



HHS Public Access

Author manuscript

Can J Stat. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:

Can J Stat. 2023 September ; 51(3): 824–851. doi:10.1002/cjs.11790.

Oscillating neural circuits: Phase, amplitude, and the complex normal distribution

Konrad N. URBAN*, Heejong BONG,

Josue ORELLANA,

Robert E. KASS

Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Abstract

Multiple oscillating time series are typically analyzed in the frequency domain, where coherence is usually said to represent the magnitude of the correlation between two signals at a particular frequency. The correlation being referenced is complex-valued and is similar to the real-valued Pearson correlation in some ways but not others. We discuss the dependence among oscillating series in the context of the multivariate complex normal distribution, which plays a role for vectors of complex random variables analogous to the usual multivariate normal distribution for vectors of real-valued random variables. We emphasize special cases that are valuable for the neural data we are interested in and provide new variations on existing results. We then introduce a complex latent variable model for narrowly band-pass-filtered signals at some frequency, and show that the resulting maximum likelihood estimate produces a latent coherence that is equivalent to the magnitude of the complex canonical correlation at the given frequency. We also derive an equivalence between partial coherence and the magnitude of complex partial correlation, at a given frequency. Our theoretical framework leads to interpretable results for an interesting multivariate dataset from the Allen Institute for Brain Science.

Abstract

Les séries temporelles à oscillations multiples sont généralement étudiées dans le domaine fréquentiel, où la cohérence est souvent considérée comme l'amplitude de la corrélation entre deux signaux à une fréquence spécifique. Cette corrélation est à valeurs complexes et présente des similitudes avec la corrélation de Pearson pour les valeurs réelles, tout en présentant des différences distinctes. Dans cette étude, les auteurs explorent la dépendance entre les séries oscillantes en utilisant la distribution normale complexe multivariée. Cette distribution est l'équivalent de la distribution normale multivariée classique, mais adaptée aux vecteurs de variables aléatoires complexes plutôt qu'aux vecteurs de variables aléatoires réelles. Les auteurs mettent l'accent sur des cas spécifiques qui revêtent une importance particulière pour les données neuronales qui les intéressent, tout en proposant de nouvelles approches et des variations des résultats existants. Ils introduisent un modèle de variables latentes complexes pour

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

*Corresponding author: kurban@andrew.cmu.edu.

Additional Supporting Information may be found in the online version of this article at the publisher's website.

les signaux filtrés en bande passante étroite à une fréquence donnée. Ils démontrent ensuite que l'estimation du maximum de vraisemblance dans ce modèle produit une cohérence latente équivalente à l'amplitude de la corrélation canonique complexe à la fréquence spécifiée. Ils établissent également une équivalence entre la cohérence partielle et l'amplitude de la corrélation partielle complexe, toujours à une fréquence donnée. Leur approche théorique conduit à des résultats interprétables pour un ensemble de données multivariées intéressant provenant de l'Allen Institute for Brain Science.

Keywords

Coherence; complex normal distribution; latent variable model; oscillations; Primary 62H20; Secondary 62P10

1. INTRODUCTION

Oscillations in neural circuits have been observed under a variety of circumstances and have provoked much speculation about their physiological function (Buzsaki & Draguhn, 2004; Fries, 2005). In the past 15 years, the role of oscillations at particular frequencies has been the subject of considerable experimental investigation, including causal manipulation (Cardin et al., 2009). Of particular interest is the intriguing possibility that oscillations facilitate purposeful communication across distinct parts of the brain, such as when an organism must retrieve and hold items from memory or direct its visual attention to a particular location (Miller, Lundqvist & Bastos, 2018; Schmidt et al., 2019). This has led to the idea that alterations of circuit oscillations could indicate brain dysfunction (Mathalon & Sohal, 2015).

From a statistical perspective, regardless of their mechanistic function, neural oscillations can be considered useful indicators of coordinated activity across brain regions. A short snippet of data, typical of those we have analyzed, is shown in the top panel of Figure 1 together with a band-pass-filtered version. In the bottom panel are band-pass-filtered series for two trials, where at many points in time the amplitudes, phases, or both are different. When series from two electrodes (in different parts of the brain) are considered, the two phases may tend to shift forward or backward together across trials, which indicates coordinated activity. Similarly, there may be trial-to-trial correlation between the amplitudes.

To quantify this form of association, an obvious question is whether it might be advantageous to consider phase and amplitude together, as defining complex numbers, and model association using multivariate complex distributions. In addition, data such as those in Figure 1 often come from many recordings in each brain region that somehow must be combined. We investigate the properties of the multivariate complex normal distribution with the goal of analyzing interactions among multiple groups of oscillating time series.

A related conceptual concern comes from the standard interpretation of coherence, which is the starting point for much frequency-based analysis of co-dependence. Under the Cramér (or Cramér–Khinchine) representation of a bivariate stationary time series, coherence is usually said to be the magnitude of the time series correlation at a particular frequency

(e.g., Shumway & Stoffer, 2017, Section 4.6). The difficulty with this interpretation is that “correlation” here refers to the complex correlation for complex random variables, which is analogous to the Pearson correlation for real-valued variables in certain respects, but not others.

This article reviews and reformulates ideas drawn from the literature and then provides several new results, including a summary of interacting groups of oscillating multivariate time series using a time-domain rendering of latent coherence, which turns out to be the magnitude of a complex-valued canonical correlation. Although we have been motivated by the analysis of neural data, we believe this article will be of general statistical interest as it concerns a basic topic in time series analysis. We also hope it represents a suitable tribute to Nancy Reid, whose work has often aimed to advance statistics through conceptual clarification and consolidation.

2. BACKGROUND AND SUMMARY

We are interested in the covariation of both phase and amplitude in two or more time series. The data we analyze in Section 6 come from experiments during which animals are shown a visual stimulus repeatedly, across many trials, while neural activity is recorded from electrodes inserted into the brain. Repeated measurements across many trials is typical of neurophysiological data, and these repetitions are helpful in dealing with the striking nonstationarity present in neural recordings (see, e.g., Kass et al., 2018). In addition, neurophysiological experiments often collect data from multiple electrodes embedded in multiple brain regions.

From oscillating series such as those in the top panel of Figure 1, the band-pass-filtered version can be obtained by Fourier analytic methods (though it can also be obtained using wavelets). As long as the band in the filtering is sufficiently narrow (surrounding some particular frequency), at any point in time the phase and amplitude of the real-valued, filtered time series can be recovered using the Hilbert transform (see the Appendix) under an assumption of local stationarity (Ombao & Van Belleghem, 2008). A phase and an amplitude define a complex number, and the Hilbert transform converts a real-valued signal into the corresponding complex-valued signal. For a pair of repeatedly observed complex-valued variables, correlation is more complicated than in the real case: covariance becomes complex-valued and there are two forms of linear covariation, called covariance and pseudo-covariance. Complex covariance is defined for two complex-valued random variables X_1 and X_2 as

$$\text{cov}(X_1, X_2) = E\{(X_1 - EX_1)(\overline{X_2 - EX_2})\},$$

where \bar{x} is the conjugate of a complex number x . Note that $\text{cov}(X_1, X_2)$ is a complex number. Complex pseudo-covariance is defined as

$$\text{pcov}(X_1, X_2) = E\{(X_1 - EX_1)(X_2 - EX_2)\}.$$

The variance–covariance and pseudo-variance–covariance matrices for a complex random vector X are defined analogously to the case of complex random variables, using the Hermitian and transpose operators, respectively:

$$\text{var}(X) = E\{(X - EX)(X - EX)^H\} \quad \text{and} \quad \text{pvar}(X) = E\{(X - EX)(X - EX)^T\},$$

where $(X - EX)^H = \overline{(X - EX)}^T$. When the pseudo-covariance is zero, a complex random vector is called *proper*.

In addition to complex correlation, dependence between phase angles can be measured through phase-locking value (PLV) (Lachaux et al., 1999) which, as we show in Section 3.2.2, can be viewed as an analogue of correlation for angular random variables. For more than two phases, a recently developed class of models called torus graphs provides a thorough rendering of multivariate phase dependence (Klein et al., 2020). A torus graph is any member of the full exponential family with means and cross-products (interaction terms) on a multidimensional torus, which is the natural home for multivariate circular data because the product of circles is a torus. Thus, torus graphs are probabilistic graphical models and represent an analogue to Gaussian graphical models for circular (angular) data.

Phase coupling is also analyzed using coherence, the magnitude of coherency, which is a frequency-domain measure that has a form resembling correlation in terms of spectral densities and cross-spectral densities. One way to understand how coherence becomes complex correlation at a particular frequency is to consider the result of filtering signals with a complex band-pass filter in a band $(\omega_0 - \delta, \omega_0 + \delta)$ for some small δ . Ombao & Van Bellegem (2008) noted that the resulting narrow-band coherence, which they called “band coherence,” is the correlation between the filtered signals. We can effectively pass to the limit as $\delta \rightarrow 0$ to get the infinitesimal version that appears in the Cramér–Khintchine representation, displayed in Section 4.1. In a similar vein, in Section 4.1 we also note that a single-frequency bivariate process with stochastic amplitudes and phases yields coherence as the magnitude of complex correlation.

Coherence has been used to understand conditional dependence relationships between time series. Dahlhaus (2000) studied so-called coherence graphs for real-valued time series and showed how conditional dependence may be inferred from the partial coherence between these signals. Partial coherence arises by inverting and rescaling a matrix of coherency values for d time series, and represents a form of conditional dependence between signals (Ombao & Pinto, 2022). Tugnait (2019a,b) analyzed conditional dependence between signals with arbitrary spectra by observing that the limiting distribution of a certain class of transformations of real-valued time series takes the form of a complex normal distribution.

Our setting, in which we observe multiple signals oscillating at the same frequency, differs from those previously studied in Dahlhaus (2000) and Tugnait (2019a,b) because our signals are oscillating only at a single frequency. In Section 4, we provide a comprehensive overview of how, for our setting, complex correlation and coherency are equivalent when we assume complex normality. Similarly, we show that complex partial correlation and

partial coherency are equivalent, and we discuss how these concepts are directly tied to the conditional dependence between the observed oscillating time series. Thus, by modelling band-pass-filtered data with the complex normal distribution, we obtain simple interpretations of estimated correlation and partial correlation matrices in terms of coherence and partial coherence.

Our interest is in recordings from multiple electrodes embedded in several brain regions. To take advantage of the multiple sources of activity observed from each region, we developed a novel complex normal latent variable model in which each of several latent variables represents activity in one of the regions (at a particular point in time). Our goal is to estimate the dependence among these latent variables. We specify the model in Section 5 and apply it to data in Section 6. The data we analyze are local field potentials (LFPs), which are voltage recordings from electrodes inserted into the brain; they are low-pass-filtered (smoothed) and typically down-sampled to 1kHz so that each second of data has 1000 observations. LFPs represent bulk activity near the electrode (roughly within 150–200 μm) involving large numbers of neurons (Buzsáki, Anastassiou & Koch, 2012; Einevoll et al., 2013; Pesaran et al., 2018). Our theoretical results enable us to obtain a simple interpretation of data from the Allen Institute for Brain Science, Seattle, WA. We examined three recordings in each of six regions of the visual cortex in response to a visual stimulus and estimated latent correlation and partial correlation matrices after band-pass-filtering at 6.5Hz to isolate the theta rhythm. There are many large latent correlations of the theta rhythm oscillations between regions (which, as we show, may be considered large values of latent coherence). However, the unique associations, that is, the latent partial correlations, between these regions are modest in size with the exception of the unique associations between AM and PM, as well as between DG and CA1, which exhibit latent partial correlations (latent partial coherence) close to 1.

Although some of the theorems in this article may be unsurprising to experts on the complex normal distribution and frequency-based analysis, they all represent at least novel extensions and reformulations, and several results are entirely new. In Section 3.3.3, we discuss the conditional distribution of the angles given the amplitudes in the polar coordinate representation of the complex normal distribution. Navarro, Frellsen & Turner (2017) showed that the multivariate generalized von Mises (mGvM) distribution can be viewed as a conditional distribution of the $2d$ -dimensional, real-valued multivariate normal distribution when all the amplitudes are equal to 1. Here, we change the setting of this result to the complex normal distribution, generalize the result to arbitrary amplitudes, and show how the conditional distribution changes when certain relevant restrictions are made on the parameters of the complex normal distribution. These results (appearing in Theorems 2 and 3 and Corollary 4) form new characterizations of torus graphs. In Section 4 we provide a general treatment of partial correlation, conditional correlation, and conditional dependence for the complex normal distribution. These results consolidate and reformulate what was previously available in the literature (see Section 4 for specific citations). The new results in Corollary 13 and Theorem 12 show that maximum likelihood estimation in our latent variable model produces estimates that are equivalent to complex-valued versions of

canonical correlation. The proofs of the theorems and corollaries we show are provided in the Supplementary Material Section S1.

3. DEPENDENCE OF COMPLEX-VALUED RANDOM VARIABLES

In this section, we first present the data-generating setting we are attempting to model. We observe complex-valued random vectors, and our goal is to measure associations between their entries. After introducing this setting, we mostly ignore the fact that we are dealing with oscillating signals, but we will return to this important point in Section 4 when we study the relationship between coherence, complex correlation, and the complex normal distribution. This important correspondence is the basis for the analysis presented in this paper.

We first review various measures of pairwise association between complex-valued random variables, including complex correlation, PLV, and amplitude correlation. We then discuss two multivariate models for association between complex-valued random vectors: the complex normal distribution, which is based on the linear association between complex random vectors, and the torus graph distribution, which is a multivariate model for phase. In Section 3.3.3, we show some new results that describe how the torus graph distribution arises by conditioning angles on amplitudes in the complex normal distribution.

3.1. Setting: Repeated Observations of Oscillating Signals

In this article, we assume we observe d time series that are oscillating at some frequency ω_0 on each of many repeated trials. We can model the data as a d -dimensional random vector $\mathbf{Y}(t)$, and assume these vectors are i.i.d. across trials. Throughout most of this article, we omit the trial number from our notation. We can first apply a local band-pass filter to the raw vector $\mathbf{Y}(t)$ around the frequency ω_0 at time t , and then apply the Hilbert transform to obtain a complex-valued random vector $\mathbf{X}(t)$ having components that are functions of the phases and amplitudes of the oscillations: $X_i(t) = R_i(t)\exp\{i\Theta_i(t)\}$ for $R_i(t) \in (0, \infty)$ and $\Theta_i(t) \in [-\pi, \pi)$ for components $i \in [d]$. Here and everywhere else in the article, $[d] = \{i \in \mathbb{N}: 1 \leq i \leq d\}$, where i represents the imaginary unit.

3.2. Pairwise Association Between Complex-Valued Random Variables

3.2.1. Complex Correlation—Linear association between complex-valued random variables is measured by complex correlation. We defined both covariance and pseudo-covariance between complex random variables and random vectors in Section 2. The corresponding complex correlations are

$$\text{corr}(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(X_1)\text{var}(X_2)}} \text{ and } \text{pcorr}(X_1, X_2) = \frac{\text{pcov}(X_1, X_2)}{\sqrt{\text{var}(X_1)\text{var}(X_2)}}.$$

3.2.2. Angular Association: Phase-Locking Value—We can study the association between two phases by considering their representations as unit-length complex random variables $X_1 = e^{i\Theta_1}$ and $X_2 = e^{i\Theta_2}$. The set of such pairs is the product of two circles, a two-dimensional torus.

The strength of the association between two angular random variables Θ_1, Θ_2 can be measured using PLV (Lachaux et al., 1999). When we have repeated observations θ_1^n, θ_2^n across N trials of angular observations, then we can define

$$\text{PLV} = \frac{1}{N} \left| \sum_{n=1}^N \exp\{l(\theta_1^n - \theta_2^n)\} \right|,$$

which is widely used to measure the angular dependence between oscillating signals (Lepage & Vijayan, 2017).

To better understand PLV, we provide a theoretical analysis of the components of phase-based association. Suppose that $c_i = E\{\cos(\Theta_i)\}$ and $s_i = E\{\sin(\Theta_i)\}$ for $i \in \{1, 2\}$. The first moment of X_i is

$$E(X_i) = E\{\cos(\Theta_i)\} + iE\{\sin(\Theta_i)\} = c_i + i s_i = r_i e^{i\theta_i},$$

where $r_i = (c_i^2 + s_i^2)^{1/2}$ and θ_i is the angle of the vector (c_i, s_i) relative to $(1, 0)$ for $i \in \{1, 2\}$. The complex covariance of X_1 and X_2 is

$$\text{cov}(X_1, X_2) = E(X_1 \bar{X}_2) - E(X_1)E(\bar{X}_2) = E\{e^{i(\Theta_1 - \Theta_2)}\} - r_1 r_2 e^{i(\theta_1 - \theta_2)},$$

where the pseudo-covariance of X_1 and X_2 is

$$\text{pcov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2) = E\{e^{i(\Theta_1 + \Theta_2)}\} - r_1 r_2 e^{i(\theta_1 + \theta_2)}.$$

Covariance and pseudo-covariance, respectively, represent *rotational* and *reflectional* association between X_1 and X_2 . Rotational covariance measures clockwise–clockwise association, and reflectional covariance measures clockwise–anticlockwise association. Both types of association are shown in Figure 2. The amplitude of the rotational covariance controls the width of the yellow high-probability band, and the phase of this complex number specifies the shift of this band, which is seen along the diagonal of the Cartesian plane; the band shifts in direction from bottom-right toward top-left (Figure 2a). The same notion applies to the reflectional covariance magnitude, but the band is positioned on the other diagonal (Figure 2b). The presence of both types of covariation is possible and can lead to concentrated marginals. However, in our experience with phases extracted from LFP neural data, we have almost always observed exclusively rotational dependence. In addition, the marginals are close to uniform. If $r_1 = r_2 = 0$ (e.g., if the Θ_i are uniform), then

$$\text{cov}(X_1, X_2) = E\{e^{i(\Theta_1 - \Theta_2)}\} \text{ and } \text{corr}(X_1, X_2) = E\{e^{i(\Theta_1 - \Theta_2)}\}.$$

The quantity $E\{e^{l(\Theta_1 - \Theta_2)}\}$ is the theoretical counterpart of PLV. Thus, in the absence of reflectional covariation, if $r_1 = r_2 = 0$, then $E\{e^{l(\Theta_1 - \Theta_2)}\}$ is the analogue of Pearson correlation for angular random variables.

3.2.3. Amplitude Correlation—For two amplitudes $R_i, R_j \in (0, \infty)$, the simplest way to characterize amplitude correlation is through the ordinary Pearson correlation of real-valued random variables $\text{corr}(R_i, R_j)$. Other approaches involve calculating the correlation between the log of the amplitudes or between the square of the amplitudes (Nolte et al., 2020).

3.3. Multivariate Models

3.3.1. The Complex Normal Distribution—A complex random vector $X \in \mathbb{C}^d$ is said to be *complex normal* (CN) if its real and imaginary parts are jointly multivariate normal (Andersen et al., 1995). Suppose that $(\Re X, \Im X) \sim \mathcal{N}(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_{22} \end{pmatrix},$$

so that $\mu_1 = E(\Re X)$, $\mu_2 = E(\Im X)$, $\Sigma_{11} = \text{var}(\Re X)$, $\Sigma_{22} = \text{var}(\Im X)$, and $\Sigma_{12} = \text{cov}(\Re X, \Im X)$. Then

$$m = \mu_1 + i\mu_2, \quad \Gamma = (\Sigma_{11} + \Sigma_{22}) + i(\Sigma_{12} - \Sigma_{12}^\top), \text{ and } C = (\Sigma_{11} - \Sigma_{22}) + i(\Sigma_{12} + \Sigma_{12}^\top),$$

giving a bijection between (m, Γ, C) and (μ, Σ) . Thus, the complex normal distribution is well defined by the parameter set (m, Γ, C) . We denote the distribution by $X \sim \mathcal{CN}(m, \Gamma, C)$. Recall that the distribution of X is said to be *proper* if the pseudo-covariance matrix C is zero. Further, X is *circularly symmetric*, meaning component-wise circularly symmetric so that its pdf satisfies $p(X) = p(e^{i\alpha}X)$ for all $\alpha \in \mathbb{R}$ if and only if X is proper and has mean zero (Adali, Schreier & Scharf, 2011).

Circularly symmetric, proper, and restriction-less CN distributions form full regular exponential families. This ensures that the maximum likelihood estimator, which involves first and second-order moment statistics, is sufficient. In addition, these families form probabilistic graphical models so that conditional independence is easily characterized. We return to this in Theorem 8 and Corollary 9. We provide an abbreviated version of this theorem here; for full details, see the Supplementary Material Section S1.

Theorem 1.: *Each of the following families forms a full and regular exponential family: (i) the family of CN distributions represented by $\mathcal{CN}(\mu, \Gamma, C)$; (ii) the family of proper CN distributions $\mathcal{CN}(\mu, \Gamma, 0)$; and (iii) the family of circularly symmetric CN distributions $\mathcal{CN}(0, \Gamma, 0)$.*

3.3.2. Torus Graphs—Multivariate models for angular random variables are most naturally defined on the torus rather than on Euclidean space. Such a model is given in Klein et al. (2020). Here, we briefly review its construction.

To define an exponential family on a torus with a given mean and covariance structure, first- and second-order sufficient statistics are needed. Using two-dimensional rectangular coordinates (involving cosines and sines), the first-order sufficient statistics are $U_1 = (\cos \Theta_1, \sin \Theta_1)$ and $U_2 = (\cos \Theta_2, \sin \Theta_2)$. Second-order behaviour is summarized by

$$U_1 U_2^\top = \begin{pmatrix} \cos \Theta_1 \cos \Theta_2 & \cos \Theta_1 \sin \Theta_2 \\ \sin \Theta_1 \cos \Theta_2 & \sin \Theta_1 \sin \Theta_2 \end{pmatrix}.$$

We can then write a natural exponential family density for $\theta = (\theta_1, \theta_2)$ with a canonical parameter η consisting of $\eta_i \in \mathbb{R}^2$ and $\eta_{ij} \in \mathbb{R}^4$ for $i, j \in \{1, 2\}$ as

$$p(\theta; \eta) \propto \exp \left\{ \eta_1^\top \begin{pmatrix} \cos \theta_1 \\ \sin \theta_1 \end{pmatrix} + \eta_2^\top \begin{pmatrix} \cos \theta_2 \\ \sin \theta_2 \end{pmatrix} + \eta_{12}^\top \begin{pmatrix} \cos \theta_1 \cos \theta_2 \\ \cos \theta_1 \sin \theta_2 \\ \sin \theta_1 \cos \theta_2 \\ \sin \theta_1 \sin \theta_2 \end{pmatrix} \right\}, \quad 0 \leq \theta_1, \theta_2 \leq 2\pi. \tag{1}$$

For $\theta = (\theta_1, \dots, \theta_d)$, (1) extends to an exponential family on a d -dimensional torus as

$$p(\theta; \eta) \propto \exp \left\{ \sum_{j=1}^d \eta_j^\top \begin{pmatrix} \cos \theta_j \\ \sin \theta_j \end{pmatrix} + \sum_{i < j} \eta_{ij}^\top \begin{pmatrix} \cos \theta_i \cos \theta_j \\ \cos \theta_i \sin \theta_j \\ \sin \theta_i \cos \theta_j \\ \sin \theta_i \sin \theta_j \end{pmatrix} \right\}, \quad 0 \leq \theta_i \leq 2\pi. \tag{2}$$

Based on simple trigonometric identities, Klein et al. (2020) reparameterized this family into a more interpretable form

$$p(\theta; \eta(\phi)) \propto \exp \left\{ \sum_{i=1}^d \phi_i^\top \begin{pmatrix} \cos \theta_i \\ \sin \theta_i \end{pmatrix} + \sum_{i < j} \phi_{ij}^\top \begin{pmatrix} \cos(\theta_i - \theta_j) \\ \sin(\theta_i - \theta_j) \\ \cos(\theta_i + \theta_j) \\ \sin(\theta_i + \theta_j) \end{pmatrix} \right\}, \tag{3}$$

which uses the first- and second-order statistics seen in the definitions of complex covariance. Klein et al. (2020) then defined a d -dimensional torus graph (TG) to be any member of the family of distributions specified by (2) or (3). If Θ is distributed according to $p(\theta; \eta)$, then we say that $\Theta \sim \mathcal{F}\mathcal{E}(\eta)$; likewise, if Θ is distributed according to $p(\theta; \eta(\phi))$, then we say that $\Theta \sim \mathcal{F}\mathcal{E}(\eta(\phi))$.

Because torus graphs form exponential families, a pair of random variables X_i and X_j will be conditionally independent given all other variables if and only if the four elements in the pairwise interaction parameter ϕ_{ij} are zero. Thus, torus graphs define probabilistic graphical models.

Torus graphs can uncover conditional dependence relationships, meaning pairwise dependence edges that are still present after conditioning on the rest of the random variables in the model. Conditional dependence is obscured with simple correlation-type measures such as PLV because for each edge, they consider only two random variables (see Figure 3, which displays a conditional independence graph from Klein et al. (2020)).

There are three parameter groups in (3): marginal concentrations, rotational covariance, and reflectional covariance. Klein et al. (2020) showed that the submodel with only rotational covariance parameters (which corresponds to a proper distribution when the statistics U_1 and U_2 are considered to be complex variables) does a good job of fitting phase angles extracted from neural LFP data. In addition, they described how several alternative families of distributions can be seen as special cases of torus graphs. They then applied their torus graph model to characterize a network graph of interactions among recordings from four brain regions during a memory task.

Although torus graphs provide interesting exponential families that could have been defined long ago, they would have been irrelevant to data analysis before recent practical developments used by Klein et al. (2020) to estimate the parameters of the model.

3.3.3. Characterization of Torus Graphs Using the Complex Normal

Distribution—We consider the polar coordinate representation of a random vector with a complex normal distribution and prove that the conditional distribution of the phases, given the amplitudes, form a torus graph. We provide a general statement and then offer additional theorems in the proper and circularly symmetric cases.

Suppose that $R_i \in [0, \infty)$ and $\Theta_i \in [-\pi, \pi)$ are the amplitude and phase of X_i for $i \in [d]$. In other words, (R_i, Θ_i) is the polar coordinate representation of X_i . We consider the conditional distribution of the phases given the amplitudes. Under certain parameter restrictions on a complex normal distribution, this conditional distribution forms a torus graph.

Theorem 2.: *Let $X \in \mathbb{C}^d$ be a complex normal random vector such that*

$(\Re X, \Im X) \sim \mathcal{N}(\mu, \Sigma)$. Let $\Omega = \Sigma^{-1}$ be the inverse covariance matrix. If $\Omega_{ii} = \Omega_{i+d, i+d}$ and $\Omega_{i, i+d} = 0$ for $i \in [d]$, then $\Theta \mid \{R = r\} \sim \mathcal{FZ}(\eta)$, where $\eta_i = r_i((\Omega\mu)_i, (\Omega\mu)_{i+d})$ and $\eta_{ij} = -r_i r_j (\Omega_{i,j}, \Omega_{i,j+d}, \Omega_{i+d,j}, \Omega_{i+d, j+d})/2$ for $i, j \in [d]$.

In Theorem 2, the components of the conditioning vector r appear as multipliers in the natural parameters but the graph is the same for every vector r .

There exists another distribution in the literature, known as the mGvM distribution, which is a more general version of the TG distribution that includes all second-moment terms (Navarro, Frellsen & Turner, 2017). Navarro et al. showed that, given a $2d$ -dimensional normal distribution analogous to the complex normal distribution, $\Theta \mid \{R = 1\}$ is a mGvM distribution without requiring any restrictions on the parameters of the original normal distribution. In the Supplementary Material Section S1, we generalize this result for any conditional distribution $\Theta \mid \{R = r\}$ and then leverage it to prove Theorem 2 (see the Supplementary Material Section S1.2).

We can obtain a more precise characterization of the conditional distribution under the parameter restrictions $C = 0$ and $m = 0$.

Theorem 3.: *If $X \sim \mathcal{CN}(m, \Gamma, 0)$, so that the complex normal distribution of X is proper, then $\Theta \mid \{R = r\} \sim \mathcal{TC}(\eta(\phi))$ as in (3), where $\phi_{i,3} = \phi_{i,4} = 0$ and $\phi_{i,j,k}$ denotes the k th component of $\phi_{i,j}$.*

Corollary 4.: *If $X \sim \mathcal{CN}(0, \Gamma, 0)$, so that the complex normal distribution of X is circularly symmetric, then $\Theta \mid \{R = r\} \sim \mathcal{TG}(\eta(\phi))$ with the same restrictions as in Theorem 3 and with the additional restriction that $\phi_i = 0$ for all $i \in [d]$.*

These theorems say that when the complex normal distribution is proper, the resulting torus graph family has only rotational dependence. The absence of reflectional dependence is intuitive from the definition in Section 3.3.2 of reflectional covariance as pseudo-covariance. Under circular symmetry, because $\phi_i = 0$, the conditional distribution $\Theta_i \mid \{R = r\}$ is uniform over $[0, 2\pi)$ and, additionally, Θ_i is marginally uniform. We repeat that the combination of uniformity and only rotational dependence is a particularly important special case, partly because this is the case in which PLV becomes a circular analogue to Pearson correlation. From a theoretical perspective, according to an argument given by Picinbono (1994), stationary band-pass-filtered signals with sufficiently narrow bands are proper. In the context of the particular neural data application reported here, we provide empirical evidence for circularity in Figure S2, which shows that the sample covariance matrix has much larger entries than the sample pseudo-covariance matrix.

4. COHERENCY AND THE COMPLEX NORMAL DISTRIBUTION

Assume we have a d -dimensional time series for which every dimension contains an oscillation at a given frequency. As discussed previously, these oscillations can be extracted by band-pass-filtering. Spectral dependence and coherency were analyzed using the complex normal distribution by Tugnait (2019a,b), who considered a more general setting in which the spectrum of the signals may be distributed across numerous frequencies. However, when band-pass-filtered signals all oscillate at the same frequency, we get simpler and more interpretable results, which we leverage using our latent variable model in Section 5.

The key to our approach is a well-known result that coherency is a form of complex correlation for stationary signals oscillating at a particular frequency (Ombao & Van Bellegem, 2008). For single-frequency signals, it further turns out that partial coherency, which is used to study conditional dependence among time series, is equivalent to partial correlation (Dahlhaus, 2000; Ombao & Pinto, 2022). For the complex normal distribution, we provide a systematic study of complex conditional correlation, complex partial correlation, and conditional independence. We also review the relationship between complex correlation and other measures of pairwise dependence.

4.1. Coherency and Pairwise Complex Correlation

Suppose we have stationary signals $X_1(t)$ and $X_2(t)$ on $t \in [0, 1]$, which may be complex-valued. Assume they have auto-covariance functions $\Sigma_{ii}(t) = \text{cov}\{X_i(t_0 + t), X_i(t_0)\}$ for $i \in \{1, 2\}$, and a cross-covariance function $\Sigma_{12}(t) = \text{cov}\{X_1(t_0 + t), X_2(t_0)\}$. These definitions do not depend on t_0 because of stationarity. For $\omega \in [-0.5, 0.5]$, the spectrum and cross-spectrum of the signals are

$$f_{ii}(\omega) = \int_0^1 \Sigma_{ii}(t)e^{-12\pi\omega t} dt \text{ and } f_{12}(\omega) = \int_0^1 \Sigma_{12}(t)e^{-12\pi\omega t} dt,$$

and the coherency is

$$\tau_{12}(\omega) = \frac{f_{12}(\omega)}{\sqrt{f_{11}(\omega)f_{22}(\omega)}}.$$

The coherency is complex-valued. Coherence is the magnitude of the coherency. The Cramér–Khinchine decomposition (see Ch. 3 of Brémaud, 2014) is

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \int_{-0.5}^{0.5} \exp(it2\pi\omega t) d \begin{pmatrix} Y_1(\omega) \\ Y_2(\omega) \end{pmatrix}.$$

where $(Y_1(\omega), Y_2(\omega))^T$ is a bivariate orthogonal-increment random process. In this case, the coherency between X_1 and X_2 at the frequency ω_0 is often considered as the complex-valued correlation coefficient between the infinitesimal increments of Y_1 and Y_2 at ω_0 , which is to say

$$\tau_{12}(\omega_0) = \text{corr}\{dY_1(\omega_0), dY_2(\omega_0)\}.$$

A quick way to see that this characterization makes sense is to consider two time series oscillating at a single frequency ω_0 , i.e.,

$$X_1(t) = R_1 \exp\{t(\Theta_1 + 2\pi\omega_0 t)\} \text{ and } X_2(t) = R_2 \exp\{t(\Theta_2 + 2\pi\omega_0 t)\},$$

where $R_i \in [0, \infty)$ and $\Theta_i \in [-\pi, \pi)$ are random variables, which we think of as representing the trial-specific amplitude and phase of X_i , respectively, for $i \in \{1, 2\}$.

Then the auto-covariance and cross-covariance kernels are $\Sigma_{ii}(t) = \text{var}(R_i e^{i\Theta_i}) e^{i2\pi\omega_0 t}$ and $\Sigma_{12}(t) = \text{cov}(R_1 e^{i\Theta_1}, R_2 e^{i\Theta_2}) e^{i2\pi\omega_0 t}$. The spectrum and cross-spectrum at the frequency ω_0 are

$$f_{kk}(\omega_0) = \int_0^1 \text{var}(R_k e^{i\Theta_k}) e^{i2\pi\omega_0 t} e^{-12\pi\omega_0 t} dt = \text{var}(R_k e^{i\Theta_k}), \quad k \in \{1, 2\}$$

(4)

and

$$f_{12}(\omega_0) = \int_0^1 \text{cov}(R_1 e^{t\Theta_1}, R_2 e^{t\Theta_2}) e^{i2\pi\omega_0 t} e^{-i2\pi\omega_0 t} dt = \text{cov}(R_1 e^{t\Theta_1}, R_2 e^{t\Theta_2}), \quad (5)$$

respectively. In this case, the coherency is the complex-valued correlation between $X_1(t)$ and $X_2(t)$ for every t :

$$\tau_{12}(\omega_0) = \frac{f_{12}(\omega_0)}{\sqrt{f_{11}(\omega_0)f_{22}(\omega_0)}} = \text{corr}(R_1 e^{t\Theta_1}, R_2 e^{t\Theta_2}) = \text{corr}\{X_1(t), X_2(t)\}.$$

As a generalization of the single-frequency case, if a pair of signals are band-pass-filtered over a small frequency band, it is possible to define “band coherency” over a narrow band of frequencies. In this case, the band coherency of the filtered signals is equal to the complex correlation of the filtered signals (Ombao & Van Bellegem, 2008).

4.2. Partial Correlation, Conditional Correlation, and Conditional Independence in the Complex Normal Distribution

In the case of real-valued random variables, conditional correlation and partial correlation have been studied extensively (Baba, Shibata & Sibuya, 2004). For such real-valued random variables, conditional correlation and partial correlation are distinct quantities, and conditional dependence between random variables need not imply that either quantity is nonzero. However, real random variables that are jointly normal have the following special properties: (i) conditional correlation and partial correlation are equal, and (ii) zero conditional correlation (or partial correlation) is equivalent to conditional independence.

In the complex case, the relationship between these quantities has been partially studied in Tugnait (2019a,b) and Andersen et al. (1995). Andersen et al. (1995) define conditional covariance for the proper complex normal distribution and show how it is related to various entries in the complex precision matrix. They further relate conditional dependence to conditional covariation for the proper complex normal distribution. Tugnait (2019a,b) discusses how to assess conditional dependence for proper and improper complex normal distributions.

In this section, we define partial correlation for complex random vectors in a manner analogous to the real case. We describe the differences between partial correlation and conditional correlation for complex random variables, and we provide results relating partial correlation and conditional independence, even in the improper complex normal case.

In what follows, assume we have some complex-valued random vector $X \in \mathbb{C}^d$ with finite first and second moments. For any subset $\mathcal{J} \subset [d]$, let $X_{\mathcal{J}}$ denote the random vector given by the elements of X whose indices are in \mathcal{J} (in order). Let $\tilde{X}_{\mathcal{J}} \in \mathbb{C}^{2|\mathcal{J}|}$ be the concatenation of $X_{\mathcal{J}}$ and its conjugate, i.e., $\tilde{X}_{\mathcal{J}} = (X_{\mathcal{J}}, \bar{X}_{\mathcal{J}})$. Let $\mathcal{L}(\tilde{X}_{\mathcal{J}}) = \{\alpha + \beta^H \tilde{X}_{\mathcal{J}} : \alpha \in \mathbb{C}, \beta \in \mathbb{C}^{2|\mathcal{J}|}\}$ be the linear space spanned by $\tilde{X}_{\mathcal{J}}$ and, for every

$i \in [d]$, let $\text{proj}_{\mathcal{X}(\tilde{x}, \mathcal{P})}(X_i) = \text{argmin}_{Y \in \mathcal{X}(\tilde{x}, \mathcal{P})} \mathbb{E}\{(Y - X_i)^H(Y - X_i)\}$ be the projection of X_i onto that space.

Definition 5.—For $i, j \in [d]$, the *partial correlation between X_i and X_j* is

$$\rho(X_i, X_j | X_{[d] \setminus \{i, j\}}) = \text{corr}\{X_i - \text{proj}_{\mathcal{X}(\tilde{x}_{[d] \setminus \{i, j\}})}(X_i), X_j - \text{proj}_{\mathcal{X}(\tilde{x}_{[d] \setminus \{i, j\}})}(X_j)\}$$

and the *partial pseudo-correlation between X_i and X_j* is

$$\tilde{\rho}(X_i, X_j | X_{[d] \setminus \{i, j\}}) = \text{pcorr}\{X_i - \text{proj}_{\mathcal{X}(\tilde{x}_{[d] \setminus \{i, j\}})}(X_i), X_j - \text{proj}_{\mathcal{X}(\tilde{x}_{[d] \setminus \{i, j\}})}(X_j)\}.$$

The *conditional correlation* is

$$\text{corr}(X_i, X_j | X_{[d] \setminus \{i, j\}}) = \frac{\text{cov}(X_i, X_j | X_{[d] \setminus \{i, j\}})}{\sqrt{\text{var}(X_i | X_{[d] \setminus \{i, j\}})\text{var}(X_j | X_{[d] \setminus \{i, j\}})}}.$$

The *conditional pseudo-correlation* is defined analogously by replacing cov with pcov.

We start by stating a correspondence between partial correlation and entries in a transformed precision matrix. The results in this section are proven in the Supplementary Material Section S1.

Theorem 6.—For any complex-valued random vector X , let $\tilde{\Gamma} = \text{var}(\tilde{X})$. Denote by $\text{diag}(A)$ the diagonal matrix, with diagonal entries the same as A . Let

$$P = \text{diag}(\tilde{\Gamma}^{-1})^{-1/2} \tilde{\Gamma}^{-1} \text{diag}(\tilde{\Gamma}^{-1})^{-1/2}.$$

Then $\rho(X_i, X_j | X_{[d] \setminus \{i, j\}}) = P_{i, j}$ and $\tilde{\rho}(X_i, X_j | X_{[d] \setminus \{i, j\}}) = P_{i, j+d}$.

Although conditional correlation and partial correlation may be different in general (Baba, Shibata & Sibuya, 2004), they coincide in the complex normal case.

Theorem 7.—If $X \sim \mathcal{CN}(m, \Gamma, C)$, then, for $i, j \in [d]$

$$\rho(X_i, X_j | X_{[d] \setminus \{i, j\}}) = \text{corr}(X_i, X_j | X_{[d] \setminus \{i, j\}})$$

and

$$\tilde{\rho}(X_i, X_j | X_{[d] \setminus \{i, j\}}) = \text{pcorr}(X_i, X_j | X_{[d] \setminus \{i, j\}}).$$

In the real-valued case, if the partial correlation is zero for jointly complex normal random variables, then the two variables are conditionally independent. Because in the complex case we have to consider both types of second-order association, we must consider both partial correlation and pseudo-partial correlation. For the following theorems, given random

variables A , B , and C , we write $A \perp\!\!\!\perp B \mid C$ if the conditional distributions $A \mid C$ and $B \mid C$ are independent.

Theorem 8.—If $X \sim \mathcal{CN}(m, \Gamma, C)$, then $\rho(X_i, X_j \mid X_{[d] \setminus \{i, j\}}) = 0$ and $\bar{\rho}(X_i, X_j \mid X_{[d] \setminus \{i, j\}}) = 0$ if and only if $X_i \perp\!\!\!\perp X_j \mid X_{[d] \setminus \{i, j\}}$.

A corollary of this theorem allows us to drop the reliance on the pseudo-correlation in the proper complex normal case since such a correlation must be zero.

Corollary 9.—If $X \sim \mathcal{CN}(m, \Gamma, 0)$, then $\rho(X_i, X_j \mid X_{[d] \setminus \{i, j\}}) = 0$ if and only if $X_i \perp\!\!\!\perp X_j \mid X_{[d] \setminus \{i, j\}}$.

By applying Theorems 6 and 8, we get the following corollary:

Corollary 10.—If $X \sim \mathcal{CN}(m, \Gamma, C)$, then $X_i \perp\!\!\!\perp X_j \mid X_{[d] \setminus \{i, j\}}$ if and only if $\tilde{\Gamma}_{ij}^{-1} = 0$ and $\tilde{\Gamma}_{i, j+a}^{-1} = 0$. In addition, if $C = 0$, then X_i, X_j are conditionally independent if and only if $\Gamma_{ij}^{-1} = 0$ for $\Gamma = \text{var}(X)$.

4.3. Partial Coherency and Partial Correlation

We now leverage the results from Section 4.2 to show a correspondence between partial coherency, a modification of coherency to account for conditional dependence, and partial correlation. Partial coherency was developed to better differentiate unique pairwise associations from associations that are shared with one or more other observed variables (Ombao & Pinto, 2022). Assume we have d signals $X_1(t), \dots, X_d(t)$ on $t \in [0, 1]$. Let $\Sigma_{ij}(t) = \text{cov}\{X_i(t + t_0), X_j(t_0)\}$, and define $f_{ij}(\omega)$ as in Section 4.1. Consider the matrix $\mathbf{f}(\omega)$ formed by assembling the entries of $f_{ij}(\omega)$. Let $\text{diag}\{\mathbf{f}(\omega)\}$ be the matrix containing the diagonal elements of $\mathbf{f}(\omega)$ and zero on the off-diagonals. Let

$$\mathbf{g}(\omega) = \text{diag}\left[\{\mathbf{f}(\omega)\}^{-1}\right]^{-1/2} \{\mathbf{f}(\omega)\}^{-1} \text{diag}\left[\{\mathbf{f}(\omega)\}^{-1}\right]^{-1/2}.$$

We call $\mathbf{g}(\omega)_{ij}$ the *partial coherency* between signals i and j . If $X_1(t), \dots, X_d(t)$ are all oscillating signals at some frequency ω_0 , we obtain the following theorem.

Theorem 11.—If $X_k(t) = R_k \exp\{i(\Theta_k + 2\pi\omega_0 t)\}$ for all $k \in [d]$ and $\text{pcorr}\{X_i(0), X_j(0)\} = 0$, then

$$\mathbf{g}(\omega_0)_{ij} = \rho\{X_i(0), X_j(0) \mid X_{[d] \setminus \{i, j\}}(0)\}.$$

This theorem follows from observing that, by our results in Section 4.1, $\Sigma_{ij}(0) = \mathbf{f}(\omega_0)_{ij}$ for such signals. We can then apply Theorem 6 and, observing that the upper-right block of $\tilde{\Gamma} = \text{cov}(\tilde{X})$ is zero because $\text{pcorr}\{X_i(0), X_j(0)\} = 0$, we obtain the desired result.

Researchers have previously studied partial coherency as a way to construct conditional independence graphs for time series (Dahlhaus, 2000). For instance, Tugnait (2019a,b) estimates conditional independence between time series by first constructing a sufficient statistic that accounts for the oscillations in a time series at all observable frequencies

and then, after observing that the limiting distribution of this statistic is a complex normal random variable, deriving conditional independence tests based on this limiting distribution. As we have seen, because we are dealing with signals oscillating at a single frequency, we can estimate partial coherency simply by computing the partial correlation between the observed variables. We will use this to interpret our latent-variable model. The Supplementary Material Section S2 details various simple examples in which PLV, amplitude correlation, and coherence are compared.

4.4. Alternative Measures of Pairwise Association Between Oscillating Signals

Thus far, we have focused on coherence as a measure of association for oscillating signals. As described in Section 3, other measures exist for studying the dependence between oscillating signals. In particular, we described PLV, which many investigators use as an alternative to coherence. Since PLV was first introduced, questions have arisen over whether to use PLV, coherence, or amplitude correlation to measure connectivity between oscillatory neural signals (e.g., Lachaux et al., 1999; Srinath & Ray, 2014; Lowet et al., 2016; Lepage & Vijayan, 2017). Our intention here is to review briefly some of these arguments to understand how coherence is situated among these other possible measures of association.

Clearly, coherence depends on both the phase and amplitude of oscillatory signals. Dependence on amplitude, however, has invited criticism of coherence as a pure representation of the degree of synchrony among the phases of oscillatory signals. PLV was introduced in part to overcome the perceived limitations of this dependence (Lachaux et al., 1999). Later work showed that, particularly in nonstationary settings, estimators of coherence are not well behaved and can fail to accurately represent synchrony (Lowet et al., 2016). Other authors have criticized coherence on the grounds that it can be biased by amplitude correlation (Srinath & Ray, 2014). However, some investigators have responded to these criticisms by pointing out that coherence may up-weight trials with larger amplitude oscillations, where information about phase is likely to be stronger (Lepage & Vijayan, 2017). Thus, the inclusion of amplitude information may be beneficial.

Further, some researchers have analyzed the relationship between PLV, amplitude correlation, and coherence when particular parametric models are assumed for data. For instance, Aydore, Pantazis & Leahy (2013) found that when data are distributed according to the complex normal distribution, PLV can be written as a function of coherence, which indicates that coherence provides all the information that PLV does. Further, Aydore, Pantazis & Leahy (2013) found that both the von Mises distribution and Gaussian models are effective for studying phase associations in LFP data, and argued that PLV and cross-correlation capture the same information. More generally, Nolte et al. (2020) made the remark that, because cross-correlation provides all possible information about the second-order statistical properties of circularly symmetric, bivariate, complex normal random variables, any pairwise measure of association can be written as a function of cross-correlation (and thus coherency). Nolte et al. found that PLV and related phase-based coupling statistics, as well as amplitude–amplitude correlational statistics, can be written as functions of coherence for complex normal random variables.

Thus, if the data are assumed to be proper and complex-normal, then complex correlation, and thus coherence, provides the most detailed measure of association available. For the oscillating neural signals that are the focus of this work, the complex normal model appears to be a reasonable approximation of the data, and so we focus on coherence as our measure of pairwise association since coherence may provide much of the information contained within other measures of association.

5. A COMPLEX-NORMAL, LATENT-VARIABLE MODEL

5.1. Model

We assume that we have recordings from K brain regions, and that the activity of each brain region is recorded by d_k electrodes ($1 \leq k \leq K$) over N repeated experimental trials. We assume that each electrode records an oscillating signal at some frequency, and we band-pass-filter the data to extract this oscillating signal as a preprocessing step to obtain a complex number that represents the phase and amplitude of the oscillation. For each trial $n \in [N]$, the observation X_k^n from brain region d_k is a d_k -dimensional complex-valued variable. We further assume that each X_k^n is driven by a univariate latent factor Z_k^n , which is also complex-variate, for the model

$$X_k^n = \beta_k Z_k^n + \epsilon_k^n, \quad \epsilon_k^n \sim \mathcal{CN}(0, \Phi_k, 0), \quad (6)$$

$$Z^n = (Z_1^n, \dots, Z_K^n) \sim \mathcal{CN}(0, \Gamma, 0), \quad (7)$$

where $\beta_k \in \mathbb{C}^{d_k}$ is a complex-valued factor loading, $\Gamma \in \mathbb{C}^{K \times K}$ is the latent covariance matrix, which is assumed to be positive definite, and $\epsilon_k^n \in \mathbb{C}^{d_k}$ is a region-specific noise with the covariance parameter $\Phi_k \in \mathbb{C}^{d_k \times d_k}$, which also is assumed to be positive definite.

Using basic properties of the complex normal distribution, the marginal distribution of the vector $X^n = (X_1^n, \dots, X_K^n)$ of all the observed signals is

$$X^n \sim \mathcal{CN}\left(0, \begin{pmatrix} \beta_1 \beta_1^H \Gamma_{11} + \Phi_1 & \beta_1 \beta_2^H \Gamma_{12} & \beta_1 \beta_3^H \Gamma_{13} & \dots \\ \beta_2 \beta_1^H \Gamma_{21} & \beta_2 \beta_2^H \Gamma_{22} + \Phi_2 & \beta_2 \beta_3^H \Gamma_{23} & \dots \\ \beta_3 \beta_1^H \Gamma_{31} & \beta_3 \beta_2^H \Gamma_{32} & \beta_3 \beta_3^H \Gamma_{33} + \Phi_3 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, 0\right). \quad (8)$$

This representation shows that the model is nonidentifiable without adding additional constraints. In particular, β_k can be scaled by an arbitrary complex number, and corresponding entries in Γ scaled by the inverse conjugate of that number, with no change to the marginal distribution. Additionally, given an arbitrary real value $a \in \mathbb{R}$, we have that

$\beta_k \beta_k^H (\Gamma_{kk} + a) + (\Phi_k - a \beta_k \beta_k^H) = \beta_k \beta_k^H \Gamma_{kk} + \Phi_k$, which indicates that, while $\Phi_k - a \beta_k \beta_k^H$ is positive semidefinite (PSD), we can add to the diagonal of Γ_{kk} and subtract $a \beta_k \beta_k^H$ from Φ_k and retain the same marginal distribution.

Thus, to ensure the identifiability of the parameters, we require that

1. $\beta_k^H \beta_k = 1$ and $\Im \mathbf{m} \beta_{k,1} = 0$, where $\beta_{k,1}$ is the first component of β_k , and
2. $\sup\{a \geq 0 : \Phi_k - a \beta_k \beta_k^H \geq 0\} = 0$.

To better understand the model, we show how it can produce maximum likelihood estimates that are related to the solutions for a complex extension of canonical correlation analysis (CCA; Hotelling, 1936). Bong et al. (2023) found that a real-valued, latent-variable model, which bears some similarity to the complex-valued, latent-variable model described here, could be viewed as a form of probabilistic CCA, first described in Bach & Jordan (2005). Because CCA has a distribution-free definition, the correspondence we show below with an analogue of CCA for complex-valued random vectors provides a similar distribution-free motivation for the estimation procedure we derive here.

If we observe a pair of real random vectors $Y_1 \in \mathbb{R}^{d_1}$ and $Y_2 \in \mathbb{R}^{d_2}$, CCA finds weights w_1, w_2 (where $w_1 \in \mathbb{R}^{d_1}, w_2 \in \mathbb{R}^{d_2}$) that maximize $|\text{corr}(w_1^T Y_1, w_2^T Y_2)|$. Kettenring (1971) extended CCA to multiple vector observations. For $(Y_1 \in \mathbb{R}^{d_1}, \dots, Y_K \in \mathbb{R}^{d_K})$, this so-called ‘‘multiset CCA’’ finds weights $w_1 \in \mathbb{R}^{d_1}, \dots, w_K \in \mathbb{R}^{d_K}$ maximizing a notion of the cross-correlation among $(w_1^T Y_1, \dots, w_K^T Y_K)$. Multiset CCA easily extends to complex-valued random vectors, $X_1 \in \mathbb{C}^{d_1}, \dots, X_K \in \mathbb{C}^{d_K}$. In particular, we can write the optimization problem solved by complex multiset CCA as

$$\arg \min_{w_k \in \mathbb{C}^{d_k}: k \in [K], \text{var}(w_k^H X_k) = 1} \det\{\text{var}(w_1^H X_1, \dots, w_K^H X_K)\}. \tag{9}$$

The solution weights $\{w_{k,cc}: k \in [K]\}$ that achieve the optimum of (9) are called canonical weights, and the resulting correlation matrix $P_{cc} = \text{corr}(w_{1,cc}^H X_1, \dots, w_{K,cc}^H X_K)$ is called a canonical correlation matrix. The sample estimates $\hat{w}_{k,cc}$ and \hat{P}_{cc} are obtained by replacing corr in (9) with the sample version $\widehat{\text{corr}}$ based on the observed signals (X_1^n, \dots, X_K^n) during experimental trials $n \in [N]$. We now state a theorem describing the equivalence between the canonical weights, the canonical correlation matrix, and the maximum likelihood estimator of the parameters in the model given in (6) and (7) under the given identifiability constraints. A proof is given in the Appendix.

Theorem 12.—*Suppose that $\hat{\beta}_k, k \in [K]$, and $\hat{\Gamma}$ are the maximum likelihood estimators of the parameters in (6) and (7) under the given identifiability constraints, labelled as 1 and 2 following (8), based on N observed tuples (X_1^n, \dots, X_K^n) . We have the following equivalence between the maximum likelihood estimators, the canonical weights, and the canonical correlation matrix:*

$$\hat{\beta}_k = \frac{\widehat{\text{var}}(X_k)\widehat{w}_{k,cc}}{\|\widehat{\text{var}}(X_k)\widehat{w}_{k,cc}\|_2} \text{ and } \text{diag}(\widehat{\Gamma}_{kk}^{-1/2}) \widehat{\Gamma} \text{diag}(\widehat{\Gamma}_{kk}^{-1/2}) = \widehat{P}_{cc}$$

for $k \in [K]$, where $\text{diag}(\widehat{\Gamma}_{kk}^{-1/2})$ is the diagonal matrix with entries $\widehat{\Gamma}_{kk}^{-1/2}$.

As a corollary of the theorem, if $K = 2$, then our model is equivalent to the pairwise CCA for two complex random vectors $X_1 \in \mathbb{C}^{d_1}$ and $X_2 \in \mathbb{C}^{d_2}$. To demonstrate this equivalence, let us write

$$w_{1,cc}, w_{2,cc} = \arg \max_{w_1 \in \mathbb{C}^{d_1}, w_2 \in \mathbb{C}^{d_2}} |\text{corr}(w_1^\top X_1, w_2^\top X_2)|,$$

let ρ_{cc} be the resulting canonical correlation coefficient, and let $\widehat{w}_{1,cc}$ and $\widehat{w}_{2,cc}$ be the corresponding sample estimators. We can then state the following corollary.

Corollary 13.—*If $\widehat{\beta}_1, \widehat{\beta}_2$, and $\widehat{\Gamma}$ are the maximum likelihood estimators of (6) and (7) under the given identifiability constraints, then they are related to the weights and canonical correlation values through*

$$\widehat{\beta}_1 = \frac{\widehat{\text{var}}(X_1)\widehat{w}_{1,cc}}{\|\widehat{\text{var}}(X_1)\widehat{w}_{1,cc}\|_2}, \quad \widehat{\beta}_2 = \frac{\widehat{\text{var}}(X_2)\widehat{w}_{2,cc}}{\|\widehat{\text{var}}(X_2)\widehat{w}_{2,cc}\|_2}, \quad \text{and} \quad \frac{\widehat{\Gamma}_{12}}{\sqrt{\widehat{\Gamma}_{11}\widehat{\Gamma}_{22}}} = \widehat{\rho}_{cc}.$$

In the general case with more than two latent factors, fitting the marginal likelihood directly is an intractable optimization problem. Instead, we fit the model using the expectation maximization (EM) algorithm; we describe our fitting procedure in the Appendix. In Figure S1, we provide evidence that, for simulated datasets similar to the real dataset we analyze in Section 6, the estimates provided by the EM fitting procedure converge to the true parameters as the size of the dataset increases. Code for performing inference and analysis with our model is available at https://github.com/urbkn7/latent_cn.

In applying Theorem 12 to the data in Section 6, we first band-pass-filter the data so that the magnitude of the complex correlation becomes the coherence (as discussed in Section 4.1). We also decompose the covariance matrix to get partial coherence (Section 4.3).

5.2. Inference

We study the strength of the associations among the estimated latent factors in our model. To do so, we develop a parametric bootstrap procedure to test for conditional dependence between brain areas. Recall from Section 4 that for vectors distributed according to the complex normal distribution, zero entries in the latent precision matrix correspond to conditional independence relationships between the associated dimensions of the random vector. Therefore, for two regions with indices $r_1, r_2 \in [R]$, we want to test the null hypothesis $\mathcal{H}_0: \Gamma_{r_1, r_2}^{-1} = 0$ against the alternative hypothesis $\mathcal{H}_a: \Gamma_{r_1, r_2}^{-1} \neq 0$.

To employ the parametric bootstrap, we first estimate the parameters of the model on the original dataset. Let $\Theta = \{\{\beta_k: k \in [K]\}, \{\Phi_k: k \in [K]\}, \Gamma\}$, and let $\mathcal{L}(\Theta; \mathcal{D})$ denote the likelihood of the parameters under the model given by (8) for the dataset $\mathcal{D} = \{X^n: n \in [N]\}$. We first estimate Θ with the EM algorithm to obtain $\hat{\Theta}(\mathcal{D})$, our approximation of the maximum likelihood estimator for the dataset \mathcal{D} . Then, we optimize the likelihood with a modified EM algorithm under the additional constraint that $\Gamma_{r_1, r_2}^{-1} = 0$ (see the Appendix for details on how we solve this optimization problem) to obtain $\hat{\Theta}_0(\mathcal{D})$. We then simulate datasets $\mathcal{D}^{(b)} \sim P_{\hat{\Theta}_0}$, where $1 \leq b \leq B$ and B is the total number of bootstrap datasets. Let

$$\lambda(\mathcal{D}) = \frac{\mathcal{L}\{\hat{\Theta}_0(\mathcal{D}); \mathcal{D}\}}{\mathcal{L}\{\hat{\Theta}(\mathcal{D}); \mathcal{D}\}}.$$

We can observe the quantile of the likelihood ratio statistic for the original data, $\lambda(\mathcal{D})$, relative to the set of likelihood ratio statistics for the bootstrapped datasets $\{\lambda(\mathcal{D}^{(b)}): b \in [B]\}$ to obtain a p -value. We can also obtain confidence intervals for parameters of interest by employing the ordinary nonparametric bootstrap. The parameters we are most interested in are contained in the latent correlation and conditional correlation matrices. We have the latent correlation matrix $S = \text{diag}(\Gamma)^{-1/2} \Gamma \text{diag}(\Gamma)^{-1/2}$ and the conditional correlation matrix $D = \text{diag}(\Gamma^{-1})^{-1/2} \Gamma^{-1} \text{diag}(\Gamma^{-1})^{-1/2}$.

6. DATA ANALYSIS

We apply the techniques we have discussed so far to a dataset of LFP data from the Allen Institute (Siegle et al., 2021). In the experiment, six electrode probes were simultaneously inserted into the mouse brain, with each probe targeting some area of the visual cortex but also recording other regions. During the experiment, the mice were presented with a variety of visual stimuli; here, we focus on presentations of drifting grating stimuli, which appear as bars moving across a screen in the mouse's field of view. In the interest of utilizing as much data as possible, we ignore differences in the direction and size of the stimuli presented. This gives us 630 trials.

The experimenters marked every electrode with the anatomical brain area in which it resided during the experiment. We observed an oscillation at 6.5Hz in many of these areas. Therefore, we band-pass-filtered the LFP signal and then used the Hilbert transform to recover the analytic signal. We selected a single time point, 2 s into the trial, which occurs well after the stimulus disappears from the screen. Selecting this time point removes the influence of trial-locked changes in brain activity. In addition, we used five electrodes from each of the six regions for which we have data to form the set of time series we analyze.

We then applied the latent variable model to this dataset. For all analyses, we used one latent variable per brain area analyzed. Each brain area has multiple electrodes embedded, which record the LFP signals; we select five electrodes from each region, and the signals from these electrodes are the observed variables in our model. In Figure 4, we show data from a single probe passing through four visual regions (AM, PM, V1, and LM) and two

hippocampal regions (CA1 and DG). We display the empirical correlation matrices of the observed data as well as the latent correlation matrices and partial correlation matrices estimated using our model. These matrices of latent correlation and partial correlations are of interest since, as discussed in Section 4, in our setting they are equivalent to the coherency and partial coherency between regions. While the estimated latent marginal correlations are relatively large between many areas, the estimated partial correlations (the $|\widehat{D}_{ij}|_s$) are strongest between two pairs of areas: between AM and PM as well as between CA1 and DG.

We perform the parametric bootstrap test discussed in Section 5.2 to study the significance of the entries in the latent partial correlation matrix. In Figure 5, we plot the results of this analysis. We observe significant correlations between the majority of pairs of areas for the 6.5Hz frequency.

To further understand how strong the correlations between these areas are, we report bootstrap confidence intervals for the significantly nonzero parameters of the correlation and partial correlation matrices in Figure 6. For nearly all pairs of regions, the partial correlations are of modest size and are substantially smaller than the corresponding marginal correlations. However, the pairs PM–AM and DG–CA1 have large partial correlations close to 1. Also of interest are the patterns of association for the LM area. Note that the strength of the correlation between the LM and the CA1 areas is larger than that between LM and any other visual region, suggesting that LM has stronger oscillatory associations at 6.5Hz frequency with hippocampal regions than with other visual regions, despite LM itself being a visual region.

7. DISCUSSION

We began with two separate goals, and ended by merging them. On the one hand, we wanted to see whether the multivariate complex normal distribution might be useful in analyzing dependence among groups of oscillating time series. On the other hand, we wished to better understand complex correlation. Our investigation, built on existing literature concerning dependence among complex random variables and complex normal variables in particular, uncovered several interesting relationships that were either not known previously or not spelled out clearly. For complex random vectors, we reviewed the distinction between partial correlation and conditional correlation (Baba, Shibata & Sibuya, 2004), which are equivalent for real, normal random vectors, and we showed that the complex partial correlation and conditional correlation also coincide under complex normality. We phrased many of our theorems in terms of the general form of complex normality, including both covariance and pseudo-covariance, and then specialized to the proper case where pseudo-covariances vanish. We showed that in the proper case, pairwise conditional independence coincides with zero partial correlation, as it does for real multivariate normal distributions, and that partial coherency may be considered a partial complex correlation at a given frequency. These facts are important for analytic interpretation, as we demonstrated in our real data example.

The scientific backdrop for our work on the analysis of neural data is one of the most pressing problems in the application of statistics to neurophysiology. Specifically, the problem lies in identifying coordinated activity across two or more regions of the brain

based on recordings of multiple time series that are highly nonstationary but are repeated across experimental trials. When there are many repetitions of multiple recorded values in two regions at a single time point, CCA provides a solution to the problem of determining their dependence. In the two-region case, we showed that maximum likelihood estimation for our latent variable model produces a form of latent coherence which is equivalent to the magnitude of the complex canonical correlation. In the multi-region case, maximum likelihood estimation for our model produces a generalized canonical correlation.

The latent coherencies and partial coherencies from our latent-variable model are not computed in the frequency domain but are obtained instead in the time domain with band-pass-filtering in a narrow band. For this we leaned on a key observation by Ombao & Van Bellegem (2008) that “band coherence” (a version of coherence written in terms of integrals over the band) could be interpreted as the magnitude of complex correlation of two signals. We used our results for the complex normal distribution to interpret the complex correlations and partial correlations we found in the data.

Part of understanding coherence requires a comparison to PLV, which can be generalized to the multivariate setting with torus graphs. The original motivation for PLV was discomfort with the dependence of coherence on amplitude (Lachaux et al., 1999). Like PLV, torus graphs ignore amplitude variation and any possible cross-covariation between amplitudes and phases. Thus, torus graphs might be considered models of covariation among phases after marginalizing over amplitudes. The complex normal results characterizing torus graphs as conditional distributions after conditioning on amplitudes show that even though the graphical structure (the conditional independence structure) of torus graphs does not depend on amplitude, calibration of the magnitude of interaction effects apparently depends on the amplitudes. When the amplitudes are roughly constant, coherence and PLV provide essentially equivalent results (see the Supplementary Material Section S2 and Lepage & Vijayan, 2017).

The results we found in the data are striking and intriguing. There are large partial coherences in two pairs of areas, one involving higher visual areas (AM and PM, the anterior and posterior parts of the medial visual area, which are further downstream than primary visual cortex, V1) and the other involving the hippocampus (CA1, a sub-region of the cornu ammonis and DG, the dentate gyrus). Both pairs involve areas that are contiguous, but that alone does not explain their interaction because all of these areas share anatomical boundaries with some other areas. In the case of CA1 and DG, these are the only hippocampal areas in the data. The coherence between the AM and PM areas could indicate close collaboration in neural processing and may be worth further investigation. In addition, we observed stronger coherences between LM and hippocampal regions than between LM and other visual regions, which could also be a subject of further investigation.

Often, there are dozens of time series in each region, and latent-variable models are attractive ways to reduce dimensionality for examining cross-region interactions. In an unpublished work (Orellana & Kass, 2023), a hierarchical model based on torus graphs has been used to describe large numbers of phase measurements made in each of several brain areas. This hierarchical structure has the advantage of reducing the total number

of parameters, which results in better statistical inference when the amount of data is small relative to its dimensionality. To assess time-varying amplitude interaction, Bong et al. (2023) developed a time-series generalization of a factor analysis model with one latent factor for each region. They allowed for nonstationarity and included all relevant time-lagged cross-correlations. They first showed how the model leads to a time-series generalization of probabilistic CCA, based on multiset CCA (analogous to Theorem 12 here). Because of nonstationarity, each combination of a time point in one region and a time point in another region could have a unique correlation across trials, which made the covariance matrix have a large number of free parameters. The authors adapted sparse estimation methods to solve the high-dimensional inference problem and showed how it produced interpretable and interesting results when applied to their data. Closely related work can be found in Bong et al. (2020).

It would be straightforward to apply the approach here at many time points, but modelling time lags, as is done in Bong et al. (2023), would require additional work. It would also be possible to estimate partial coherence, even at a single point, using a latent, multivariate time series model, but this would require either a very high dimensional formulation along the lines of Bong et al. (2023) or a specific time series model, both of which present their own challenges.

To simplify interpretation, it was important for us to assume that the complex normal distributions were proper. In our data, this seemed to be a reasonable assumption, and in all the data we have examined using torus graphs, reflectional dependence is, similarly, either absent or difficult to detect. Perhaps future uses of the complex normal distribution, along the lines outlined here, will reveal situations in which pseudo-covariance needs to be considered. The substantial additional complication of such cases would likely present new challenges, but we hope the framework we have summarized here would provide a useful starting point.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported in part by the National Institute of Mental Health (grant RO1 MH064537). Urban was supported in part by the National Institute on Drug Abuse (grant 5T90DA022762).

APPENDIX

A NOTE ON THE HILBERT TRANSFORM

Here we provide an overview of the Hilbert transform, which is commonly used to recover phase and amplitude from oscillating signals (Cohen, 2014). Let $X(t) = \Re X(t) + i\Im X(t)$ be the output of a complex signal that has been filtered in a band $(\omega_0 - \delta, \omega_0 + \delta)$.

We only observe $\Re X(t)$, and the problem is to recover $X(t)$, which is possible when δ is sufficiently small. The Hilbert transform operates on $\Re X(t)$ to produce $\Im X(t)$ according to

$$\Im X(t) = \text{p. v.} \int \frac{\Re X(s)}{\pi(t-s)} ds, \quad (\text{A1})$$

where the integral is over the domain of t . The notation $\text{p. v.} \int f(t) dt$ denotes the Cauchy principal value; this formula is given in numerous sources (e.g., Pandey, 2011), but because we have not seen a concise derivation, we provide one here.

Let us write the Fourier transforms of $X(t)$ and its complex conjugate $\bar{X}(t)$ as $\mathcal{F}(X)$ and $\mathcal{F}(\bar{X})$, and the evaluation of such transforms at a frequency ω by $\mathcal{F}(X)(\omega)$, etc. From the general relations

$$\Re X(t) = \frac{1}{2} \{X(t) + \bar{X}(t)\} \quad \text{and} \quad \Im X(t) = -\frac{i}{2} \{X(t) - \bar{X}(t)\},$$

we have the Fourier transform

$$\mathcal{F}(\Re X) = \frac{1}{2} \{ \mathcal{F}(X) + \mathcal{F}(\bar{X}) \} \quad \text{and} \quad \mathcal{F}(\Im X) = -\frac{i}{2} \{ \mathcal{F}(X) - \mathcal{F}(\bar{X}) \}.$$

Note that $\mathcal{F}(X)(\omega)$ is, from the band-pass filtering, concentrated around ω_0 , and $\mathcal{F}(\bar{X})(\omega)$ is concentrated around $-\omega_0$. Therefore, when $\omega > 0$ the formula above gives $\mathcal{F}(\Re X)(\omega) = \mathcal{F}(X)(\omega)/2$ and, similarly, when $\omega < 0$, $\mathcal{F}(\Re X)(\omega) = \mathcal{F}(\bar{X})(\omega)/2$. Thus, using $\text{sgn}(\omega)$ to denote the sign of ω , when we multiply $\mathcal{F}(\Re X)(\omega)$ by $-i \text{sgn}(\omega)$, we get

$$\{-i \text{sgn}(\omega)\} \mathcal{F}(\Re X)(\omega) = \mathcal{F}(\Im X)(\omega).$$

As a result, the convolution $\Re X(t) * \mathcal{F}^{-1}\{-i \text{sgn}(\omega_0)\}$ satisfies

$$\Re X(t) * \mathcal{F}^{-1}\{-i \text{sgn}(\omega)\} = \Im X(t).$$

Inserting the formula for $\mathcal{F}^{-1}\{-i \text{sgn}(\omega_0)\}$ into the definition of convolution produces (A1), giving the Hilbert transform of $\Re X(t)$. The resulting $X(t)$ is called the analytic signal.

FITTING PROCEDURE FOR THE LATENT FACTOR MODEL IN SECTION 5

In this section, we discuss the procedure for fitting the latent factor model (Eqs. 6 and 7) in Section 5 to the data $\{X_1^n, \dots, X_k^n\}_{n=1, \dots, N}$ from N independent repeated observations. The procedure provides estimates of the model parameters Γ , β_k , and Φ_k using the EM algorithm (Dempster, Laird & Rubin, 1977). The EM algorithm starts with initial estimates for the parameters, denoted by $\Gamma^{(0)}$, $\beta_k^{(0)}$, and $\Phi_k^{(0)}$, and iteratively updates the estimates to optimize the likelihood through alternating E-steps and M-steps, which we will describe next. We denote the estimates after the r th update by $\Gamma^{(r)}$, $\beta_k^{(r)}$, and $\Phi_k^{(r)}$.

At the $(r + 1)$ th iteration, the E-step calculates the sufficient statistics of the latent factors $\{Z_1^n, \dots, Z_k^n\}$ conditional on the observed data $\{X_1^n, \dots, X_k^n\}$ and the parameter estimates $\Gamma^{(r)}$, $\beta_k^{(r)}$, and $\Psi_k^{(r)}$ after the r th iteration. The full joint probability density of the observed data and the latent factors, given the r th parameter estimates, is

$$p(X, Z; \Gamma^{(r)}, \beta_k^{(r)}, \Phi_k^{(r)}) \propto \frac{1}{\det(\Gamma^{(r)})} \exp\left(-Z^H \{\Gamma^{(r)}\}^{-1} Z\right) \times \prod_{k=1}^K \left(\frac{1}{\det(\Phi_k^{(r)})} \exp\left\{-(X_k - \beta_k^{(r)} Z_k)^H \{\Phi_k^{(r)}\}^{-1} (X_k - \beta_k^{(r)} Z_k)\right\} \right),$$

where $Z = (Z_1, \dots, Z_k)$ is the concatenation of the latent factors. Further,

$$E(Z | X, \Gamma^{(r)}, \beta_k^{(r)}, \Phi_k^{(r)}) = \left(\{\Gamma^{(r)}\}^{-1} + D^{(r)} \right)^{-1} V^{(r)},$$

$$\text{var}(Z | X, \Gamma^{(r)}, \beta_k^{(r)}, \Phi_k^{(r)}) = \left(\{\Gamma^{(r)}\}^{-1} + D^{(r)} \right)^{-1}, \text{ and}$$

$$\text{pvar}(Z | X, \Gamma^{(r)}, \beta_k^{(r)}, \Phi_k^{(r)}) = O_K,$$

where $D^{(r)}$ is a diagonal matrix with entries $D_{kk}^{(r)} = \beta_k^{(r)H} \{\Phi_k^{(r)}\}^{-1} \beta_k^{(r)}$, $V^{(r)}$ is a vector of elements $V_k^{(r)} = \beta_k^{(r)H} \{\Phi_k^{(r)}\}^{-1} X_k$, and O_K is the $K \times K$ zero matrix. For brevity, we denote $E(Z^n | X^n, \Gamma^{(r)}, \beta_k^{(r)}, \Phi_k^{(r)})$ and $\text{var}(Z^n | X^n, \Gamma^{(r)}, \beta_k^{(r)}, \Phi_k^{(r)})$ by $E^{n(r)}$ and $\text{var}^{(r)}$. (We note that $\text{var}^{(r)}$ is invariant over $n \in [N]$.)

In the M-step, we find the parameters maximizing the expectation of the full log-likelihood function with respect to both $\{X^n\}_n$ and $\{Z^n\}_n$ conditional on $\{X_1^n, \dots, X_k^n\}_n$, $\Gamma^{(r)}$, $\beta_k^{(r)}$, and $\Phi_k^{(r)}$. The conditional expectation of the full log-likelihood function for the model is

$$E(\ell(\Gamma, \beta_k, \Phi_k; \{X^n, Z^n\}_n) | X, \Gamma^{(r)}, \beta_k^{(r)}, \Phi_k^{(r)}) \stackrel{+C}{=} \sum_{n=1}^N \left[-\ln \det(\Gamma) - \text{tr} \left\{ \Gamma^{-1} \left(\text{var}^{(r)} + E^{n(r)} E^{n(r)H} \right) \right\} \right. \\ \left. + \sum_{k=1}^K \left\{ -\ln \det(\Phi_k) - (X_k^n)^H \Phi_k^{-1} X_k^n + \bar{E}_k^{n(r)} \beta_k^H \Phi_k^{-1} X_k^n + (X_k^n)^H \Phi_k^{-1} \beta_k E_k^{n(r)} - \beta_k^H \Phi_k^{-1} \beta_k \left(\text{var}_{kk}^{(r)} + E_k^{n(r)} \bar{E}_k^{n(r)} \right) \right\} \right].$$

The parameters are complex-valued, so we use Wirtinger calculus to take derivatives. Using formulas from Adali, Schreier & Scharf (2011), we take derivatives and set them equal to zero to achieve the update steps

$$\Gamma^{(r+1)} = \text{var}^{(r)} + \frac{1}{N} \sum_{n=1}^N E^{n(r)} E^{n(r)H},$$

$$\beta_k^{(r+1)} = \frac{\sum_n X_k^n \bar{E}_k^{n(r)}}{\sum_n (\text{var}_{kk}^{(r)} + E_k^{n(r)} \bar{E}_k^{n(r)})}, \text{ and}$$

$$\Phi_k^{(r+1)} = \text{var}_{kk}^{(r)} \beta_k^{(r+1)} (\beta_k^{(r+1)})^H + \frac{1}{N} \sum_{n=1}^N (X_k^n - \beta_k^{(r+1)} E_k^{n(r)}) (X_k^n - \beta_k^{(r+1)} E_k^{n(r)})^H.$$

We also have the identifiability constraints $\beta_k^H \beta_k = 1$, $\Im \mathbf{m} \beta_k(1) = 0$, and $\sup\{a \geq 0: \phi_k - a \beta_k \beta_k^H \geq 0\} = 0$. To produce estimates that satisfy these constraints, we utilize the following procedure. First, we adjust the elements of $\hat{\Gamma}$ and $\hat{\beta}_k$ so that $\hat{\beta}_k^H \hat{\beta}_k = 1$, and $\Im \mathbf{m} \hat{\beta}_k(1) = 0$; this can be accomplished by multiplying the rows and columns in $\hat{\Gamma}$ and the $\hat{\beta}_k$ vectors by a complex-valued scalar. To ensure that $\sup\{a \geq 0: \phi_k - a \beta_k \beta_k^H \geq 0\} = 0$, we solve a convex optimization problem by maximizing α under the constraint that $\phi_k - a \beta_k \beta_k^H \geq 0$. We then subtract $a \hat{\beta}_k \hat{\beta}_k^H$ from $\hat{\phi}_k$ and add α to Γ_{kk} .

We use a procedure based on the method of moments to obtain starting values for the EM algorithm, a strategy that has been shown theoretically to often result in near-optimal parameter estimates (Balakrishnan, Wainwright & Yu, 2017). To do so, we start by computing the sample covariance matrix over the data. Then, we estimate the parameter vectors β_k by examining the submatrices given by $\widehat{\text{cov}}(X_k^n, X_i^n)$. In particular, because under the correct model specification, we have that $\text{cov}(X_k^n, X_i^n) = \Gamma_{ki} \beta_k \beta_i^H$, we can take any column of $\widehat{\text{cov}}(X_k^n, X_i^n)$ and normalize it according to the identifiability constraints to obtain an estimate of β_k . We do this and average over all the corresponding columns to get an overall estimate of β_k . Once β_i and β_j have been estimated, it is then straightforward to obtain estimates of Γ_{ij} by dividing the entries in $\widehat{\text{cov}}(X_i^n, X_j^n)$ by those in $\beta_i \beta_j^H$.

Now we briefly address our strategy for parameter estimation under a null hypothesis, as is done in the inference method we introduce. Under the null hypothesis, we are assuming that some entry of the latent precision matrix, Γ_{ij}^{-1} , is zero. To satisfy this assumption, we need to modify our estimation procedure. Observe that the conditional expectation of the log-likelihood depends on Γ through

$$-\ln \det \Gamma - \text{tr} \left\{ \Gamma^{-1} \left(\text{var}^{(r)} + E^{n(r)} E^{n(r)H} \right) \right\}.$$

Thus, we need to optimize this expression over Γ under the constraints that $\Gamma_{ij}^{-1} = 0$ while ensuring that Γ remains a Hermitian PSD matrix. Fortunately, this is a convex optimization problem, and while there is no explicit expression for the optimal Γ under these constraints,

we can solve this problem numerically using a convex optimization package. Then, we can estimate the remaining variables using the explicit formulas given above.

PROOF OF THEOREM 12

Let $u_k = S_{kk}^{-1/2} \beta_k \Gamma_{kk}^{1/2}$ and $\Psi_k = S_{kk}^{-1/2} \Phi_k S_{kk}^{-1/2}$, where $S_{k\ell} = \beta_k \Gamma_{k\ell} \beta_\ell^H + \Phi_k \delta_{k\ell}$ is a submatrix of the marginal covariance matrix S of (X_1, \dots, X_K) and $\delta_{k\ell} = 1$ if $k = \ell$ and $\delta_{k\ell} = 0$ otherwise, for $k, \ell \in [K]$. Then the second identifiability constraint can be rewritten as $u_k^H u_k = \beta_k^H S_{kk}^{-1} \beta_k \Gamma_{kk} = 1$ and

$$\begin{aligned} u_k^H \Psi_k u_k &= \Gamma_{kk}^{1/2} \beta_k^H S_{kk}^{-1} \Phi_k S_{kk}^{-1} \beta_k \Gamma_{kk}^{1/2} \\ &= \Gamma_{kk}^{1/2} \beta_k^H S_{kk}^{-1} (S_{kk} - \beta_k \Gamma_{kk} \beta_k^H) S_{kk}^{-1} \beta_k \Gamma_{kk}^{1/2} \\ &= \beta_k^H S_{kk}^{-1} \beta_k \Gamma_{kk} - (\beta_k^H S_{kk}^{-1} \beta_k \Gamma_{kk})^2 \\ &= 1 - 1^2 = 0, \quad k \in [K]. \end{aligned}$$

That is, u_k is orthogonal to Ψ_k . Denoting the block diagonal matrix of $\{S_{kk} : k \in [K]\}$ by V , we have that $R = V^{-1/2} S V^{-1/2}$ has submatrices

$$R_{k\ell} = S_{kk}^{-1/2} S_{k\ell} S_{\ell\ell}^{-1/2} = u_k \Gamma_{kk}^{-1/2} \Gamma_{k\ell} \Gamma_{\ell\ell}^{-1/2} u_\ell^H + \Psi_k \delta_{k\ell}.$$

Because of the orthogonality between u_k and Ψ_k , the calculation of $\det(R)$ and R^{-1} is straightforward: $\det(R) = \det(\Omega) / \prod_k \text{pdet}(\Psi_k)$ and $Q = R^{-1}$ consists of submatrices

$$Q_{kl} = u_k \Omega_{kl} u_\ell^H + \Psi_k^+ \delta_{k\ell},$$

where $\Omega = \text{diag}(\Gamma_{kk}^{1/2}) \Gamma^{-1} \text{diag}(\Gamma_{kk}^{1/2})$ is the inverse correlation matrix and $\text{pdet}(A)$ and A^+ are the pseudo determinant and Moore–Penrose pseudo-inverse of a PSD matrix A . Notice that $\Psi_k = I - u_k u_k^H = \Psi_k^+$ and hence $\text{pdet}(\Psi_k) = 1$. In turn, the negative log-likelihood under the model (Eqs. 6 and 7) of the parameter set $\{\Gamma\} \cup \{\beta_k, \Phi_k : k \in [K]\}$ with respect to the observed time series $\{X_1^n, \dots, X_K^n\}_{n=1, \dots, N}$ is

$$\begin{aligned} \text{nll}(\Gamma, \beta_k, \Phi_k; \{X_1^n, \dots, X_K^n\}_{n=1, \dots, N}) &= \ln \det(S) + \text{tr}(S^{-1} \hat{S}) \\ &= -\ln \det(\Omega) + \sum_k \ln \text{pdet}(\Psi_k) + \sum_k \ln \det(S_{kk}) + \text{tr}(\Omega \hat{P}) + \sum_k \text{tr}(\Psi_k^+ S_{kk}^{-1/2} \hat{S}_{kk} S_{kk}^{-1/2}) \\ &= -\ln \det(\Omega) + \text{tr}(\Omega \hat{P}) + \sum_k \left[\ln \det(S_{kk}) + \text{tr} \left\{ (S_{kk}^{-1} - w_k w_k^H) \hat{S}_{kk} \right\} \right], \end{aligned}$$

where

$$\hat{P} = \widehat{\text{var}}[w_k^H X_k : k \in [K]], \quad \hat{S}_{kl} = \widehat{\text{cov}}[X_k, X_l], \quad w_k = S_{kk}^{-1/2} u_k,$$

and $\widehat{\text{var}}$ and $\widehat{\text{cov}}$ indicate the sample variance and covariance operators, for $k, \ell \in [K]$. The maximum likelihood estimator minimizes $\text{nll}(\theta; \{X_1^n, \dots, X_K^n\}_{n=1, \dots, N})$ with respect to $S_{kk}^{-1} : w_k^H S_{kk} w_k = 1$. That is,

$$\nabla_{S_{kk}^{-1}} \text{nll} = S_{kk} - \hat{S}_{kk} = S_{kk} w_k \lambda_k w_k^H S_{kk},$$

for some $\lambda_k \in \mathbb{R}$ and for all $k \in [K]$. Because $w_k^H S_{kk} w_k = 1$,

$$1 - w_k^H \hat{S}_{kk} w_k = w_k^H S_{kk} w_k - w_k^H \hat{S}_{kk} w_k = w_k^H S_{kk} w_k \lambda_k w_k^H S_{kk} w_k = \lambda_k.$$

Therefore, the two terms $\ln \det(S_{kk})$ and $\text{tr}((S_{kk}^{-1} - w_k w_k^H) \hat{S}_{kk})$ may be rewritten as

$$\ln \det(S_{kk}) = -\ln(1 - \lambda_k) + \ln \det(\hat{S}_{kk})$$

and

$$\text{tr}((S_{kk}^{-1} - w_k w_k^H) \hat{S}_{kk}) = \text{tr}((S_{kk}^{-1} - w_k w_k^H)(S_{kk} - S_{kk} w_k \lambda_k w_k^H S_{kk})) = d_k - 1.$$

The maximum likelihood estimation problem then reduces to minimizing

$$\text{nll}(\Omega, w_k, \lambda_k; \{X_1^n, \dots, X_K^n\}_{n=1, \dots, N}) = -\ln \det(\Omega) - \sum_k \ln(1 - \lambda_k) + \text{tr}(\Omega \hat{P}).$$

with the restriction that $\text{diag}(\Omega^{-1}) = 1$. Let $w'_k = w_k / \sqrt{1 - \lambda_k}$,

$\Omega' = \text{diag}(\sqrt{1 - \lambda_k}) \Omega \text{diag}(\sqrt{1 - \lambda_k})$, and $\hat{P}' = \widehat{\text{var}}[w_k^H X_k; k \in [K]]$. The likelihood can be rewritten as

$$\text{nll}(\Omega', w'_k; \{X_1^n, \dots, X_K^n\}_{n=1, \dots, N}) = -\ln \det(\Omega') + \text{tr}(\Omega' \hat{P}'),$$

which is maximized when $\Omega' = \hat{P}'^{-1}$, given that w'_k is fixed for $k \in [K]$. Thus, maximum likelihood estimation is equivalent to finding w'_k minimizing $\ln \det(\hat{P}')$ under $w_k^H \hat{S}_{kk} w_k = 1$ for $k \in [K]$, which is the GENVAR procedure of Kettenring (1971).

REFERENCES

- Adali T., Schreier PJ., & Scharf LL. (2011). Complex-valued signal processing: The proper way to deal with impropriety. *IEEE Transactions on Signal Processing*, 59(11), 5101–5125.
- Andersen HH, Hojbjerg M, Sorensen D, & Eriksen PS (1995). *Linear and Graphical Models: For the Multivariate Complex Normal Distribution*, Springer, New York.
- Aydore S, Pantazis D, & Leahy RM (2013). A note on the phase locking value and its properties. *Neuroimage*, 74, 231–244. [PubMed: 23435210]
- Baba K, Shibata R, & Sibuya M (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics*, 46(4), 657–664.
- Bach FR & Jordan MI (2005). *A Probabilistic Interpretation of Canonical Correlation Analysis*, Technical Report 688, University of California, Berkeley.
- Balakrishnan S, Wainwright MJ, & Yu B (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *The Annals of Statistics*, 45(1), 77–120.

- Bong H, Liu Z, Ren Z, Smith M, Ventura V, & Kass RE (2020). Latent dynamic factor analysis of high-dimensional neural recordings. In *Advances in Neural Information Processing Systems*, Vol. 33, Neural Information Processing Systems Foundation Inc., San Diego, 16446–16456. [PubMed: 36605231]
- Bong H, Ventura V, Yttri EA, Smith MA & Kass RE (2023). Cross-population amplitude coupling in high-dimensional oscillatory neural time series. arXiv preprint, arXiv:2105.03508.
- Brémaud P (2014). *Fourier Analysis and Stochastic Processes*, Springer, Cham, Switzerland.
- Buzsáki G, Anastassiou CA, & Koch C (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6), 407–420. [PubMed: 22595786]
- Buzsáki G & Draguhn A (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926–1929. [PubMed: 15218136]
- Cardin JA, Carlén M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai L-H, & Moore CI (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature*, 459(7247), 663–667. [PubMed: 19396156]
- Cohen MX (2014). *Analyzing Neural Time Series Data: Theory and Practice*, MIT Press, Cambridge, MA.
- Dahlhaus R (2000). Graphical interaction models for multivariate time series. *Metrika*, 51(2), 157–172.
- Dempster AP, Laird NM, & Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39, 1–22.
- Einevoll GT, Kayser C, Logothetis NK, & Panzeri S (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11), 770–785. [PubMed: 24135696]
- Fries P (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10), 474–480. [PubMed: 16150631]
- Hotelling H (1936). Relations between two sets of variates. *Biometrika*, 28(3–4), 321–377.
- Kass RE, Amari S-I, Arai K, Brown EN, Diekmann CO, Diesmann M, Doiron B et al. (2018). Computational neuroscience: Mathematical and statistical perspectives. *Annual Review of Statistics and its Application*, 5, 183–214.
- Kettenring JR (1971). Canonical analysis of several sets of variables. *Biometrika*, 58(3), 433–451.
- Klein N, Orellana J, Brincat SL, Miller EK, & Kass RE (2020). Torus graphs for multivariate phase coupling analysis. *The Annals of Applied Statistics*, 14(2), 635–660. [PubMed: 36605359]
- Lachaux J-P, Rodriguez E, Martinerie J, & Varela FJ (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4), 194–208. [PubMed: 10619414]
- Lepage KQ & Vijayan S (2017). The relationship between coherence and the phase-locking value. *Journal of Theoretical Biology*, 435, 106–109. [PubMed: 28918333]
- Lowet E, Roberts MJ, Bonizzi P, Karel J, & De Weerd P (2016). Quantifying neural oscillatory synchronization: A comparison between spectral coherence and phase-locking value approaches. *PLoS One*, 11(1), e0146443.
- Mathalon DH & Sohal VS (2015). Neural oscillations and synchrony in brain dysfunction and neuropsychiatric disorders: It's about time. *Journal of the American Medical Association Psychiatry*, 72(8), 840–844. [PubMed: 26039190]
- Miller EK, Lundqvist M, & Bastos AM (2018). Working memory 2.0. *Neuron*, 100(2), 463–475. [PubMed: 30359609]
- Navarro A., Frellsen J., & Turner R. (2017). The multivariate generalised von Mises distribution: Inference and applications. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31, Association for the Advancement of Artificial Intelligence, Washington, DC.
- Nolte G, Galindo-Leon E, Li Z, Liu X, & Engel AK (2020). Mathematical relations between measures of brain connectivity estimated from electrophysiological recordings for Gaussian distributed data. *Frontiers in Neuroscience*, 14, 577574. <https://www.frontiersin.org/articles/10.3389/fnins.2020.577574/>
- Ombao H & Pinto M (2022). Spectral dependence. *Econometrics and Statistics*. 10.1016/j.ecosta.2022.10.005

- Ombao H & VanBellegem S. (2008). Evolutionary coherence of nonstationary signals. *IEEE Transactions on Signal Processing*, 56(6), 2259–2266.
- Orellana J & Kass RE (2023). Latent Torus Graphs for Dense Recordings and Cross-Region Phase Coupling Analysis (unpublished manuscript).
- Pandey JN (2011). *The Hilbert Transform of Schwartz Distributions and Applications*, Wiley, New York.
- Pesaran B, Vinck M, Einevoll GT, Sirota A, Fries P, Siegel M, Truccolo W, Schroeder CE, & Srinivasan R (2018). Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation. *Nature Neuroscience*, 21(7), 903–919. [PubMed: 29942039]
- Picinbono B (1994). On circularity. *IEEE Transactions on Signal Processing*, 42(12), 3473–3482.
- Schmidt R, Ruiz MH, Kilavik BE, Lundqvist M, Starr PA, & Aron AR (2019). Beta oscillations in working memory, executive control of movement and thought, and sensorimotor function. *Journal of Neuroscience*, 39(42), 8231–8238. [PubMed: 31619492]
- Shumway RH & Stoffer DS (2017). *Time Series Analysis and its Applications: With R Examples*, Springer, Cham, Switzerland.
- Siegle JH, Jia X, Durand S, Gale S, Bennett C, Graddis N, Heller G et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852), 86–92. [PubMed: 33473216]
- Srinath R & Ray S (2014). Effect of amplitude correlations on coherence in the local field potential. *Journal of Neurophysiology*, 112(4), 741–751. [PubMed: 24790174]
- Tugnait JK (2019a). Edge exclusion tests for graphical model selection: Complex Gaussian vectors and time series. *IEEE Transactions on Signal Processing*, 67(19), 5062–5077.
- Tugnait JK (2019b). Edge exclusion tests for improper complex Gaussian graphical model selection. *IEEE Transactions on Signal Processing*, 67(13), 3547–3560.

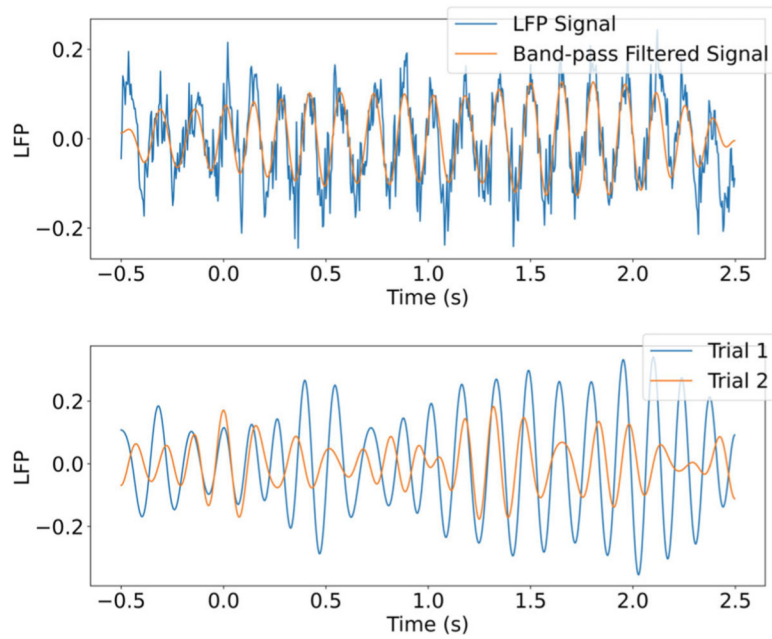


Figure 1:

Three seconds of local field potential (LFP) data filtered at 6.5Hz with a window from 6 to 7Hz. In the top figure, we plot both the raw LFP signal, which consists of a noisy oscillation around 6.5Hz, and the filtered signal, which removes the noise. In the bottom figure, we plot filtered data from the same electrode across two trials. At many points in time, the two trials have different phases or amplitudes.

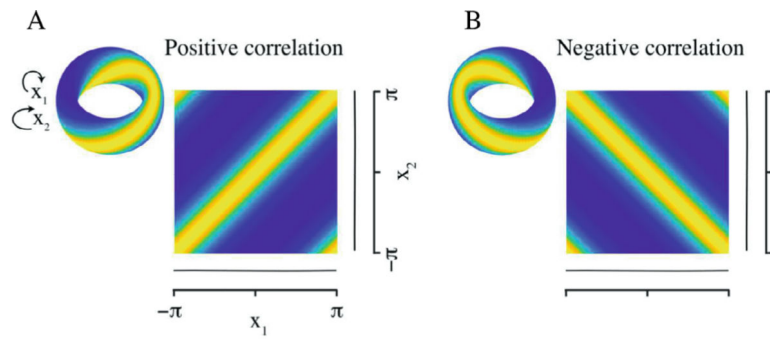


Figure 2:

The torus is the natural domain for a pair of circular random variables. Illustration of both types of circular covariance in a torus graph bivariate density, with uniform marginal densities, plotted side by side on a two-dimensional torus and a Cartesian plane. (a) Positive rotational dependence. (b) Negative reflectional dependence. The figure has been adapted from a figure in Klein et al. (2020).

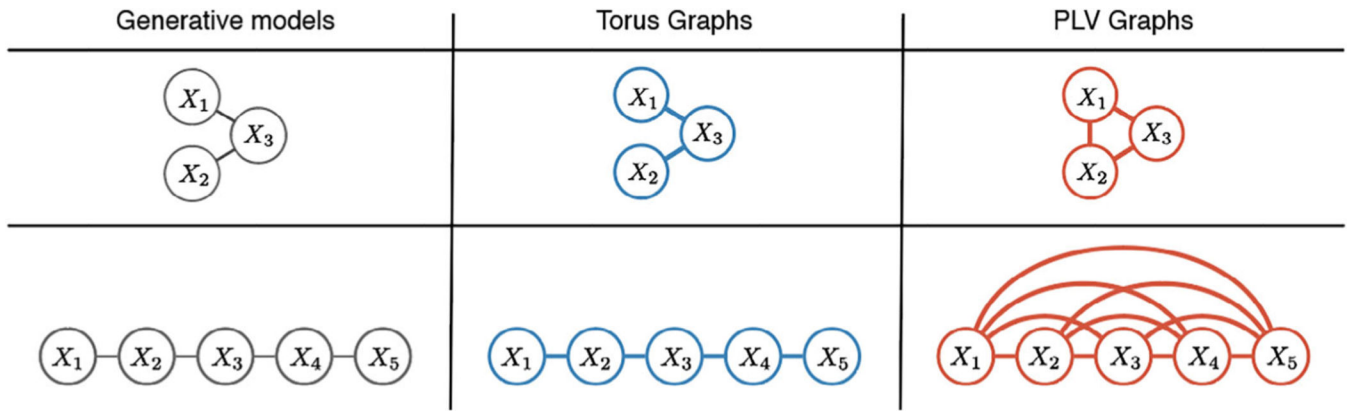


Figure 3: Torus graphs can recover conditional dependence graphs when PLV fails in simulated data; see Klein et al. (2020) for details. The figure has been adapted from one in Klein et al. (2020).

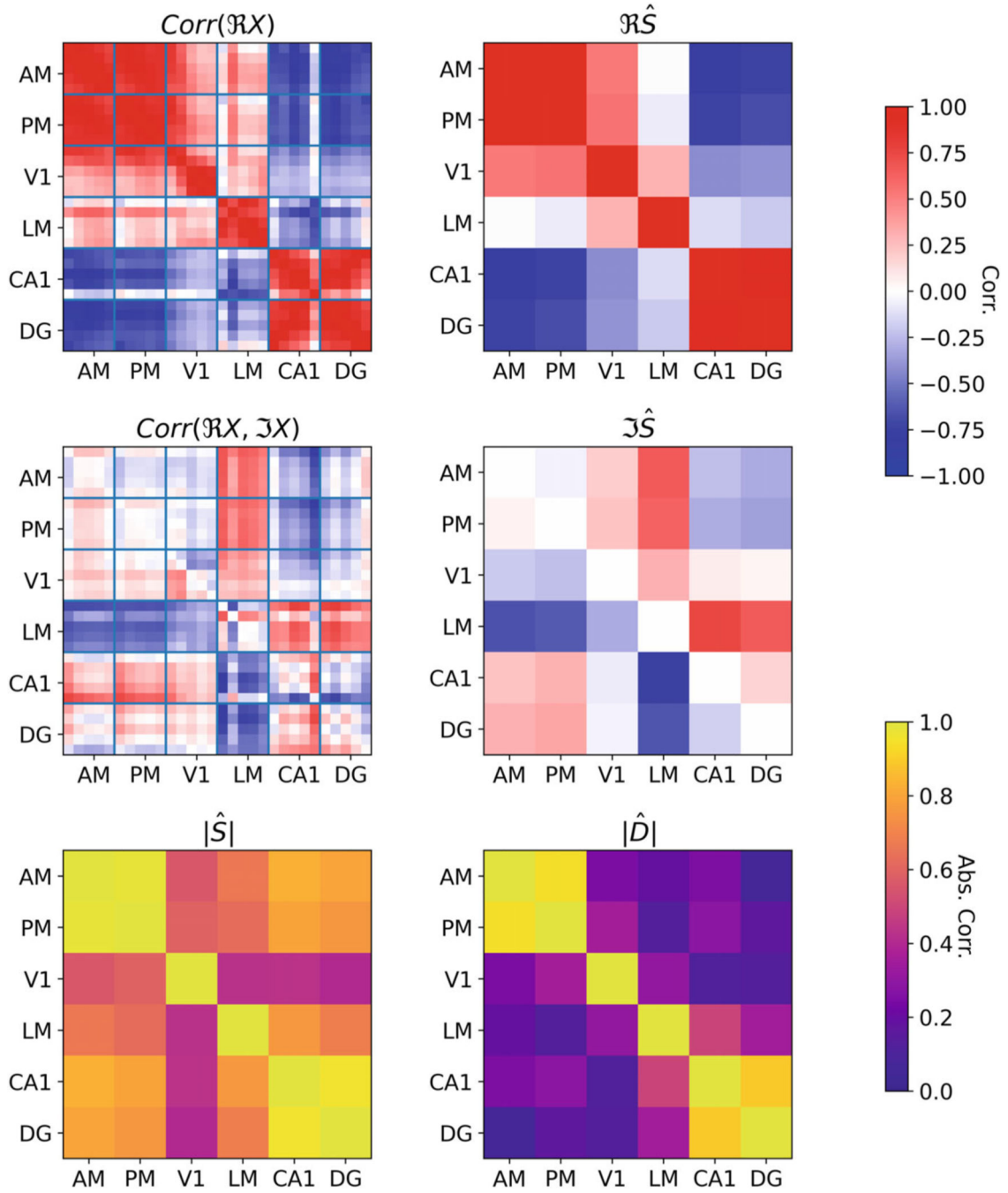


Figure 4: An example of an application of the latent variable model to LFP data from the Allen Institute. The entries in the correlation matrices are arranged consecutively according to their vertical position in the inserted probe. The sample correlation for the real part of the data and the sample correlation between the real and imaginary parts of the data are denoted by $\text{corr}(\Re X)$ and $\text{corr}(\Re X, \Im X)$. The estimated latent correlation matrix is denoted by \hat{S} , and the estimated latent partial correlation matrix is denoted by \hat{D} . The notations \hat{S} and \hat{D} denote the absolute value of each entry in the corresponding matrices. Electrodes are labelled by the

anatomical region in which they resided during the experiment. We observe that while the estimated latent correlation matrix \hat{S} has large values between many regions, the estimated latent partial correlation primarily has large values between two pairs of regions: between DG and CA1 as well as between AM and PM.

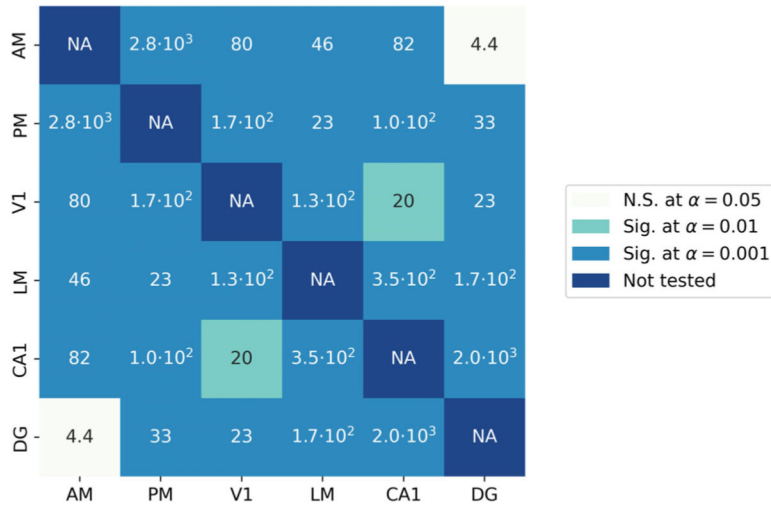


Figure 5: Results of parametric bootstrap test performed on all pairs of regions to test if a significant conditional correlation exists between each pair of areas. Numbers inside the boxes are likelihood ratio statistics testing the hypothesis that $|D_{ij}| > 0$. The colour of each box represents the results of significance tests for three values of α , the false-positive rate, where each significance test is done with a Bonferroni correction for multiple comparisons. The level $\alpha = 0.001$ is the smallest we are able to test given our simulation settings. The likelihood ratio values indicate that some pairs would be significant at much smaller values of α . There are significant partial correlations between most pairs of areas at $\alpha = 0.001$, except between DG and AM as well as between V1 and CA1.

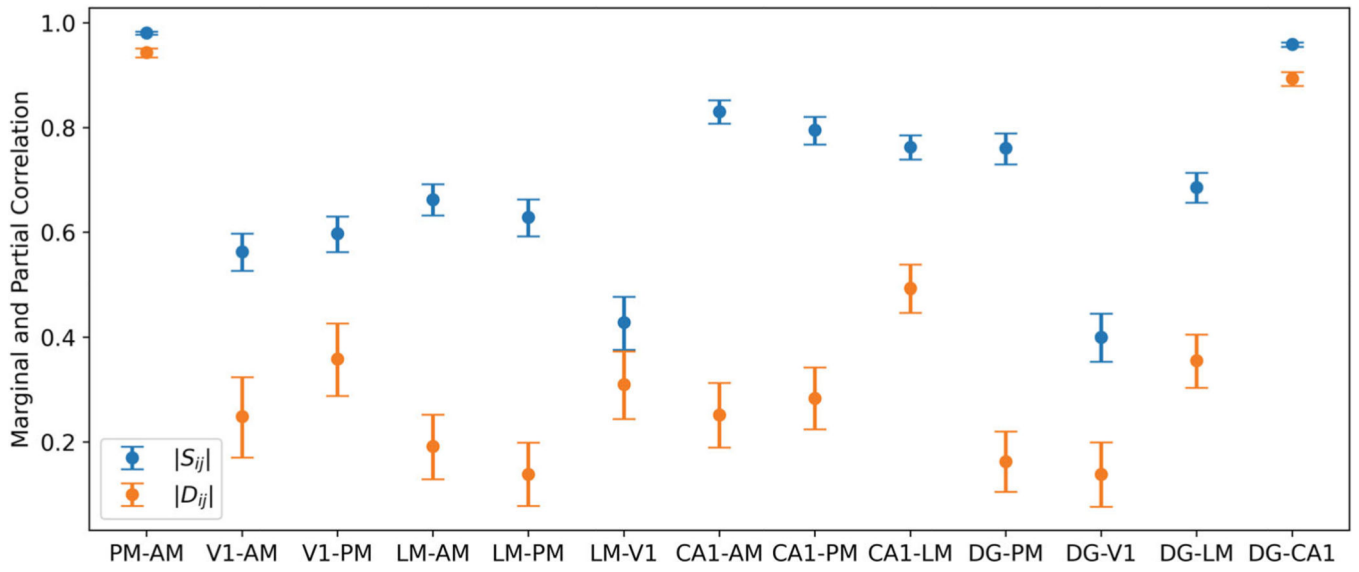


Figure 6: Point estimates and confidence intervals of correlations (blue) and partial correlations (orange) for all pairs of regions with significantly nonzero latent partial correlations. All confidence intervals have 95% coverage. The partial correlations are smaller than the corresponding marginal correlations, with the exception of the large partial correlations between PM and AM and between DG and CA1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript