rangle \\
&= \langle \mathbf{e}_b, \mathbf{e}_b \rangle - 2\langle \mathbf{e}_b, \mathbf{e}_f \rangle + \langle \mathbf{e}_f, \mathbf{e}_f \rangle \\
&= \|\mathbf{e}_b\|_2^2 - 2\langle Y - \hat{a}X, Y - \hat{a}X - \hat{b}Z \rangle + \|\mathbf{e}_f\|_2^2 \\
&= \|\mathbf{e}_b\|_2^2 - 2\langle \hat{b}Z + \mathbf{e}_f, \mathbf{e}_f \rangle + \|\mathbf{e}_f\|_2^2 \\
&= \|\mathbf{e}_b\|_2^2 - \cancel{2\hat{b}\langle Z, \mathbf{e}_f \rangle} - 2\langle \mathbf{e}_f, \mathbf{e}_f \rangle + \|\mathbf{e}_f\|_2^2 \\
&= \|\mathbf{e}_b\|_2^2 - \|\mathbf{e}_f\|_2^2.
\end{aligned}$$

Therefore, we can re-write the F-statistic in (31) as

$$F = \frac{\|\mathbf{e}_b\|_2^2 - \|\mathbf{e}_f\|_2^2}{\|\mathbf{e}_f\|_2^2/(N-3)} = (N-3)\left(\frac{\|\mathbf{e}_b\|_2^2}{\|\mathbf{e}_f\|_2^2} - 1\right).$$

Furthermore, we know that $\|\mathbf{e}_b\|_2^2 \geq \|\mathbf{e}_f\|_2^2$, since

$$\begin{aligned}
\|\mathbf{e}_f\|_2^2 &= \langle \mathbf{e}_f, \mathbf{e}_f \rangle \\
&= \langle \mathbf{e}_b - \hat{b}Z, \mathbf{e}_b - \hat{b}Z \rangle \\
&= \|\mathbf{e}_b\|_2^2 - 2\langle \mathbf{e}_b, \hat{b}Z \rangle + \hat{b}^2 \|Z\|_2^2 \\
&= \|\mathbf{e}_b\|_2^2 - 2\langle \mathbf{e}_f + \hat{b}Z, \hat{b}Z \rangle + \hat{b}^2 \|Z\|_2^2 \\
&= \|\mathbf{e}_b\|_2^2 - \cancel{2\langle \mathbf{e}_f, \hat{b}Z \rangle} - 2\hat{b}^2 \langle Z, Z \rangle + \hat{b}^2 \|Z\|_2^2 \\
&= \|\mathbf{e}_b\|_2^2 - \hat{b}^2 \|Z\|_2^2.
\end{aligned}$$

Thus, $\|\mathbf{e}_b\|_2^2 \geq \|\mathbf{e}_f\|_2^2$, where equality holds iff $\hat{b} = 0$, which implies $F \geq 0$. Therefore, our test-statistic will always be positive. In fact, under the null hypothesis, $\mathrm{E}(\hat{b}) = 0$, so we expect that if y does not influence x, then this F-ratio should be close to zero. However, if the null hypothesis is wrong, and y does influence x, then including y_k in the regressive model should give us a better prediction for x_{k+1}, meaning $\|\mathbf{e}_f\|_2^2$ will be smaller than $\|\mathbf{e}_b\|_2^2$. Also, we note that $\|\mathbf{e}_b\|_2^2$ is a scaled estimate for the variance of the residual terms in the base model, and $\|\mathbf{e}_f\|_2^2$ is a scaled estimate for the variance of the residual terms in the full model. Therefore, we are simply comparing how much the variances of these two sets of residuals differ from each other with this statistic. Note, we can also re-write the F-statistic using the above relation as:

$$F = \frac{\|\mathbf{e}_b\|_2^2 - \|\mathbf{e}_f\|_2^2}{\|\mathbf{e}_f\|_2^2/(N-3)} = \frac{\hat{b}^2 \|Z\|_2^2}{\|\mathbf{e}_f\|_2^2/(N-3)}.$$

There is an equivalent test statistic used to test the same null hypothesis stated above. This test involves comparing the estimate for b, $\hat{b}$, to zero. The way to do this is by a t-test, and the statistic used is t_{score}, defined as

$$t_{\text{score}} = \frac{\hat{b}}{SE_{\hat{b}}} \sim \mathcal{T}_{N-3}, \tag{32}$$

where $SE_{\hat{b}}$ is the standard error of $\hat{b}$, and the t_{score} has a Student's t-distribution with $N-3$ degrees of freedom. The standard error is defined as

$$SE_{\hat{b}} = \sqrt{\text{Var}(\hat{b})}.$$

We need to estimate the standard error using the given data. To do this, we use the assumptions made in the null hypothesis. We assume that $b = 0$, and $Y = Xa + \varepsilon$, where ε is a $N-1 \times 1$ vector of independent, identically distributed Gaussians. Then, under these assumptions, our estimator $\hat{b}$ will be unbiased! To see this, we calculate its expected value

$$\begin{aligned}
\text{E}(\hat{b}) &= \text{E}\left((Z^T Z)^{-1}(Z^T Y - Z^T Xa)\right) \\
&= \text{E}\left((Z^T Z)^{-1} Z^T (Y - Xa)\right) \\
&= \text{E}\left((Z^T Z)^{-1} Z^T \varepsilon\right) \\
&= (Z^T Z)^{-1} \text{E}(Z^T \varepsilon) \\
&= 0.
\end{aligned}$$

We were able to pull out the term $(Z^T Z)^{-1}$ in the second-to-last line because they can be considered as non-random weights, depending only on the given values of the y_k's, independent of the noise terms, σG_k. Now, still working under the null hypothesis, we can estimate the variance of $\hat{b}$ as follows, using the fact that $\text{E}(\hat{b}) = 0$,

$$
\begin{aligned}
\mathrm{Var}(\hat{b}) &= \mathrm{E}\left((\hat{b}-\mathrm{E}(\hat{b}))(\hat{b}-\mathrm{E}(\hat{b}))^T\right) \\
&= \mathrm{E}\left(\hat{b}\hat{b}^T\right) \\
&= \mathrm{E}\left(((Z^TZ)^{-1}Z^T\varepsilon)((Z^TZ)^{-1}Z^T\varepsilon)^T\right) \\
&= \mathrm{E}\left((Z^TZ)^{-1}Z^T\varepsilon\varepsilon^TZ(Z^TZ)^{-1}\right) \\
&= (Z^TZ)^{-1}Z^T\mathrm{E}\left(\varepsilon\varepsilon^T\right)Z(Z^TZ)^{-1} \\
&= \sigma^2(Z^TZ)^{-1}Z^TZ(Z^TZ)^{-1} \\
&= \sigma^2(Z^TZ)^{-1} \\
&= \frac{\sigma^2}{\|Z\|_2^2}
\end{aligned}
$$

Since we do not know what σ^2 is, we have to use the estimator which we get from taking the 2-norm of the residuals from the full model, and dividing by $N-3$,

$$\hat{\sigma}^2 = \frac{1}{N-3}\|\mathbf{e}_f\|_2^2.$$

Therefore, the value we get for the t_{score} is

$$t_{\text{score}} = \frac{\hat{b}}{\sqrt{\mathrm{Var}(\hat{b})}} = \frac{\hat{b}}{\sqrt{\frac{\frac{1}{N-3}\|\mathbf{e}_f\|_2^2}{\|Z\|_2^2}}} = \frac{\|Z\|_2\hat{b}}{\|\mathbf{e}_f\|/\sqrt{N-3}}.$$

This t_{score} is precisely the square root of the F-statistic which we calculated previously. That is, $t_{\text{score}} = \sqrt{F}$. Furthermore, the F-distribution with degrees of freedom 1 and degrees of freedom $N-3$ shares exactly the same distribution as the Student's t-distribution with degrees of freedom $N-3$. Thus, whenever we reject the null hypothesis that $b=0$ using the F-statistic, we will get exactly the same result as using the t-statistic. This proves an interesting point, namely, that in Granger causality, we can either compare the variance

of the residuals in the full vs. the reduced model to see if one variable Granger causes another, or we can compare the estimated coefficient to zero. However, we have used the assumptions here that x_{k+1} can only depend *linearly* on x_k and y_k, and that the noise terms are independent, identically distributed Gaussians. If there are non-linear dynamics or non-Gaussian noise, then this analysis will not be applicable.

When using the above procedure to determine if x_k influences y_k and vice-versa at the $\alpha = 0.05$-significance level in 1000 simulations, each simulation using 100 data points of the original problem where

$$C = \begin{bmatrix} 0.30 & -0.10 \\ 0.20 & 0.40 \end{bmatrix},$$

we are correct in establishing causality from y to x in 21.0% of simulations, and we are correct in establishing causality from x to y in 53.6% of simulations. Why such a discrepancy? Part of the reason is that b is closer to zero than c, and our N is too small in each simulation. Letting $N = 500$ in each simulation, we are able to detect causality from y to x in 69.9% of simulations, and we are correct in establishing causality from y to x in 99.4% of cases. When $N = 1000$, we are correct in establishing causality from y to x in 92.6% of cases, and we are correct in establishing causality from y to x in 100% of cases. So, when the linear causal effect of one variable on another is greater, we have a better chance of detecting that relationship in Granger analysis. Furthermore, when we have a larger set of observations, then our accuracy increases as well.

Now, suppose we have rejected the null hypothesis that $b = 0$ and we have rejected the null hypothesis that $c = 0$ using the above statistical tests. Then, our estimates a, b, c, and d will be biased, since we were working under the assumption of the null hypothesis, that $b = 0$ and $c = 0$, when estimating these quantities. So, we need a way to estimate the coefficient matrix which will produce an unbiased estimator for the coefficient matrix. We

can do this using the Yule-Walker equations or the Multivariate Least-Squares estimation technique, which we present in the following sections.

3.3 Coefficient Estimation using Yule-Walker equation

One method for estimating the entries of C from (27) uses the *Yule-Walker equations*, which we describe as follows. Multiply both sides of equation (27) by $\mathbf{x}_k^T$ and take expectations of both sides of the equation to get

$$\begin{aligned}
\mathrm{E}\left(\mathbf{x}_{k+1}\mathbf{x}_k^T\right) &= \mathrm{E}\left(C\mathbf{x}_k\mathbf{x}_k^T + \mathbf{G}_{k+1}\mathbf{x}_k^T\right) \\
&= C\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_k^T\right) + \cancel{\mathrm{E}\left(\mathbf{G}_{k+1}\mathbf{x}_k^T\right)} \\
&= C\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_k^T\right)
\end{aligned}$$

We were able to drop the term $\mathrm{E}\left(\mathbf{G}_{k+1}\mathbf{x}_k^T\right)$ in the above calculation since G_{k+1} and H_{k+1} are independent of $\mathbf{x}_k$, and both have mean zero. The resulting equation,

$$\mathrm{E}\left(\mathbf{x}_{k+1}\mathbf{x}_k^T\right) = C\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_k^T\right) \tag{33}$$

is called the Yule-Walker equation. Here, note that $\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_k^T\right)$ is the variance-covariance matrix of $\mathbf{x}_k$. The variance-covariance matrix of two random variables is symmetric semi-positive definite and never singular, unless one random variable is a constant multiple of the other, or one of the random variables is a constant. Therefore, we can solve the Yule-Walker equation for C from Equation (33) to get

$$C = \mathrm{E}\left(\mathbf{x}_{k+1}\mathbf{x}_k^T\right)\left(\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_k^T\right)\right)^{-1}. \tag{34}$$

There is a natural way of using (34) to estimate C. First, average the products $\mathbf{x}_{k+1}\mathbf{x}_k^T$ from the computed realization of (27). This will give an estimate of $\mathrm{E}(\mathbf{x}_{k+1}\mathbf{x}_k^T)$. Then, average $\mathbf{x}_k\mathbf{x}_k^T$, which gives an estimate of $\mathrm{E}(\mathbf{x}_k\mathbf{x}_k^T)$. Finally solve for the inverse of the estimate for $\mathrm{E}(\mathbf{x}_k\mathbf{x}_k^T)$, and multiply the inverse to the left of our estimate for $\mathrm{E}(\mathbf{x}_{k+1}\mathbf{x}_k^T)$ to get an estimate of C. The matrix from the above example was

$$C = \begin{bmatrix} 0.30 & -0.10 \\ 0.20 & 0.40 \end{bmatrix}.$$

Using the 100 data points shown in Figure 27, we get the following estimate for C using the Yule-Walker estimate:

$$\hat{C} = \begin{bmatrix} .37 & -0.05 \\ 0.22 & 0.27 \end{bmatrix}.$$

Note that this is not a very good estimate of the actual coefficient matrix. This is due to effect of two small a segment of the sequence $\{\mathbf{x}_k\}$. We only used 100 data points in the estimate. If we let the simulation run for 10,000 data points, then we get the following estimate:

$$\hat{C} = \begin{bmatrix} 0.32 & -0.11 \\ 0.19 & 0.40 \end{bmatrix}.$$

We can also solve for the variance-covariance matrix of the white noise process $\mathbf{G}_k$, which is $\sigma^2 I$, by multiplying (27) on the right side of the equation by $\mathbf{x}_{k+1}^T$ and taking expectations to get

$$
\begin{aligned}
\mathrm{E}\left(\mathbf{x}_{k+1}\mathbf{x}_{k+1}^T\right) &= C\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_{k+1}^T\right) + \sigma\mathrm{E}\left(\mathbf{G}_{k+1}\mathbf{x}_{k+1}^T\right) \\
\mathrm{E}\left(\mathbf{x}_{k+1}\mathbf{x}_{k+1}^T\right) &= C\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_{k+1}^T\right) + \sigma\mathrm{E}\left(\mathbf{G}_{k+1}\left(C\mathbf{x}_k + \sigma\mathbf{G}_{k+1}\right)^T\right) \\
\mathrm{E}\left(\mathbf{x}_{k+1}\mathbf{x}_{k+1}^T\right) &= C\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_{k+1}^T\right) + \sigma^2\mathrm{E}\left(\mathbf{G}_{k+1}\mathbf{G}_{k+1}^T\right) + \cancel{\sigma^2\mathrm{E}\left(\mathbf{G}_{k+1}\mathbf{x}_k^T\right)C^T} \\
\sigma^2 I &= \mathrm{E}\left(\mathbf{x}_{k+1}\mathbf{x}_{k+1}^T\right) - C\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_{k+1}^T\right) \qquad (35)
\end{aligned}
$$

This provides a method to estimate the variance-covariance matrix of the noise process from a given set of observations by replacing C, $\mathrm{E}\left(\mathbf{x}_k\mathbf{x}_k^T\right)$, and $\mathrm{E}\left(\mathbf{x}_{k+1}\mathbf{x}_k^T\right)$ by their respective estimates in the above equation.

3.4 Coefficient Estimation using Least Squares

We begin with the same system as described in the previous section, and consider a given set of observations from this process, $\{\mathbf{x}_k\}_{k=1}^N$. Now, define the following matrices:

$$
A = \begin{bmatrix} x_N & x_{N-1} & \cdots & x_2 \\ y_N & y_{N-1} & \cdots & y_2 \end{bmatrix}, \quad B = \begin{bmatrix} x_{N-1} & x_{N-2} & \cdots & x_1 \\ y_{N-1} & y_{N-2} & \cdots & y_1 \end{bmatrix}, \quad D = \sigma \begin{bmatrix} G_N & G_{N-1} & \cdots & G_2 \\ H_N & H_{N-1} & \cdots & H_2 \end{bmatrix}.
$$

Then, we can then write

$$
A = CB + D.
$$

Multiplying both sides of this equation by B^T, we get

$$
AB^T = CBB^T + DB^T.
$$

Taking expectations of both sides of this equation, and using the fact that G_k and H_k have mean zero and are uncorrelated with x_j and y_j for $j < k$, we find

$$\mathrm{E}(AB^T) = C\mathrm{E}(BB^T) \implies C = \mathrm{E}(AB^T)\mathrm{E}(BB^T)^{-1}.$$

Finally, we can replace $\mathrm{E}(AB^T)$ and $\mathrm{E}(BB^T)^{-1}$ by their estimates using the set of observations to get

$$\hat{C} = \frac{1}{N-1}AB^T\left(\frac{1}{N-1}BB^T\right)^{-1}$$
$$\hat{C} = AB^T(BB^T)^{-1}. \tag{36}$$

We can also solve for the variance covariance matrix of the white-noise process using the Least Squares estimation technique by re-writing (27) as $\mathbf{G}_{k+1} = \mathbf{x}_{k+1} - C\mathbf{x}_k$, and solving for the variance of both sides of the equation. Replacing these quantities by their estimates, we find that the least squares estimate for the variance-covariance matrix of the white noise process is

$$\sigma^2 I \approx \frac{1}{N-2}(A - \hat{C}B)(A - \hat{C}B)^T. \tag{37}$$

Using the least squares estimation technique to estimate C from the set of observations from the previous section, with $N = 100$, we get

$$\hat{C} = \begin{bmatrix} 0.37 & -0.05 \\ 0.22 & 0.27 \end{bmatrix}$$

When $N = 10,000$, we get

$$\hat{C} = \begin{bmatrix} 0.32 & -0.11 \\ 0.19 & 0.40 \end{bmatrix}$$

These estimates are extremely close to the Yule-Walker estimates. In fact, the Yule-Walker estimate and Least Squares estimate for C share the same normal asymptotic distribution [14]. If we define $\text{vec}(C)$ to be the vector formed by placing the columns of the matrix $\hat{C}$ on top of each other. That is,

$$\text{vec}(C) = \text{vec}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a \\ c \\ b \\ d \end{bmatrix}$$

.

Then, the asymptotic distribution of the estimates for $\hat{C}$ is

$$\sqrt{N}\text{vec}\left(\hat{C} - C\right) \xrightarrow[N\to\infty]{d} \mathcal{N}\left(0, \text{E}(\mathbf{x}_k\mathbf{x}_k^T)^{-1} \otimes \Sigma\right)$$

where $\Sigma = \sigma^2 I$, where $I \in \mathbb{R}^{2\times 2}$ is the identity matrix, is the variance-covariance matrix of the noise vector $\sigma\,[G_k, H_k]^T$, and the $\otimes$ denotes the Kronecker product of the matrices $\text{E}(\mathbf{x}_k\mathbf{x}_k^T)^{-1}$ and $\sigma^2 I$. Using the asymptotic normal distributions of the estimates of the matrix C, we can also form a hypothesis test to determine if $b = 0$ or $c = 0$, using a t-score. This hypothesis test is asymptotically equivalent to the tests described above, however, for small N, these tests will not always agree.

4 Adding Noise to Continuous-Time Deterministic Dynamics

There are generally speaking two types of noise that can be added to a dynamical model. When taking measurements of a dynamical system, noise may be added to these measurements based on intrinsic properties of the measuring device. This type of noise will not influence the dynamics of the system, but will have a blurring effect on the data collected. This type of noise is known as *observational noise*. The other type of noise is known as *dynamical noise*, in which unmeasured variables may have an overall random effect on the dynamics of a variable of interest, thereby changing the properties of the underlying dynamics of the system. An example of this noise which occurs in neuroscience is a pyramidal cell receiving synaptic inputs from thousands of other cells. The large number of synaptic inputs may seem to have an overall random effect (ie. noise) on the behavior of the pyramidal cell. Other types of dynamical noise in neuroscience include conductance fluctuations of ion channels in the cell membrane (intrinsic) and fluctuations in ionic concentrations in the extracellular water (extrinsic) [15]. These examples illustrate a very important point, that knowledge of the basic physical properties governing a dynamical system are needed in order to model noise accurately.

Typically, an additive noisy driving force in a physical model should reflect some underlying physical process occurring. One of the difficulties in modeling noise, and adding it to a deterministic process, is that we are often not able to measure the noise directly in the physical process. However, we can add different types of stochastic processes to a deterministic model, and understand how the presence of these stochastic processes affect the dynamics of the system we are studying.

4.1 The Ornstein-Uhlenbeck process

One of the simplest types of continuous-time stochastic processes is the Wiener process. A *Wiener process*, also known as *Brownian motion*, $\{W_t\}_{t\geq 0}$, is defined by the following three properties:

1. $W_0 = 0$
2. W_t has independent increments with $W_t - W_s$ distributed normally with mean zero and variance $t - s$ for $0 \leq s < t$
3. The random path, $t \mapsto W_t$ is almost surely continuous and almost surely nowhere differentiable, that is, continuous and nowhere differentiable with probability 1

Here are some examples of a Wiener process with 95% confidence interval bounds.

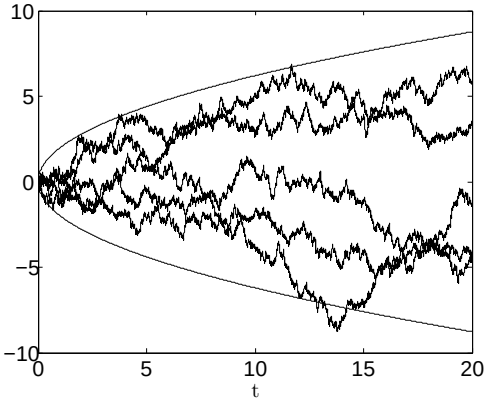

Figure 32: Wiener Process with 95% confidence interval bounds

The probability density function of a Gaussian random variable, X, with mean μ and variance σ^2, where $\sigma > 0$, is defined as

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{38}$$

At a fixed point in time, $t > 0$, the Wiener process is normally distributed with mean 0 and variance t, since $W_t - W_0 = W_t$ is distributed normally with mean 0 and variance

$t - 0 = t$, by definition of the Wiener process. Thus, the probability density function at a fixed time t of a Wiener process is

$$f_{W_t}(w) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{w^2}{2t}\right). \tag{39}$$

Furthermore, we can use the property that a Wiener process has independent increments to calculate its covariance. Let $0 < s < t$. Then,

$$\begin{aligned}
\mathrm{Cov}(W_t, W_s) &= \mathrm{E}\left[(W_t - E(W_t))(W_s - E(W_s))\right] \\
&= \mathrm{E}\left[W_t W_s\right] \\
&= \mathrm{E}\left[((W_t - W_s) + (W_s - W_0))(W_s - 0)\right] \\
&= \cancel{\mathrm{E}\left[(W_t - W_s)(W_s - 0)\right]} + \mathrm{E}\left[(W_s - 0)(W_s - 0)\right] \\
&= \mathrm{Var}(W_s) \\
&= s.
\end{aligned}$$

The cancellation in the calculation is due to the fact that $W_t - W_s$ and $W_s - W_0$ are independent increments of the Wiener process, which implies $\mathrm{E}\left[(W_t - W_s)(W_s - 0)\right] = \mathrm{E}[W_t - W_s]\,\mathrm{E}\left[W_s\right] = 0$.

Note that the variance of the Wiener process grows as time increases, and although the expected value of the Wiener process is always zero, the Wiener process wanders far away from its expected value of zero as time progresses, and in fact the Wiener process can be thought of as the scaling limit of a random walk. A type of mean-reverting stochastic process, in which the process tends to drift towards its long-term mean is the *Ornstein-Uhlenbeck process*, which is defined by the following stochastic differential equation.

$$dX = -\rho(X - \mu)dt + \sigma dW, \tag{40}$$

where $\rho > 0$, $\sigma > 0$, and W denotes the Wiener process. ρ represents the rate at which the Ornstein-Uhlenbeck process reverts towards its mean, μ is the long-term mean of the process, and σ^2 is the variance of the Wiener process. Note that if $\rho = 0$, then X is simply the Wiener process with a variance of σ^2. Note that this equation is given in differential form, since the Wiener process, W_t, is continuous but not differentiable. To solve the Ornstein-Uhlenbeck process numerically, we can discretize the system according to Euler's method as follows

$$X_k - X_{k-1} = -\rho(X_{k-1} - \mu)\Delta t + \sigma(W_k - W_{k-1})$$

$$X_k = (1 - \rho\Delta t)X_{k-1} + \rho\mu\Delta t + \sigma\sqrt{\Delta t}G_k, \;\; k = 1, 2, 3, \ldots \tag{41}$$

where $\{G_k\}_{k\in\mathbb{Z}}$ represents an independent sequence of normally distributed random variables with mean zero and variance 1. In this context, Euler's method is also called the *Euler-Maruyama method*. Note: If we set $\Delta t = 1$ and $\mu = 0$, then this equation is precisely the same equation for the univariate discrete-time stochastic process which Granger analyzed! So, what Granger really studied was how to measure causality between two discretized Ornstein-Uhlenbeck processes. The invariant distribution of the discretized Ornstein-Uhlenbeck process will be Gaussian due to the Gaussian noise. We can solve for the invariant distribution by solving for the mean and variance of X_k as follows.

$$\mathrm{E}(X_k) = (1 - \rho\Delta t)\mathrm{E}(X_k) + \rho\mu\Delta t \implies \mathrm{E}(X_k) = \mu.$$

$$\mathrm{Var}(X_k) = (1 - \rho\Delta t)^2\mathrm{Var}(X_k) + \sigma^2\Delta t \implies \mathrm{Var}(X_k) = \frac{\sigma^2}{2\rho - \rho^2\Delta t}.$$

In the limit as $\Delta t \to 0$, we see that the variance of the invariant distribution approaches $\frac{\sigma^2}{2\rho}$. The invariant distribution of the Ornstein-Uhlenbeck process itself has a Gaussian distribution with mean μ and variance $\frac{\sigma^2}{2\rho}$. Here is a sample solution of the Ornstein-Uhlenbeck process using the Euler-Maruyama method with time step $\Delta t = 0.01$, and parameters $\rho = 1$, $\mu = 0$, and $\sigma = 1$.

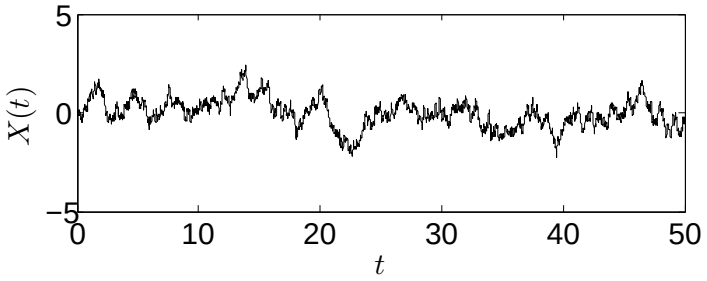

Figure 33: Solution to Ornstein-Uhlenbeck process

As one can see, the process drifts about its mean, but is always pushed back towards its mean. As is the case for the Wiener process, the Ornstein-Uhlenbeck process is almost surely continuous, and almost surely nowhere differentiable.

Using Itō calculus, the Ornstein-Uhlenbeck process can be represented as

$$X(t) = X_0 - \rho \int_0^t (X_s - \mu) ds + \sigma \int_0^t dW_s.$$

The final solution to this equation which uses Itō calculus will be

$$X(t) = X_0 e^{-\rho t} + \mu(1 - e^{-\rho t}) + \int_0^t \sigma e^{\rho(s-t)} dW_s,$$

where the integral with respect to the Wiener process is an *Itō integral*, defined as the left-sided Riemann integral of $\sigma e^{\rho(s-t)}$ with respect to the Wiener measure,

$$\int_0^t \sigma e^{\rho(s-t)} dW_s = \lim_{\Delta t \to 0} \sum_{i=1}^{N} \sigma e^{\rho(t_{i-1}-t)} \Delta W_i,$$

where $\Delta W_i = W_{t_i} - W_{t_{i-1}}$ is a step of the Wiener process across the interval $[t_{i-1}, t_i]$, and $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = t$. The Euler-Maruyama discretization of the Ornstein-Uhlenbeck process evaluated at time t is in fact equivalent to this solution as $\Delta t \to 0$. To see this, recall from Equation (41) that the Euler-Maruyama discretization of the Ornstein-Uhlenbeck process is

$$X_k = (1 - \rho\Delta t)X_{k-1} + \rho\mu\Delta t + \sigma\sqrt{\Delta t}G_k, \;\; k = 1, 2, 3, \ldots$$

Assuming $X(0) = X_0$ is given, we can write X_k in terms of X_0 as follows

$$X_k = (1 - \rho\Delta t)^k X_0 + \mu \sum_{j=0}^{k-1} (1 - \rho\Delta t)^j \rho\Delta t + \sigma\sqrt{\Delta t} \sum_{j=1}^{k} (1 - \rho\Delta t)^{k-j} G_j.$$

We want to evaluate this discretization at time t, where $t = k\Delta t$. Thus, we can replace Δt in the above equation by t/k to get

$$X(t) = \left(1 - \frac{\rho t}{k}\right)^k X_0 + \mu \sum_{j=0}^{k-1} \left(1 - \frac{\rho t}{k}\right)^j \frac{\rho t}{k} + \sigma \sum_{j=1}^{k} \sqrt{\frac{t}{k}} \left(1 - \frac{\rho t}{k}\right)^{k-j} G_j. \tag{42}$$

If we take the limit of the solution in (42) as $k \to \infty$ (Note: This is in effect taking the limit as $\Delta t \to 0$), then the first term in (42) becomes

$$\lim_{k\to\infty} \left(1 - \frac{\rho t}{k}\right)^k X_0 = X_0 e^{-\rho t}.$$

This matches the first term in the Itō integral solution. Furthermore, as $k \to \infty$, the second term in (42) becomes

$$\lim_{k\to\infty} \mu \sum_{j=0}^{k-1} \left(1 - \frac{\rho t}{k}\right)^j \frac{\rho t}{k} = \mu \lim_{k\to\infty} \left(\frac{1 - \left(1 - \frac{\rho t}{k}\right)^k}{1 - \left(1 - \frac{\rho t}{k}\right)} \right) \frac{\rho t}{k}$$

$$= \mu \lim_{k \to \infty} \left(1 - \left(1 - \frac{\rho t}{k} \right)^k \right) = \mu(1 - e^{-\rho t}).$$

This is precisely the second term in the Itō integral solution. Finally, the last term in (42) is by definition a discretization of the Itō integral. Thus, in the limit as $\Delta t \to 0$, the Euler-Maruyama method is equivalent to the Itō integral solution of the Ornstein-Uhlenbeck process.

4.2 Continuous Linear System Driven by Ornstein-Uhlenbeck Noise

Suppose we have a one-dimensional linear system described by the following equation,

$$\frac{dx}{dt} = -ax$$

where $a \geq 0$. Let's assume we are given a single solution to this system and we want to determine if $a > 0$ or $a = 0$. In other words, we want to determine if x self-influences or not. Assuming $a > 0$, and $x(0) = 0$, the solution to this system will be zero for all time. That is, if the system starts at its stable fixed point, then we cannot determine if it self-influences. Now, let's add a noisy driving force to the system in the form of a mean zero Ornstein-Uhlenbeck process with $\rho = 1$, call it $\omega(t)$, to get the following system of equations

$$\frac{dx}{dt} = -ax + \omega \tag{43}$$

$$d\omega = -\omega dt + \sigma dW \tag{44}$$

where W is a Wiener process. Here is a sample solution to this system with $x(0) = 0$, $\omega(0) = 0$, $a = 1$, $\sigma = 1$, calculated using a time step of $\Delta t = 0.001$.

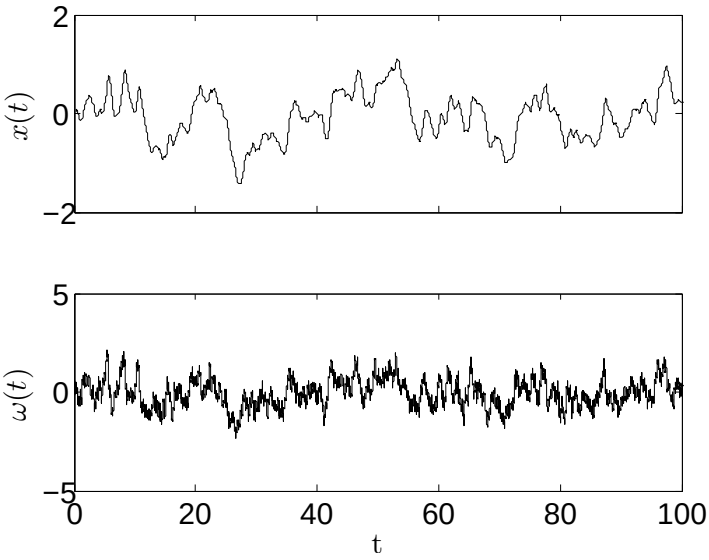

Figure 34: Sample Solution to $x(t)$ and $\omega(t)$

Note that the process $x(t)$ will be almost surely differentiable, since ω is almost surely continuous.

Assume we are given a set of observations of x only, $\{x_k\}_{k=0}^{N}$, where $x_k = x(\Delta t k)$, and from these observations, we want to determine if x self-influences, that is, if $a \neq 0$. A natural way to solve this process would be to fix a value, α, and look at times when $x = \alpha$. This will be a sequence of times. We can estimate $\frac{dx}{dt}$ at those times, and afterwards, take the average of the estimated quantity $\frac{dx}{dt}$ at all points in time in that sequence. Let's call this estimate $\hat{c}$. This equates to estimating the following expectation:

$$\mathrm{E}\left(\frac{dx}{dt}\,\middle|\, x = \alpha\right).$$

One would naturally presume that for $\alpha > 0$, when $x(t)$ cross the point α, it will be driven back towards 0 by the deterministic dynamics in such a way that on the average, it will overpower the random influences of the Ornstein-Uhlenbeck process. However, this is not the case. It is easy to make the mistake of replacing $\frac{dx}{dt}$ with $-ax + \omega$ in the above equation to get

$$\mathrm{E}\left(\frac{dx}{dt}\,\middle|\, x=\alpha\right)=\mathrm{E}\left(-ax+\omega\,\middle|\, x=\alpha\right)=-a\alpha$$

However, this is wrong! From the above example in Figure 34, which shows a segment of a solution to this system from time $t=0$ to $t=100$, (the full simulation was calculated from time $t=0$ to $t=500$), for $\alpha=0.1$, our estimate for $\mathrm{E}\left(\frac{dx}{dt}\,\middle|\, x=0.1\right)$ is $\hat{c}=0.0013$. Dividing this number by $-\alpha=-0.1$ gives -0.013. This is not even close to the actual value of a, which is $a=1$. So, what went wrong? The fallacy lies in the fact that conditioning $\frac{dx}{dt}$ on $x=\alpha$ is in fact not independent of the Ornstein-Uhlenbeck process, even though the Ornstein-Uhlenbeck process itself, $\omega(t)$, is completely independent of the dynamics of x. To see why this is true, we need to calculate the distribution of the conditional random variable $(\omega(t)|x(t)=\alpha)$, and in order to do this, we need to calculate the distribution of the bivariate random variable $(x(t),\omega(t))$.

We know the Euler-Maruyama discretization to the Ornstein-Uhlenbeck process, ω, is

$$\omega_{k+1}=\omega_k-\Delta t\omega_k+\sigma\sqrt{\Delta t}G_{k+1}. \tag{45}$$

We can also discretize the differential equation for x, (43), as follows:

$$\frac{x_{k+1}-x_k}{\Delta t}=-ax_k+\omega_k \tag{46}$$

In order to solve for the probability distribution of the bivariate random variable (x,ω), we will need to determine how the mean and variance-covariance matrix evolve over time with given initial conditions. Since $(x(t),\omega(t))$ will be a bivariate Gaussian random variable, these are the quantities we need to know to fully describe its distribution. Define the mean of $x(t)$ and $\omega(t)$ as $\mathrm{E}(x(t))=m_x(t)$, and $\mathrm{E}(\omega(t))=m_\omega(t)$. Similarly, $\mathrm{Var}(x(t))=V_x(t)$, $\mathrm{Var}(\omega(t))=V_\omega(t)$, and $\mathrm{Cov}(x(t),\omega(t))=C_{x\omega}(t)$. Thus,

$$E\begin{bmatrix} x(t) \\ \omega(t) \end{bmatrix} = \begin{bmatrix} m_x(t) \\ m_\omega(t) \end{bmatrix}, \quad \text{cov}\begin{bmatrix} x(t) \\ \omega(t) \end{bmatrix} = \begin{bmatrix} V_x(t) & C_{x\omega}(t) \\ C_{x\omega}(t) & V_\omega(t) \end{bmatrix}$$

In order to solve for the evolution of $m_x(t)$ and $m_\omega(t)$, we first take expectations of both sides of Equation (45) to get

$$\mathrm{E}(\omega_{k+1}) = \mathrm{E}\left(\omega_k - \Delta t\omega_k + \sigma\sqrt{\Delta t}G_{k+1}\right) \implies \mathrm{E}(\omega_{k+1}) - \mathrm{E}(\omega_k) = -\Delta t\mathrm{E}(\omega_k).$$

Here, we have used the fact that G_k is a mean zero Gaussian for all k. Dividing through by Δt, and taking the limit as $\Delta t \to 0$, we find

$$\frac{\mathrm{E}(\omega_{k+1}) - \mathrm{E}(\omega_k)}{\Delta t} = -\mathrm{E}(\omega_k) \implies \boxed{\frac{dm_\omega}{dt} = -m_\omega}$$

Doing the same to Equation (46), we find

$$\frac{\mathrm{E}(x_{k+1}) - \mathrm{E}(x_k)}{\Delta t} = -a\mathrm{E}(x_k) + \mathrm{E}(\omega_k) \implies \boxed{\frac{dm_x}{dt} = -am_x + m_\omega}$$

Therefore, the time evolution of the means of this process, $m_x(t)$ and $m_\omega(t)$, are

$$m_\omega(t) = m_\omega(0)e^{-t}. \tag{47}$$

$$m_x(t) = m_x(0)e^{-at} + \frac{m_\omega(0)}{a-1}\left(e^{-t} - e^{-at}\right). \tag{48}$$

We can solve for the time evolution of the $V_x(t)$, $C_{x\omega}(t)$ and $V_\omega(t)$ in a similar way. First square both sides of Equation (46) to get

$$(x_{k+1}-x_k)^2 = (-ax_k\Delta t + \omega_k\Delta t)^2$$

$$x_{k+1}^2 - 2x_{k+1}x_k + x_k^2 = a^2\Delta t^2 x_k^2 - 2a\Delta t^2 x_k\omega_k + \Delta t^2\omega_k^2$$

Taking expectations of both sides of this equation and dropping all the terms with Δt^2, we get

$$\mathrm{E}(x_{k+1}^2) - 2\mathrm{E}(x_{k+1}x_k) + \mathrm{E}(x_k^2) = 0$$

$$\mathrm{E}\left(x_{k+1}^2\right) - 2\mathrm{E}\left((x_k - ax_k\Delta t + \Delta t\omega_k)x_k\right) + \mathrm{E}(x_k^2) = 0$$

$$\mathrm{E}(x_{k+1}^2) - 2\mathrm{E}(x_k)^2 + 2a\Delta t\mathrm{E}(x_k^2) - 2\Delta t\mathrm{E}(x_k\omega_k) + \mathrm{E}(x_k^2) = 0.$$

$$\mathrm{E}\left(x_{k+1}^2\right) - \mathrm{E}(x_k^2) = -2a\Delta t\mathrm{E}(x_k^2) + 2\Delta t\mathrm{E}(x_k\omega_k)$$

Dividing through by Δt and taking the limits as $\Delta t \to 0$, we find

$$\frac{\mathrm{E}\left(x_{k+1}^2\right) - \mathrm{E}(x_k^2)}{\Delta t} = -2a\mathrm{E}(x_k^2) + 2\mathrm{E}(x_k\omega_k)$$

$$\boxed{\frac{dV_x}{dt} = -2aV_x + 2C_{x\omega}}$$

Similarly, we can solve for the differential equation defining the time evolution of V_ω and $C_{x\omega}$ in the following way. Multiplying both sides of Equation (46) by ω_{k+1},we get

$$\left(\frac{x_{k+1}-x_k}{\Delta t}\right)\omega_{k+1} = (-ax_k + \omega_k)\omega_{k+1}$$

$$\left(\frac{x_{k+1}\omega_{k+1} - x_k\omega_{k+1}}{\Delta t}\right) = (-ax_k + \omega_k)\omega_{k+1}$$

$$\left(\frac{x_{k+1}\omega_{k+1} - x_k\left(\omega_k + \Delta t\omega_k - \sigma\sqrt{\Delta t}G_{k+1}\right)}{\Delta t}\right) = (-ax_k + \omega_k)\left(\omega_k - \Delta t\omega_k + \sigma\sqrt{\Delta t}G_{k+1}\right)$$

$$\frac{x_{k+1}\omega_{k+1} - x_k\omega_k}{\Delta t} = (-ax_k + \omega_k)\left(\omega_k - \Delta t\omega_k + \sigma\sqrt{\Delta t}G_{k+1}\right) - x_k\omega_k + \sigma\frac{x_k G_{k+1}}{\sqrt{\Delta t}}$$

Taking expectations of both sides of this equation, and dropping all the terms that involve $x_k G_{k+1}$ and $\omega_k G_{k+1}$, since G_{k+1} is uncorrelated with x_k and ω_k,

$$\frac{\mathrm{E}(x_{k+1}\omega_{k+1}) - \mathrm{E}(x_k\omega_k)}{\Delta t} = -a\mathrm{E}(x_k\omega_k) + a\Delta t\mathrm{E}(x_k\omega_k) + \mathrm{E}(\omega_k^2) - \Delta t\mathrm{E}(\omega_k^2) - \mathrm{E}(x_k\omega_k).$$

In the limit as $\Delta t \to 0$, this becomes

$$\boxed{\frac{dC_{x\omega}}{dt} = -(1+a)C_{x\omega} + V_\omega}$$

Finally, we solve for dV_ω/dt by multiplying Equation (45) by ω_{k+1},

$$\omega_{k+1}\omega_{k+1} = (\omega_k - \Delta t\omega_k + \sigma\sqrt{\Delta t}G_{k+1})(\omega_k - \Delta t\omega_k + \sigma\sqrt{\Delta t}G_{k+1})$$

$$\omega_{k+1}^2 = \omega_k^2 - 2\Delta t\omega_k^2 + 2\sigma\sqrt{\Delta t}\omega_k G_{k+1} + \Delta t^2\omega_k^2 - 2\Delta t\sqrt{\Delta t}\omega_k G_{k+1} + \sigma^2\Delta t G_{k+1}^2$$

$$\frac{\omega_{k+1}^2 - \omega_k^2}{\Delta t} = -2\omega_k^2 + 2\sigma\frac{\omega_k G_{k+1}}{\sqrt{\Delta t}} + \Delta t\omega_k^2 - 2\sqrt{\Delta t}\omega_k G_{k+1} + \sigma^2 G_{k+1}^2$$

Again, taking expectations, and dropping all terms with $\omega_k G_{k+1}$, and using the fact that $\mathrm{E}(G_{k+1}^2) = 1$, we get

$$\frac{\mathrm{E}(\omega_{k+1}^2) - \mathrm{E}(\omega_k^2)}{\Delta t} = -2\mathrm{E}(\omega_k^2) + \Delta t\mathrm{E}(\omega_k^2) + \sigma^2.$$

In the limit as $\Delta t \to 0$, this becomes

$$\boxed{\frac{dV_\omega}{dt} = -2V_\omega + \sigma^2}.$$

Suppose we are given deterministic initial conditions to this problem, where $x(0) = 0$, and $\omega(0) = 0$. Then, $m_x(t) = 0$ and $m_\omega(t) = 0$ for all t. Also, $V_\omega(0) = V_x(0) = C_{x\omega}(0) = 0$. Solving the above differential equations, we find

$$V_\omega(t) = \frac{\sigma^2}{2}\left(1 - e^{-2t}\right) \tag{49}$$

$$C_{x\omega}(t) = \frac{\sigma^2}{2(a^2-1)}\left(a - 1 - (a+1)e^{-2t} + 2e^{-(a+1)t}\right) \tag{50}$$

$$V_x(t) = \frac{\sigma^2}{2a(a-1)(a^2-1)}\left((a-1)^2 - a(a+1)e^{-2t} + 4ae^{-(a+1)t} + (2a^2-5a-1)e^{-2at}\right). \tag{51}$$

At any given time $t > 0$, the solution to the system of equations in (43) and (44) will be a bivariate Gaussian random variable, since the noisy force driving the process in the Ornstein-Uhlenbeck process was Gaussian. Thus, the solution to this system of equations with initial condition $x(0) = 0, \omega(0) = 0$ is Gaussian for all time $t > 0$ with mean zero and variance-covariance matrix at time t

$$\begin{bmatrix} V_x(t) & C_{x\omega}(t) \\ C_{x\omega}(t) & V_\omega(t) \end{bmatrix}.$$

However, we are only really interested in the steady state distribution, after time has got very large. As t gets large, this variance-covariance matrix approaches the following steady-state covariance matrix

$$\sigma^2 \begin{bmatrix} \frac{1}{2a(a+1)} & \frac{1}{2(a+1)} \\ \frac{1}{2(a+1)} & \frac{1}{2} \end{bmatrix} \tag{52}$$

Note that the variance of the Ornstein-Uhlenbeck process in the steady-state is independent of a. It will aways be $\frac{\sigma^2}{2}$. This is because the Ornstein-Uhlenbeck process itself is

independent of the dynamical process $x(t)$.

Our original problem was the following. What is the distribution of the conditional random variable $(\omega(t)|x(t)=\alpha)$? We know the distribution of $(x(t),\omega(t))$ is bivariate Gaussian. Thus, from section 3.1, we know the conditional distribution of $(\omega(t)|x(t)=\alpha)$ is Gaussian with mean

$$\mu_c = m_\omega + C_{x\omega}V_x^{-1}(\alpha - m_x)$$

and variance

$$V_c = V_\omega - C_{x\omega}V_x^{-1}C_{x\omega}.$$

Let's calculate $\mu_c = \mathrm{E}(\omega(t)|x(t)=\alpha)$. Since $m_\omega = \mathrm{E}(\omega) = 0$, and $m_x = \mathrm{E}(x) = 0$,

$$\mu_c = \frac{\sigma^2}{2(a+1)}\frac{2a(a+1)}{\sigma^2}\alpha = a\alpha.$$

Therefore, during the steady state invariant distribution of the process,

$$\begin{aligned}
\mathrm{E}\left(\frac{dx}{dt}\middle| x(t)=\alpha\right) &= \mathrm{E}\left(-ax(t)+\omega(t)|x(t)=\alpha\right) \\
&= \mathrm{E}(-ax(t)|x(t)=\alpha) + \mathrm{E}(\omega(t)|x(t)=\alpha) \\
&= -a\alpha + a\alpha \\
&= 0.
\end{aligned}$$

This will be true regardless of the value of α! This is a very strange fact. This was

only during the steady state invariant distribution of the process. However, the process will approach its steady state invariant distribution very quickly. For example, when $a = 1$, we have $C_{x\omega}(t) = \frac{1}{4}\sigma^2\left(1 - e^{-2t}(2t+1)\right)$ and $V_x(t) = \frac{1}{4}\sigma^2\left(1 - e^{-2t}(2t^2+t+1)\right)$. So,

$$\mu_c = \frac{1 - e^{-2t}(2t+1)}{1 - e^{-2t}(2t^2+t+1)}\alpha.$$

Let's take a look at the graphs of $1 - e^{-2t}(2t+1)$ and $1 - e^{-2t}(2t^2+t+1)$.

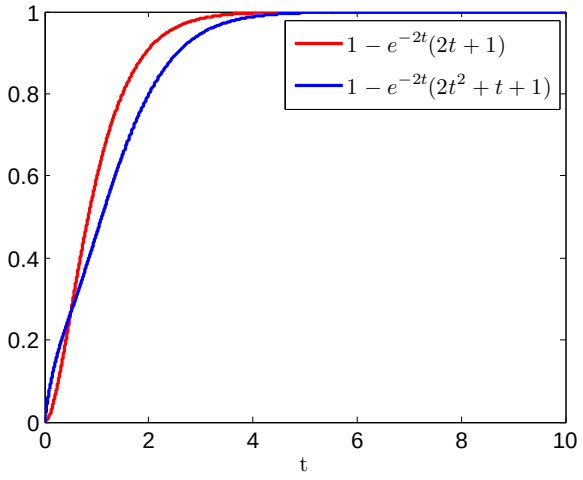

Figure 35: Function Comparison

Both of these functions approach 1 quite rapidly, and are indistinguishable after about time $t = 5$. For larger values of a, the process will approach the steady state even more quickly. This is a very counter-intuitive result, and illustrates an interesting point, namely, that when working with continuous-time stochastic processes, conditioning the derivative of a stochastic dynamic variable on its present value will not be causality revealing. When using Granger causality for a discrete-time stochastic process, we worked from the mathematical fact that conditioning the present on the past was independent of the noise, that is, if $x_{k+1} = ax_k + \sigma G_k$, where G_k is a standard normal Gaussian process with mean zero,

$$\mathrm{E}\left(x_{k+1}|x_k = x\right) = \mathrm{E}\left(ax_k + \sigma G_k|x_k = x\right) = \mathrm{E}(ax_k|x_k = x) + \cancel{\mathrm{E}(\sigma G_k|x_k = x)} = ax.$$

For the continuous-time stochastic process driven by Ornstein-Uhlenbeck noise, conditioning the derivative of x on the present state of x is not independent of the noise process. The point is that there is a difference between white noise (which does not remember its past) and Ornstein-Uhlenbeck noise (which forgets its past rapidly, but not immediately). To express this point slightly differently, we write the discretized linear system driven by Ornstein-Uhlenbeck noise as

$$\frac{x_{k+1} - x_k}{\Delta t} = -a x_k + \omega_k$$

Solving for x_{k+1} and conditioning on x_k, we find

$$x_{k+1}|x_k = (x_k - a\Delta t x_k + \Delta t \omega_k)\,|x_k = (1 - a\Delta t x_k)x_k + \Delta t \omega_k | x_k.$$

Now, as we showed earlier, ω_k is not independent of x_k in the invariant distribution. Therefore, regressing the present on the past is not independent of the noise process, and we cannot use this extension of Granger analysis for this problem.

So, we have to consider a different method for determining if x influences itself, having only observations of the process $x(t)$. What if we again fix $x(t) = \alpha$, assuming the process has reached the steady state, and look at the variance of the random variable $\left(\frac{dx}{dt}\middle| x(t) = \alpha\right)$ and compare this to the variance of the random variable $\frac{dx}{dt}$, without conditioning? With conditioning, we get:

$$
\begin{aligned}
\mathrm{Var}\left(\frac{dx}{dt}\middle| x(t)=\alpha\right) &= \mathrm{Var}\left(-ax(t)+\omega(t)|x(t)=\alpha\right) \\
&= a^2\mathrm{Var}(x(t)|x(t)=\alpha)+\mathrm{Var}(\omega(t)|x(t)=\alpha) \\
&\quad -2a\mathrm{Cov}(x(t)|x(t)=\alpha,\omega(t)|x(t)=\alpha) \\
&= V_\omega - C_{x\omega}V_x^{-1}C_{x\omega} \\
&= \frac{\sigma^2}{2}-\frac{\cancel{\sigma^2}}{2(a+1)}\frac{2a(a+1)}{\cancel{\sigma^2}}\frac{\sigma^2}{2(a+1)} \\
&= \frac{\sigma^2}{2}-\frac{a\sigma^2}{2(a+1)} \\
&= \frac{\sigma^2}{2(a+1)}.
\end{aligned}
$$

Without conditioning, we get:

$$
\begin{aligned}
\mathrm{Var}\left(\frac{dx}{dt}\right) &= \mathrm{Var}\left(-ax(t)+\omega(t)\right) \\
&= a^2\mathrm{Var}\left(x(t)\right)+\mathrm{Var}(\omega(t))-2a\mathrm{Cov}(x(t),\omega(t)) \\
&= a^2\frac{\sigma^2}{2a(a+1)}+\frac{\sigma^2}{2}-2a\frac{\sigma^2}{2(a+1)} \\
&= \frac{\sigma^2}{2a(a+1)}\left(a^2+a(a+1)-2a^2\right) \\
&= \frac{\sigma^2}{2a(a+1)}a \\
&= \frac{\sigma^2}{2(a+1)}.
\end{aligned}
$$

This shows that conditioning the random variable $\frac{dx}{dt}$ on a fixed value of $x(t)$ does not provide any insight to determining if x self-influences. However, we know the variance of $x(t)$ in the steady state, as well as the variance of $\frac{dx}{dt}$ in the steady state, namely

$$\mathrm{Var}(x(t)) = \frac{\sigma^2}{2a(a+1)}, \quad \mathrm{Var}\left(\frac{dx}{dt}\right) = \frac{\sigma^2}{2(a+1)}.$$

Combining these two equations, and solving for a, we find

$$a = \frac{\mathrm{Var}\left(\frac{dx}{dt}\right)}{\mathrm{Var}(x(t))}. \tag{53}$$

This provides a method to estimate a, only having information about the process $x(t)$! Back to the original example, with $a = 1$. Using the same set of observations, $\{x_k\}_{k=0}^{N}$, where $x_k = x(\Delta t k)$, $\Delta t = 0.001$ and $N = 500000$ (so, the simulation ran until $t = 500$), we can first estimate the derivative $\frac{dx}{dt}$ using a second-order difference scheme, and from there estimate the variance of $\frac{dx}{dt}$ and $x(t)$. When doing so, we get the estimate $\hat{a} = 0.9263$. In order to understand how accurate these estimates are, we need to know the distribution of the estimate $\hat{a}$. However, we will not discuss how to do this.

Granger studied how to analyze causality between two discretized stochastic dynamical systems. His approach to measuring causality lies in the assumption that if one variable is linearly dependent on the past of another variable, then the correlation between the current value of that variable and the past value of the other should be non-zero. However, if the underlying causal relationship between two variables is not linear, then we cannot use the traditional method of Granger causality for discrete-time stochastic processes. There exist non-linear Granger causality tests [16] [17], as well other methods entirely, for example dynamical causal modeling [18]. Here, however, we will only present some basic ideas for how one could measure causality in noise-driven oscillators, and what complications one can run into while doing so.

4.3 Phase Oscillators Driven by Continuous-Time Noise Processes

Suppose we have two phase oscillators, θ_1 and θ_2, which start in a phase-locked state. From the first section, we know that it is impossible to detect causal interactions between these two phase oscillators when they are in their phase-locked state. Thus, we need some type of noisy driving force to knock the system of the phase-locked state. One way to do this is to add a Poisson stream of delta functions to both the dynamics of θ_1 and θ_2. In the general context of neuroscience, Poisson processes are often used to model incoming signals from neighboring neurons, which arrive at highly irregular time intervals. The model is defined by the following system,

$$\frac{d\theta_1}{dt} = \omega_1 + c_{12}\sin(\theta_2 - \theta_1), \quad \theta_1(t) = \theta_1(t) + \frac{\pi}{10}\sum_{k=1}^{\infty}\delta\left(t - T_k^{(1)}\right) \tag{54}$$

$$\frac{d\theta_2}{dt} = \omega_2 + c_{21}\sin(\theta_1 - \theta_2), \quad \theta_2(t) = \theta_2(t) + \frac{\pi}{10}\sum_{k=1}^{\infty}\delta\left(t - T_k^{(2)}\right), \tag{55}$$

where $T_1^{(1)}, T_2^{(1)}, \ldots$ and $T_1^{(2)}, T_2^{(2)}, \ldots$ are arrival times of the Poisson process, and as before $c_{12}, c_{21} \geq 0$. Whenever $t = T_k^{(i)}$, for $k = 1, 2, \ldots$, $i = 1, 2$ the phase of the corresponding oscillator is increased by a factor of $\frac{\pi}{10}$. Here is a solution to the above system which begins in the phase-locked state, as well as the corresponding plots for $\frac{d\cos(\theta_1)}{dt}$ and $\frac{d\cos(\theta_2)}{dt}$.

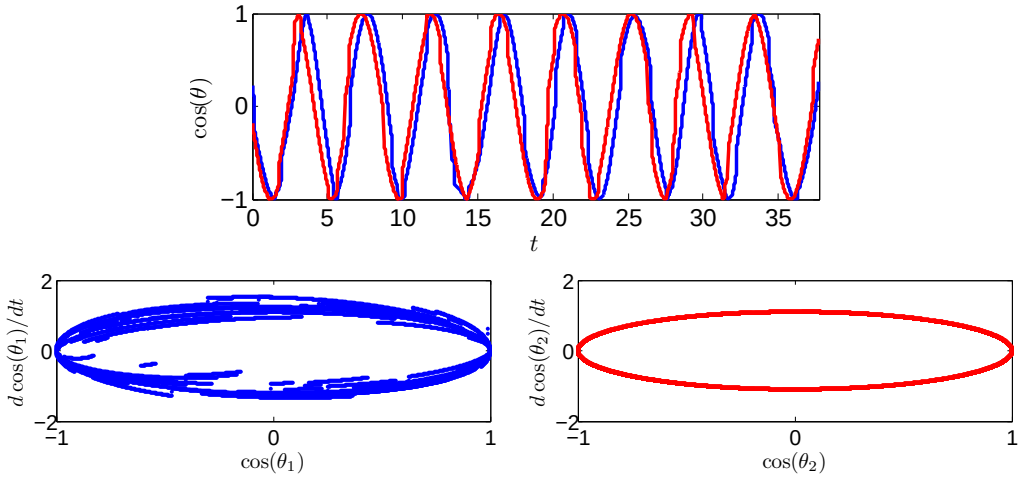

Figure 36: Phase Oscillators Driven by Poisson Noise

During the inter-arrival times of the Poisson process, we clearly are able to detect that $\cos(\theta_2)$ influences $\cos(\theta_1)$, but not vice-versa. The Poisson process continuously knocks the system out of the stable phase-locked state, and the approach back to the phase-locked state reveals who is influencing whom in this system.

Now, we make the noisy driving force for coupled phase oscillators an Ornstein-Uhlenbeck process, and model the stochastic system as follows.

$$\frac{d\theta_1}{dt} = \omega_1(t) + c_{12}\sin(\theta_2 - \theta_1) \tag{56}$$

$$\frac{d\theta_2}{dt} = \omega_2(t) + c_{21}\sin(\theta_1 - \theta_2) \tag{57}$$

where $\omega_1(t)$ and $\omega_2(t)$ are Ornstein-Uhlenbeck processes, and as before $c_{12}, c_{21} \geq 0$. Suppose the Ornstein-Uhlenbeck processes, $\omega_1(t)$ and $\omega_2(t)$ have mean $\hat{\omega}_1$ and $\hat{\omega}_2$, variances σ_1^2 and σ_2^2, and parameters, $\rho_1 = 1$ and $\rho_2 = 1$. That is,

$$d\omega_1 = -(\omega_1 - \hat{\omega}_1)dt + \sigma_1 dW_1, \tag{58}$$

$$d\omega_2 = -(\omega_2 - \hat{\omega}_2)dt + \sigma_2 dW_2, \tag{59}$$

where $\sigma_1 > 0$, $\sigma_2 > 0$, and W_1 and W_2 are independent Wiener processes. Suppose $c_{12} + c_{21} \geq |\hat{\omega}_2 - \hat{\omega}_1|$, so that if $\sigma_1 = \sigma_2 = 0$, the system would reach a stable phase-locked state. Here is a sample solution to such a system.

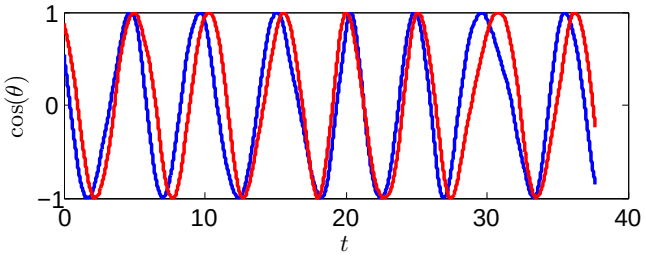

Figure 37: Phase Oscillators Driven by OU noise, $c_{12} = 0.5$, $c_{21} = 0$, $\hat{\omega}_1 = 1.5$, $\hat{\omega}_2 = 1.1$, $\sigma_1 = \sigma_2 = 0.3$

The corresponding plots of the estimated derivatives of $\cos(\theta_1)$ and $\cos(\theta_2)$ are pictured here.

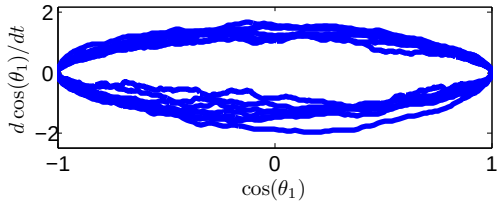

Figure 38: $\cos(\theta_1)$ vs. $d\cos(\theta_1)/dt$

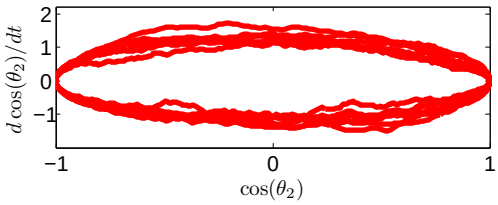

Figure 39: $\cos(\theta_2)$ vs. $d\cos(\theta_2)/dt$

It appears that the plot of $d\cos(\theta_1)/dt$ varies more than $d\cos(\theta_2)/dt$, however it is not immediately clear if this implies causality from $\cos(\theta_2)$ to $\cos(\theta_1)$. Due to the noisy driving forces of the Ornstein-Uhlenbeck processes, we are no longer able to deduce with any certainty the direction of causality simply by looking at these two graphs. So, we have to consider a different method of deducing directional causality in this case. To do so, we can try to incorporate some of the ideas of Granger causality in our analysis. For linear discrete-time stochastic processes, Granger causality rests upon the idea of regressing the present of one variable on its own past, and then on the past of another variable. If the second regression provides a more accurate prediction, then the other variable is said to Granger cause the first variable. Let's see if we can apply this idea to the phase oscillators.

The first idea is to fix $\cos(\theta_1(t))$, however, when we fix $\cos(\theta_1(t))$ at a value $\alpha \in (-1,1)$, and look at $\frac{\cos(\theta_1(t))}{dt}$ at times where $\cos(\theta_1(t))$ is close to α, then the histogram plot will be bimodal. For example, when letting the above simulation run for $10{,}000\pi$ units of time, and plotting the density histogram of $\frac{d\cos(\theta_1(t))}{dt}$ at times where $\cos(\theta_1(t))$ is near $\alpha = 0.5$, then we get the following histogram plot

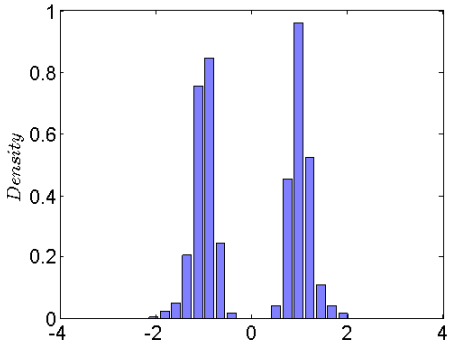

Figure 40: Density Histogram of $d\cos(\theta_1(t))/dt$ whenever $\cos(\theta_1(t)) \approx 0.5$

We can also fix $\cos(\theta_1(t))$ at times where $\cos(\theta_1(t))$ is near $\alpha = 0.5$, and $\frac{d\cos(\theta_1(t))}{dt} < 0$. When we do this, we get the following density histogram plot.

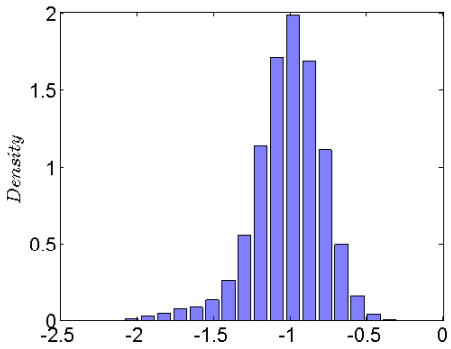

Figure 41: Density Histogram of $d\cos(\theta_1(t))/dt$ whenever $\cos(\theta_1(t)) \approx 0.5$ and $d\cos(\theta_1(t))/dt < 0$

Now, we can also pick times within this set whenever $\cos(\theta_2)$ is near a certain value, call it β. What we are doing here, in effect, is regressing $d\cos(\theta_1(t))/dt$ on $\cos(\theta_1(t))$, and then regressing $d\cos(\theta_1(t))/dt$ on both $\cos(\theta_1(t))$ and $\cos(\theta_2(t))$. For $\beta = -0.5$, we get the following histogram density plot, which we overlay on the above plot.

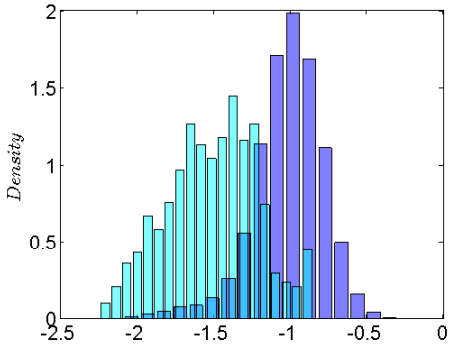

Figure 42: Density Histogram of $d\cos(\theta_1(t))/dt$ whenever $\cos(\theta_1(t)) \approx 0.5$ and $d\cos(\theta_1(t))/dt < 0$ (in blue) and the subset where $\cos(\theta_2) \approx -0.5$ (in cyan)

Here, we see a clear difference in the distribution of the density histogram plots, which leads us to believe that $\cos(\theta_2)$ influences $\cos(\theta_1)$. Now, we do the same analysis to see if $\cos(\theta_1)$ influences $\cos(\theta_2)$. First, we fix $\cos(\theta_2(t))$ at times where $\cos(\theta_2(t))$ is near -0.5 and $\frac{d\cos(\theta_2)}{dt} < 0$. When doing this, we get the following density histogram plot.

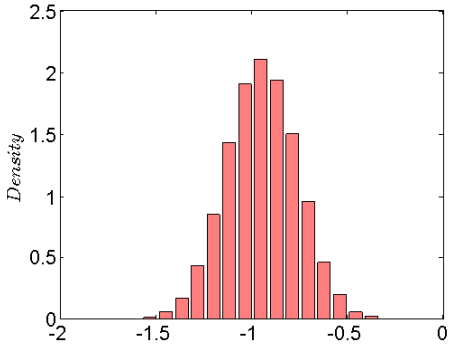

Figure 43: Density Histogram of $d\cos(\theta_2(t))/dt$ whenever $\cos(\theta_2(t)) \approx -0.5$ and $d\cos(\theta_2)/dt < 0$.

As before, we pick times within this set whenever $\cos(\theta_1)$ is near a certain value, let it be 0.5, and overlay the histogram density plot on the above histogram plot to get the following.

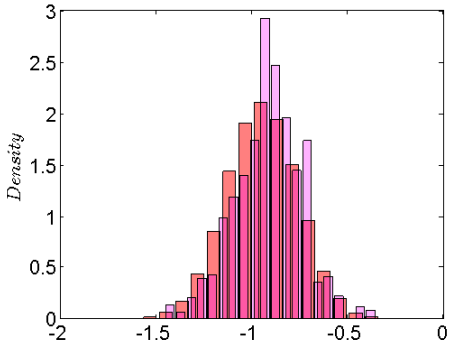

Figure 44: Density Histogram of $d\cos(\theta_2(t))/dt$ whenever $\cos(\theta_2(t)) \approx -0.5$ and $d\cos(\theta_2)/dt < 0$ (in red) and the subset where $\cos(\theta_1) \approx 0.5$ (in magenta)

These two density histogram plots look very similar. Hence, these plot correctly suggest that $\cos(\theta_2)$ influences $\cos(\theta_1)$ but $\cos(\theta_1)$ does not influence $\cos(\theta_1)$. However, in order to test this, we would need to know the underlying distribution of these dynamic processes, which we do not. Also, in order for this type of analysis to be possible, we need to let the simulation run for a long enough time, so that the trajectory passes through the target range enough times, and we need strong enough noise, so that the process gets kicked out of the attracting phase-locked state enough.

What if there is mutual causation in the phase oscillators model? Then, we would expect that both of the density histogram plots should be different. Let's try it. Setting $c_{12} = 0.2$ and $c_{21} = 0.2$ and leaving all other parameters the same as they were in the above simulation, we get the following density histogram plots.

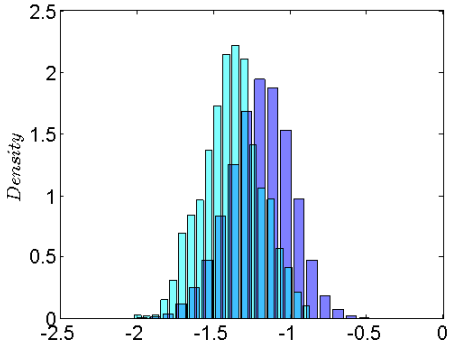

Figure 45: Density Histogram of $d\cos(\theta_1(t))/dt$ whenever $\cos(\theta_1(t)) \approx 0.5$ and $d\cos(\theta_1(t))/dt < 0$ (in blue) and the subset where $\cos(\theta_2) \approx -0.5$ (in cyan)

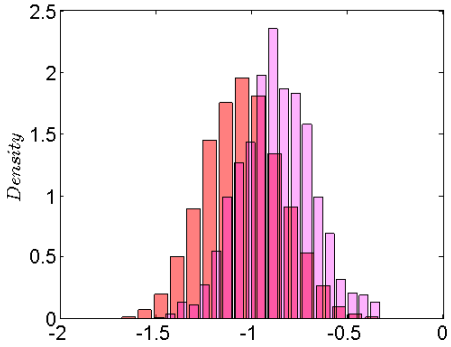

Figure 46: Density Histogram of $d\cos(\theta_2(t))/dt$ whenever $\cos(\theta_2(t)) \approx -0.5$ and $d\cos(\theta_2)/dt < 0$ (in red) and the subset where $\cos(\theta_1) \approx 0.5$ (in magenta)

Here, there is a clear difference in both of the histogram plots! When setting $c_{12} = 0$ and $c_{21} = 0.5$, and leaving all other parameters the same as they in the above simulation, we get the following density histogram plots.

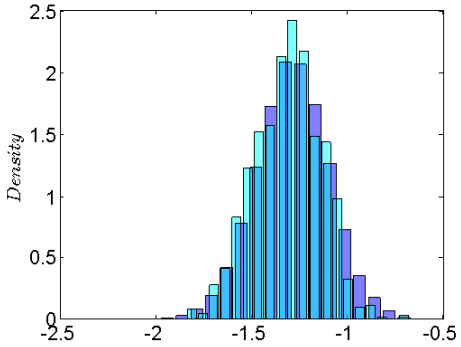

Figure 47: Density Histogram of $d\cos(\theta_1(t))/dt$ whenever $\cos(\theta_1(t)) \approx 0.5$ and $d\cos(\theta_1(t))/dt < 0$ (in blue) and the subset where $\cos(\theta_2) \approx -0.5$ (in cyan)

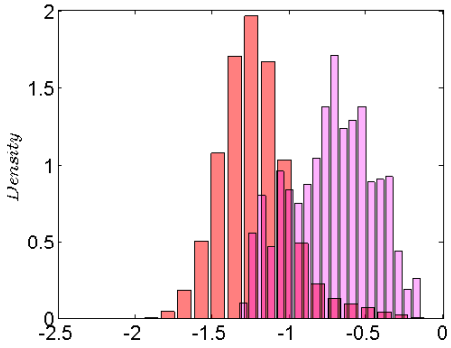

Figure 48: Density Histogram of $d\cos(\theta_2(t))/dt$ whenever $\cos(\theta_2(t)) \approx -0.5$ and $d\cos(\theta_2)/dt < 0$ (in red) and the subset where $\cos(\theta_1) \approx 0.5$ (in magenta)

Here, we see a clear difference in the histogram density plots in Figure 48, however, we

do not see a difference in the histogram density plots in Figure 47. This correctly suggests that $\cos(\theta_1)$ influences $\cos(\theta_2)$, but $\cos(\theta_2)$ does not influences $\cos(\theta_1)$. So, it would seem that this extension of Granger causality works, however, without knowing the underlying distribution of these estimates, we cannot set up a statistical test to determine causality.

4.4 Wilson-Cowan Oscillators Driven by Continuous-Time Noise Processes

We would like to use this same approach to measure causality in a coupled Wilson-Cowan oscillator driven by continuous-time noise. We use the same model as presented in section 2.4, and now add Ornstein-Uhlenbeck drive to all four variables of the system, $E^{(1)}$, $I^{(1)}$, $E^{(2)}$, and $I^{(2)}$. That is,

$$\begin{aligned}
\frac{dE^{(1)}}{dt} &= \frac{E_\infty^{(1)} - E^{(1)}}{5} + \omega_1(t), \\
\frac{dI^{(1)}}{dt} &= \frac{I_\infty^{(1)} - I^{(1)}}{10} + \omega_2(t), \\
\frac{dE^{(2)}}{dt} &= \frac{E_\infty^{(2)} - E^{(2)}}{5} + \omega_3(t), \\
\frac{dI^{(2)}}{dt} &= \frac{I_\infty^{(2)} - I^{(2)}}{10} + \omega_4(t),
\end{aligned}$$

where $\omega_1(t)$, $\omega_2(t)$, $\omega_3(t)$, and $\omega_4(t)$ are independent Ornstein-Uhlenbeck processes. Assume we are given only observations of the excitatory components of this network, $E^{(1)}$ and $E^{(2)}$, and want to determine if there is causal relationship between these two components. So, we are only observing two variables of a phase space which consists of eight variables, $E^{(1)}$, $I^{(1)}$, $E^{(2)}$, $I^{(2)}$, ω_1, ω_2, ω_3, and ω_4. But, our situation is not hopeless. We can try to use the same approach from the last section to measure causality between the excitatory components of this network. Here is a sample solution of $E^{(1)}(t)$ and $E^{(2)}(t)$ for

$t \in [0ms, 500ms]$, where the independent Ornstein-Uhlenbeck processes ω_1, ω_2, ω_3, and ω_4 all have mean zero and variance $\sigma^2 = 9.0$, and the same mean-reverting rate, $\rho = 1.0$.

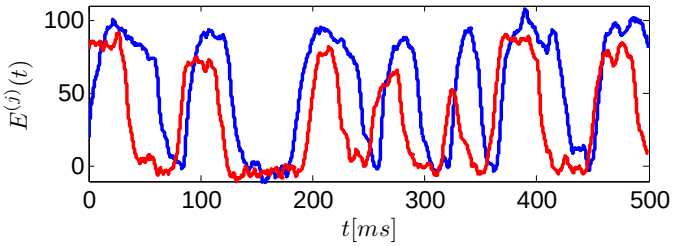

Figure 49: Excitatory components of two Noisy Wilson-Cowan Oscillators

Can we detect if $E^{(1)}$ influences $E^{(2)}$ or if $E^{(2)}$ influences $E^{(1)}$ just from looking at this plot. Well, this segment is too short to detect causality in either direction, but what if we had a segment of the dynamical system for a length of time of 100,000 ms? This would only be 100 seconds, not an unrealistic length of time for a segment of an EEG. Would it be possible to apply the extension of Granger causality we used on the phase oscillators here? Let's try it!

Perhaps we can use some of our understanding of the Wilson-Cowan model to develop a good method for detecting causality. As usual, we build our test to look for causality based on the null hypothesis that there is no causality. Now, we know that if the system is on a stable limit cycle, then we cannot detect directional causality. However, with the added Ornstein-Uhlenbeck noise, the system is constantly being pushed away from the stable limit cycle. If the system is far enough away from the stable limit cycle, and there is no causality from $E^{(2)}$ to $E^{(1)}$ (that is, $\varphi_1 = 0$), then we would expect $E^{(1)}$ to be indifferent if $E^{(2)}$ is large or small. However, if $\varphi_1 \neq 0$, then whenever $E^{(2)}$ is large, it should have a positive influence on $E^{(1)}$, that is, $E^{(1)}$ should increase. There is a physical interpretation of this as well, namely, if two neuronal networks are connected from excitatory cells to other excitatory cells, then more activity in the excitatory cells in one of the networks should cause an increase in the activity of the excitatory cells in the other network.

For the following example, we set $\varphi_1 = 0.5$ and $\varphi_2 = 0$. So, $E^{(2)}$ influences $E^{(1)}$, but $E^{(1)}$ does not influence $E^{(2)}$. We can test this as follows. Fix $E^{(1)}(t)$ at times where $E^{(1)}(t)$ is near the value $\alpha = 10$ (so there is minimal excitatory activity in the network), and $\frac{dE^{(1)}(t)}{dt} > 0$, that is, $E^{(1)}(t)$ is in the middle of an upward trajectory. When we do this, we get the following density histogram plot.

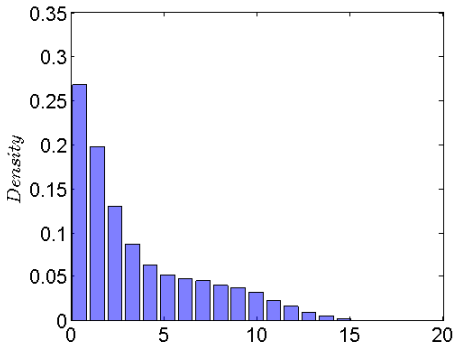

Figure 50: Density Histogram of $dE^{(1)}/dt$ whenever $E^{(1)}(t) \approx 10$ and $dE^{(1)}/dt > 0$

Now, we can also pick times within this set whenever $E^{(2)}(t)$ is very large, so there is a lot of excitatory activity in network 2. We restrict ourselves to values $E^{(2)}(t) > 70$, and plot a density histogram over the previous figure. Now, we would expect that if $\varphi_1 = 0$, then this histogram plot should look similar to the one above. However, if $\varphi_1 \neq 0$, then we would expect that the histogram plot should be shifted to the right. That is, whenever we regress on large values of $E^{(2)}(t)$, we expect an increase in the mean of the distribution of the derivative $dE^{(1)}(t)/dt$. When doing this, we get the following histogram plot, overlaid over the previous plot.

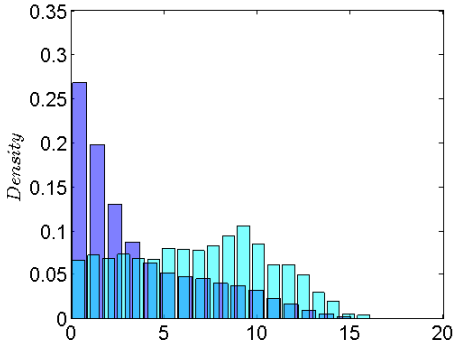

Figure 51: Density Histogram of $dE^{(1)}/dt$ whenever $E^{(1)}(t) \approx 10$ and $dE^{(1)}/dt > 0$ (in blue) and the subset where $E^{(2)}(t) > 70$ (in cyan)

It looks like the cyan density histogram plot is shifted to the right, so this plot correctly suggests that $E^{(2)}$ influences $E^{(1)}$. Now, we can do the same analysis to see if $E^{(1)}$ influences $E^{(2)}$, that is, if $\varphi_2 \neq 0$. First, we fix $E^{(2)}(t)$ at times where $E^{(2)}(t)$ is near 10 and $\frac{dE^{(2)}}{dt} > 0$. When doing this, we get the following density histogram plot.

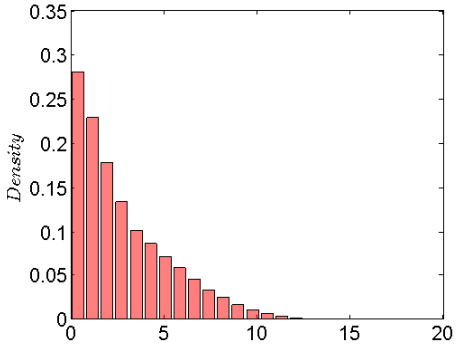

Figure 52: Density Histogram of $dE^{(2)}/dt$ whenever $E^{(2)}(t) \approx 10$ and $dE^{(2)}/dt > 0$

As before, pick times within this set whenever $E^{(1)}(t) > 70$, and overlay the histogram density plot on the above plot.

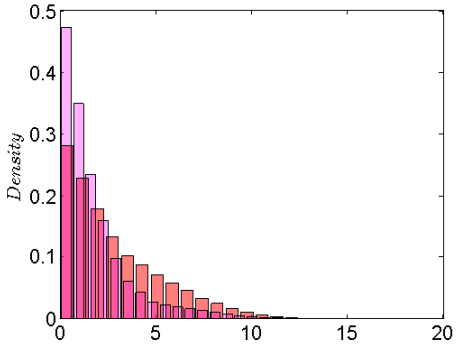

Figure 53: Density Histogram of $dE^{(2)}/dt$ whenever $E^{(2)}(t) \approx 20$ and $dE^{(2)}/dt > 0$ (in red) and the subset where $E^{(1)}(t) > 70$ (in magenta)

The restricted histogram plot on large values of $E^{(1)}(t)$ seems to be shifted to the left. This certainly does not suggest that $E^{(2)}$ influences $E^{(1)}$, since otherwise, we would have expected a shift to the right. So, this plot correctly suggests that $E^{(1)}$ does not influence $E^{(2)}$.

Now, we do the same analysis for $\varphi_2 = 0.5$ and $\varphi_1 = 0$, so there is one-way causality from $E^{(1)}$ to $E^{(2)}$, but not from $E^{(2)}$ to $E^{(1)}$. Doing the same analysis as above we get the following density histogram plots.

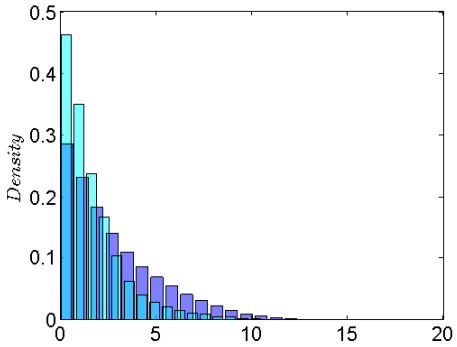

Figure 54: Density Histogram of $dE^{(1)}/dt$ whenever $E^{(1)}(t) \approx 10$ and $dE^{(1)}/dt > 0$ (in blue) and the subset where $E^{(2)}(t) > 70$ (in cyan)

Here, we see a clear shift to the right in the density histogram plot in Figure 55, which correctly suggests $E^{(1)}$ influences $E^{(2)}$. However, we see no shift to the right in Figure 54, which correctly suggests $E^{(2)}$ does not influence $E^{(1)}$.

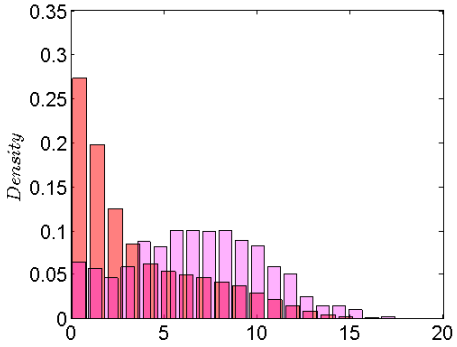

Figure 55: Density Histogram of $dE^{(2)}/dt$ whenever $E^{(2)}(t) \approx 20$ and $dE^{(2)}/dt > 0$ (in red) and the subset where $E^{(1)}(t) > 70$ (in magenta)

What if there is mutual causation in the Wilson-Cowan oscillator? Let $\varphi_1 = 0.2$ and $\varphi_2 = 0.2$, so there is causality in both directions. Are we still able to detect the mutual causation using this method? Here is a short segment of $E^{(1)}(t)$ and $E^{(2)}(t)$ for the first 500 ms of the simulation.

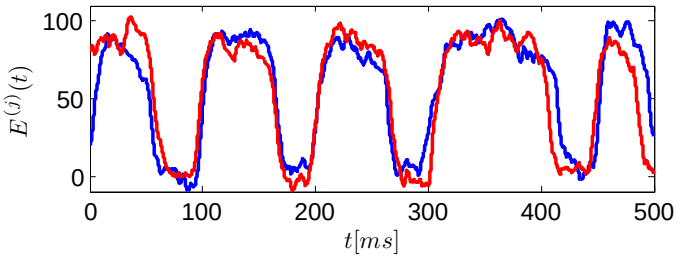

Figure 56: Excitatory components of two Noisy Wilson-Cowan Oscillators with mutual causality

Note how correlated the two components are, so we do not expect too many instances where $E^{(1)}(t)$ and $E^{(2)}(t)$ differ by a lot. However, over time, the noise will eventually push them further away from the attracting limit cycle. When doing the analysis as before, and letting the simulation run until 100,000 ms, we get the following density histogram plots.

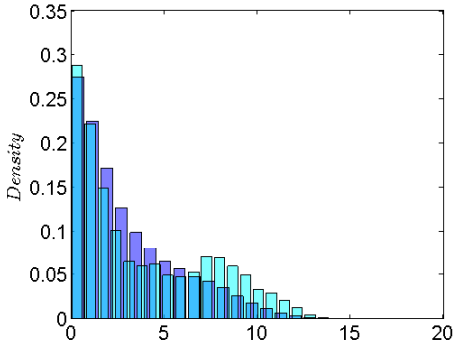

Figure 57: Density Histogram of $dE^{(1)}/dt$ whenever $E^{(1)}(t) \approx 10$ and $dE^{(1)}/dt > 0$ (in blue) and the subset where $E^{(2)}(t) > 70$ (in cyan)

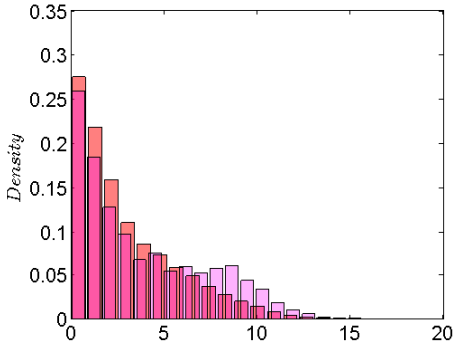

Figure 58: Density Histogram of $dE^{(2)}/dt$ whenever $E^{(2)}(t) \approx 20$ and $dE^{(2)}/dt > 0$ (in red) and the subset where $E^{(1)}(t) > 70$ (in magenta)

We do see a slight shift in both density histogram plots to the right, suggesting mutual causality. Thus, this simple extension of Granger causality to the noise-driven Wilson-Cowan oscillators works.

5 Conclusion

In this work, we have looked at how to detect causality in continuous-time dynamical systems in the absence of noise. We saw that without any given information about the dynamical system itself, it is impossible to detect causal interactions between components of a system while it is in a steady state – stable fixed point for the linear system, phase-locked state for the phase oscillators, and the stable limit cycle for coupled Wilson-Cowan oscillators. However, it is possible to detect causal interactions during the approach to the steady state.

We saw how and why Granger causality works for a linear discrete-time stochastic process. The main idea of Granger causality in this setting is to regress the present of one dynamic variable on its own past, and then include into that regression the past of another dynamic variable. If the second regression is more accurate, that is the variance of the error terms in the regression is smaller than in the first regression, then the other dynamic variable is said to Granger cause the first. When the underlying dynamics of the system are linear, and the system has a stationary invariant distribution, this technique works very well.

When introducing continuous-time noise processes to a dynamical system, however, the idea of Granger causality becomes murky. As we saw with the linear system driven by an Ornstein-Uhlenbeck process, we cannot condition the present of one dynamic variable onto its past and ignore the noise terms as we did in the discrete time stochastic process. This is mainly because the Ornstein-Uhlenbeck process is not uncorrelated with its own past and the past of the variable it influences, as it was in the case of the white noise process in the discrete case. However, for the linear system driven by Ornstein-Uhlenbeck noise, we can calculate the invariant distribution of the process, and using that, estimate causality.

We also saw that for the phase oscillator, shocks to the system in the form of a Poisson

process lead to easy detection of causal interactions. However, when introducing Ornstein-Uhlenbeck drive, it is not so clear how to detect causal interactions. However, our extension of Granger causality lead us to believe that it is possible to detect causality from one phase oscillators to the other by simple heuristics.

Lastly, we showed that for coupled Wilson-Cowan oscillators driven by Ornstein- Uhlenbeck noise, our extension of Granger causality worked. This shows that it is still possible to detect causality in a more complicated model, where not all of the elements of that dynamic system are observable. However, we were only using simple heuristics. The question of how to set up a statistical test to test for causality in more complicated, nonlinear dynamical systems is still open-ended, and there are still many interesting questions to be answered. For example, if we introduce other types of stochastic processes besides the Ornstein-Uhlenbeck process, are we still able to detect causality? What if the noisy driving forces have different parameters, that is, different variances and mean-reverting rates in the case of Ornstein-Uhlenbeck noise? How do we deal with time-delays, if the affect from one dynamic variable to another is delayed rather than instantaneous?

We have shown that a simple extension of Granger causality from discrete-time linear stochastic processes to continuous-time non-linear stochastic processes is effective. In the context of neuroscience, people are interested in building the effective connectivity of different brain regions from collected neuronal data. Using this extension, it may be possible to detect causality between different brain regions from collected data.

References

[1] Wiener, N. 1956 “The theory of prediction,” *Modern Mathematics for Engineers* Vol. 1 (ed. E. F. Beckenbach), New York: McGraw-Hill.

[2] Granger, C. W. J. 1969 “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica* Vol. 37, pp. 424-438.

[3] Wilson, H. R., Cowan, J. D. 1972 “Excitatory and inhibitory interactions in localized populations of model neurons”, *Biophysical Journal* Vol. 12, pp. 1 - 24.

[4] Wilson, H. R., 1999 *Spikes, Decision, and Actions: The Dynamical Foundations of Neurosciences*. Oxford University Press.

[5] Ueta, T., Chen, G. 2001 “On Synchronization and Control of Coupled Wilson-Cowan Neural Oscillators”, *International Journal of Bifurcation and Chaos* Vol. 13, No. 1, pp. 163 - 175.

[6] Hoppensteadt, F.C., Izhikevich, E. M. 1997 “Weakly Connected Neural Networks”, *AMS* Vol. 126, Springer.

[7] Strogatz, S. H. 1994 *Nonlinear Dynamics and Chaos With Applications to Physics, Biology, Chemistry, and Engineering*. Reading, MA: Perseus Books.

[8] Rockland K. S., Lund J. S. 1982 “Intrinsic laminate lattice connection in primate visual cortex”, *Science* Vol. 215, pp. 1532 - 1534.

[9] Gilbert C. D., Wiesel T. N. 1983 “Clustered intrinsic connections in cat visual cortex”, *Journal of Neuroscience* Vol. 3, pp. 1116 - 1133.

[10] Kisvarday Z. F., Martin K. A. C., Fuend T. F., Magloczky Z., Witteridge D., Somogyi P. 1986 “Synaptic targets of HRP-filled layer II pyramidal cells in the cat striate cortex”, *Experimental Brain Research* Vol. 64, pp. 541 - 552.

[11] McGuire B. A., Gilbert C. D., Rivlin P. K., Wiesel T. N. 1991 “Targets of horizontal connections in macaque primary visual cortex”, *Journal of Comparative Neurology* Vol. 305, pp. 370 - 392.

[12] Brockwell, P. J., Davis, R. A. 1987 *Time Series: Theory and Methods*. New York, Springer-Verlag.

[13] Kuntsevich, V. M., Lychak, M. M. 1980 “Solution of discrete matrix lyapunov and riccati equations and their generalizations”, *Cybernetics* Vol. 16, pp. 321-327.

[14] Lütkepohl, H. 2005 *A New Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg: Springer-Verlag.

[15] Beuter, A., Glass, L., Mackey, M.C., Titcombe, M. S. 2003 *Nonlinear Dynamics in Physiology and Medicine*. New York, Springer-Verlag.

[16] Faes, L., Nollo, G., Erla, S., Papadelis, C., Braun, C., Porta, A. 2010 “Detecting nonlinear causal interactions between dynamical systems by non-uniform embedding of multiple time series”, *Conf Proc IEEE End Med Biol Soc.*

[17] Bai, Z., Wong, W. K., Zhang, B. 2010 “Multivariate linear and nonlinear casuality tests”, *Mathematics and Computers in Simulation* Vol. 81, pp. 5 - 17.

[18] Friston, K. J., Harrison, L., Penny, W. 2003 “Dynamic causal modelling”, *Neuroimage* Vol. 4, pp/ 1273 - 1302.