Discovery of Functional Predictivity across Brain Regions from Local Field Potentials

Heejong Bong

July 2022

Department of Statistics and Data Sciences
Dietrich College of Humanities and Social Sciences
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Robert E. Kass, Co-chair
Valérie Ventura, Co-chair
Aaditya Ramdas
Liam Paninski (Columbia University, New York City)

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

Copyright © 2022 Heejong Bong

Keywords: Neuroscience, Functional Connectivity, Local Field Potential, High-dimensional time-series, Latent Factor Model, Multi-set Canonical Correlation Analysis, Matrix-variate Gaussian Graphical Model, Complex Gaussian Model

To my family,
for their endless love, making me feel home 6937 miles apart

Abstract

Neural recordings from high-density microelectrode arrays yield high-dimensional time-series observations, simultaneously recorded from multiple brain regions, with good spatial and temporal resolution. A careful study of such time-series covariance structure can uncover functional associations between the regions, and in particular, lead-lag relationships could indicate possible directional flows of neural information. Because such relationships in an alert animal's brain are observed transiently, we use repeated trials to estimate the non-stationary covariance structure. In this thesis work, we develop methods for estimating non-stationary lead-lag relationships among high-dimensional time series.

The first approach is to treat high-dimensional time-series recordings as matrix-variate observations and model them by matrix-variate Gaussian graphical models. In these models, two covariance matrices describe spatial association between the LFP channels and autocorrelation across time, respectively. We discover cross-regional connectivity using statistical inference on the spatial covariance graph estimate. We proved the theoretical validity of the proposed bootstrap test, based on high-dimensional central limit theorems. We showed that the proposed method increases statistical power by incorporating the shared covariance structures in multiple recording sessions. We also demonstrate the efficacy of the new method through both simulated datasets and multi-session LFP recordings from the same experiment as in LaDyS.

Another approach is based on extensions of probabilistic canonical correlation analysis (pCCA) to the time-series setting. Starting with recordings from only two brain regions, the model assumes that all of the time series within a given region are driven by the same latent univariate time series; the resulting latent bivariate time series then defines the time-varying cross-region dependence we seek to identify. By leaving the correlation matrix unspecified instead of assuming a parametric structure for the cross-dependence, the model provides a model-based interpretation of multiset CCA. These generalizations come at a cost: we now have a high-dimensional time series problem within each brain region, involving a high-dimensional covariance structure. We solved these high-dimensional problems by imposing sparsity of the dominant effects within a range of possible interesting lead-lag effects and developed Latent Dynamic Analysis via Sparse Banded Graphs (LaDynS).

We also developed and studied inferential procedures to decide time epochs of significant cross-region association from the LaDynS estimate. In particular, an autoregressive model, based on the LaDynS estimate, enables statistical inferences of cross-regional association in terms of Granger causality. LaDynS performed well in simulations designed to mimic real-data situations. We also applied it to 192 simultaneous local field potential (LFP) recordings from the prefrontal cortex (PFC) and visual cortex (area V4) during a visual memory task. We found lead-lag relationships that are highly plausible and consistent with related results in the literature.

Furthermore, we developed an improved method, Latent Dynamic Factor Analysis of High-dimensional Time-series (LDFA-H), by incorporating the two previous methods. LDFA-H uses the LaDynS latent factor models to describe cross-regional connectivity, while imposing a spatio-temporal matrix-variate graphical model on auto-correlations within brain areas. This hierarchical structure allows LDFA-H to provide better estimates than LaDynS and other known methods, even in the presence of high noise.

Last, we revisit covariance structure in LFP recordings in context of the oscillatory nature of the data mode. We considered phase and amplitude together using the complex normal distribution, for which the distinction between covariance and pseudo-covariance is important. This provides a new characterization of standard oscillatory correlation measures, based on conditional distributions of phase given amplitude. We defined a complex Gaussian latent variable model for evaluating the strength of associations across multiple brain areas and applied it to data involving multivariate LFP time series that exhibit pronounced oscillations.

Acknowledgments

It would be impossible for my dissertation to take its current form without the student-friendly environment shaped by all the staff and faculty members in the Department of Statistics and Data Science at Carnegie Mellon University. They have always provided endless support to doctoral students. In particular, I want to express my sincere gratitude to my advisors, professors Robert E. Kass and Valerie Ventura, for believing in my potential to be a professional researcher. They have encouraged me to cultivate critical rigor while providing enthusiasm in the face of difficult situations. The support and instruction from both the department and my advisors have provided ground for training my work ethic. It was the main thrust of my academic journey and allowed me to overcome the many impediments to completing the dissertation.

I am also intellectually indebted to professors Larry Wasserman, Alessandro Rinaldo, Sivaraman Balakrishnan, Aaditya Ramdas, Zhao Ren, and Arun Kuchibhotla when I was finding my way after the doctoral program. Their advice inspired me to participate in a flurry of projects, some of which developed into parts of the dissertation and also several upcoming works. My special thanks go to the fellow students with whom I have collaborated during those projects. The list includes but is not limited to Wanshan Li, Shamindra Shrotrya, Zongge Liu, Josue Orellana, Konrad Urban, and the members of the NeuroStat research group. I was pleased to share fruitful discussions, encouragement, and even simple chats at coffee shops. These friendly interactions comprised some of the best moments during my academic journey.

I cannot forget to mention the continuous support from my family and friends. The Ph.D. class of 2022 in Statistics and Data Science at CMU was the best cohort I can imagine. I love our group's diversity and warm-hearted atmosphere, through which we established wonderful synergy without unnecessary fierce competition. I would also like to thank all the other fellow students in the department. In particular, the following previous and current Korean graduate students have provided me with much-needed emotional support: Jisu Kim, Justin (Sangwon) Hyun, Ilmun Kim, Jaehyuk Shin, Kwangho Kim, Beomjo Park, YJ Choe, and Woonyoung Chang. I also received enormous assistance from my Korean friends both within the Pittsburgh area and abroad. My gratitude goes to Khidong Kim, Kyungwon Lee, Minkyu Kim, Taeyong Kwon, Junewoo Park, Sungwon Heo, the Club '93 (Korean CMU students born in 1993), and the Korean Graduate Student Association.

Finally, I want to appreciate my friends formed through an online community in which we are fighting against discrimination on the basis of gender identity, sexual orientation, race, ethnicity, nation, and disability. Despite our different backgrounds, we gathered together because we shared the experience of being excluded from mainstream society. This invaluable experience allowed me to be myself and become involved in the broader world outside the university.

Contents

1 Introduction 1

2 Simultaneous Inference in Multiple Matrix-Variate Graphs for High-Dimensional Neural Recordings 5

2.1 Introduction 5

2.1.1 Other Related Work 7

2.1.2 Organization of the Paper and Notations 8

2.2 Simultaneous Inference Framework 8

2.2.1 Estimation of Spatial Covariance Matrix 9

2.2.2 Simultaneous Test by Parametric Bootstrap 10

2.2.3 Estimation of Temporal Covariance Matrix 13

2.3 Theoretical Properties 14

2.3.1 Non-asymptotic error bound for the group Lasso estimate 16

2.3.2 Theoretical justification for the simultaneous edge testing 17

2.3.3 Non-asymptotic error bound for the temporal covariance matrix estimate 18

2.4 Numerical Studies 19

2.4.1 Simulation Studies 19

2.4.2 Experimental Data Analysis 22

2.5 Conclusion 26

3 Cross-population Analysis of High-dimensional Non-stationary Neural Oscillations 29

3.1 Introduction 29

3.2 Methods 31

3.2.1 Probabilistic CCA for two random vectors 31

3.2.2 Probabilistic CCA for two time series of random vectors 32

3.2.3 Latent Dynamic Analysis via Sparse Banded Graphs (LaDynS) 35

3.2.4 Inference for associations between two vector time series 37

3.2.5 Locally Stationary State-space Model and Local Granger Causality . . . 39

3.3 LaDynS performance on simulated data 40

3.3.1 Simulated datasets with known canonical correlation matrix 40

3.3.2 LaDynS estimation details 41

3.3.3 Results 41

3.4 Experimental Data Analysis 44

3.5 Discussion 46

4 Latent Dynamic Factor Analysis of High-Dimensional Neural Recordings 49
4.1 Introduction 49
4.2 Latent Dynamic Factor Analysis of High-dimensional time series 50
4.3 Results 54
4.3.1 LDFA-H retrieves cross-correlations even when noise auto-correlations dominate 54
4.3.2 Experimental Data Analysis from Memory-Guided Saccade Task 55
4.4 Conclusion 56

5 Oscillating neural circuits: Phase, amplitude, and the complex normal distribution 61
5.1 Introduction 61
5.2 Torus Graphs 64
5.3 Latent Dynamical Model 68
5.4 The Complex Normal Distribution 69
5.5 Coherence and PLV 70
5.5.1 Coherence and Complex Correlation 70
5.5.2 PLV as an Analogue of Pearson Correlation 71
5.5.3 Comparing Measures of Covariation 72
5.6 A Latent Variable Model for the Proper Complex Normal Distribution 73
5.6.1 Model 73
5.6.2 Synthetic Data Example 74
5.7 Data Analysis 74
5.7.1 Analysis of Propriety 76
5.7.2 Application of the Latent Variable Model 77

A Thesis flow chart with model details. 79

B Appendix for Chapter 2 81
B.1 Additional Figures 81
B.2 Proofs 82
B.2.1 Preliminary Lemmas 82
B.2.2 Proof of Theorem 2.3.1 83
B.2.3 Proof of Proposition 2.3.2 87
B.2.4 Proof of Proposition 2.3.4 89
B.2.5 Proof of Theorem 2.3.5 91
B.2.6 Proof of Theorem 2.3.6 92
B.2.7 Proof of Proposition 2.3.7 94
B.2.8 Proof of Theorem 2.3.8 97
B.2.9 Proof of Corollary 2.3.9 98
B.3 Proof of the Lemmas 99
B.3.1 Proof of Lemma B.2.1 99
B.3.2 Proof of Lemma B.2.2 99
B.3.3 Proof of Lemma B.2.3 100
B.3.4 Proof of Lemma B.2.4 101

B.3.5 Proof of Lemma B.2.5 . . . 101
B.3.6 Proof of Lemma B.2.6 . . . 102
B.3.7 Proof of Lemma B.2.7 . . . 103
B.4 Derivation of the asymptotic covariance between the edge-wise test statistics . . . 104

C Appendix for Chapter 3 **107**
C.1 Proof of Theorem 3.2.3 . . . 107
C.2 Fitting LaDynS . . . 109
C.3 Inference on the change of the factor loadings . . . 109

D Appendix for Chapter 4 **113**
D.1 EM-algorithm to fit LDFA-H (Section 2) . . . 113
D.2 Simulation details (Section 3) . . . 116
D.3 Experimental data analysis details (Section 3.2) . . . 117

E Appendix for Chapter 5 **119**
E.1 A Note on the Hilbert Transform . . . 119
E.2 Theorems and Proofs . . . 120
E.2.1 Proof of Theorem 5.4.1 . . . 120
E.2.2 The Multivariate Generalized Von Mises Distribution . . . 123
E.2.3 Proof of Theorem 5.4.2 . . . 125
E.2.4 Structure of Proper Matrices . . . 125
E.2.5 Proof of Theorem 5.4.3 . . . 127
E.2.6 Proof of Corollary 5.4.4 . . . 127
E.3 Fitting Procedure for Latent Variable Model . . . 128
E.4 Table of Examples Comparing PLV, Amplitude Correlation, and Complex Correlation . . . 129

Bibliography **131**

List of Figures

1.1 **Flow chart of research projects in this dissertation.** *White boxes code for existing methods by other researchers, orange boxes for completed projects in this thesis. Arrows indicate conceptual and methodological dependence. See Fig. A.1 for a detailed version.* 3

2.1 Simulated spatial graphs. 20

2.2 Simulation results under different graph configurations and temporal dimensions. We fix $m = 5$, $n = 5$, and $q = 30$. Rows change with types of graphs, and columns correspond to different temporal dimensions. Blue curve corresponds to our method (M0) while other colors correspond to baseline methods. Our method is consistently better than baselines while our advantage is very obvious for large p, thanks to our temporal covariance estimation procedure. 21

2.3 (a) Primate cortical areas of the attention network (Sapountzis and Gregoriou, 2017). Pink areas are the approximate locations of attention control areas while blue areas are the approximate locations of visual areas. (b) Utah array with 10x10 recording electrodes with $400\mu m$ interval (McKee, Matthew, 2019). For each region, one Utah array with 96 electrodes is implanted. (c) LFPs were recorded simultaneously from V4 and PFC. The X axis is time in millisecond, while Y is the electrode. Time is aligned at $t = 0$ for each trial when the circular target just appeared. The total length is 750ms covering all experimental stages. 23

2.4 The average test statistic vs physical distance during late delay period in V4. Notice that the test statistic declines as the physical distance increase. This phenomenon is consistently identified over all experimental stages, both in PFC and V4 24

2.5 Connectivity strength distribution over 2D spatial array for PFC and V4 across the experimental stages. The connectivity in both area decays during the late delay stage, while V4 seems to be more influenced than PFC. 25

2.6 Significant cross-region edges for PFC and V4 over various experimental stages. X and Y axes are spatial coordinates of the electrodes on each array. Lower left pane shows electrodes in PFC, while upper right shows electrodes in V4. We only show significant cross-area edges in red color. 26

3.1 **(a)** *Graphical representation of the pCCA model of Bach and Jordan (2005), where X_1 and X_2 are random vectors and Z is a random variable.* **(b)** *A variation on (a) that facilitates extension to the case in which X_1 and X_2 are multivariate time series and (Z_1, Z_2) is a bivariate time series.* 31

3.2 **Extended pCCA model for two multivariate time series $X_1^{(t)}$ and $X_2^{(s)}$, $t, s = 1, \dots, T$.** **(a)** *Dynamic associations between vectors $X_1^{(t)}$ and $X_2^{(s)}$ are summarized by the dynamic associations between their associated 1D latent variables $Z_1^{(t)}$ and $Z_2^{(s)}$, and estimated by their cross-precision matrix Ω_{12}.* **(b)** *When a significant cross-precision entry is identified, e.g., the red star in the expanded view of Ω_{12}, its coordinates and distance from the diagonal indicate at what time in the experiment connectivity between two brain areas occurs, and at what lead or lag. Here the red star is in the upper diagonal of Ω_{12}, which means that, at this particular time, region 1 leads region 2, or $Z_1 \to Z_2$ in short (a non-zero entry in the lower diagonal would mean $Z_2 \to Z_1$). We represent this association by the red arrow on the right-most plot, with a lag of two units of time for illustration.* . . 34

3.3 *The elements of Ω_{kk}, $k = 1, 2$, and Ω_{12} are set to zero outside of the gray bands of widths $(1 + 2d_{\text{auto}})$ and $(1 + 2d_{\text{cross}})$, respectively.* 36

3.4 **Output and inference of LaDynS applied to a dataset simulated as described in Section 3.3.1**. **(a)** *Cross-precision estimate $\widehat{\Omega}_{12}$ of the true Ω_{12} shown in* **(d)**. **(b)** *Bootstrap p-values for the de-sparsified estimate $\widetilde{\Omega}_{12}$.* **(c)** *Discovered non-zero cross-precision estimates by the BH procedure at nominal FDR 5% followed by the excursion test at significance level 5%.* 42

3.5 **Null distributions of three representative entries of $\widetilde{\Omega}_{12}^{(t,s)}/\sqrt{\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]}$.** *The null distributions are obtained from $R = 60$ simulated datasets (Section 3.3.1) and compared to the standard Gaussian distribution via QQ-plots. There is good agreement.* 43

3.6 **Standard deviations of desparsified precision elements.** *Variance obtained* **(a)** *from samples from the ground-truth generative multiset pCCA model and* **(b)** *from permutation bootstrapped samples.* **(c)** *F-statistics of ratios between the two variances for null entries of Ω_{12}, showing good agreement.* 43

3.7 **False Discovery Rate control of LaDynS' inference.** **(a)** *Estimated false discovery rate and* **(b)** *false non-discovery rate for target FDR $\in [0, 10]\%$, under the connectivity scenario in Fig. 3.4d, for connectivity intensities $r = 0.2$, 0.4, 0.6, 0.8 and 1.0 in Eq. (3.27). The dotted line is a (0,1) line.* 43

3.8 **LaDynS output and inference for experimental dataset.** (**a**) *LaDynS cross-precision estimate* $\widehat{\Omega}_{12}$*. The light gray area shows the region of time considered (one area leading the other by at most 0.1 second).* (**b**) *Bootstrap p-values for the de-sparsified estimate* $\widetilde{\Omega}_{12}$ *(Eq.* (3.18) *with bootstrap simulation size* $B = 200$*; see Section 3.2.4).* (**c**) *Discovered regions of cross-precision using STAR at nominal FDR 5%, suggesting that activity in PFC preceded that in V4 immediately post-stimulus and around* 0.1 *seconds thereafter (blue blobs), and that activity in V4 preceded that in PFC around time points* 0.2 *and* 0.4 *seconds post-stimulus (red blobs). These four epochs had excursion test* p*-values of* $p = .024$, $p = .0015$, $p = .001$ *and* $p < .0005$*, respectively. The two smaller blobs were not significant* ($p > 0.5$) *and are therefore likely spurious.* 45

3.9 **Estimated partial R^2 from locally stationary state-space model.** *For* $Z_1 \rightarrow Z_2$*, partial* R^2 *was computed corresponding to linear regression of* $Z_1^{(t)}$ *against* $Z_2^{(t-0.015\text{sec.}:t-0.03\text{sec.})}$ *conditional to the other covariates, including* Z_2 *at the other lags, for every time point* t *at sampling rate 200 Hz. The black solid lines indicate the estimated partial* R^2 *for* (**a**) *V4* $\rightarrow$ *PFC and* (**b**) *PFC* $\rightarrow$ *V4. The pink shaded area shows 95 percentile area of null partial* R^2 *under independence between V4 and PFC, obtained from the permutation test.* 45

3.10 **Estimated variance of electrophysiological activity in V4 and PFC driven by latent factors.** *The* ℓ_2 *norm of the factor loading vector* $\beta_k^{(t)}$ *(representing the total variance of the data* $X_k^{(t)}$ *attributable to the latent time series at time* t*, see Eq.* (3.7)*) is plotted across time* t *for V4 (blue) and PFC (orange).* 46

3.11 *Factor loadings of V4 and PFC, spatially smoothed, normalized, and color coded over the electrode coordinates (*μm*) at* (**left**) *the first bump of Fig. 3.9b (0.15 sec) and* (**right**) *the second bump of Fig. 3.9b or the first bump of Fig. 3.9a (0.40 sec). Contours at fractions .25, .5, and .75 of maximal power have been added.* 47

4.1 **LDFA-H model.** (**a**) *Dynamic associations between vectors* $X^1_{:,t}$ *and* $X^2_{:,s}$ *are summarized by the dynamic associations between their associated 1D latent variables* $Z^1_{:,t}$ *and* $Z^2_{:,s}$*.* (**b**) *When a significant cross-precision entry is identified, e.g., the red star in the expanded view of* Π_f^{12}*, its coordinates and distance from the diagonal indicate at what time in the experiment connectivity between two brain areas occurs, and at what lead or lag. Here the red star is in the upper diagonal of* Π_f^{12}*, which means that, at this particular time, region 1 leads region 2, or* $Z_f^1 \rightarrow Z_f^2$ *in short (a non-zero entry in the lower diagonal would mean* $Z_f^2 \rightarrow Z_f^1$*). We represent this association by the red arrow on the right-most plot, with a lag of two units of time for illustration.* 51

4.2 **Simulation settings.** **(a)** *(Left to right panels) True correlation matrix* P_1 *for latent factors* $Z^1_{1,:}$ *and* $Z^2_{1,:}$ *from model in Eq.* (4.2)*; close-up of the cross-correlation matrix; corresponding precision matrix* $\Pi_1 = \mathrm{P}_1^{-1}$*; and close-up of cross-precision matrix* Π_1^{12} *(Eq.* (4.3)*). Matrix axes represent the duration,* $T = 50$ *ms, of the time series. Factors* Z^1 *and* Z^2 *are associated in two epochs:* Z^2 *precedes* Z^1 *by* 7*ms from* $t = 13$ *to* 19*ms, and* Z^1 *precedes* Z^2 *by* 7*ms from* $t = 33$ *to* 42*ms.* **(b)** *Noise auto-correlation matrices (Eq.* (4.5)*) for pairs of simulated time series at four strength levels.* $\log\det$ *in (a) and (b) measure correlation strengths.* 55

4.3 **Simulation results: LDFA-H cross-precision matrix estimates.** *Estimates of* Π_1^{12}*, shown in the right-most panel of Fig. 4.2a, using LDFA-H, for the four noise auto-correlation strengths shown in Fig. 4.2b. LDFA-H identified the true cross-area connections at all noise strengths.* 56

4.4 **Simulation results: cross-correlation matrix estimates.** *Estimates of* Σ_1^{12} *under four noise correlation levels using* **(a)** *averaged pairwise correlation (APC),* **(b)** *correlation of averaged signal (CAS),* **(c)** *canonical correlation analysis (CCA, Hotelling (1936)),* **(d)** *dynamic kernel CCA (DKCCA, Rodu et al. (2018)),* **(e)** *LaDynS (Bong et al. (2020c)), and* **(f)** *LDFA-H. Only LDFA-H successfully recovered the true cross-correlation at all noise auto-correlation strengths.* 58

4.5 **Experimental data results for the top 3 factors.** *(a) Cross-precision matrices. Blue represents positive precision matrix entries, corresponding to negative association. Factors have different connectivity patterns over the experimental trials.* $\log_{10}\|\Sigma_f\|_F^2$*, written atop the panels, measures the strength of each factor. The first factor is more than 6 times larger than the second and third, and displays activity in V4 leading PFC centered around 200 milliseconds and activity in PFC leading V4 centered around 320 milliseconds post cue disappearance. This is also shown in panel (d). (b,c) Factor loadings, smoothed and color coded, plotted on the electrode coordinates* (μm)*. Here, positivity is arbitrary, due to identifiability. Panels (b) and (c) display loadings for the V4 and PFC arrays, respectively. The first factor has activity in V4 centered in two distinct subregions of the array, while activity in PFC is more broadly distributed. (d) Dynamic information flow in the directions* $V4 \to PFC$ *(blue) and* $PFC \to V4$ *(orange).* 59

5.1 Three seconds of data from one electrode (a local field potential or LFP in blue) and a band-pass filtered version of the data (orange). 62

5.2 Scatter plot showing rotational (positive) dependence. The magenta-diamond marker is at (0,0), black and blue markers are positioned along the diagonal band of points which looks similar to bivariate normal data, except this pattern is not elliptical and the margins are pretty uniform. Also the points near the red marker, on the upper-left corner, look surprising. They appear because, in Cartesian (x, y) coordinates, the angles must jump from π to $-\pi$ as they wrap around the circular domain. Figure has been adapted from a figure in Klein et al. (2020). 65

5.3 The torus is the natural domain for a pair of circular random variables. Illustration of both type of circular covariance in a torus graph bivariate density, with uniform marginal densities, plotted side-by-side on the 2d-torus and cartesian plane. A. Positive rotational dependence. B. Negative reflectional dependence. Figure has been adapted from a figure in Klein et al. (2020). 65

5.4 Torus Graphs is able to recover conditional dependence graphs but PLV fails. The generative model used simulated data using an independent model; see Klein et al. (2020) for details. Figure has been adapted from a figure in Klein et al. (2020). . 68

5.5 Covariance matrices for synthetic example described in the test. Real and complex parts of the entries in the true Γ matrix used to generate the latent factors are shown in the top row. Diagonal entries are omitted. In the middle row are the empirical covariance matrices of the observed data. Finally, in the bottom row are the real and complex parts of the estimated latent covariance matrix. 75

5.6 Elements of the covariance and pseudo-covariance matrices for LFP data. The covariance matrices (denoted by Γ) were computed, along with the pseudo-covariance matrices (denoted by C). These are plotted above, and appear to show generally larger entries for Γ than for C. 76

5.7 An example of the application of the latent variable model to LFP data from the Allen Institute is shown here. The entries in the covariance matrices are arranged consecutively according to their vertical position in the inserted probe. The sample covariance matrices for the real part of the data and the covariance between the real and imaginary parts of the data are shown on the left. Then, the latent variable model was fit to the data, and the estimated latent covariance matrix is denoted by Γ (diagonal entries are omitted). Electrodes are labeled by the anatomical region in which they resided during the experiment. 77

5.8 This figure was obtained using the same setup as in figure 5.7, except that electrodes were obtained from multiple probes and therefore are only ordered by vertical position within each region. 78

A.1 **Flow chart of research projects in this dissertation.** *White boxes code for existing methods by other researchers, orange boxes for completed projects in this thesis. Arrows indicate conceptual and methodological dependence.* 80

B.1 ROC curve under 3 graph types with $n = 5$, $q = 30$ and $d = 5$ at tuning parameters $\lambda \in (1, 0.01, 1e-4, 1e-6)$. Our method is consistently better regardless of tuning parameters. 81

C.1 **Inference on the the factor loading change from 150ms to 400ms.** *The estimated difference in the factor loadings for* **(a)** *V4 and* **(b)** *PFC from 150ms to 400ms alongside the 97.5 and 2.5 percentiles of the bootstrap distribution. The .05 and -.05 contours have been added to the plots to show regions, blue above .05, red below -.05, where the differences may be distinguishable from zero. The differences are more dramatic in PFC: there is a region in the upper left corner of the array where the 2.5 percentile is well above .05, and thus separated from zero, while the differences in the patch of* $(1000, 1500) \times (500, 1500)$ *have 97.5 percentile well below -.05.* 110

D.1 *Squared Frobenius norms of covariance matrix estimates,* $\widehat{\Sigma}_f$*, for all factors* $f = 1, \ldots, 10$*. Notice that the amplitudes of the top four factors dominate the others.* 118

D.2 **Information flow by partial R^2 for the top three factors.** *In this figure, we characterize dynamic information flow in terms of partial* R^2*. We show dynamic information flow from* $V4 \rightarrow PFC$ *(blue) and* $PFC \rightarrow V4$ *(orange). The results in the first panel are consistent with those in the first panel of Fig. 4.5d.* 118

List of Tables

2.1 Average of empirical coverages and their standard deviations for $m = 3, p = 50, q = 30$. 22

2.2 Average of empirical coverages and their standard deviations for $m = 5, p = 50, q = 30$. 22

E.1 Examples of various ways in which complex covariance and means can capture associations not captured by PLV or amplitude correlation. In all cases, assume that variables are independent unless otherwise stated. We define $X_1 = R_1 \exp(i\Theta_1)$ and $X_2 = R_2 \exp(i\Theta_2)$. 130

Chapter 1

Introduction

Contemporary technologies for recording electrical activity generated by networks of neurons have created many opportunities but also challenges for statistical machine learning. For example, local field potential (LFP) data recorded by high-density microelectrode arrays in recent *in vivo* animal experiments (e.g., Jun et al., 2017; Steinmetz et al., 2018) consist of dozens to hundreds of time series in each of two or more brain regions, recorded simultaneously across many experimental trials. These LFP recordings result from neural currents generated near the electrode (roughly within 150-200 microns), involving large numbers of neurons (Buzsáki et al., 2012; Einevoll et al., 2013; Pesaran et al., 2018), and they have been shown to correlate substantially with the BOLD fMRI signal (Logothetis et al., 2001; Magri et al., 2012a). Under behavioral tasks, the LFP recordings exhibit correlated trial-to-trial variability among the brain regions, indicating their cross-regional interactions relevant to behavior (Buesing et al., 2014; Gallagher et al., 2017; Sarnthein et al., 1998). In the data we have examined, the LFPs are first band-pass filtered, making the time-series smooth.

In some situations, the covariation takes place with lead-lag, and the lead-lag relationship could indicate functional and directional predictivity across the regions (Liebe et al., 2012). If we observe univariate stationary time-series from each brain region, this is the problem solved by frequency-domain analysis and Granger causality (Geweke, 1982). Frequency-domain analysis decomposes the covariance structure into its spectrum and cross-spectrum under a range of frequencies. Then, the complex-valued coherency and its magnitude, coherence, are used to represent the coordinated activity between two brain regions. (See Section 5.5.1 for details.) Vector auto-regressive (VAR) models can be established from the coherency structure among multiple brain regions, and the estimated parameters can be used for Granger causality inference (Ombao and Pinto, 2021). When multi-electrode recordings are made from an alert animal during a behavioral task, however, the data from each brain region are multivariate nonstationary time-series. The statistical challenge is how concise and interpretable summaries can be made to assess the covariation and lead-lag relationship across groups of time-series observations. One approach, in Liu et al. (2021), is to extend the Granger causality to multivariate non-stationary time series. In this thesis, we develop and study other spatio-temporal statistical methods for estimating non-stationary functional associations among high-dimensional time series. The relationship between the methods is described in Fig. 1.1, where the arrows indicate conceptual and methodological dependence.

In Chapter 2, we treat high-dimensional time-series recordings as matrix-variate observations and model them by matrix-variate Gaussian graphical models (Dawid, 1981). In the model, two covariance matrices describe spatial association between the LFP channels and autocorrelation across time, respectively. We discover cross-regional connectivity using statistical inference on the spatial covariance graph estimate. We prove the theoretical validity of a proposed bootstrap test, based on high-dimensional central limit theorems. We show that the proposed method increases statistical power by incorporating the shared covariance structures in multiple recording sessions (on different days). We also demonstrate the efficacy of the new method through both simulated datasets and multi-session LFP recordings from the same experiment as in LaDyS.

Another approach in Chapter 3 is based on extensions of probabilistic canonical correlation analysis (pCCA, Bach and Jordan, 2005) to the time-series setting. Starting with recordings from only two brain regions, the model assumes that all of the time series within a given region are driven by the same latent univariate time series; the resulting latent bivariate time series then defines the time-varying cross-region dependence we seek to identify. By leaving the correlation matrix unspecified instead of assuming a parametric structure for the cross-dependence, the model provides a model-based interpretation of multiset CCA (Kettenring, 1971). These generalizations come at a cost: we now have a high-dimensional time series problem within each brain region, involving a high-dimensional covariance structure. We solved these high-dimensional problems by imposing sparsity of the dominant effects within a range of possible interesting lead-lag effects and developed Latent Dynamic analysis via Sparse banded graphs (LaDynS).

We also developed and studied inferential procedures to decide time epochs of significant cross-region association from the LaDynS estimate. In particular, an autoregressive model, based on the LaDynS estimate, enables statistical inferences of cross-regional association in terms of Granger causality. (See Section 3.2.4.) LaDynS performed well in simulations designed to mimic real-data situations. We also applied it to 192 simultaneous local field potential (LFP) recordings from the prefrontal cortex (PFC) and visual cortex (area V4) during a visual memory task. We found lead-lag relationships that are highly plausible and consistent with related results in the literature.

In Chapter 4, we developed an improved method, Latent Dynamic Factor Analysis of High-dimensional Time-series (LDFA-H), by incorporating the two previous methods. As an extension of Gaussian process factor analysis (GPFA, Yu et al., 2009), LDFA-H uses the LaDynS latent factor models to describe cross-regional connectivity, while imposing a spatio-temporal matrix-variate graphical model on auto-correlations within brain areas. This hierarchical structure allows LDFA-H to provide better estimates than LaDynS and other known methods, even in the presence of high noise.

In Chapter 5, we revisit covariance structure in LFP recordings in the context of the oscillatory nature of the data. We considered phase and amplitude together using the complex normal distribution, for which the distinction between covariance and pseudo-covariance is important. This provides a new characterization of existing oscillatory correlation measures, based on conditional distributions of phase given amplitude. We defined a complex Gaussian latent variable model as an extension of probabilistic CCA model for evaluating the strength of associations across multiple brain areas and applied it to the Allen Institute dataset (Siegle et al. (2021)), which consists of LFP recordings from 6 multi-electrode probe inserted in 6 brain regions.

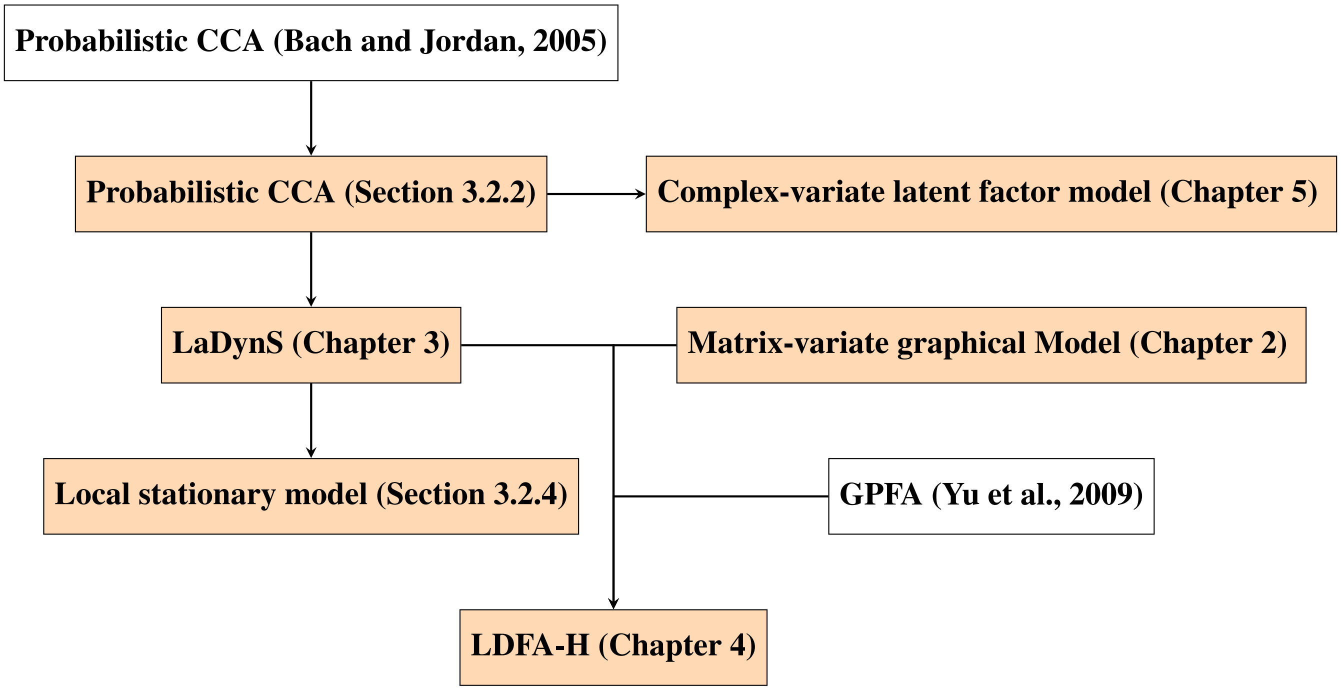

Probabilistic CCA (Bach and Jordan, 2005)
Probabilistic CCA (Section 3.2.2)
Complex-variate latent factor model (Chapter 5)
LaDynS (Chapter 3)
Matrix-variate graphical Model (Chapter 2)
Local stationary model (Section 3.2.4)
GPFA (Yu et al., 2009)
LDFA-H (Chapter 4)

Figure 1.1: **Flow chart of research projects in this dissertation.** *White boxes code for existing methods by other researchers, orange boxes for completed projects in this thesis. Arrows indicate conceptual and methodological dependence. See Fig. A.1 for a detailed version.*

Chapter 2

Simultaneous Inference in Multiple Matrix-Variate Graphs for High-Dimensional Neural Recordings

This chapter is a collaboration with Zongge Liu, Zhao Ren, Matthew A. Smith and Robert E. Kass. This work is currently in progress, and in particular, we will further modify the experimental data results in the final journal submission.

2.1 Introduction

As recent progress in technology enables neural measurements with finer resolution and larger volumes (Siegle et al., 2019), it also poses new challenges for developing tools to analyze the data in high-dimensional multi-variate time-series (Jun et al., 2017; Steinmetz et al., 2018). A motivating dataset of this paper is local field potential recordings (LFPs) from prefrontal cortex (PFC) and visual area V4 during a visual working memory task. Visual cortical area V4 has been reported to retain higher order information (e.g. color and shape) and attention to visual objects (Orban, 2008; Fries et al., 2001), while prefrontal cortex (PFC) is considered to exert cognitive control in working memory (Miller and Cohen, 2001). Despite their spatial distance and functional differentiation, these regions have been presumed to cooperate during visual working memory tasks (Sarnthein et al., 1998; Liebe et al., 2012). Johnston et al. (2020) implanted two Utah arrays (McKee, Matthew, 2019) in the visual (V4) and prefrontal cortices (PFC) of a macaque monkey and recorded LFPs from 96 electrodes dimension in each brain area during a memory-guided saccade task. The data consist of measurements at different time points and electrodes spaced in a two-dimensional grid. Designating rows and columns to correspond to time points and electrodes, respectively, we treat the data as a matrix-variate observation (Ding and Dennis Cook, 2018), which allows us to treat temporal and spatial effects as distinct. In this paper, we develop and study matrix-variate statistical methods for discovering the functional association between the two brain areas from the spatio-temporal LFP recordings, across several sets of recordings made during different sessions (on different days).

If we fix a single time point, the LFP measurement is vector-variate, where each component is the measurement from the corresponding electrode. In this case, we discover the cross-regional association by looking into the coordinated activity between the pairs of electrodes from different brain areas. The vector-variate Gaussian graphical model has been widely applied to high dimensional data in scientific studies to explore the conditional dependence relationships among entries of a random vector, including the neural data (Fornito et al., 2013; Vinci et al., 2018). For a random vector $X \in \mathbb{R}^q$, we can define an undirected graph $G = (V, E)$ associated with X, where the set V contains q nodes each of which corresponds to an entry in X, and the set E consists of edges among V. Specifically, there is no edge between two nodes in V if and only if the corresponding two variables are conditionally independent given the rest of variables in Z. For a Gaussian vector, one can assess the graph structure, i.e., the set E, in terms of the precision matrix, the inverse of covariance matrix of X (Meinshausen and Bühlmann, 2006) because two variables are conditionally independent if and only if their partial correlation is zero.

Because our LFP data $X (\in \mathbb{R}^{p \times q})$ involve both time ($t = 1, \ldots, p$) and space (electrodes $i = 1, \ldots, q$ on the Utah arrays), instead of a vector-variate Gaussian Graphical model, we consider a matrix-variate extension. A simple extension is imposing the vector-variate model on the flattened matrix-variate observation, but the resulting model requires prohibitive over-parametrization by $p^2 q^2$ entries in the covariance matrix and often compromises statistical efficiency by losing dependence structure in the matrix formulation. We instead use matrix-variate Gaussian Graphical Model (MGGM) with Kronecker product covariance. The model consists of two covariance parameters: temporal (row) covariance matrix $\Sigma^{(\mathcal{T})} \in \mathbb{R}^{p \times p}$ and spatial (column) covariance matrix $\Sigma^{(\mathcal{S})} \in \mathbb{R}^{q \times q}$. Then, a random matrix $X \in \mathbb{R}^{p \times q}$ is said to follow the model, denoted by $\mathrm{N}(\mu, \Sigma^{(\mathcal{T})}, \Sigma^{(\mathcal{S})})$, if and only if $\mathrm{vec}(X) \sim \mathrm{N}(\mathrm{vec}(\mu), \Sigma^{(\mathcal{S})} \otimes \Sigma^{(\mathcal{T})})$, where $\mu \in \mathbb{R}^{p \times q}$ is the matrix mean, $\mathrm{vec}(X) \in \mathbb{R}^{pq}$ denotes the flattened matrix X, and $\otimes$ denotes the Kronecker product. It imposes a strong assumption, implying, for instance, that the auto-correlation of measurement in every electrode is proportional to $\Sigma^{(\mathcal{T})}$, and the spatial covariance at every time point is proportional to $\Sigma^{(\mathcal{S})}$. But, as long as it is reasonable to separate temporal and spatial covariation, it has the advantage of reducing the number of parameters to $p^2 + q^2$ from $p^2 q^2$ of the vector-variate formulation, and it has been applied in analyzing spatio-temporal data of biomedical imaging and financial markets (Zhou, 2014; Chen and Liu, 2015; Zhu and Li, 2018). If we denote the temporal (row) and column (spatial) precision matrices by $\Omega^{(\mathcal{T})} = \Sigma^{(\mathcal{T})}$ and $\Omega^{(\mathcal{S})} = \Sigma^{(\mathcal{T})-1}$, respectively, the inverse of $\Sigma^{(\mathcal{S})} \otimes \Sigma^{(\mathcal{T})}$ has a simple analytic form: the precision matrix of $\mathrm{vec}(X)$ is $(\Sigma^{(\mathcal{S})} \otimes \Sigma^{(\mathcal{T})})^{-1} = \Omega^{(\mathcal{S},l)} \otimes \Omega^{(\mathcal{T},l)}$. The cross-regional association is encoded in $\Omega^{(\mathcal{S})}$; the spatial partial correlation between electrodes i and j is defined as

$$\rho_{ij}^{(\mathcal{S})} := -\frac{\Omega_{ij}^{(\mathcal{S})}}{\sqrt{\Omega_{ii}^{(\mathcal{S})} \Omega_{jj}^{(\mathcal{S})}}}. \tag{2.1}$$

For the rest of this paper, we propose a three-step estimation and inference procedure for the spatial partial-correlation. First, we estimate our spatial partial-correlation matrix by node-wise regression (Ren et al., 2019). Second, we follow the modified Cholesky factor decomposition Bickel and Levina (2008) to estimate the temporal covariance matrix. Finally, based on the results from the previous two steps, we construct a new linear-functional test statistic, which effectively tests significance of single edges or collections of multiple edges. The new inference framework is motivated by the linear functional-based tests by Ren et al. (2019), the matrix-variate graph inference by Chen and Liu (2015), and the global testing framework for a collection of edges in vector-variate graphical models by Chang et al. (2018). The rest of the contributions of the paper are summarized as follows:

- We develop and study a method of using multiple session data collectively so that we can discover the common spatial correlation graph across the sessions with improved sensitivity. To do so, we use group Lasso to borrow information across the sessions and achieve more accurate spatial covariance matrix estimates, compared to naive approaches of fitting the model separately to each session. (See Section 2.2.1.) We demonstrate the gain in the rate of convergence in theoretical (Section 2.3) and simulation studies (Section 2.4).
- Although group Lasso is not new, its application and consistency on auto-correlated samples have never been theoretically studied. In Section 2.3, we provide a self-contained analysis and derive the rates of convergence of estimation as well as prediction. It might be of independent interest for high-dimensional statistics on correlated data.
- We provide theoretical guarantees on the proposed inference procedures using a Gaussian approximation bootstrap method, based on Chernozhukov et al. (2012). It is the first attempt to formulate a global significance test of an edge group in a matrix-variate graph, rather than FDR control as in Chen and Liu (2015).

2.1.1 Other Related Work

Methods to estimate the sparse precision matrix for normal distribution are well-studied in recent years (Meinshausen and Bühlmann, 2006; Yuan and Lin, 2006). There are two main categories of method for sparse precision matrix estimation: penalized likelihood method (Allen and Tibshirani, 2010) and regression based method (Liu, 2013; Ren et al., 2015). The penalized likelihood method is a popular method for Gaussian Graphical Model estimation and inference. For example, in Graphical LASSO (Friedman et al., 2008), the goal is to maximize the log-likelihood function from the raw data and L_1 penalty term. For statistical inference, Jankova and Van De Geer (2015) proposed a method to estimate confidence band for entries in sparse precision matrices, which guaranteed asymptotic normality for each edge. However, since the objective function is non-convex, the optimization approach is theoretically less appealing; moreover, methods based on optimization will usually require the irrepresentability condition, which is difficult to validate. We adapted the regression-based approach in our work, and estimate the entries in precision matrix with the regression residuals and coefficients. For MGGM, there are many interesting works focusing on the graph estimation (Leng and Tang, 2012; Yin and Li, 2012; Zhou, 2014) and inference (Chen and Liu, 2015; Ye et al., 2019). For example, the work of Chen and Liu (2015) developed a multiple testing framework for support recovery in MGGM, and provided theoretical analysis for asymptotic normality and false discovery rate (FDR) control. The work of Ye et al. (2019) developed a paired test of matrix graphs to infer brain connectivity with correlated samples, and similarly, their testing procedure was based on multiple testing and FDR control. We are interested, however, in deriving a single statistic to simultaneously test the strength of conditional dependence between subgroups of nodes across multiple graphs.

Inference under multiple matrix-variate graphs is missing in literature. However, estimation of multiple ordinary Gaussian graphical models based on optimization method was widely studied

(Danaher et al., 2014; Cai et al., 2016a; Lee and Liu, 2015). Moreover, Peterson et al. (2015) proposed Bayesian inference on multiple Gaussian graphical models. A Markov random field prior is implemented to encourages common edges across graphs, and a spike-and-slab prior is placed on the parameters to learn the groups which have a shared structure. Therefore, this model can learn the information between sample groups, and measure the relative network similarity across groups effectively. The most relevant work to our topic is by Zhu and Li (2018), where the main focus is on multiple matrix-variate graph estimation. The authors established a non-convex optimization method with sparse and group Lasso penalization to estimate multiple matrix-variate Gaussian graphs for matrix-normal distributed data. They further designed an efficient optimization algorithm, and established the asymptotic properties of the estimator under special scenarios with sparse penalty or group penalty only. However, since this work was based on non-convex optimization method, stronger assumptions are needed to guarantee convergence of estimation, and inference remains to be unknown under such a formulation.

2.1.2 Organization of the Paper and Notations

The paper is organized as follows. In Section 2.2, we propose the procedure for simultaneous test while leaving theoretical justifications for our procedure in Section 2.3. In Section 2.4.1, we present simulation studies to support our theoretical findings, and compare the performance of multiple matrix-variate graphs with several baseline methods. The simulation results demonstrate the advantage of our method over the state-of-art multi-graph estimation procedures. In Section 2.4.2, we apply our method on the experimental data and reveals change of within-area and cross-area connectivity with respect to the behavioral stages in the experimental design. Our test results provide new insights into understanding the neural activity in PFC and V4 during saccade task.

We adopt the following notation throughout this paper. For a euclidean vector X, we write $X_{i:j}$ for its sub-vector $(X_i, X_{i+1}, \cdots, X_j)'$. Moreover, let $\|X\|_p$ denote the vector ℓ_p-norm of X. For a real matrix X, let $X_{i,\cdot}$ denote the i-th row of X, $X_{\cdot,j}$ denote the j-th column of X. Let $\|X\|_p$ denote the matrix p-norm of X, while $\|X\|_2$ is also the spectral norm or operator norm. We use $\|X\|_F$ to denote its Frobenius norm and $\|X\|_\infty$ to denote the entry-wise sup-norm. For any set A we denote its cardinality by $|A|$. In our theoretical arguments, $C(\dots)$ means a constant with implicit dependency to the parameters in the parentheses, whose value is changing across lines. For universal constants without any dependency, we omit the parentheses and denote them by C.

2.2 Simultaneous Inference Framework

In this section, we develop a graph inference framework for matrix-variate observations and leave all theoretical studies in Section 2.3. The full procedure is summarized in Algorithm 1 at the end of this section. For each dataset on session $l = 1, \dots, m$, we have n_l i.i.d $p \times q$ matrix-variate samples, $X^{(1,l)}, \dots, X^{(n_l,l)}$, following matrix-variate Gaussian distribution $\mathrm{N}(0, \Sigma^{(\mathcal{T},l)}, \Sigma^{(\mathcal{S},l)})$, where $\Sigma^{(\mathcal{T},l)} \in \mathbb{R}^{p\times p}$ and $\Sigma^{(\mathcal{S},l)} \in \mathbb{R}^{q\times q}$ are the temporal and spatial covariance matrices in session l, respectively. We denote the temporal and spatial precision matrices by $\Omega^{(\mathcal{T},l)} := (\Sigma^{(\mathcal{T},l)})^{-1}$ and $\Omega^{(\mathcal{S},l)} := (\Sigma^{(\mathcal{S},l)})^{-1}$, respectively.

2.2.1 Estimation of Spatial Covariance Matrix

We first discuss the model $X^{(k,l)} \sim \mathrm{N}(0, \Sigma^{(\mathcal{T},l)}, \Sigma^{(\mathcal{S},l)})$ for each graph/session $l = 1, \dots, m$ as a motivation for our approach. Node-wise regression (Meinshausen and Bühlmann (2006); Liu (2013)) has been a popular approach for graphical model analysis. Indeed, it is well-known that the conditional distribution of $X_{ti}^{(k,l)}$ (i.e., electrode i at time t) against the rest variables in $X_{t,\cdot}^{(k,l)}$ follows a linear regression,

$$X_{ti}^{(k,l)} = X_{t,\cdot}^{(k,l)} \beta_{\cdot,i}^{(\mathcal{S},l)} + \epsilon_{ti}^{(\mathcal{S},k,l)}, \tag{2.2}$$

where it follows from the linear regression theory that

$$\beta_{ji}^{(\mathcal{S},l)} = \begin{cases} -\Omega_{ji}^{(\mathcal{S},l)} / \Omega_{ii}^{(\mathcal{S},l)}, & \text{if } i \neq j, \\ 0, & \text{if } i = j, \end{cases} \tag{2.3}$$

and $\mathbb{E}[\epsilon_{ti}^{(\mathcal{S},k,l)}] = 0$. Due to this connection between the coefficient $\beta_{ti}^{(\mathcal{S},l)}$ and the spatial precision matrix $\Omega^{(\mathcal{S},l)}$, the sparse partial-correlation assumption naturally implies sparsity in the regression coefficient.

Defining $\Phi^{(\mathcal{S},l)}$ to be the matrix of elements, $\Phi_{ij}^{(\mathcal{S},l)} := \Omega_{ij}^{(\mathcal{S},l)} / (\Omega_{ii}^{(\mathcal{S},l)} \Omega_{jj}^{(\mathcal{S},l)})$, the covariance matrix between the node-wsie spatial regression residuals, i.e., $\epsilon_{t,\cdot}^{(\mathcal{S},k,l)} := (\epsilon_{t1}^{(\mathcal{S},k,l)}, \dots, \epsilon_{tp}^{(\mathcal{S},k,l)})^\top$, is $\Sigma_{tt}^{(\mathcal{T},l)} \cdot \Phi^{(\mathcal{S},l)}$. Under the identifiability constraint that $\mathrm{tr}(\Sigma^{(\mathcal{T},l)}) = p$,

$$\Phi^{(\mathcal{S},l)} = \mathbb{E}\Big[\frac{1}{n_l p} \sum_{k=1}^{n_l} \sum_{t=1}^{p} \epsilon_{t,\cdot}^{(\mathcal{S},k,l)} \epsilon_{t,\cdot}^{(\mathcal{S},k,l)\top}\Big]. \tag{2.4}$$

Therefore, testing $\Phi_{ij}^{(\mathcal{S},l)} = 0$ is equivalent to testing if the partial correlation between i and j is 0. In the regression, we treat each row $X_{t,\cdot}^{(\mathcal{S},l)}$ as a q-dimensional sample, which leaves us p correlated vector-variate samples for a sparse linear regression model. The covariance among these "row samples" is characterized by $\Sigma^{(\mathcal{T},l)}$. Note that

$$\rho_{ij}^{(\mathcal{S},l)} = -\frac{\Omega_{ij}^{(\mathcal{S},l)}}{\sqrt{\Omega_{ii}^{(\mathcal{S},l)} \Omega_{jj}^{(\mathcal{S},l)}}} = -\Phi_{ij}^{(\mathcal{S},l)} \sqrt{\Omega_{ii}^{(\mathcal{S},l)} \Omega_{tjj}^{(\mathcal{S},l)}}. \tag{2.5}$$

In what follows, we build our test statistics based on some accurate estimation of $\Phi_{ij}^{(\mathcal{S},l)}$ and the equivalence between $\rho_{ij}^{(\mathcal{S},l)} = 0$ and $\Phi_{ij}^{(\mathcal{S},l)} = 0$.

Having discussed the model for each session above, we are in a position to consider all m sessions together to improve the estimation accuracy of each $\rho_{ij}^{(\mathcal{S},l)}$. The fact that all sessions tend to share the same support among the column precision matrices $\Omega_{ij}^{(\mathcal{S},l)}$'s implies the m coefficient vectors $\beta_{\cdot,i}^{(\mathcal{S},l)}$'s share the same support as well. To this end, we naturally treat $\beta_{ij}^{(\mathcal{S},1)}, \dots, \beta_{ij}^{(\mathcal{S},m)}$ as the group of the coefficient parameters for pair (i, j) across the m sessions. Under a joint sparsity structure of m graphs, we stack the m linear models and use the group Lasso (Yuan and Lin

(2006)) to take the advantage of this group structure as follows:

$$\begin{aligned}
&\{\widehat{\beta}^{(\mathcal{S},l)}_{\cdot,i}\}_{l=1,\dots,m} \\
&:= \underset{\{b^{(l)}\in\mathbb{R}^p\}_{l=1,\dots,m}}{\operatorname{argmin}} \left\{ \frac{1}{2n_0p}\sum_{l=1}^{m}\sum_{k=1}^{n_l}\|X^{(k,l)}_{\cdot,i} - X^{(k,l)}b^{(l)}\|_2^2 + \lambda_i \sum_{j:j\neq i}\sqrt{\sum_{l=1}^{m}\frac{\|\sum_k X^{(k,l)}_{\cdot,j}\|_2^2}{n_lp}b_j^{(l)2}} \right\} \\
&\text{w.r.t } \ b_i^{(l)} = 0,
\end{aligned} \tag{2.6}$$

where $n_0 = \min_{1\leq l\leq m} n_t$. The parameter λ_i can be tuned using cross-validation or other model selection methods. In our theoretical analysis, a data-driven yet conservative choice of each λ_i can be picked.

With estimated regression coefficients, the fitted residuals can be calculated as

$$\widehat{\epsilon}^{(\mathcal{S},k,l)}_{ti} = X^{(k,l)}_{ti} - X^{(k,l)}_{t,\cdot}\widehat{\beta}^{(\mathcal{S},l)}_{\cdot,i}. \tag{2.7}$$

Estimating the population covariance among residuals, the empirical covariance matrix of $\widehat{\epsilon}^{(\mathcal{S},k,l)}$ has a larger bias compared to the expected $\sqrt{n_lp}$ rate due to the Lasso penalty, as demonstrated in vector-Gaussian graphical models (Liu, 2013). We introduce a bias-correction term in the covariance estimate to remedy the large bias:

$$\widehat{\Phi}^{(\mathcal{S},l)}_{ij} := \begin{cases} -\frac{1}{n_lp}\sum_{k=1}^{n_l}\sum_{t=1}^{p}\left(\widehat{\epsilon}^{(\mathcal{S},k,l)}_{ti}\widehat{\epsilon}^{(\mathcal{S},k,l)}_{tj} + \widehat{\epsilon}^{(\mathcal{S},k,l)2}_{tj}\widehat{\beta}^{(\mathcal{S},l)}_{ji} + \widehat{\epsilon}^{(\mathcal{S},k,l)2}_{ti}\widehat{\beta}^{(\mathcal{S},l)}_{ij}\right), & \text{if } \ i\neq j, \\ \frac{1}{n_lp}\sum_{k=1}^{n_l}\sum_{t=1}^{p}\widehat{\epsilon}^{(\mathcal{S},k,l)}_{ti}\widehat{\epsilon}^{(\mathcal{S},k,l)}_{tj}, & \text{if } \ i = j. \end{cases} \tag{2.8}$$

Then, we estimate $\Omega^{(\mathcal{S},l)}$ and $\rho^{(\mathcal{S},l)}$ by

$$\widehat{\Omega}^{(\mathcal{S},l)}_{ij} := \frac{\widehat{\Phi}^{(\mathcal{S},l)}_{ij}}{\widehat{\Phi}^{(\mathcal{S},l)}_{ii}\widehat{\Phi}^{(\mathcal{S},l)}_{jj}} \text{ and } \widehat{\rho}^{(\mathcal{S},l)}_{ij} := -\frac{\widehat{\Phi}^{(\mathcal{S},l)}_{ij}}{\sqrt{\widehat{\Phi}^{(\mathcal{S},l)}_{ii}\widehat{\Phi}^{(\mathcal{S},l)}_{jj}}}, \text{ respectively.} \tag{2.9}$$

Our estimator for the partial-correlation resembles the form proposed by Chen and Liu (2015) for a single matrix-variate Gaussian graphical model. However, we further leverage the information in multiple sessions using the group Lasso estimate and obtain a faster convergence rate, which is summarized in Theorem 2.3.1 and Remark 2.3.3. This results into the improved testing accuracy for our goals by the factor of $\sqrt{m}$, which is demonstrated in Theorems 2.3.5 and 2.3.6.

2.2.2 Simultaneous Test by Parametric Bootstrap

Single Edge Test

Borrowing information across m sessions not only enables us to estimate d graphs efficiently by the group Lasso Eq. (2.6) but constructs powerful tests to discover significant associations. We first focus on the single edge test: for a pair (i, j), we are interested in testing the following null hypothesis,

$$H_{0,ij} : \rho^{(\mathcal{S},l)}_{ij} = 0, \ \ \forall l = 1, \dots, m. \tag{2.10}$$

In light of the group sparsity structure, we construct a test statistic by aggregating the partial-correlation estimates $\widehat{\rho}_{ij}^{(\mathcal{S},l)}$ from all sessions $l = 1, \ldots, m$. In our application, we assume that the signs of the associations do not change across the sessions, so that the sum of the estimates is close to $\sum_{l=1}^{m} |\rho_{ij}^{(\mathcal{S},l)}|$, which is zero under the null hypothesis. we moreover address the heterogeneous uncertainty of $\widehat{\rho}_{ij}^{(\mathcal{S},l)}$ across l a by summing the estimates weighted by $\sqrt{n_l p}$ and propose the following test statistic

$$\widehat{T}_{ij} := \frac{1}{\sqrt{m}} \sum_{l=1}^{m} \sqrt{n_l p}\, \widehat{\rho}_{ij}^{(\mathcal{S},l)}. \tag{2.11}$$

In Proposition 2.3.2 of Section 2.3, we show that our test statistic is asymptotically close to a normal distribution with mean

$$T_{ij} := \frac{1}{\sqrt{m}} \sum_{l=1}^{m} \sqrt{n_l p}\, \rho_{ij}^{(\mathcal{S},l)}, \tag{2.12}$$

so $\widehat{T}_{ij}$ should be significantly larger than zero when the alternative is true. Once the asymptotic variance can be consistently estimated, it is straightforward to construct the confidence interval for $T_{ij}^{(\mathcal{S},l)}$ and the associate p-value for the single edge. We do not pursue this direction immediately here. Instead, we move to the more challenging multiple edge test scenario, and leave one the single edge test as a trivial special case.

Remark 2.2.1. More generally, an additional sign information on elements of the alternative $\rho_{ij}^{(\mathcal{S},l)}$ may be available. With this additional knowledge, we present a test statistic based on a linear combination of those $\widehat{\rho}_{ij}^{(\mathcal{S},l)}$ for $l = 1, \ldots, m$, which is closely related to its ℓ_1 norm. More specifically, with an edge-specific sign vector $\sigma_{ij} := (\sigma_{ij}^{(1)}, \ldots, \sigma_{ij}^{(m)})^\top \in \{-1, 1\}^m$, we replace $\widehat{T}_{ij}$ with the following sign-addressed test statistic,

$$\widehat{T}_{ij,\sigma} := \frac{1}{\sqrt{m}} \sum_{l=1}^{m} \sigma_{ij}^{(l)} \sqrt{n_l p} \widehat{\rho}_{ij}^{(\mathcal{S},l)}. \tag{2.13}$$

The normal approximation we establish in Section 2.3 also applies to the sign-addressed test statistic.

Simultaneous Test

In the multiple edge test scenario, for a cross-region set E, we aim to test whether there is no edge at all in E, which is stated as the following null hypothesis,

$$H_{0,E} : \rho_{ij}^{(\mathcal{S},l)} = 0, \;\; \forall (i,j) \in E, \;\; \forall l = 1, \ldots, m. \tag{2.14}$$

Notice that the single edge test can be treated as a special case of multiple edge test when $E = \{(i,j)\}$. In practical neuroscience applications, E is a collection of edges connecting different brain areas, of which the cardinality can increase as large as the rate of q^2. Thus, even if we establish asymptotic normality of a single T_{ij}, the theory cannot apply to the high-dimensional setting of testing $H_{0,E}$ with growing $|E|$. One may remedy the multiple testing issue via Bonferroni correction, but it is known to exert overly conservative decision. In this paper, we propose a simultaneous testing based on the supremum norm of $\widehat{T}_E := (\widehat{T}_{ij} : (i,j) \in E)$:

$$\|\widehat{T}_E\|_\infty := \max_{(i,j)\in E} |\widehat{T}_{ij}|. \tag{2.15}$$

We notice that the asymptotic mean and coviariance matrix of $\widehat{T}_E$ is given by T_E and S_{EE}, respectively, where T is the groundtruth counterpart of $\widehat{T}$

$$T_{ij} := \frac{1}{\sqrt{m}} \sum_{l=1}^{m} \sqrt{n_l p}\, \rho_{ij}^{(\mathcal{S},l)}, \tag{2.16}$$

and S is the matrix consisting of asymptotic covariance between $T_{i_1 j_1}$ and $T_{i_2 j_2}$ for $(i_1, j_1), (i_2, j_2) \in [d] \times [d]$, given by

$$\begin{aligned} &S_{(i_1,j_1),(i_2,j_2)} \\ &:= \frac{1}{m} \sum_{l=1}^{m} \frac{\|\Sigma^{(\mathcal{T},l)}\|_F^2}{p} \left[\begin{aligned} &\rho_{i_1 i_2}^{(\mathcal{S},l)} \rho_{j_1 j_2}^{(\mathcal{S},l)} + \rho_{i_1 j_2}^{(\mathcal{S},l)} \rho_{i_2 j_1}^{(\mathcal{S},l)} + \frac{1}{2} \rho_{i_1 j_1}^{(\mathcal{S},l)} \rho_{i_2 j_2}^{(\mathcal{S},l)} \left(\rho_{i_1 i_2}^{(\mathcal{S},l)2} + \rho_{j_1 j_2}^{(\mathcal{S},l)2} + \rho_{i_1 j_2}^{(\mathcal{S},l)2} + \rho_{i_2 j_1}^{(\mathcal{S},l)2} \right) \\ &- \rho_{i_1 i_2}^{(\mathcal{S},l)} \rho_{i_2 j_2}^{(\mathcal{S},l)} \rho_{i_2 j_1}^{(\mathcal{S},l)} - \rho_{i_1 i_2}^{(\mathcal{S},l)} \rho_{i_1 j_1}^{(\mathcal{S},l)} \rho_{i_1 j_2}^{(\mathcal{S},l)} - \rho_{j_1 j_2}^{(\mathcal{S},l)} \rho_{i_2 j_2}^{(\mathcal{S},l)} \rho_{i_1 j_2}^{(\mathcal{S},l)} - \rho_{j_1 j_2}^{(\mathcal{S},l)} \rho_{i_2 j_1}^{(\mathcal{S},l)} \rho_{i_1 j_1}^{(\mathcal{S},l)} \end{aligned} \right]. \end{aligned} \tag{2.17}$$

The key idea follows the theory of high-dimensional central limit theorems (e.g., Chernozhukov et al., 2012): although the entire vector $\widehat{T}_E$ is not asymptotically normal, $\|\widehat{T}_E - T_E\|_\infty$ has the same limiting behavior as $\|Z\|_\infty$, where $Z \sim \mathrm{MVN}(0, S_{EE})$.We later elucidate this idea in Section 2.3.

The distribution of $\|Z\|_\infty$ involves a supremum norm of a normal random vector, of which the quantile is not analytically available, and also unknown parameters in Eq. (2.17). We replace S_{EE} with the plug-in estimator $\widehat{S}_{EE}$ based on the estimates $\widehat{\Sigma}^{(\mathcal{T},l)}$ and $\widehat{\rho}^{(\mathcal{S},l)}$, and approximate the distribution of $\|Z\|_\infty$ by the bootstrap distribution of $\|\widehat{Z}\|_\infty$. Based on bootstrap samples $\widehat{Z} \sim \mathrm{MVN}(0, \widehat{S}_{EE})$, we construct a $(1-\alpha)$ confidence region

$$\mathcal{C}_E(1-\alpha) := \left\{ \{\rho^{(\mathcal{S},l)}\}_{l=1,\dots,m} : \max_{(i,j)\in E} \left| \frac{1}{\sqrt{m}} \sum_{l=1}^{m} \sqrt{n_l p} (\widehat{\rho}_{ij}^{(\mathcal{S},l)} - \rho_{ij}^{(\mathcal{S},l)}) \right| \le \widehat{q}_{\|\widehat{Z}\|_\infty, 1-\alpha} \right\}, \tag{2.18}$$

where $\widehat{q}_{\|\widehat{Z}\|_\infty, 1-\alpha}$ is the bootstrap $(1-\alpha)$-quantile of $\|\widehat{Z}\|_\infty$. Theorem 2 provides the validity of the confidence region in Eq. (2.18), and Theorem 3 demonstrates the complementary power analysis. Based on the estimated confidence region, we can obtain the trivial testing procedure extension for c-level testings.

Remark 2.2.2. The row samples we fit the group Lasso Eq. (2.6) are correlated within each session $l = 1, \dots, m$, and the correlation is characterized by the temporal covariance matrix $\Sigma^{(\mathcal{T},l)}$. We present an estimate of $\Sigma^{(\mathcal{T},l)}$ based on the Kronecker product covariance structure in the matrix-variate Gaussian graphical models and plug it in Eq. (2.18) to obtain a consistent confidence region. It is worthwhile to mention that under other dependence structures among samples, one may construct different procedures for a vector-variate Gaussian graph. For instance, Chang et al. (2018) implemented a kernel estimator for estimation of $\Sigma^{(\mathcal{T},l)}$ in multi-variate time series data.

2.2.3 Estimation of Temporal Covariance Matrix

Our testing framework developed in Section 2.2.2 requires to estimate the Frobenius norm of the temporal covariance matrix, $\Sigma^{(\mathcal{T},l)}$. The temporal covariance matrix can be estimated by a modified Cholesky decomposition (Bickel and Levina, 2008; Liu and Ren, 2020). The Cholesky decomposition reveals the natural auto-regressive relationship between the signal at a particular time point and the past signals from the matrix-variate Gaussian distribution with Kronecker product covariance. Suppose that the Cholesky decomposition of $\Omega^{(\mathcal{T},l)}$ is

$$\Omega^{(\mathcal{T},l)} = L^{(\mathcal{T},l)} D^{(\mathcal{T},l)} L^{(\mathcal{T},l)\top}, \tag{2.19}$$

where $D^{(\mathcal{T},l)}$ is the diagonal matrix so that $L^{(\mathcal{T},l)}$ is lower triangular and has diagonal entries 1. Let $\beta^{(\mathcal{T},l)} := I - L^{(\mathcal{T},l)\top}$ and $\Phi^{(\mathcal{T},l)} = \frac{\mathrm{tr}(\Sigma^{(\mathcal{S},l)})}{q} D^{(\mathcal{T},l)}$ so that the Cholesky decomposition of $\Omega^{(\mathcal{T},l)}$ is rewritten by

$$\Omega^{(\mathcal{T},l)} = \frac{\mathrm{tr}(\Sigma^{(\mathcal{S},l)})}{q} (I - \beta^{(\mathcal{T},l)})^\top \Phi^{(\mathcal{T},l)-1} (I - \beta^{(\mathcal{T},l)}). \tag{2.20}$$

Then, $\beta^{(\mathcal{T},l)}_{\cdot,t}$ is the solution of the linear regression model

$$X^{(k,l)}_{ti} = X^{(k,l)\top}_{\cdot,i} b + \epsilon^{(\mathcal{T},k,l)}_{ti}, \tag{2.21}$$

for each time t, and $\Phi^{(\mathcal{T},l)}$ is the residual variance

$$\Phi^{(\mathcal{T},l)}_{tt} = \frac{1}{n_l q} \sum_{k=1}^{n_l} \sum_{i=1}^{q} \mathrm{Var}\left[\epsilon^{(\mathcal{T},k,l)}_{ti}\right]. \tag{2.22}$$

Despite time-aligned experimental stages, the temporal alignment of neural response is not guaranteed due to the existence of response latencies of neurons to stimulus (Ventura, 2004). For example, the neural response time for saccade tasks might depend on a variety of factors in previous studies (Armington and Bloom, 1974; Dandekar et al., 2012; Ossandón et al., 2010), such as saccade amplitude, direction and change in luminance. Hence, we do not assume $\Sigma^{(\mathcal{T},l)}$ to be the same or very close across l but impose a rather weaker assumption that the all temporal precision matrices share bandedness. In physiological time-series signals, the dependence across time points exhibits natural decay as the time lag increases, so a reasonable assumption is that $\beta^{(\mathcal{T},l)}_{st}$ decreases to 0 as $t - s \to \infty$. For example, the popular auto-regressive $\mathrm{AR}(h)$ model satisfies the assumption by $\beta^{(\mathcal{T},l)}_{st} = 0$ where $t - s > h$. Indeed, the $\mathrm{AR}(h)$ model provides with an efficient regularization method in estimating auto-regressive coefficients for time-series data (Seth et al., 2015; Brincat and Miller, 2015, 2016; Liu and Ren, 2020). The idea is to obtain the optimal bias-variance trade-off in the estimation of $\beta^{(\mathcal{T},l)}_{\cdot,t}$ by ignoring weak dependences between far time points. Here, we follow the procedure in Liu and Ren (2020) on the $n_l q$ observations of time-series, $(X^{(k,l)}_{ti} : t = 1, \ldots, T)$, for $i = 1, \ldots, q$ and $k = 1, \ldots, n_l$, which appear as the columns of $X^{(k,l)}$'s. That is, under the model, a linear regression model at each time point t is formulated as

$$\begin{aligned} \widehat{\beta}^{(\mathcal{T},l)}_{\cdot,t} &:= \underset{b \in \mathbb{R}^p}{\mathrm{argmin}} \frac{1}{2 n_l q} \sum_{k=1}^{n_l} \| X^{(k,l)}_{t,\cdot} - X^{(k,l)\top} b \|_2^2 \\ &\text{w.r.t } \; b_s = 0 \; \text{ where } \; s < t - h_l \; \text{ or } \; s \geq t. \end{aligned} \tag{2.23}$$

We note that, unlike the spatial case, the estimation of the temporal covariance matrix is done individually for each graph/session.

Having the empirical regression coefficients obtained, we estimate $\Phi^{(\mathcal{T},l)}$ given by

$$\widehat{\Phi}_{tt}^{(\mathcal{T},l)} = \frac{1}{n_l q} \sum_{k=1}^{n_l} \left\| X_{t,\cdot}^{(k,l)} - X^{(k,l)\top} \beta_{\cdot,t}^{(\mathcal{T},l)} \right\|_2^2. \tag{2.24}$$

For technical issues in the following theoretical analyses, we truncate the eigenvalues of $I - \widehat{\beta}$ such as in Liu and Ren (2020). For a square matrix A, let

$$P_\eta(A) := U \ \max\{\min\{\Lambda, \eta\}, \eta^{-1}\} \ V^\top, \tag{2.25}$$

where A has a singular value decomposition $A = U\Lambda V^\top$ and the $\min$ and $\max$ above are element-wise operations. Let $\overline{\Omega}^{(\mathcal{T},l)}$ be the precursor estimator of $\Omega^{(\mathcal{T},l)}$ given by

$$\overline{\Omega}^{(\mathcal{T},l)} = P_\eta(I - \widehat{\beta}^{(\mathcal{T},l)})^\top \widehat{\Phi}^{(\mathcal{T},l)-1} P_\eta(I - \widehat{\beta}^{(\mathcal{T},l)}). \tag{2.26}$$

Then, one might consider to plug $\overline{\Omega}^{(\mathcal{T},l)}$ and $\widehat{\Sigma}^{(\mathcal{S},l)}$ from Section 2.2.1 in Eq. (2.20). However, this approach involves the uncertainties from both the group Lasso estimate in Eq. (2.6) and banded Cholesky factor estimate in Eq. (2.23). Instead, we notice that $\Sigma^{(\mathcal{T},l)}$ is the standardized $\overline{\Sigma}^{(\mathcal{T},l)} := \overline{\Omega}^{(\mathcal{T},l)-1}$ by a scalar multiplication so that the $\mathrm{tr}(\Sigma^{(\mathcal{T},l)}) = p$. Our proposed estimate is

$$\widehat{\Sigma}^{(\mathcal{T},l)} = \frac{p}{\mathrm{tr}(\overline{\Sigma}^{(\mathcal{T},l)})} \overline{\Sigma}^{(\mathcal{T},l)} \ \text{ and } \ \widehat{\Omega}^{(\mathcal{T},l)} = \frac{\mathrm{tr}(\overline{\Sigma}^{(\mathcal{T},l)})}{p} \overline{\Omega}^{(\mathcal{T},l)}. \tag{2.27}$$

2.3 Theoretical Properties

Throughout the subsequent theoretical arguments, we use $X^{(\mathcal{S},l)}$ (or $X^{(\mathcal{T},l)}$) to notate stacked spatial (or temporal) observation over temporal (or spatial) components and samples within session l:

$$\begin{aligned} X^{(\mathcal{S},l)} &= \left[X^{(1,l)\top}, \ldots, X^{(n_l,l)\top}\right]^\top \in \mathbb{R}^{n_l p \times q} \\ X^{(\mathcal{T},l)} &= \left[X^{(1,l)}, \ldots, X^{(n_l,l)}\right]^\top \in \mathbb{R}^{n_l q \times p}. \end{aligned} \tag{2.28}$$

These notations help us with connecting matrix-variate Gaussian graphical models to node-wise regression models:

$$X_{ti}^{(\mathcal{S},l)} = X_{t,\cdot}^{(\mathcal{S},l)} \beta_{\cdot,i}^{(\mathcal{S},l)} + \epsilon_{ti}^{(\mathcal{S},l)} \tag{2.29}$$

for $t = 1, \ldots, n_l p$ where $X_t^{(\mathcal{S},l)}$ are now dependent and not identically distributed due to the temporal association, unlike the usual regression regime. Also, for any vector $b^{(\cdot)} := \{b^{(l)} \in \mathbb{R}^p\}_{l=1,\ldots,m}$, we use $\underline{b}^{(\cdot)}$ to denote the collection of standardized elements, i.e.,

$$\underline{b}_j^{(\cdot)} := \left\{ \frac{\|X_{\cdot,j}^{(\mathcal{S},l)}\|_2}{\sqrt{n_l p}} b_j^{(l)} \right\}_{l=1,\ldots,m} \tag{2.30}$$

Algorithm 1 Simultaneous Testing for Multiple MGGMs

```
Input: Multi-session data $\mathcal{D}$, edge set $\mathcal{S}$, test level $\alpha$
Output: Confidence region $\mathcal{C}_E(1-\alpha)$
Spatial precision matrix estimation:
for $i = 1 : q$ do
    Estimate the regression coefficient $\beta^{(\mathcal{S},l)}_{\cdot,i}$ and the residual $\epsilon^{(\mathcal{S},k,l)}_{\cdot,i}$ using Eq. (2.6).
end for
for $l = 1 : m, i = 1 : q, j = 1 : q$ do
    Estimate the de-biased residual variance $\Phi^{(\mathcal{S},l)}_{ij}$ using Eq. (2.8).
    Estimate the spatial precision $\Omega^{(\mathcal{S},l)}_{ij}$ and partial correlation $\rho^{(\mathcal{S},l)}_{ij}$ using Eq. (2.9).
end for
Temporal precision matrix estimation:
for $t = 1 : d$ do
    Estimate the temporal regression coefficient $\beta^{(\mathcal{T},l)}$ and residual variance $\Phi^{(\mathcal{T},l)}$ using
      Eqs. (2.23) and (2.24).
end for
Hypothesis testing based on bootstrap:
Estimate the covariance matrix $S_{EE}$ of the test statistic $T_E$ using the plug-in estimator $\widehat{S}_{EE}$ in
  Eq. (2.17).
Sample $\{\widehat{Z}_i\}_{i=1,\dots,B} \sim N(0, \widehat{S}_{EE})$ and calculate the confidence region inc Eq. (2.18).
return Confidence region $\mathcal{C}_E(1-\alpha)$
```

so that Eq. (2.6) is expressed in a more canonical form:

$$\widehat{\beta}^{(\mathcal{S},\cdot)}_{\cdot,i} := \operatorname*{argmin}_{\{b^{(l)}\in\mathbb{R}^p\}_{l=1,\dots,m}} \left\{ \frac{1}{2n_0 p} \sum_{l=1}^{m} \|X^{(\mathcal{S},l)}_{\cdot,i} - X^{(\mathcal{S},l)} b^{(l)}\|_2^2 + \lambda_i \sum_{j:j\neq i} \left\| \underline{b}^{(\cdot)}_{j} \right\|_2 \right\} \tag{2.31}$$

$$\text{w.r.t } \; b^{(l)}_i = 0.$$

We make the following assumptions on the observed dataset.

Assumption 1. $\max_{l=1,\dots,m} \frac{n_l}{n_0} \le \kappa_1$ *for some positive constant* κ_1 *where* $n_0 = \min_{l=1,\dots,m} n_l$.

Assumption 2. $\mathrm{tr}(\Sigma^{(\mathcal{T},l)}) = p$, $\forall l = 1, \dots, m$.

Assumption 3. *For* $l = 1, \dots, m$, *let* $\{\lambda^{(\mathcal{T},l)}_i\}_{i=1,\dots,p}$ *are the eigenvalues of* $\Sigma^{(\mathcal{T},l)}$ *while* $\frac{1}{\kappa_3} \le \lambda^{(\mathcal{T},l)}_1 \le \lambda^{(\mathcal{T},l)}_2 \le \cdots \le \lambda^{(\mathcal{T},l)}_p \le \kappa_3$ *for some constant* $\kappa_3 > 0$*; define and assume* $\{\lambda^{(\mathcal{S},l)}_i\}_{i=1,\dots,q}$ *similarly for* $\Sigma^{(\mathcal{S},l)}$.

Assumption 4. *Let* d *be the group-wise maximum node degree, i.e.,*

$$d := \max_i \left| \left\{ j \in [p] \backslash \{i\} : \Omega^{(\mathcal{S},l)}_{ij} \neq 0 \,\textit{for some}\; l \in [m] \right\} \right|. \tag{2.32}$$

We assume group sparsity of the partial correlation graph in spatial association by

$$d \cdot \frac{\max\{m, \log^2(qmn_0p)\}}{(mn_0p)^{1/2}} \to 0 \tag{2.33}$$

as $n_0 \to \infty$.

Assumption 5. *We assume the temporal precision matrix* $\Omega^{(\mathcal{T},l)}$ *for each* $l = 1, \ldots, m$ *has Cholesky decomposition as in Eq.* (2.20) *where* $\beta^{(\mathcal{T},l)}$ *satisfies*

$$|\beta_{st}^{(\mathcal{T},l)}| < \kappa_5 (t-s)^{-\alpha_l - 1} \textit{ for } t, s : s < t, \tag{2.34}$$

for $\alpha_l > 0$. *We further assume that, for* $\alpha_0 = \min_{l=1,\ldots,m} \alpha_l$,

$$\frac{\log(qmn_0p)}{(n_0q)^{1-1/(\alpha_0+1)}} \to 0 \tag{2.35}$$

as $n_0 \to \infty$.

Assumption 1 suggests that the sample size from each graph is balanced and we use n_0 to represent the common level. Assumption 2 is for the identifiability. Assumption 3 is a standard eigenvalue assumption in covariance estimation (Cai et al., 2016b). Assumption 4 indicates that the column precision matrix is sparse, and limits the spatial dimension of the matrix variable given the number of samples, the temporal dimension and the number of graphs. The first part of Assumption 5 is a fair assumption for neural time series, since neural data, especially LFPs, are usually modeled as an auto-regressive process with limited order, which is also widely considered in literature (Bickel and Levina, 2008; Liu and Ren, 2020); the second part of is similar to Assumption 4 and limits the temporal dimension.

2.3.1 Non-asymptotic error bound for the group Lasso estimate

We first provide a theoretical justification for our group Lasso procedure proposed in Section 2.2.1. Although with correlated rows, our results below demonstrate that the optimal rates of convergence for estimation and prediction can be still obtained compared to the case with i.i.d. samples. The proof is provided in Appendix B.2.2.

Theorem 2.3.1. *Suppose that* $\lambda_i = \sqrt{\frac{m+\log(qmn_0p)}{n_0p}}$. *Then,*

$$\mathbb{P}\left[\begin{array}{l} \max\limits_i \sum\limits_{j:j\neq i} \left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2^2 \leq C(\kappa_1,\kappa_3)\, d\, \dfrac{m+\log(qmn_0p)}{n_0p}, \\ \max\limits_i \sum\limits_{j:j\neq i} \left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2 \leq C(\kappa_1,\kappa_3)\, d\, \sqrt{\dfrac{m+\log(qmn_0p)}{n_0p}}, \\ \max\limits_i \dfrac{1}{2n_0p} \sum\limits_{l=1}^{m} \left\|X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l))}\right\|_2^2 \leq C(\kappa_1,\kappa_3)\, d\, \dfrac{m+\log(qmn_0p)}{n_0p} \end{array}\right] \geq 1 - C(qmn_0p)^{-1/2}. \tag{2.36}$$

for a sufficiently large n_0.

2.3.2 Theoretical justification for the simultaneous edge testing

We now turn to theoretical results from Section 2.2.2. In the following proposition, we show that the error of our partial-correlation estimate $\widehat{\rho}_{ij}^{(\mathcal{S},l)}$ in Eq. (2.9) is driven by the leading term

$$\Theta_{ij}^{(\mathcal{S},l)} := \frac{\widetilde{\delta}_{ij}^{(\mathcal{S},l)}}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} - \frac{\Phi_{ij}^{(\mathcal{S},l)}\widetilde{\delta}_{jj}^{(\mathcal{S},l)}}{2\Phi_{jj}^{(\mathcal{S},l)}\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} - \frac{\Phi_{ij}^{(\mathcal{S},l)}\widetilde{\delta}_{ii}^{(\mathcal{S},l)}}{2\Phi_{ii}^{(\mathcal{S},l)}\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} \tag{2.37}$$

where $\widetilde{\delta}_{ij}^{(\mathcal{S},l)} := \widetilde{\Phi}_{ij}^{(\mathcal{S},l)} - \Phi_{ij}^{(\mathcal{S},l)} = \frac{1}{n_l p}\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}\epsilon_{\cdot,j}^{(\mathcal{S},l)} - \Phi_{ij}^{(\mathcal{S},l)}$. We further provide a non-asymptotic bound for $O_{ij}^{(\mathcal{S},l)}$, which is the error not governed by $\Theta_{ij}^{(\mathcal{S},l)}$. That is, $\widehat{\rho}_{ij}^{(\mathcal{S},l)} - \rho_{ij}^{(\mathcal{S},l)} = \Theta_{ij}^{(\mathcal{S},l)} + O_{ij}^{(\mathcal{S},l)}$. We provide the proof in Appendix B.2.3.

Proposition 2.3.2. *Suppose that $\widehat{\rho}^{(\mathcal{S},l)}$ is estimated based on λ_i's given as in Theorem 2.3.1. Then,*

$$\mathbb{P}\left[\max_{i,j:i\neq j}\sum_{l=1}^{m}\left|O_{ij}^{(\mathcal{S},l)}\right| \geq C(\kappa_1,\kappa_3)\, d\,\frac{m+\log^2(qmn_0p)}{n_0p}\right] \leq C(qmn_0p)^{-1/2}. \tag{2.38}$$

for a sufficiently large n_0.

Remark 2.3.3. For the simultaneous multiple edge test, in order to make $\|\sum_{l=1}^{m}\frac{\sqrt{n_l p}}{\sqrt{m}}O_E^{(\mathcal{S},l)}\|_\infty$ converging to 0 in probability, the sample size requirement is $n_0p = \omega(\frac{d^2(m+\log^2(qmn_0p))^2}{m})$ as $n_0 \to \infty$. In comparison, one can also naively apply the estimation procedure for each graph separately as in Chen and Liu (2015), and compute a similar test statistic to perform single edge and multiple edge test following our procedure. However, such a naive method will require a much stronger sample size assumption, which is $n_0p = \omega(d^2 m\log^3 q)$.

Built upon the idea from Chernozhukov et al. (2012), we establish the Gaussian approximation result for our test statistic $\|\widehat{T}_E\|_\infty$.

Proposition 2.3.4. *Let $Z \sim \mathrm{N}(0, S_{EE})$ where the elements of S_{EE} are given as in Eq.* (2.17). *Then,*

$$\begin{aligned}&\sup_{x>0}\left|\mathbb{P}[\|\widehat{T}_E - T_E\|_\infty > x] - \mathbb{P}[\|Z\|_\infty > x]\right|\\&\leq C(\kappa_1,\kappa_3)\max\left\{\frac{\log^{7/8}(qmn_0p)}{(mn_0p)^{1/8}},\ d\log^{1/2}|E|\frac{m+\log^2(qmn_0p)}{(mn_0p)^{1/2}}\right\}\end{aligned} \tag{2.39}$$

for a sufficiently large n_0.

The following theorem parallels to the previous proposition except that we replace the population covariance S_{EE} with its plug-in estimator $\widehat{S}_{EE}$. At a high level, as long as the covariance S_{EE} can be estimated well under the $\|\cdot\|_\infty$-norm, the Gaussian approximation results remains valid. Its proof relies on Propositions Propositions 2.3.2 to 2.3.4 and is provided in Appendix B.2.5.

Theorem 2.3.5. *Let $\widehat{Z} \sim \mathrm{N}(0, \widehat{S}_{EE})$ where the elements of $\widehat{S}_{EE}$ is a plug-in estimator of S_{EE} in*

Eq. (2.17) *given observed data $\mathcal{D}$. Then,*

$$\sup_{x>0}\Big|\mathbb{P}[\|\widehat{T}_E - T_E\|_\infty > x] - \mathbb{P}[\|\widehat{Z}\|_\infty > x|\mathcal{D}]\Big|$$
$$\leq C(\kappa_1,\kappa_3,\kappa_5)\max\left\{\begin{matrix}\dfrac{\log^{7/8}(qmn_0p)}{(mn_0p)^{1/8}},\ d\log^{1/2}|E|\dfrac{m+\log^2(qmn_0p)}{(mn_0p)^{1/2}}, \\ \log^{2/3}|E|\dfrac{\log^{1/3}(qmn_0p)}{(n_0q)^{1/3-1/(6\alpha_0+6)}},\ \log^{2/3}|E|\left(\dfrac{m+\log(qmn_0p)}{mn_0p}\right)^{1/6}\end{matrix}\right\} \tag{2.40}$$

with probability at least $1-C(qmn_0p)^{-1/2}$ for a sufficiently large n_0.

The above theorem establishes the theoretical foundation for the simultaneous multiple edge testing procedure in Algorithm 1. Next, we formally state the validity of our testing procedure as well as a power analysis. The proof is given in Appendix B.2.6

Theorem 2.3.6. *Suppose that n_0 increases at a faster rate than*

$$\frac{1}{mp}\max\left\{\begin{matrix}\log^7(qmn_0p),\ \log^4|E|(m+\log(qmn_0p)), \\ d^2\log|E|(m+\log^2(qmn_0p))^2\end{matrix}\right\} \tag{2.41}$$

and

$$\frac{1}{q}(\log^2|E|\log(qmn_0p))^{1+1/(2\alpha_0+1)}. \tag{2.42}$$

Under the null $H_{0,E}$,

$$\mathbb{P}[\mathbf{0}\notin\mathcal{C}_E(1-\alpha)]\xrightarrow{p}\alpha. \tag{2.43}$$

On the other hand, as an alternative, if

$$\|T_E\|_\infty \geq C(\kappa_1,\kappa_3,\kappa_5)\sqrt{(\log q+\log(1/\alpha))\max_{(i,j)\in E}S_{(i,j),(i,j)}}, \tag{2.44}$$

then we have $\mathbb{P}[\mathbf{0}\notin\mathcal{C}_E(1-\alpha)]\xrightarrow{p}1$.

The theorem implies that as long as the sum of the partial correction vector $\rho_{ij}^{(\mathcal{S},l)}$ across l is above the order of $\sqrt{m\log q/(n_0p)}$, the power converges to 1 as $n_0\to\infty$. Assuming the same order $\rho_{ij}^{(\mathcal{S},l)}$ across all graphs $1\leq l\leq m$, i.e., $\rho_{ij}^{(\mathcal{S},1)}\asymp\cdots\asymp\rho_{ij}^{(\mathcal{S},m)}$, the power of the test converges to 1 as $n_0\to\infty$, if $\max_{(i,j)\in E}|\rho_{ij}^{(\mathcal{S},l)}|$ is larger than $\sqrt{\log q/(mn_0p)}$. In contrast, the corresponding detection boundary is $\sqrt{\log q/(n_0p)}$ if we do not aggregate multiple graphs. Therefore, by borrowing the information from multiple graphs/sessions, we are able to reduce the detection accuracy by a factor of root m.

2.3.3 Non-asymptotic error bound for the temporal covariance matrix estimate

Proposition 2.3.7. *Suppose that $h_l=\lfloor(n_lq)^{1/(1+\alpha)}\rfloor$. Then, following the procedure defined in Section 2.2.3,*

$$\mathbb{P}\left[\max_l\max_{1\leq t\leq p}\|\widehat{\beta}_{\cdot,t}^{(\mathcal{T},l)}-\beta_{\cdot,t}^{(\mathcal{T},l)}\|_2\geq C(\kappa_1,\kappa_3,\kappa_5)\sqrt{\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}}}\right]\leq C(qmn_0p)^{-1/2}, \tag{2.45}$$

$$\mathbb{P}\left[\max_{l}\max_{1\le t\le p}\left|\widehat{\Phi}_{tt}^{(\mathcal{T},l)} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Phi_{tt}^{(\mathcal{T},l)}\right| \ge C(\kappa_1,\kappa_3,\kappa_5)\sqrt{\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}}}\right] \le C(qmn_0p)^{-1/2}. \tag{2.46}$$

In the end, we summarize the estimation bounds under the Frobenius norm for individual temporal covariance and precision matrices obtained in Section 2.2.3. Although exiting results for i.i.d. sample are available in Liu and Ren (2020), there is no result for correlated samples as derived in our model. We thus provide a self-contained analysis, which might be of independent interest. The proof is given in Appendix B.2.8.

Theorem 2.3.8. *Suppose that* $h_l = \lfloor (n_l q)^{1/(1+\alpha)} \rfloor$ *and* $\eta = C(\kappa_3)$ *satisfies* $\eta \le \lambda_1(I - \beta^{(\mathcal{T},l)})$ *for* $l = 1,\dots,m$, *where* $\lambda_1(I - \beta^{(\mathcal{T},l)})$ *is the smallest eigenvalue of* $I - \beta^{(\mathcal{T},l)}$. *Then,*

$$\mathbb{P}\left[\begin{array}{l} \displaystyle\max_{l}\frac{1}{p}\left\|\overline{\Sigma}^{(\mathcal{T},l)} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Sigma^{(\mathcal{T},l)}\right\|_F^2 \ge C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}}, \\ \displaystyle\max_{l}\frac{1}{p}\left\|\overline{\Omega}^{(\mathcal{T},l)} - \frac{q}{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}\Omega^{(\mathcal{T},l)}\right\|_F^2 \ge C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}} \end{array}\right] \le C(qmn_0p)^{-1/2}. \tag{2.47}$$

Consequently, their scaled Frobenius norms can be consistently estimated, which is sufficient for our main result Theorem 2.3.5. The proof is given in Appendix B.2.9.

Corollary 2.3.9. *Suppose that* $h_l = \lfloor (n_l q)^{1/(1+\alpha)} \rfloor$ *and* $\eta = C(\kappa_3)$ *satisfies* $\eta \le \lambda_1(I - \beta^{(\mathcal{T},l)})$ *for* $l = 1,\dots,m$. *Then,*

$$\mathbb{P}\left[\max_{l}\left|\frac{\|\widehat{\Sigma}^{(\mathcal{T},l)}\|_F^2 - \|\Sigma^{(\mathcal{T},l)}\|_F^2}{p}\right| \ge C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}},\right] \le C(qmn_0p)^{-1/2} \tag{2.48}$$

for sufficiently large n_0.

2.4 Numerical Studies

2.4.1 Simulation Studies

We study the performance of our method under three spatial precision matrix structures, which are shown in Figure 2.1: (1) an random graph, where the edges between each nodes are randomly generated, with probability $\sqrt{\frac{3}{q}}$ of having an edge between i and j; (2) a hub graph, where the nodes are divided into $\lceil \frac{q}{20} \rceil$ hub groups; and (3) a chain graph, which is a special case of banded graphs with bandwidth equal to 1. Once the common graph structure is fixed, we assign the precision value for the non-zero edges randomly from Unif$(0, \frac{0.3}{2^{l-1}})$ for each session l. For the temporal precision matrix, we generate $\Sigma^{(\mathcal{T},l)}$ by Eq. (2.20). Following Assumption 5, we set $\beta_{st}^{(\mathcal{T},l)} = \kappa_5(t-s)^{-\alpha_l-1}$ for $1 \le s < t \le p$ with $M = 0.2$ and $\alpha_l = 1$. For $\Phi^{(\mathcal{T},l)}$, we simply use the $p \times p$ identity matrix, for $1 \le t \le d$.

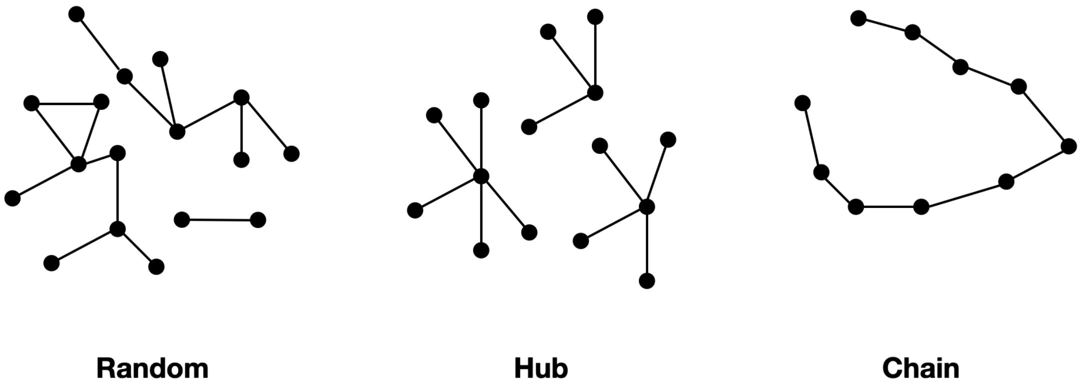


Figure 2.1: Simulated spatial graphs.

In Section 2.4.1, we compare the performance of our method with existing others in edge detection and precision entry estimation. In turn, in Section 2.4.1, we demonstrate the accuracy of our proposed bootstrap procedure in approximating the $\|T_E\|_\infty$ as theoretically shown in Theorem 2.3.5.

Edge-wise Estimation Comparison

We compare our method (M0) with the following Gaussian Graph estimation methods:

- (M1): matrix-variate Gaussian multi-graph estimation method by Zhu and Li (2018)
- (M2): regression based Gaussian graphical model estimation method by Ren et al. (2019)
- (M3 & M4): optimization based Gaussian graphical model estimation methods Cai et al. (2016a) and Lee and Liu (2015), respectively

Methods (M0) and (M1) are based on the matrix-variate Gaussian model, while the others use the vector-variate model which does not incorporate temporal correlation. Before applying the methods based on vector-variate model, we pre-process the data with whitening over temporal dimension and treat signal at each time point as i.i.d sample.

The edge detection performances of the methods are evaluated by receiver operating characteristic (ROC) curves. An ROC curve is a plot of the true positive rate (TPR) against the false positive rate (FPR) while detection threshold varies by changing hyperparameters. For (M0) and (M2), we fix the sparsity penalty hyperparameter at $\lambda \in \{1e-2, 1e-4, 1e-6\}$ and obtain the ROC curves across different values of α. For the other methods, we obtain the ROC curves across different values of the sparsity penalty hyperparameter while fixing the other hyperparameters. The results with $\lambda = 1e-4$, $n = 5$, $q = 30$, and $m = 5$ are shown in Fig. 2.2. The ROC curves show that our method recovers the underlying graph structure accurately and outperforms the other methods. Moreover, comparing methods designing for ordinary Gaussian graph, our method is much better when temporal dimension p is large, thanks to the efficient use of spatial observations and temporal precision estimation based on Cholesky decomposition in Section 2.2.3. Our method shows moderate sensitivity to the choice of the group lasso hyper-parameter; it performs uniformly better than the baseline methods, compared in Section 2.4, across the studied range of tuning parameters, especially under high temporal and spatial dimensions. See Fig. B.1 for the ROC curve results at different values of the tuning parameter.

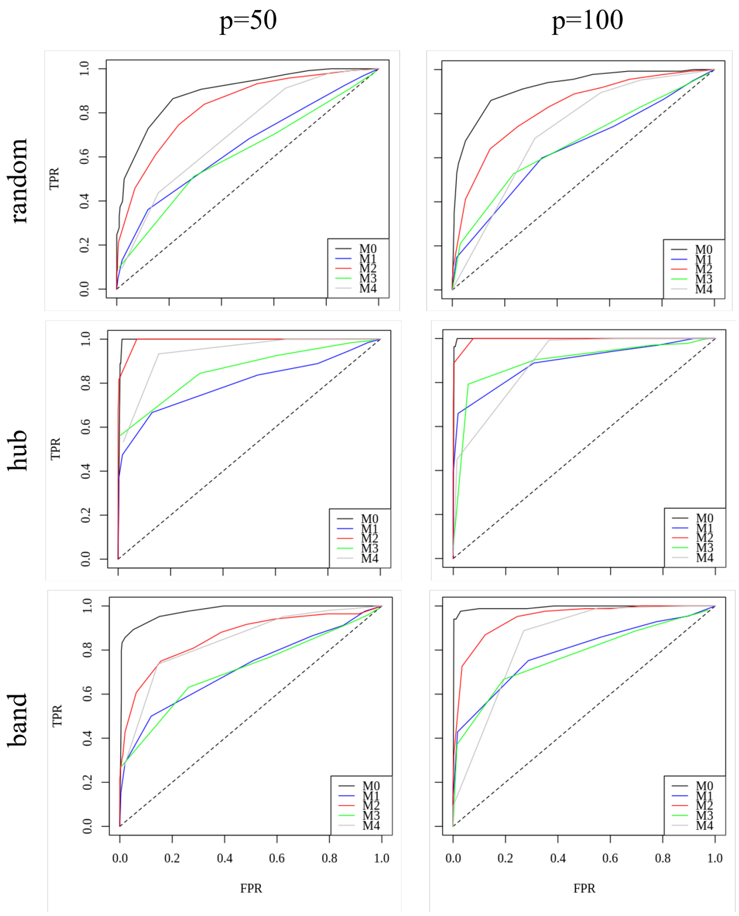


Figure 2.2: Simulation results under different graph configurations and temporal dimensions. We fix $m = 5$, $n = 5$, and $q = 30$. Rows change with types of graphs, and columns correspond to different temporal dimensions. Blue curve corresponds to our method (M0) while other colors correspond to baseline methods. Our method is consistently better than baselines while our advantage is very obvious for large p, thanks to our temporal covariance estimation procedure.

Simultaneous Test

In this section, we evaluate the performance of the proposed bootstrap test in Eq. (2.18) and verify the correctness of Theorem 2.3.5 by simulated data. Under fixed temporal dimension $p = 50$ and spatial dimension $q = 30$, we examined the accuracy of estimated quantiles of $\|Z\|_\infty$ in Eq. (2.18) from 3000 bootstrap samples, under two edge sets: $E_{\text{off}} = \{(i, j) : i \neq j\}$ and $E_{\text{zero}} = \{(i, j) : \Omega_{ij}^{(\mathcal{S},l)} = 0, \, \forall l = 1, \ldots, m\}$. For the groundtruth, we estimated the corresponding empirical quantiles of $\|\widehat{T}_E - T_E\|_\infty$ from 1000 simulated datasets, following the matrix-variate

Gaussian distribution with Kronecker product covariance and the three previously studied spatial partial-correlation graph (random, hub, and chain). Tables 2.1 and 2.2 show the mean and standard deviation of the empirical coverage by 100 bootstrap quantiles $\widehat{q}_{\|\widehat{Z}\|_\infty, 1-\alpha}$ over simulated $\|\widehat{T}_E - T_E\|_\infty$. We observe that the empirical coverages are near to the nominal values at $n = 5$ and converge further as the number of sample n increases. This result demonstrates the asymptotic validity of the bootstrap confidence interval, which has theoretical foundation on Theorem 2.3.5. Comparing Table 2.1 and Table 2.2, we observe that our method performs better for smaller number of sessions, which agrees with our theory.

n	Quantile	Random		Hub		Band	
		E_{off}	E_{zero}	E_{off}	E_{zero}	E_{off}	E_{zero}
5	0.925	0.897(0.012)	0.898(0.009)	0.908(0.010)	0.908(0.009)	0.903(0.009)	0.907(0.010)
	0.95	0.935(0.006)	0.932(0.010)	0.939(0.007)	0.939(0.007)	0.934(0.008)	0.939(0.007)
	0.975	0.962(0.006)	0.969(0.002)	0.970(0.004)	0.971(0.004)	0.971(0.004)	0.971(0.003)
10	0.925	0.926(0.005)	0.924(0.006)	0.923(0.005)	0.923(0.006)	0.929(0.005)	0.927(0.006)
	0.95	0.944(0.003)	0.949(0.004)	0.945(0.004)	0.945(0.005)	0.953(0.005)	0.950(0.005)
	0.975	0.967(0.004)	0.970(0.002)	0.978(0.003)	0.977(0.005)	0.980(0.003)	0.977(0.003)
20	0.925	0.926(0.005)	0.921(0.004)	0.926(0.005)	0.926(0.006)	0.928(0.004)	0.925(0.005)
	0.95	0.954(0.005)	0.944(0.003)	0.948(0.004)	0.948(0.004)	0.951(0.005)	0.943(0.005)
	0.975	0.978(0.001)	0.975(0.004)	0.977(0.005)	0.975(0.004)	0.978(0.003)	0.975(0.003)

Table 2.1: Average of empirical coverages and their standard deviations for $m = 3, p = 50, q = 30$.

n	Quantile	Random		Hub		Band	
		E_{off}	E_{zero}	E_{off}	E_{zero}	E_{off}	E_{zero}
5	0.925	0.906(0.009)	0.900(0.009)	0.901(0.006)	0.904(0.006)	0.886(0.008)	0.888(0.007)
	0.95	0.935(0.006)	0.938(0.006)	0.932(0.006)	0.934(0.007)	0.922(0.007)	0.923(0.008)
	0.975	0.971(0.004)	0.962(0.005)	0.962(0.004)	0.963(0.004)	0.959(0.005)	0.959(0.005)
10	0.925	0.908(0.006)	0.913(0.005)	0.931(0.005)	0.928(0.006)	0.931(0.006)	0.928(0.007)
	0.95	0.934(0.004)	0.937(0.004)	0.951(0.003)	0.950(0.003)	0.953(0.004)	0.953(0.005)
	0.975	0.961(0.004)	0.961(0.003)	0.971(0.003)	0.971(0.003)	0.974(0.003)	0.975(0.003)
20	0.925	0.917(0.007)	0.920(0.004)	0.934(0.004)	0.930(0.006)	0.931(0.006)	0.928(0.007)
	0.95	0.947(0.006)	0.948(0.005)	0.959(0.004)	0.955(0.004)	0.953(0.004)	0.953(0.005)
	0.975	0.978(0.002)	0.975(0.004)	0.985(0.002)	0.984(0.002)	0.974(0.003)	0.975(0.002)

Table 2.2: Average of empirical coverages and their standard deviations for $m = 5, p = 50, q = 30$.

2.4.2 Experimental Data Analysis

In this paper, we analyze multiple LFP recordings from prefrontal cortex (PFC) and visual area V4. PFC has been considered to mediate control of attention and response modulation in higher-order visual areas, such as V4. The Smith Lab in Carnegie Mellon Neuroscience Institute collected the dataset from a monkey performing a memory-guided saccade task (Johnston et al., 2020). One trial of the task consists of the following timeline:

- The animal fixated on a point at the center of the screen for 200ms.
- A circular target appeared at one of eight randomly chose locations of the screen for 50ms.

- The animal had to remember the location of the target while maintaining fixation for a delay period of 500ms.
- After the delay period, the fixation point was turned off, and the monkey had to make a saccade to remember the location of the target.

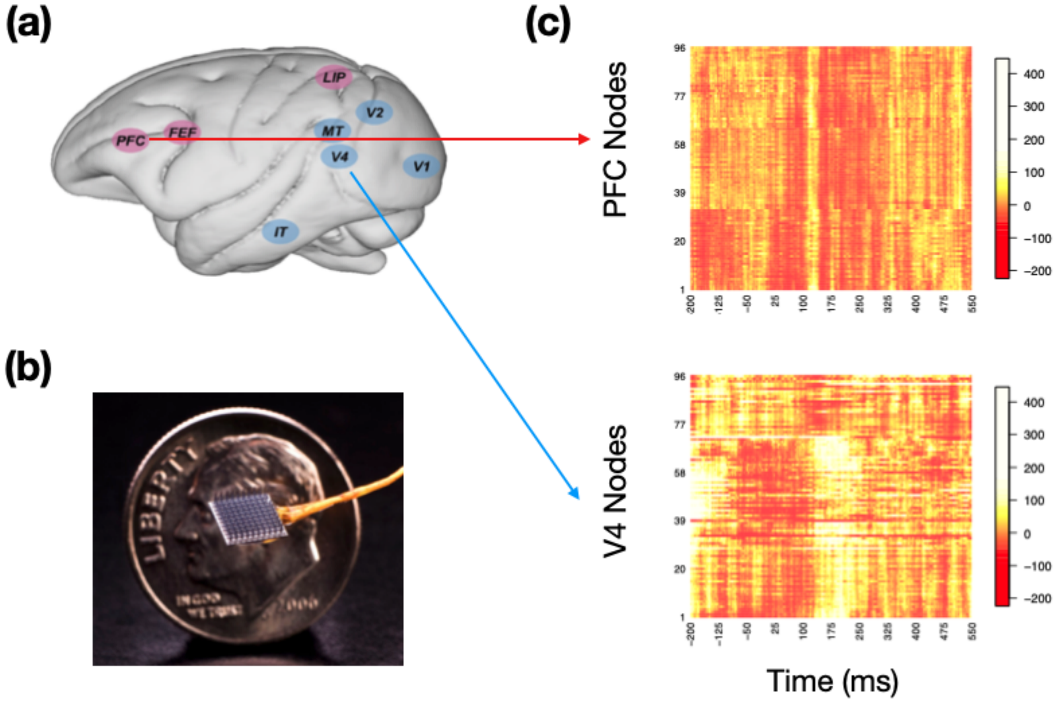


Figure 2.3: (a) Primate cortical areas of the attention network (Sapountzis and Gregoriou, 2017). Pink areas are the approximate locations of attention control areas while blue areas are the approximate locations of visual areas. (b) Utah array with 10x10 recording electrodes with $400\mu m$ interval (McKee, Matthew, 2019). For each region, one Utah array with 96 electrodes is implanted. (c) LFPs were recorded simultaneously from V4 and PFC. The X axis is time in millisecond, while Y is the electrode. Time is aligned at $t = 0$ for each trial when the circular target just appeared. The total length is 750ms covering all experimental stages.

Local field potentials in the two brain regions were simultaneously measured by two Utah arrays, each of which consists of 96 electrodes, while the same monkey was performing the task over $m = 5$ sessions. During data preprocessing, only successful trials were kept in the dataset. For each session $l = 1, \ldots, m$, we observe 1000 matrix-variate observations (n_l) with 192 spatial channels (q) and 750 time points (q) at sampling rate 1kHz. From each session, we leave 100 trials to cross-validate the group lasso penalty tuning parameter, λ. A subset of the data containing the first trial is shown in Fig. 2.3. The color in the heatmap corresponds to the intensity of LFP signal.

Our objective is to detect changes in the spatial correlation structure within and between the brain regions across four experimental stages: fixation stage (200ms), cue stage (50ms), early delay stage (the first 250ms of delay stage) and late delay stage (the last 250ms of delay stage).

Correlated neural connectivity vs. Physical distance

First, we apply our method in PFC and V4 separately. We set λ_i to be the same value λ and use the cross-validation method to determine the group Lasso tuning parameter λ.

Using the spatial location of the electrodes, we can infer the relationship between neural connectivity and physical distance. Because the magnitude of the test statistic T represents the connectivity intensity significance of the connectivity is encoded in our test statistic, we look into the average of the test statistic estimates $\widehat{T}$ on groups of physically equidistant edges. Fig. 2.4 demonstrates the monotonic decreasing relationship of the average test statistic with the physical distance at late delay period in V4. We can see the same trends in the other experimental stages and PFC, which echo with previous studies (Goris et al., 2014; Vinci et al., 2018) about the strong dependency of correlated neural activity on the physical distance. This also serves as a sanity check about the implication of our test statistic for the strength of connectivity.

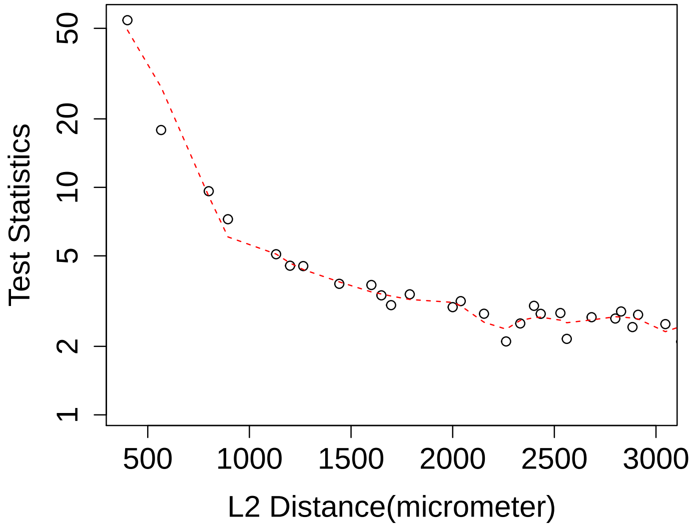


Figure 2.4: The average test statistic vs physical distance during late delay period in V4. Notice that the test statistic declines as the physical distance increase. This phenomenon is consistently identified over all experimental stages, both in PFC and V4

Within-area Inference

Next, for each area and experimental stage, we apply our method on the corresponding data segments. For each electrode i, we evaluate its overall connectivity within the same region by $\frac{1}{d}\sum_{j:(i,j)\in E_*}|\sum_{t=1}^{d}\rho_{tij}|$, where E_* is the set of significant edges at level $\alpha = 0.05$ under single edge test (Eq. (2.10)). Fig. 2.5 shows the distributions of within-areal connectivity across area and experimental stages. It suggests that the within-areal connectivity are strongest during fixation and cue stage, while it declines during delay stage. We also observe that PFC exhibits more within-areal connectivity than V4. In particular, the connectivity within V4 winds down at the delay stage, compared to PFC, when the animal needs to process the visual signals.

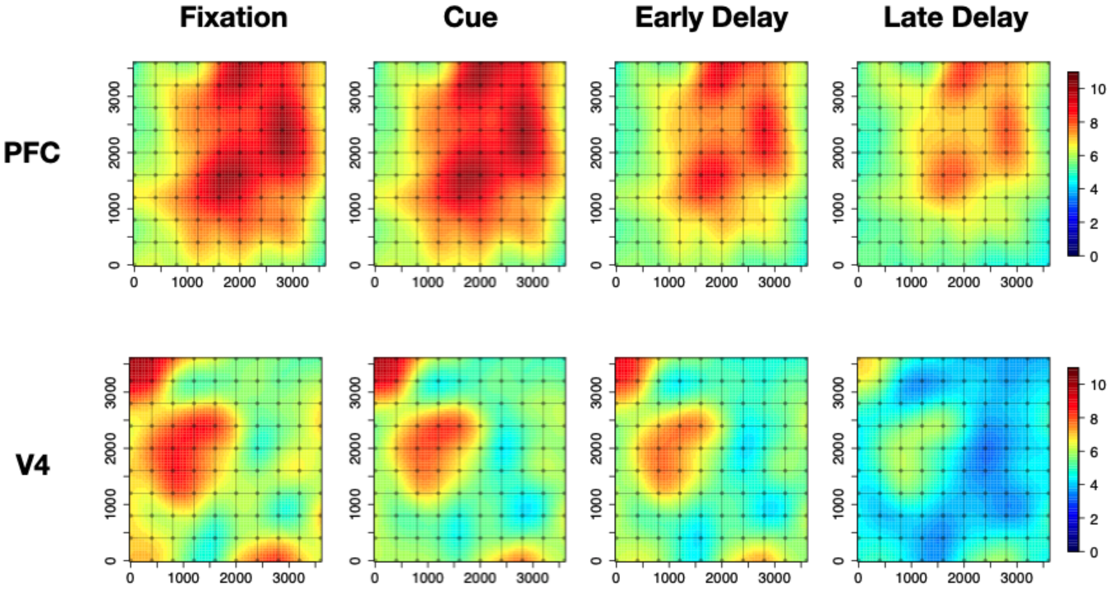


Figure 2.5: Connectivity strength distribution over 2D spatial array for PFC and V4 across the experimental stages. The connectivity in both area decays during the late delay stage, while V4 seems to be more influenced than PFC.

Cross-area Inference

Now we apply our method to the recordings from the both area and identify significant cross-area connectivity. For limited computation resource and collinearity, we subsample the electrodes in each area by taking every other node along the physical dimension, reducing the spatial dimension from 192 to 50 in total. It reduces the number of cross-area edges to 625 cross-area edges. Fig. 2.6 shows the significant cross-area edges at $\alpha = 0.05$. We identify the least number of edges during the cue stage, whereas the number increases during the early delay stage. The late delay stage, when the animal is about to make a choice, exhibited the most identified cross-area edges.

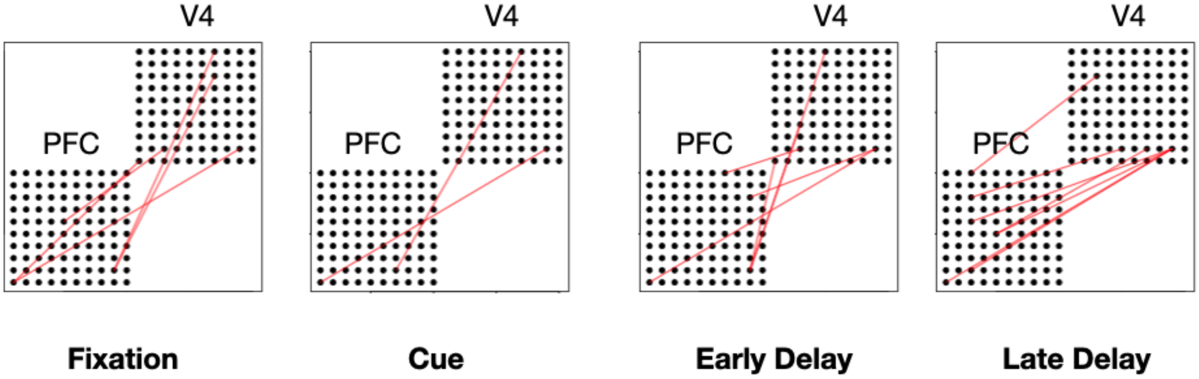


Figure 2.6: Significant cross-region edges for PFC and V4 over various experimental stages. X and Y axes are spatial coordinates of the electrodes on each array. Lower left pane shows electrodes in PFC, while upper right shows electrodes in V4. We only show significant cross-area edges in red color.

We analyzed overall cross-region connectivity at the four exerimental stages, using the multiple edge test statistic $T_{E_{\text{cross}}}$, where E_{cross} consists of cross-area edges. We identified that the two area are most strongly connected during late delay period, and there is few connectivity between two areas during cue period.

Our inference results support the previous studies that neural variability in the spiking of neurons declines during the stimulus onset (Churchland et al., 2010), and visual stimuli causes a substantial decrease in correlation of cortical neurons (Smith and Kohn, 2008). We also discovered robust sustained wihtin-area connectivity in PFC during the delay stage, compared to V4, which was also reported by Leavitt et al. (2017).

2.5 Conclusion

In this paper, we propose a linear-functional based test using partial correlation estimator to detect sparse edges and infer existence and strength of connectivity between two groups of nodes in multiple matrix-variate Gaussian Graphical Models. The spatial dimension, temporal dimension and number of graphs are allowed to diverge and even exceed the number of samples.

Both our model and our assumptions are driven by the practical concerns in neural data analysis. In real data, we observe the within-area connectivity and cross-area connectivity changes accordingly, as the animal entered different experimental stages. Especially, within-area connectivity peaks during early experimental stages, while cross-area connectivity grows when the animal processes the visual signal during late delay stage. Our inference results are illuminating for scientists to understand the activity and connectivity of PFC and V4 during visual tasks.

Our method is the first attempt to address the simultaneous test problem in multiple matrix-variate Gaussian graphs. It would be interesting to extend our method to other popular yet non-Gaussian type of graphs such as Poisson networks. Besides, we currently implemented group Lasso for our regression model which involves one tuning parameter; in the future, a tuning-free or scale-free method such as self-tuned Dantzig selector and scaled Lasso is desirable to handle the issue of heterogeneity and correlation in regression with data from multiple matrix-variate Gaussian graphical models. These directions are beyond the scope of this work and will be interesting for future directions.

Chapter 3

Cross-population Analysis of High-dimensional Non-stationary Neural Oscillations

This chapter is a collaboration with Valérie Ventura, Eric Yttri, Matthew A. Smith, and Robert E. Kass. This work was submitted to Annals of Applied Statistics and is currently under revision for resubmission. The simulation study and experimental data result is subject to change in the final journal submission.

3.1 Introduction

Contemporary technologies for recording neural activity can produce multiple time series in each of two or more brain regions (e.g., Jun et al., 2017; Steinmetz et al., 2018), enabling identification of interactions across regions that are relevant to behavior. In some situations, it may be possible to find lead-lag relationships, which could indicate directional flow of information. For stationary time series, this is the problem solved by Granger causality (Geweke, 1982). When recordings are made from an alert animal during a behavioral task, however, the neural activity is constantly evolving, and the statistical challenge is to find a method that allows for non-stationarity.

Various non-stationary methods have been applied to discovering simultaneous associations within brain regions (Buesing et al., 2014; Zhao and Park, 2017) or between regions (Gallagher et al., 2017), but their ability to find lead-lag relationships has not yet been established. The DKCCA method of Rodu et al. (2018) uses a dynammic kernel canonical correlation analysis to establish cross-region interaction, and it can estimate non-stationary lead-lag relationships, but it is unable to assess statistical significance of those relationships. In this paper we define and develop a framework that is able to detect lead-lag relationships across two brain regions, prefrontal cortex (PFC) and visual area V4, during a visual memory task, based on multiple time series of local field potential recordings (LFPs).

Visual cortical area V4 has been reported to retain higher order information (e.g. color and shape) and attention to visual objects (Orban, 2008; Fries et al., 2001), while prefrontal cortex (PFC) is considered to exert cognitive control in working memory (Miller and Cohen, 2001).

Despite their spatial distance and functional differentiation, these regions have been presumed to cooperate during visual working memory tasks. For example, Sarnthein et al. (1998) discovered enhanced oscillatory activity in both regions during visual memory retention and Liebe et al. (2012) found delay-specific oscillatory coupling between V4 and lateral PFC. The nature and exact timing of this interaction, however, have remained unknown. One leading possibility is that PFC exerts control using oscillations in the beta range (16-30 Hz) that are coupled across brain regions (Miller et al., 2018). The previous work of Klein et al. (2020) developed methods for analyzing multivariate phase coupling of such oscillations. Here, we turn our attention to coupling of the oscillatory power, while the subject was remembering the location of a visual target, with the goal being to identify lead-lag relationships in beta-band power between PFC and V4, observed in LFPs recorded simultaneously from the two regions.

LFPs result from neural currents generated near the electrode (roughly within 150-200 microns), involving large numbers of neurons (Buzsáki et al., 2012; Einevoll et al., 2013; Pesaran et al., 2018), and they have been shown to correlate substantially with the BOLD fMRI signal (Logothetis et al., 2001; Magri et al., 2012a). The recordings in this study come from a pair of 96-electrode arrays implanted in PFC and V4. The data were sampled every millisecond across a 500 millisecond time interval, with 3000 replications (3000 experimental trials). Each of the 3000 observations forms a 192-dimensional time series, in two groups of 96-dimensional time series. The method we developed, Latent Dynamic analysis via Sparse banded graphs (LaDynS), was able to identify non-stationary cross-area interactions from the PFC and V4 data. Our data-analytic findings are summarized in Figs. 3.8 to 3.11, and presented in full in Section 3.4.

According to the LaDynS model, each of the two multivariate time series is driven by a latent univariate time series, with the resulting latent bivariate time series potentially representing the time-varying cross-area dependence we seek to identify. Instead of assuming a parametric structure for the cross-dependence of the latent bivariate time series, however, we have left the correlation matrix unspecified. As we show in Section 3.2.2, assuming each multivariate observation, at each time point, depends linearly on a latent normal random variable, having mean 0 and variance 1, we then get a time series generalization of probabilistic CCA (pCCA; Bach and Jordan (2005)), as depicted graphically in Fig. 3.1. Furthermore, maximum likelihood applied to the model gives the same data summary as multiset CCA (Kettenring, 1971). This provides a nice interpretation, but it isn't yet practical because the latent time series correlation matrix, which must be estimated, has $2T(2T-1)/2$ parameters, where T is the number of time points. We therefore resorted to sparse estimation methods.

As we describe in Section 2, LaDynS uses L1 regularization on the cross-precision parameters within a band defined by a maximal lead or lag of the partial cross-correlations. There is also a maximal lead or lag of the partial auto-correlations, but the remaining partial auto-correlations (at less than the maximum lead or lag) are not regularized. Because the data we analyzed were band-pass filtered, the time series were unusually smooth. This led us to consider and apply additional regularization along the diagonals of the auto-correlation matrices. Statistical inference in this context can be based on false discovery rate (FDR) based on p-values computed from asymptotics for a desparsified version of the estimated precision matrix (Jankova and Van De Geer, 2015). Simulations in Section 3.3 show that LaDynS is able to correctly identify the timing, relative to behavior, of interactions between regions, as illustrated in Fig. 3.2b, when applied to artificial data designed to be similar to those we analyzed. The simulations make credible our

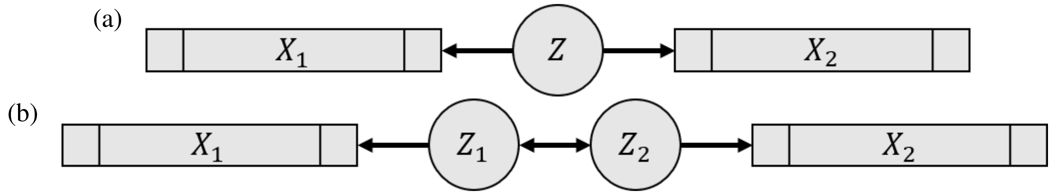


Figure 3.1: **(a)** *Graphical representation of the pCCA model of Bach and Jordan (2005), where* X_1 *and* X_2 *are random vectors and* Z *is a random variable.* **(b)** *A variation on (a) that facilitates extension to the case in which* X_1 *and* X_2 *are multivariate time series and* (Z_1, Z_2) *is a bivariate time series.*

data-analytic results in Section 3.4. We add some discussion in Section 3.5.

3.2 Methods

We begin by reviewing and reformulating probabilistic CCA (pCCA) in Section 2.1, then generalize to time series, in Section 2.2, obtaining to form a dynamic version of pCCA. Theorem 2.2 establishes an equivalence between the GENVAR version of multi-set CCA (Kettenring, 1971) and maximum likelihood applied to our dynamic pCCA. We define LaDynS based on the loglikelihood function in Equation (16) of Section 2.3. We go over choice of regularization parameters in Section 2.3.1, application to smooth time series in Section 2.3.2, and our algorithm for solving the penalized maximum likelihood problem in Section 2.3.3. We discuss statistical inference in Section 2.4.

3.2.1 Probabilistic CCA for two random vectors

Given two random vectors $X_1 \in \mathbb{R}^{d_1}$ and $X_2 \in \mathbb{R}^{d_2}$, canonical correlation analysis (CCA) (Hotelling, 1992) finds the sets of weights $w_1 \in \mathbb{R}^{d_1}$ and $w_2 \in \mathbb{R}^{d_2}$ that maximize Pearson's correlation between linear combinations $w_1^\top X_1$ and $w_2^\top X_2$. This can be rewritten as

$$\sigma_{cc} = \max_{w_k, k=1,2: w_k^\top \Sigma_{kk} w_k = 1} \left| w_1^\top \Sigma_{12} w_2 \right| \tag{3.1}$$

where $\Sigma_{kk} = \mathrm{Var}(X_k)$ is the covariance matrix of X_k, $k = 1, 2$, and $\Sigma_{12} = \mathrm{Cov}(X_1, X_2)$ the cross-covariance matrix between X_1 and X_2. The sample estimator $\widehat{\sigma}_{cc}$ is obtained by replacing Σ_{kk} and Σ_{12} with their sample analogs $\overline{\Sigma}_{kk}$ and $\overline{\Sigma}_{12}$ respectively. The maximizing weights $\widehat{w}_k$ and linear combinations $\widehat{w}_k^\top X_k$ are referred to as the canonical weights and canonical variables, respectively.

Probabilistic CCA assumes that X_1 and X_2 are driven by a common one dimensional latent variable Z:

$$\begin{aligned} X_k | Z &= \mu_k + Z \cdot \beta_k + \epsilon_k, \; k = 1, 2, \\ Z &\sim N(0, 1) \end{aligned} \tag{3.2}$$

where $\mu_k \in \mathbb{R}^{d_k}$ and $\beta_k \in \mathbb{R}^{d_k}$ are mean vectors and factor loadings, respectively, and $\epsilon_k \overset{\text{indep}}{\sim} \mathrm{MVN}(0, \Phi_k)$ (Bach and Jordan, 2005). Fig. 3.1a depicts the dependence of X_1 and X_2 on Z. The parameters in Eqs. (3.1) and (3.2) relate as follows:

Theorem 3.2.1 (Bach and Jordan, 2005, Theorem 2)**.** *The maximum likelihood estimators (MLEs) $(\widehat{\beta}_1, \widehat{\beta}_2)$ in Eq.* (3.2) *based on N observed vector pairs $\left\{X_{1,[n]}, X_{2,[n]}\right\}_{n=1,2,\ldots,N}$ are equivalent to the CCA solution $(\widehat{w}_1, \widehat{w}_2, \widehat{\sigma}_{cc})$ in Eq.* (3.1) *according to:*

$$\widehat{\beta}_k = \overline{\Sigma}_{kk}\widehat{w}_k m_k, \text{ where } m_1 m_2 = \widehat{\sigma}_{cc} \text{ and } |m_k| \leq 1,\ k = 1,2. \tag{3.3}$$

Theorem 3.2.1 proves that the original CCA setting and the generative pCCA model both yield the same estimate of σ_{cc}. Here, we introduce an alternative pCCA extension that allows distinct latent variables for X_1 and X_2, as depicted in Fig. 3.1b. Specifically, we assume that

$$X_k | Z_k = \mu_k + Z_k \cdot \beta_k + \epsilon_k \tag{3.4}$$

where $\mu_k \in \mathbb{R}^{d_k}$, $\beta_k \in \mathbb{R}^{d_k}$ and $\epsilon_k \overset{\text{indep}}{\sim} \mathrm{MVN}(0, \Phi_k)$ are defined as in Eq. (3.2), and (Z_1, Z_2) are bivariate normally distributed:

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathrm{MVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix} \right). \tag{3.5}$$

We now state an equivalence similar to Theorem 3.2.1 between the original CCA and the alternative pCCA model.

Theorem 3.2.2. *The MLEs $(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}_{12})$ in Eqs.* (3.4) *and* (3.5) *based on N observed vector pairs $\left\{X_{1,[n]}, X_{2,[n]}\right\}_{n=1,2,\ldots,N}$ are equivalent to the CCA solution $(\widehat{w}_1, \widehat{w}_2, \widehat{\sigma}_{cc})$ according to:*

$$\widehat{\beta}_k = \overline{\Sigma}_{kk}\widehat{w}_k m_k, \text{ where } m_1 m_2 \widehat{\sigma}_{12} = \widehat{\sigma}_{cc} \text{ and } |m_k| \leq 1,\ k = 1,2. \tag{3.6}$$

Theorem 3.2.2 is a corollary of Theorem 3.2.3, below. In practice, out of all possible solutions, we take $m_1 = m_2 = 1$ because then $Z_k | X_k = \widehat{w}_k^\top X_k$ is the canonical variable almost surely, as proven in Theorem 3.2.3, and $\sigma_{12} = \mathrm{Cov}[Z_1, Z_2]$ equals the canonical correlation σ_{cc}. This means that σ_{12} is an interpretable parameter, and one for which inference is simpler than for the canonical correlation in model (3.2), because in that case the MLE $\widehat{\sigma}_{cc}$ is an indirect function of the model parameters (see Theorem 3.2.1). The interpretability property also carries on when we extend model (3.5) to model (3.8) below, to capture lagged association between two vector time series. Finally, the choice $m_1 = m_2 = 1$ also implies that the MLEs $(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}_{12})$ do not depend on the Gaussian assumption in Eq. (3.4) (see Theorem 3.2.3), which may be questionable if, for example, the X's are positive variables like LFP power envelopes or discrete like spike counts.

3.2.2 Probabilistic CCA for two time series of random vectors

Suppose now that we are interested in the correlation dynamics between two times series of random vectors $X_1^{(t)} \in \mathbb{R}^{d_1}$ and $X_2^{(t)} \in \mathbb{R}^{d_2}$, $t = 1, 2, \ldots, T$. We use Eq. (3.4) to model the dependence of $X_k^{(t)}$ on its associated latent variable $Z_k^{(t)}$ at each time t:

$$X_k^{(t)} | Z_k^{(t)} = \mu_k^{(t)} + \beta_k^{(t)} \cdot Z_k^{(t)} + \epsilon_k^{(t)},\ k = 1,2, \tag{3.7}$$

where $\mu_k^{(t)}$, $\beta_k^{(t)}$ and $\epsilon_k^{(t)} \overset{\text{indep}}{\sim} \text{MVN}(0, \Phi_k^{(t)})$ are defined as in Eq. (3.4). Then for each t we could define a parameter $\sigma_{12}^{(t)}$ as in Eq. (3.5) to capture population-level association between $X_1^{(t)}$ and $X_2^{(t)}$ at t. But because we are also interested in lagged associations between $X_1^{(t)}$ and $X_2^{(s)}$ for $s \neq t$, we replace bivariate model (3.5) for $Z_1^{(t)}$ and $Z_2^{(t)}$ for a given t by a global model for all $t = 1, \ldots, T$:

$$\left(\left(Z_1^{(t)} \right)_{t=1,\ldots,T}, \left(Z_2^{(t)} \right)_{t=1,\ldots,T} \right)^\top \sim \text{MVN}(0, \Sigma), \quad \text{diag}(\Sigma) = \mathbf{1}, \tag{3.8}$$

where Σ captures jointly all simultaneous and lagged associations within and between the two time series. Fig. 3.2a illustrates the dependence structure of this model. We decompose Σ and its inverse Ω as

$$\Sigma = \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{12}^\top & \Sigma_{22} \end{array} \right) \text{ and } \Omega = \left(\begin{array}{c|c} \Omega_{11} & \Omega_{12} \\ \hline \Omega_{12}^\top & \Omega_{22} \end{array} \right) \tag{3.9}$$

to highlight the auto-correlations Σ_{11} and Σ_{22} within and cross-correlations Σ_{12} between the time series, and denote by $\Sigma_{kl}^{(t,s)}$, $(t, s) \in [T]^2$, the elements of Σ_{kl}. Then $\Sigma_{12}^{(t,t)}$ for some fixed t has the same interpretation as σ_{12} in Eq. (3.2). Further, $-\Omega_{12}^{(t,s)} / \sqrt{\Omega_{11}^{(t,t)} \Omega_{22}^{(s,s)}}$ is the partial correlation between the two time series at times t and s. Thus, when an element of Ω_{12} is non-null, depicted as the red star in the expanded display in Fig. 3.2b, its coordinates (t, s) and distance $(t - s)$ from the diagonal indicate at what time in the trial a connectivity happens between two time series, and at what lead or lag, respectively. In our neuroscience application, they represent the timing of connections and direction of information flow between two brain regions.

Theorem 3.2.2 provided an equivalence between a non-distributional method (CCA) and its probabilistic representation (pCCA). We now derive a similar connection between the multi-set generalization of CCA introduced by Kettenring (1971) and the dynamic pCCA model in Eqs. (3.7) and (3.8). Multi-set CCA applied to $2T$ random vectors $\{X_1^{(t)}, X_2^{(t)} : t = 1, \ldots, T\}$ finds weights $\{w_1^{(t)}, w_2^{(t)} : t = 1, \ldots, T\}$ that maximize a notion of correlation among linear combinations $\{w_1^{(t)\top} X_1^{(t)}, w_2^{(t)\top} X_2^{(t)} : t = 1, \ldots, T\}$. In particular, the `GENVAR` extension minimizes the generalized variance of these linear combinations, defined as the determinant of their correlation matrix (Wilks, 1932), which we refer to as the canonical correlation matrix:

$$\widehat{w}_1^{(1)}, \ldots, \widehat{w}_2^{(T)} = \underset{w_1^{(1)},\ldots,w_2^{(T)}}{\text{argmin}} \det \left(\overline{\text{Var}} \left[\left(w_1^{(t)\top} X_1^{(t)} \right)_{t=1,\ldots,T}, \left(w_2^{(t)\top} X_2^{(t)} \right)_{t=1,\ldots,T} \right] \right) \tag{3.10}$$

where $\overline{\text{Var}}$ denotes the sample variance-covariance matrix and the weights $w_k^{(t)}$ are scaled so that every diagonal entry of the matrix is 1.

Theorem 3.2.3. *Suppose that* $\widehat{\beta}_k^{(t)}$, $k = 1, 2$, $t = 1, \ldots, T$, *and* $\widehat{\Sigma}$ *are the MLE in Eqs.* (3.7) *and* (3.8) *satisfying minimum conditional entropy of* $\{X_1^{(t)}, X_2^{(t)}\}$ *given* $\{Z_1^{(t)}, Z_2^{(t)}\}$*, based on* N *observed pairs of vector time series* $\left\{ X_{1[n]}^{(t)}, X_{2[n]}^{(t)} : t = 1, \ldots, T \right\}$, $n = 1, \ldots, N$. *Then, they minimize the negative log-likelihood*

$$\log\det(\Sigma) + \text{tr}\left(\Sigma^{-1} \overline{\Sigma} \right), \tag{3.11}$$

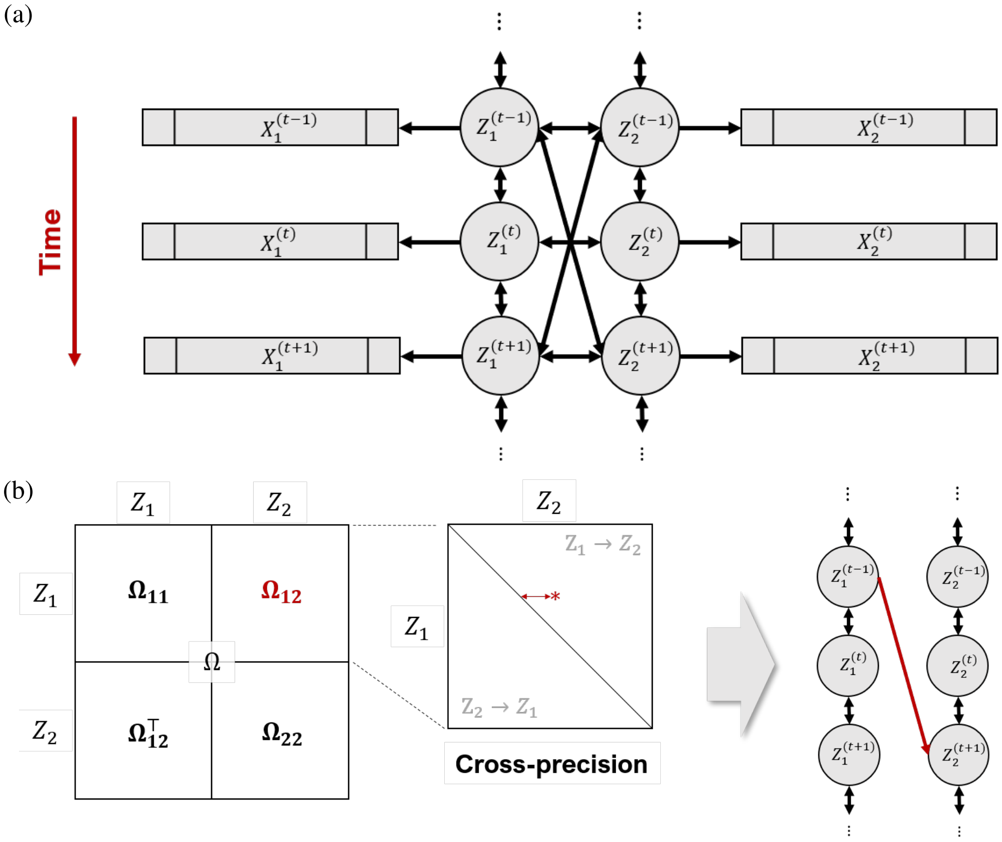

(a)
Time
$X_1^{(t-1)}$
$Z_1^{(t-1)}$
$Z_2^{(t-1)}$
$X_2^{(t-1)}$
$X_1^{(t)}$
$Z_1^{(t)}$
$Z_2^{(t)}$
$X_2^{(t)}$
$X_1^{(t+1)}$
$Z_1^{(t+1)}$
$Z_2^{(t+1)}$
$X_2^{(t+1)}$
(b)
Z_1
Z_2
Z_1
Z_2
Ω_{11}
Ω_{12}
Ω
$\Omega_{12}^\top$
Ω_{22}
Z_2
$Z_1 \to Z_2$
Z_1
$Z_2 \to Z_1$
Cross-precision
$Z_1^{(t-1)}$
$Z_2^{(t-1)}$
$Z_1^{(t)}$
$Z_2^{(t)}$
$Z_1^{(t+1)}$
$Z_2^{(t+1)}$

Figure 3.2: **Extended pCCA model for two multivariate time series $X_1^{(t)}$ and $X_2^{(s)}$, $t, s = 1, \dots, T$.** **(a)** *Dynamic associations between vectors $X_1^{(t)}$ and $X_2^{(s)}$ are summarized by the dynamic associations between their associated 1D latent variables $Z_1^{(t)}$ and $Z_2^{(s)}$, and estimated by their cross-precision matrix Ω_{12}.* **(b)** *When a significant cross-precision entry is identified, e.g., the red star in the expanded view of Ω_{12}, its coordinates and distance from the diagonal indicate at what time in the experiment connectivity between two brain areas occurs, and at what lead or lag. Here the red star is in the upper diagonal of Ω_{12}, which means that, at this particular time, region 1 leads region 2, or $Z_1 \to Z_2$ in short (a non-zero entry in the lower diagonal would mean $Z_2 \to Z_1$). We represent this association by the red arrow on the right-most plot, with a lag of two units of time for illustration.*

where $\overline{\Sigma} = \overline{\mathrm{Var}}\left[\left(\beta_1^{(t)\top}\overline{\mathrm{Var}}^{-1}[X_1^{(t)}]X_1^{(t)}\right)_{t=1,\dots,T}, \left(\beta_2^{(t)\top}\overline{\mathrm{Var}}^{-1}[X_2^{(t)}]X_2^{(t)}\right)_{t=1,\dots,T}\right]$, *with* β_1 *and* β_2 *scaled such that* diag $\overline{\Sigma} = \mathbf{1}$, *and they are equivalent to Kettenring's* GENVAR *multiset solution*

$$\widehat{\beta}_k^{(t)} = \overline{\mathrm{Var}}[X_k^{(t)}]\widehat{w}_k^{(t)} \quad \textit{and} \quad \widehat{\Sigma} = \overline{\mathrm{Var}}\left[\left(\widehat{w}_1^{(t)\top}X_1^{(t)}\right)_{t=1,\dots,T}, \left(\widehat{w}_2^{(t)\top}X_2^{(t)}\right)_{t=1,\dots,T}\right]. \tag{3.12}$$

The proof is in Appendix C.1. We have two remarks. *Remark 1:* Theorem 3.2.3 generalizes Theorem 3.2.2. To see that, let $T = 1$ so that $X_1 \equiv X_1^{(1)}$, $X_2 \equiv X_2^{(1)}$ and $\Sigma =$

$\begin{pmatrix} 1 & \Sigma_{12}^{(1,1)} \\ \Sigma_{12}^{(1,1)\top} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}$. The `GENVAR` procedure solves

$$\underset{w_1,w_2}{\operatorname{argmin}} \det\left(\overline{\mathrm{Var}}\left[w_1^\top X_1, w_2^\top X_2\right]\right) \equiv \underset{w_1,w_2}{\operatorname{argmin}} \det\left(\begin{pmatrix} 1 & \overline{\sigma}_{12} \\ \overline{\sigma}_{12} & 1 \end{pmatrix}\right),$$

where $\overline{\sigma}_{12} = \frac{w_1^T \overline{\Sigma}_{12} w_2}{\sqrt{w_1^T \overline{\Sigma}_1 w_1}\sqrt{w_2^T \overline{\Sigma}_2 w_2}}$. This minimization problem is equivalent to the CCA problem in Eq. (3.1) and $\overline{\mathrm{Var}}\left[\widehat{w}_1^T X_1, \widehat{w}_2^T X_2\right] = \begin{pmatrix} 1 & \widehat{\sigma}_{cc} \\ \widehat{\sigma}_{cc} & 1 \end{pmatrix}$, which implies $\widehat{\sigma}_{12} = \widehat{\sigma}_{cc}$, as in Theorem 3.2.2.

Remark 2: Eq. (3.11) does not involve any component of Eq. (3.7) so the MLEs do not depend on the Gaussian assumption of $X_k^{(t)}$ given $Z_k^{(t)}$. This is confirmed by simulation in Section 3.3. The MLEs do however depend on the Gaussian distribution in Eq. (3.8), but this assumption is more easily justifiable: the latent factors are the canonical variables $\widehat{w}_k^{(t)\top} X_k^{(t)}$, which are weighted sums of the observations and thus likely to be Gaussian based on some central limit theorem.

3.2.3 Latent Dynamic Analysis via Sparse Banded Graphs (LaDynS)

Our goal is to estimate the association dynamics between two multivariate time series using the covariance matrix Σ of their associated latent time series in Eq. (3.11). However, the prohibitive number of parameters in Σ means its estimation is prone to errors, especially when T is large. We reduce their number by regularizing $\Omega = \Sigma^{-1}$ in Eq. (3.11), rewriting $\log\det(\Sigma) = \log\det(\Omega^{-1}) = -\log\det(\Omega)$, and assuming that Ω has the banded structure depicted in Fig. 3.3.

Definition 3.2.4 (LaDynS). Given N simultaneously recorded pairs of multivariate time series $\{X_1[n], X_2[n]\}_{n=1,\dots,N}$, and a $2T \times 2T$ sparsity matrix Λ with element $\Lambda_{kl}^{(t,s)}$ regularizing $|\Omega_{kl}^{(t,s)}|$, $k, l = 1, 2$, LaDynS finds weights $\left\{\widehat{w}_k^{(t)}, t = 1, 2, \dots, T, k = 1, 2\right\}$ and precision matrix $\widehat{\Omega}$ that minimize the penalized negative log-likelihood:

$$-\log\det(\Omega) + \mathrm{tr}(\Omega\overline{\Sigma}) + \|\Lambda \odot \Omega\|_1, \tag{3.13}$$

where $\overline{\Sigma} = \overline{\mathrm{Var}}\left[w_1^{(1)\top} X_1^{(1)}, \dots, w_2^{(T)\top} X_2^{(T)}\right]$ satisfies $\mathrm{diag}(\overline{\Sigma}) = \mathbf{1}$, $\odot$ denotes the Hadamard product operator such that $(A \odot B)_{ij} = A_{ij} \times B_{ij}$, $\|A\|_1 = \sum_{i,j} |A_{ij}|$, and

$$\Lambda_{kl}^{(t,s)} = \begin{cases} \lambda_{\text{cross}}, & k \neq l \text{ and } 0 < |t-s| \leq d_{\text{cross}}, \\ \lambda_{\text{auto}}, & k = l \text{ and } 0 < |t-s| \leq d_{\text{auto}}, \\ \lambda_{\text{diag}}, & t = s, \\ \infty, & \text{otherwise}, \end{cases}$$

which constrains auto-precision and cross-precision elements within a specified range.

In our neuroscience application, in particular, it is reasonable to assume that lead-lag relationships occur with delay less than temporal bandwidth d_{cross}, which can be determined by the maximal transmission time in synaptic connections between two brain regions under study. We

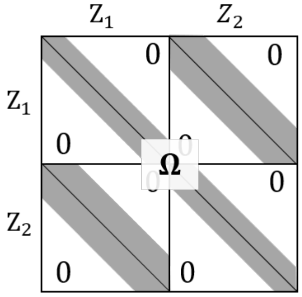


Figure 3.3: *The elements of* Ω_{kk}, $k = 1, 2$, *and* Ω_{12} *are set to zero outside of the gray bands of widths* $(1 + 2d_{\text{auto}})$ *and* $(1 + 2d_{\text{cross}})$, *respectively.*

thus set $\Lambda_{12}^{(t,s)} = \infty$ when $|t - s| > d_{\text{cross}}$ to force the corresponding cross-precision elements to zero and thus impose a banded structure on Ω_{12}. We apply sparsity constraint $\Lambda_{12}^{(t,s)} = \lambda_{\text{cross}} > 0$ on the remaining off-diagonals of Ω_{12} to focus our discovery on sparse dominant associations and reduce the effective parameter size. We proceed similarly with the auto-precision matrices Ω_{11} and Ω_{22}, using penalty λ_{auto} and temporal bandwidth d_{auto}. Unless domain knowledge is available, we recommend that d_{auto} be set to the largest significant auto-correlation across all observed time series $X_{k,i}^{(t)}$, $k = 1, 2$, $i = 1, \ldots, N$, and impose no further sparsity ($\lambda_{\text{auto}} = 0$) unless there is reason to expect it.

Notice that to facilitate the choice of Λ, we grouped its elements into diagonal and off-diagonal elements and assigned the same penalties, λ_{cross}, $\lambda_{\text{auto}} = 0$ and λ_{diag}, within each group.

Choosing regularization parameters

In graphical LASSO (gLASSO) problems, where the aim is to recover correct partial correlation graphs, penalties are often chosen to minimize the predictive risk (Shao, 1993; Zou et al., 2007; Tibshirani and Taylor, 2012). Our aim is different: only the partial cross-precision matrix Ω_{12} is of substantive interest, and because minimizing the predictive risk does not select models consistently (Shao, 1993; Zhu and Cribben, 2018) and may thus fail to retrieve non-zero elements of Ω_{12}, we choose instead a value of λ_{cross} that controls the number of false cross-precision discoveries. We proceed by permuting the observed time series in one brain region to create a synthetic dataset that contains no cross-region correlation, then applying LaDynS to that data for a range of values of λ_{cross} and recording the resulting number of significant partial correlation estimates, which are necessarily spurious. We use the smallest λ_{cross} that yields fewer false discoveries than a chosen threshold. We expect this regularization to make similarly few false discoveries on experimental data.

Finally, if $\widehat{\Sigma}$ cannot be inverted, as is the case for the band-pass filtered experimental data we analyze in Section 3.4, we penalize its diagonal by $\lambda_{\text{diag}} > 0$. We explain the specific calibration we used for the analyzed datasets in Section 3.3.2 and study the properties in Section 3.3.3.

Fitting LaDynS

Eq. (3.13) is not a convex function of the weights and precision elements (although it is not impossible that it may be for some particular Σ) and its convex relaxation is unknown, so it is difficult to find its global minimum. The following coordinate descent algorithm finds a minimum, possibly local.

Assuming that all canonical weights $w_k^{(t)}$ are fixed, Eq. (3.13) reduces to the gLASSO problem:

$$\underset{\Omega}{\operatorname{argmin}} - \log\det(\Omega) + \operatorname{tr}(\Omega\overline{\Sigma}) + \|\Lambda \odot \Omega\|_1, \tag{3.14}$$

which we can solve efficiently using a number of existing algorithms; here we use the P-gLASSO algorithm of Mazumder and Hastie (2012). Then assuming that all parameters are fixed but a single weight $w_k^{(t)}$, Eq. (3.13) can be re-arranged as the linear problem:

$$\underset{w_k^{(t)}}{\operatorname{argmin}} \sum\nolimits_{(l,s)\neq(k,t)} w_k^{(t)\top} \overline{\operatorname{Cov}}\left[X_k^{(t)}, X_l^{(s)}\right] w_l^{(s)} \Omega_{kl}^{(t,s)} \text{ s.t. } w_k^{(t)\top}\overline{\operatorname{Var}}(X_k^{(t)}) w_k^{(t)} = 1, \tag{3.15}$$

for which an analytical solution is available. That is, our algorithm alternates between updating Ω and the weights $w_k^{(t)}$ until the objective function in Eq. (3.13) converges. Its computational cost is inexpensive: a single iteration on our cluster server (with 11 Intel(R) Xeon(R) CPU 2.90GHz processors) took in average less than 0.8 seconds, applied to the experimental data in Section 3.4. A single fit on the same data took 47 iterations for around 33.57 seconds until the objective function converged at threshold 0.001. See Algorithm 2 in Appendix C.2 for details and Python package `ladyns` on `github.com/HeejongBong/ladyns`.

3.2.4 Inference for associations between two vector time series

Let $\widehat{\Omega}$ and $\widehat{w}_k^{(t)}$, $t = 1, \dots, T$, $k = 1, 2$, be the LaDynS estimates of canonical precision matrix and canonical weights, and $\overline{\Sigma} = \overline{\operatorname{Var}}\left[\widehat{w}_1^{(1)\top} X_1^{(1)}, \dots, \widehat{w}_2^{(T)\top} X_2^{(T)}\right]$ be the empirical covariance of the estimated latent variables, defined in Eq. (3.13). Note that $\widehat{\Omega} \neq \overline{\Sigma}^{-1}$ since Ω is constrained to be sparse. Based on these estimates, we want to identify the non-zero partial cross-correlations in Ω_{12}, that is to identify the epochs of association between the two time series.

Formal inference methods for Ω based on its LaDynS estimate (Eq. (3.13)) are not available, but because LaDynS reduces to graphical LASSO (gLASSO) when the weights $w_k^{(t)}$ in Eq. (3.14) are fixed, we co-opt gLASSO inference methods. Specifically, Jankova and Van De Geer (2015) suggested de-sparsifying the gLASSO estimate $\widehat{\Omega}$ according to

$$\widetilde{\Omega} = 2\widehat{\Omega} - \widehat{\Omega}(\overline{\Sigma} + \lambda_{\text{diag}} I_T)\widehat{\Omega}, \tag{3.16}$$

and proved that, under mild assumptions and as $n \to \infty$, each entry of $\widetilde{\Omega}$ satisfies the Central Limit Theorem with center the true precision Ω:

$$\forall (t,s), \quad \frac{\left(\widetilde{\Omega}_{12}^{(t,s)} - \Omega_{12}^{(t,s)}\right)}{\sqrt{\operatorname{Var}[\widetilde{\Omega}_{12}^{(t,s)}]/N}} \xrightarrow{d} \mathrm{N}(0,1). \tag{3.17}$$

We applied this result to the de-sparsified LaDynS estimate of Ω, even though we do not quite have a gLASSO setup, and we verified by simulation that its elements are indeed approximately normal in Section 3.3.3. Jankova and Van De Geer (2015) also proposed an estimator of $\sigma_{12}^{(t,s)}$, but it is likely to be downward biased in our framework since estimating the canonical weights $w_k^{(t)}$ induces extra randomness. Instead, we use the bootstrap estimate $\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]$ described at the end of this section, and rely on Eq. (3.17) to obtain p-values:

$$p^{(t,s)} = 2 - 2\Phi\left(|\widetilde{\Omega}_{12}^{(t,s)}| \Big/ \sqrt{\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]/N}\right) \tag{3.18}$$

to test $H_0^{(t,s)} : \Omega_{12}^{(t,s)} = 0$, for each $(t,s) \in [T]^2$ within d_{cross} of the diagonal of Ω_{12}.

Permutation bootstrap estimate of $\mathrm{Var}[\widetilde{\Omega}_{12}^{(t,s)}]$*:* A permutation bootstrap sample $\{X_{1,[n]}^*, X_{2,[n]}^*\}_{n=1,\ldots,N}$ is generated by sampling a random permutation of N trials independently from each of $\{X_{1,[n]}\}_{n=1,\ldots,N}$ and $\{X_{2,[n]}\}_{n=1,\ldots,N}$. The resulting sample contains no correlated activity. Hence, applying LaDynS to the bootstrap sample yields estimates of canonical precision matrix $\widehat{\Omega}^*$, canonical weights $\widehat{w}_k^*(t)$s, empirical covariance of the estimated latent variables $\overline{\Sigma}^* = \overline{\mathrm{Var}}\left(\widehat{w}_1^{*(1)\top} X_1^{*(1)}, \ldots, \widehat{w}_2^{*(T)\top} X_2^{*(T)}\right)$, and de-sparsified precision matrix estimate $\widetilde{\Omega}^*$ (Eq. (3.16)) under the global null hypothesis of no correlated activity. Repeating the bootstrap simulation B times produces B bootstrap values $\widehat{\Omega}^b$, $\widehat{w}_k^{b(t)}$, $\overline{\Sigma}^b$, and $\widetilde{\Omega}^b$, $b = 1, \ldots, B$. We estimate $\mathrm{Var}[\widetilde{\Omega}_{12}^{(t,s)}]$ with $\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]$, the sample standard error of $\{\widetilde{\Omega}_{12}^{b,(t,s)}\}_{b=1,2,\ldots,B}$. Notice that $\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]$ is obtained under the global null hypothesis – i.e. under $H_0^{(t,s)} : \Omega_{12}^{(t,s)} = 0$ simultaneously for all (t,s) – because it is not trivial to simulate bootstrap data that satisfy a specific $H_0^{(t,s)}$ without assuming that all other elements of Ω_{12} are also null. We garnered from simulations that $\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]$ is thus likely to slightly underestimate $\mathrm{Var}[\widetilde{\Omega}_{12}^{(t,s)}]$, which makes for slightly sensitive p-values.

Control of false discoveries: Because we perform tests for many entries of Ω_{12}, we cap the false discovery rate

$$\text{FDR} = \mathbb{E}\left[\text{FDP}\right], \quad \text{where } \text{FDP} = \frac{\#\{\text{falsely discovered entries}\}}{\#\{\text{discovered entries}\} \vee 1} \tag{3.19}$$

below a pre-specified level α_{BH} using the procedure of Benjamini and Hochberg (1995) (BH). To proceed, let $p^{[1]} \leq \cdots \leq p^{[n_{\text{roi}}]}$ denote the ordered permutation bootstrap p-values $p^{(t,s)}$ that correspond to n_{roi} cross-precision elements in the region of interest. Then, we find the maximum k_{BH} satisfying $p^{[k_{\text{BH}}]} \leq \frac{k_{\text{BH}}}{n_{\text{roi}}}\alpha_{\text{BH}}$ and reject $H_0^{(t,s)}$ with $p^{(t,s)}$ smaller than $\frac{k_{\text{BH}}}{n_{\text{roi}}}\alpha_{\text{BH}}$. The FDR guarantee is established by Benjamini and Hochberg (1995) as long as the $p^{(t,s)}$'s are independent and valid p-values.

Cluster-wise inference by excursion test: As a further safeguard against falsely detecting correlated activity between brain areas, we obtain p-values for each identified connectivity epoch using the excursion test of Ventura et al. (2005), as follows. For each cluster k identified by the BH procedure, we calculate the test statistic:

$$T_k := -2 \sum_{(t,s)\in\text{cluster } k} \log p^{(t,s)}, \tag{3.20}$$

which is reminiscent of Fisher's method for testing the global significance of multiple hypotheses. We calculate the corresponding p-value as $\int_{T_k}^{\infty} f_0(u)\, du$, since large values of T_k provide evidence against cross-area connectivity in cluster k, where f_0 is the null distribution of $\max_j T_j$ under the global null hypothesis of no connectivity anywhere. We use f_0 rather than the respective null distributions of each T_k to control the family-wise type I error rather than the type I error for each cluster. We approximate f_0 by the previous permutation bootstrap: for each permuted dataset $b = 1, \dots, B$, we estimate the cross-precision matrix and corresponding p-values, identify all clusters of p-values below $\frac{k_{\text{BH}}}{n_{\text{roi}}}\alpha_{\text{BH}}$, calculate the corresponding test statistics in Eq. (3.20), and let S_b be their maximum. The B values S_b are a sample from f_0, which we use to approximate the p-value for cluster k by the sampling proportion:

$$\frac{1}{B}\sum_{b=1}^{B} \mathbb{I}(S_b \geq T_k).$$

3.2.5 Locally Stationary State-space Model and Local Granger Causality

Our model in Eqs. (3.7) and (3.8) can be formulated as a state-space model by rewriting the joint multivariate Gaussian model for the latent vectors in Eq. (3.8) as the set of all conditional distributions

$$\begin{aligned} Z_1^{(t)} &= \sum_{s=1}^{d_{\text{auto}}} \alpha_{11,s}^{(t)} Z_1^{(t-s)} + \sum_{s=1}^{d_{\text{cross}}} \alpha_{12,s}^{(t)} Z_2^{(t-s)} + \eta_1^{(t)}, \\ Z_2^{(t)} &= \sum_{s=1}^{d_{\text{auto}}} \alpha_{22,s}^{(t)} Z_2^{(t-s)} + \sum_{s=1}^{d_{\text{cross}}} \alpha_{21,s}^{(t)} Z_1^{(t-s)} + \eta_2^{(t)}, \end{aligned} \tag{3.21}$$

where d_{auto} and d_{cross} are the bandwidth parameters in Section 3.2.3, $\eta_k^{(t)}$, $k = 1, 2$, are independent $N(0, \phi_k^{(t)})$ noise random variables, and the $\alpha_{kl,s}^{(t)}$'s are vector auto-regressive coefficient parameters for the auto-correlation within region if $k = l$, $k = 1, 2$, and the cross-correlation between regions if $k \neq l$ with time lag s.

This state-space formulation is convenient to impose local stationarity on the latent time series, since the functional connectivity within and between brain regions relatively changes slowly over time, and to calculate the Granger causality between regions. We achieve the former by fitting stationary state-space models in moving windows of time. For the latter, Z_2 is said to Granger-cause Z_1 at time t if some $\alpha_{12,s}^{(t)}$ are non-zero (Ombao and Pinto, 2021) (and conversely if some $\alpha_{21,s}^{(t)}$ are non-zero). The partial coefficient of determination (partial R^2) between $\left(Z_2^{(t-d_{\text{cross}})}, \dots, Z_2^{(t-1)}\right)$ and $Z_1^{(t)}$, conditional on $Z_1^{(t-d_{\text{auto}})}, \dots, Z_1^{(t-1)}$, calculated as

$$R^2_{2\to 1}(t) = 1 - \frac{\text{Var}[\text{residual of Regression 1}]}{\text{Var}[\text{residual of Regression 2}]}, \tag{3.22}$$

where

$$\begin{aligned} &\text{Regression 1}: Z_1^{(t)} \sim Z_1^{(t-d_{\text{auto}})} + \cdots + Z_1^{(t-1)} + Z_2^{(t-d_{\text{cross}})} + \cdots + Z_2^{(t-1)} \\ &\text{Regression 2}: Z_1^{(t)} \sim Z_1^{(t-d_{\text{auto}})} + \cdots + Z_1^{(t-1)} \end{aligned} \tag{3.23}$$

is therefore a measure of local Granger causality at time t. To allow a physiologically meaningful minimum connection time τ_1 from brain regions 2 to 1, we can also replace the second regression by

$$\begin{aligned}\text{Regression 2}: Z_1^{(t)} \sim & Z_1^{(t-d_{\text{auto}})} + \cdots + Z_1^{(t-1)} + Z_2^{(t-\tau_1+1)} + \cdots + Z_2^{(t-1)} \\ & + \mathbf{1}\{\tau_2 > d_{\text{cross}}\} \left(Z_2^{(t-d_{\text{cross}})} + \cdots + Z_2^{(t-\tau_2-1)} \right),\end{aligned}$$

where $\tau_2 = d_{\text{cross}}$, unless there is reason to consider a shorter connection epoch. A plug-in estimator of $R^2_{2\to1}(t)$ is easily obtained from the estimated covariance matrix of

$$\left(Z_1^{(t)}, \ldots, Z_1^{(t-d_{\text{auto}})}, Z_2^{(t-1)}, \ldots, Z_2^{(t-d_{\text{cross}})} \right), \tag{3.24}$$

without actually running the regressions.

Autocorrelations in the latent time series can inflate R^2 values. We therefore test the statistical significance of $R^2_{2\to1}(t)$ (or $R^2_{1\to2}(t)$) by comparing its observed value to its null distribution, obtained by repeatedly permuting the trials in one region and calculating $R^2_{2\to1}(t)$ in the permuted data. The permuted data satisfy the null hypothesis of no cross-region connection and exhibit the same autocorrelation structure as the original latent time series.

3.3 LaDynS performance on simulated data

We have introduced LaDynS to estimate the dynamic connectivity between two or more multivariate time series, and proposed inference procedures to identify when connectivity is statistically significant. We apply LaDynS to experimental data in Section 3.4, but first we examine its performance on simulated datasets that have properties similar to the experimental data. We check the validity of the proposed p-value estimation and FDR procedure using simulated datasets with known canonical correlation matrices. The reproducible code scripts for the simulations and experimental data analyses are provided at `github.com/HeejongBong/ladyns`.

3.3.1 Simulated datasets with known canonical correlation matrix

One simulated dataset consists of $N = 1000$ i.i.d. vector time-series X_1 and X_2 of dimensions $d_1 = d_2 = 25$ and durations $T = 50$, simulated from Eq. (3.7). The latent time series Z_1 and Z_2 in Eqs. (3.7) and (3.8) have zero mean vectors and covariance matrix $\Sigma = \Omega^{-1}$, with

$$\Omega = \begin{bmatrix} (\Sigma_{0,1} + \lambda I_T)^{-1} & \Omega_{12} \\ \Omega_{12}^\top & (\Sigma_{0,2} + \lambda I_T)^{-1} \end{bmatrix}, \tag{3.25}$$

where Ω_{12} is the cross-precision matrix of interest. The elements of the auto-precision matrices were simulated from the squared exponential function:

$$\Sigma_{0,k}^{(t,s)} = \exp\left(-c_{0,k}(t-s)^2\right), \; k = 1, 2 \tag{3.26}$$

with $c_{0,1} = 0.148$ and $c_{0,2} = 0.163$ chosen to match the LFPs autocorrelations in the experimental dataset. The diagonal regularizer λI_T was added to ensure that $\Sigma_{0,1}$ and $\Sigma_{0,2}$ are invertible, and we set $\lambda = 1$. For Ω_{12}, we considered the connectivity scenario depicted in Fig. 3.4d, where the two latent times series connect in three epochs, the first with no latency, the second with series 2 preceding series 1, and the third with series 1 preceding series 2. We accordingly set the cross-precision matrix elements to

$$\Omega_{12}^{(t,s)} = \begin{cases} -r, & \text{if } (t,s) \text{ is colored blue,} \\ 0, & \text{elsewhere,} \end{cases} \tag{3.27}$$

where r measures the intensity of the connection. Finally, we rescaled Σ to have diagonal elements equal to one.

Once the latent time series Z_1 and Z_2 were generated, we simulated a pair of observed time series according to

$$X_k^{(t)} = Y_k^{(t)} - \beta_k^{(t)} \left(w_k^{(t)\top} Y_k^{(t)} - \widehat{m}_k^{(t)} \right) + \beta_k^{(t)} Z_k^{(t)}, \tag{3.28}$$

for $k = 1, 2$ and $t = 1, \ldots, T$, where $Y_1^{(t)}$ and $Y_2^{(t)}$ are uncorrelated baseline time series, $\beta_k^{(t)}$ are factor loadings that change smoothly over time, $w_k^{(t)}$ are canonical weights that satisfy the relationship with $\beta_k^{(t)}$ in Eq. (3.12), and $\widehat{m}_k^{(t)}$ is the empirical mean of $\beta_k^{(t)\top} Y_k^{(t)}$, for $k = 1, 2$. We subtracted $\beta_k^{(t)} \left(w_k^{(t)\top} Y_k^{(t)} - \widehat{m}_k^{(t)} \right)$ to ensure that $X_k^{(t)}$ has canonical correlation matrix Σ and the same mean as $Y_k^{(t)}$. We took $Y_1^{(t)}$ and $Y_2^{(t)}$ to be the two multivariate time-series of neural recordings analyzed in Section 3.4, which we permuted to remove all cross-correlations. To reduce auto-correlation across time, we added space-correlated white noise to the baseline time series. The amount of noise was set to be comparable to the diagonal regularization λI_T introduced in Eq. (3.25). Finally we set $\beta_k \in \mathbb{R}^{d_k \times T}$ to be the factor loadings estimated in Section 3.4. The resulting latent time series $\beta_k^{(t)} Z_k^{(t)}$ and noise baseline vector $Y_k^{(t)}$ in Eq. (3.28) have comparable scales and auto-correlations by construction, for $k = 1, 2$, to ensure that the simulated observed time series $X_k^{(t)}$ have scales and auto-correlations comparable to the experimental data in Section 3.4.

3.3.2 LaDynS estimation details

For the simulated dataset, we do not need the regularization on the diagonal of Σ, so λ_{diag} was set 0. The other hyperparameters were set to $d_{\text{auto}} = 10$, $d_{\text{cross}} = 10$, and $\lambda_{\text{auto}} = 0$. The penalty on the cross-correlation elements, λ_{cross}, was automatically tuned at every run to control false discoveries (see Section 3.2.3).

3.3.3 Results

Figure 3.4a displays the LaDynS cross precision estimate $\widehat{\Omega}_{12}$ fitted to one dataset simulated as in Section 3.3.1 under the connectivity scenario depicted in Fig. 3.4d, with connection strength $r = 0.4$ in Eq. (3.27). Figure 3.4b shows the permutation bootstrap p-values for the entries of the desparsified cross-precision estimate $\widetilde{\Omega}_{12}$ (Eq. (3.18) with permutation bootstrap simulation size $B = 200$; see Section 3.2.4). Small p-values concentrate near the locations of true non-zero

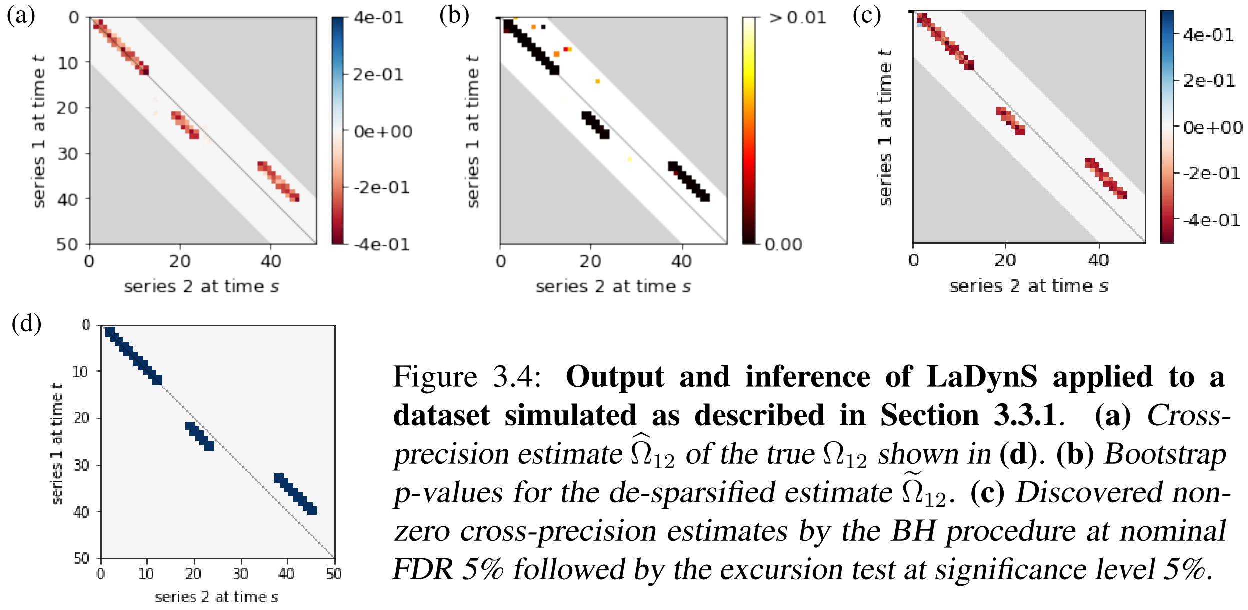


Figure 3.4: **Output and inference of LaDynS applied to a dataset simulated as described in Section 3.3.1**. **(a)** *Cross-precision estimate* $\widehat{\Omega}_{12}$ *of the true* Ω_{12} *shown in* **(d)**. **(b)** *Bootstrap p-values for the de-sparsified estimate* $\widetilde{\Omega}_{12}$. **(c)** *Discovered non-zero cross-precision estimates by the BH procedure at nominal FDR 5% followed by the excursion test at significance level 5%.*

cross-precision entries and are otherwise scattered randomly. We applied first the BH procedure with target FDR 5% (Section 3.2.4) and subsequently the excursion test at the 5% significance level to all discovered clusters. The significant clusters ($p < 0.005$) are plotted in Fig. 3.4c .

Normal approximation for the p-values in Eq. (3.18)*:* We investigate the validity of the Normal assumption by comparing the empirical distribution of $R = 60$ repeat estimates $\widetilde{\Omega}_{12}^{(t,s)}/\sqrt{\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]}$ (Eq. (3.17)) to the standard normal distribution using QQ-plots. Fig. 3.5 shows QQ-plots for three randomly chosen representative time pairs (t, s) with $\Omega_{12}^{(t,s)} = 0$, which confirms that the null distribution of $\widetilde{\Omega}_{12}^{(t,s)}/\sqrt{\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]}$ is close to standard Gaussian. We further check the validity of the permutation bootstrap variance estimates $\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]$, shown in Fig. 3.6b, by comparing it to the empirical variance of the $R = 60$ estimates $\widetilde{\Omega}_{12}^{(t,s)}$, shown in Fig. 3.6a. There is good agreement for the entries that have precision value zero, $\Omega_{12}^{(t,s)} = 0$. Fig. 3.6c further displays the Q-Q plot of the repeat ratios of permutation bootstrap over empirical estimates of $\mathrm{Var}[\Omega_{12}^{(t,s)}]$ for these entries, with $F(B-1, R-1)$ being the reference distribution. The good agreement suggests that the bootstrap estimate of $\mathrm{Var}[\Omega_{12}^{(t,s)}]$ is reliable.

FDR control: Fig. 3.7 shows estimated FDR and FNR for a range of simulated connection strengths (r in Eq. (3.27)) and nominal FDR values. The estimated and nominal FDRs agree when the connection is weak but the former slightly exceeds the latter otherwise. This is likely due to the well-known sensitivity of the BH procedure to positively correlated p-values, consistent with the fact that, in our experience, the false discoveries are mostly adjacent to correct discoveries (discoveries that have $\Omega_{12}^{(t,s)} \neq 0$).

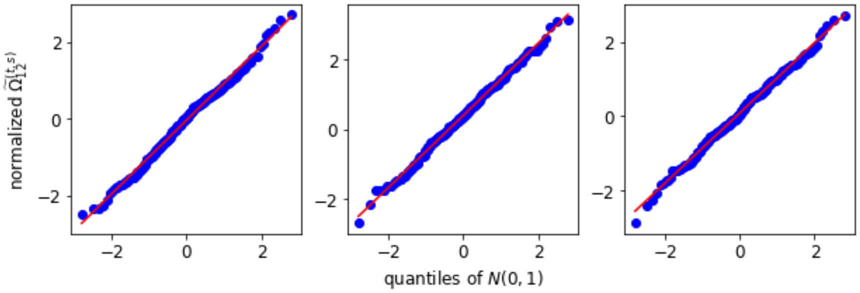


Figure 3.5: **Null distributions of three representative entries of $\widetilde{\Omega}_{12}^{(t,s)}/\sqrt{\widehat{\mathrm{Var}}[\widetilde{\Omega}_{12}^{(t,s)}]}$.** *The null distributions are obtained from $R = 60$ simulated datasets (Section 3.3.1) and compared to the standard Gaussian distribution via QQ-plots. There is good agreement.*

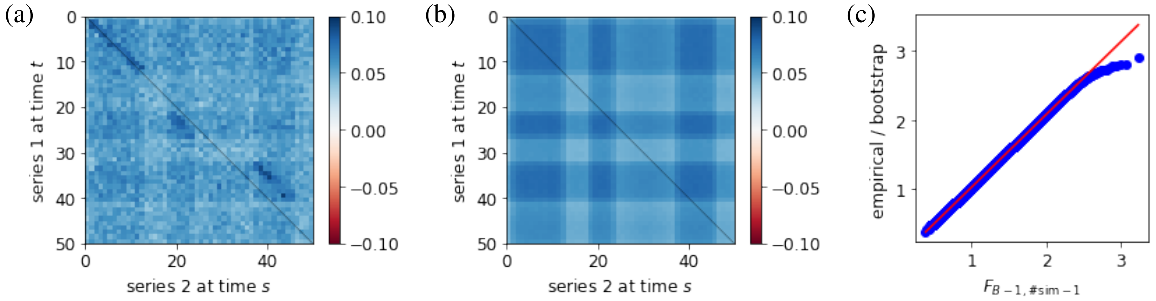


Figure 3.6: **Standard deviations of desparsified precision elements.** *Variance obtained* **(a)** *from samples from the ground-truth generative multiset pCCA model and* **(b)** *from permutation bootstrapped samples.* **(c)** *F-statistics of ratios between the two variances for null entries of Ω_{12}, showing good agreement.*

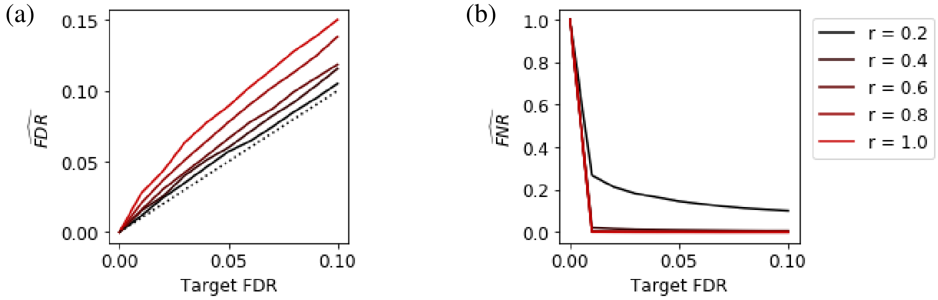


Figure 3.7: **False Discovery Rate control of LaDynS' inference. (a)** *Estimated false discovery rate and* **(b)** *false non-discovery rate for target FDR $\in [0, 10]\%$, under the connectivity scenario in Fig. 3.4d, for connectivity intensities $r = 0.2$, 0.4, 0.6, 0.8 and 1.0 in Eq. (3.27). The dotted line is a (0,1) line.*

3.4 Experimental Data Analysis

We applied LaDynS to local field potentials (LFPs), collected as described in Khanna et al. (2020), from two Utah arrays implanted in a Macaque monkey's prefrontal cortex (PFC) and V4 during a memory-guided saccade task. Each trial of the task started with a monkey fixating its eyes toward the center of the screen. A visual cue was given for 50ms to indicate a target location, which was randomly chosen from eight locations. The monkey had to remember the target location while maintaining eye fixation for a delay period of 500ms. After the delay period, the monkey reoriented its eyes toward the remembered position, and reward was given on successful trials. As in Khanna et al. (2020), we analyzed the time series during the delay period, based on 3000 successful trials. Because beta oscillations are often associated with communication across brain areas (Klein et al., 2020; **?**), we first filtered LFP recordings using complex Morlet wavelets at beta oscillation frequency 18Hz (which was the most active frequency in the magnitude scalograms), and obtained the beta oscillation power envelopes as the absolute values of the filtered signals. We note that the target frequencies were different from those in Bong et al. (2020a), yielding different data and somewhat different results. After downsampling them to 200Hz, we applied LaDynS. Because the filtered data were very smooth, we used a regularization $\lambda_{\text{diag}} > 0$ on the diagonal of Σ, as in Section 3.3.2.

Figure 3.8 displays the LaDynS cross precision estimate $\widehat{\Omega}_{12}$, the bootstrap p-values for the entries of the desparsified cross-precision estimate $\widetilde{\Omega}_{12}$, and the epochs of connectivity identified by STAR at target FDR 5%. Time $t = 0$ corresponds to when the visual stimulus (the memory cue) was turned off. One highly significant contiguous region of the precision matrix ($p \approx .01$ or smaller), discovered by BH, is shown in Fig. 3.8. It corresponds to beta power in PFC being correlated with earlier beta power in V4 (after conditioning on beta power at all other times and lags), about 200 and 400 milliseconds post-stimulus (red blobs). It is important to keep in mind that the sparse estimation method finds a relatively small number of large effects; smaller effects, some of which may remain interesting, do not appear.

To better understand the lead-lag relationships, we used the estimated latent time-series to compute a series of partial R^2 values from the locally stationary state-space model, as described in Section 3.2.4. We filtered LFP signals at beta frequency (18 Hz, Period: 0.055 sec.), and we expect non-stationarity at a finer time scale than few periods is hard to detect. For each t_0, we set a local time window of length 0.2 sec., which encompasses less than 4 beta wave cycles. We used $\tau_1 = 0.015$ and $\tau_2 = 0.035$ seconds. The null distribution was based on 200 trial permutations. Fig. 3.9 displays two sets of partial R^2 values, the first set, in panel (a), corresponds to the local Granger causality of $Z_2^{(t-.015:t-.035)}$ over $Z_1^{(t)}$, where Z_1 was the latent time series for V4 and Z_2 the latent time series for PFC. Thus, PFC leads V4 by about 45 milliseconds in these regressions. Panel (b) displays partial R^2 values of the cross-correlation at the opposite direction.

There is a large excursion of R^2 above the null values in the PFC $\to$ V4 plot and one in the V4 $\to$ PFC plot. The excursion in the V4 $\to$ PFC plot centered around 400 ms, are consistent with the results in Fig. 3.8. The excursion in the PFC $\to$ V4 plot, combined with the excursion in the V4 $\to$ PFC plot, shows mutually predictive activity.

These results are consistent with the idea that V4 responds strongly to visual stimuli and PFC receives input from the visual processing stream, which includes V4, but we would expect the directional influences to differ across time, during the task. Fig. 3.10 displays the estimated total

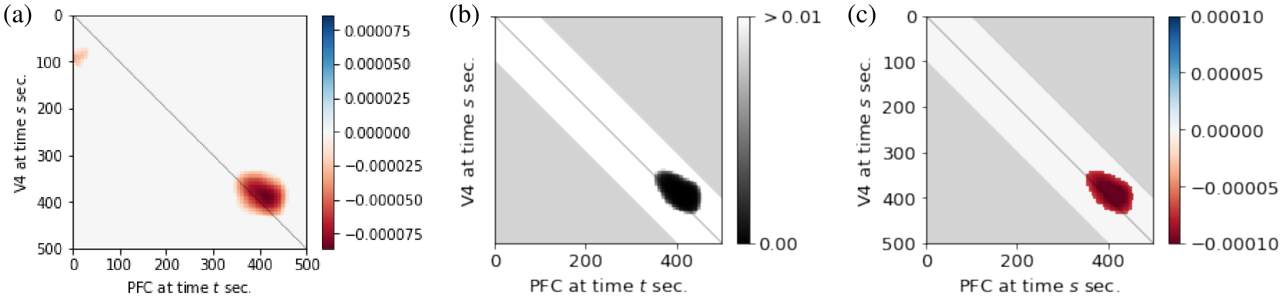


Figure 3.8: **LaDynS output and inference for experimental dataset.** **(a)** *LaDynS cross-precision estimate* $\widehat{\Omega}_{12}$*. The light gray area shows the region of time considered (one area leading the other by at most 0.1 second).* **(b)** *Bootstrap p-values for the de-sparsified estimate* $\widetilde{\Omega}_{12}$ *(Eq.* (3.18) *with bootstrap simulation size* $B = 200$*; see Section 3.2.4).* **(c)** *Discovered regions of cross-precision using STAR at nominal FDR 5%, suggesting that activity in PFC preceded that in V4 immediately post-stimulus and around* 0.1 *seconds thereafter (blue blobs), and that activity in V4 preceded that in PFC around time points* 0.2 *and* 0.4 *seconds post-stimulus (red blobs). These four epochs had excursion test* p*-values of* $p = .024$, $p = .0015$, $p = .001$ *and* $p < .0005$*, respectively. The two smaller blobs were not significant* ($p > 0.5$) *and are therefore likely spurious.*

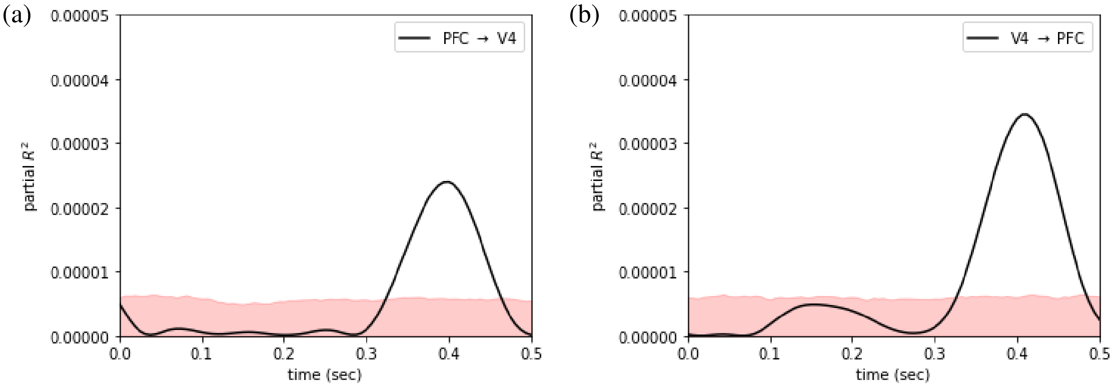


Figure 3.9: **Estimated partial R^2 from locally stationary state-space model.** *For* $Z_1 \to Z_2$*, partial* R^2 *was computed corresponding to linear regression of* $Z_1^{(t)}$ *against* $Z_2^{(t-0.015\text{sec.}:t-0.03\text{sec.})}$ *conditional to the other covariates, including* Z_2 *at the other lags, for every time point* t *at sampling rate 200 Hz. The black solid lines indicate the estimated partial* R^2 *for* **(a)** *V4* $\to$ *PFC and* **(b)** *PFC* $\to$ *V4. The pink shaded area shows 95 percentile area of null partial* R^2 *under independence between V4 and PFC, obtained from the permutation test.*

beta power due to the latent time series, across time, for both PFC and V4.

The latent beta power in PFC is largest at the beginning of the task, then drops to a roughly constant amplitude, while the latent beta power in V4 initially drops slightly but then starts to rise just before 100 ms, peaks around 250 ms, and then declines but remains large. Taking Fig. 3.8b and Fig. 3.10 together, these results show that starting a little after 0.3 sec, each of the two latent beta power envelopes is able to predict the other, 0.03 to 0.06 sec in the future, suggesting the two

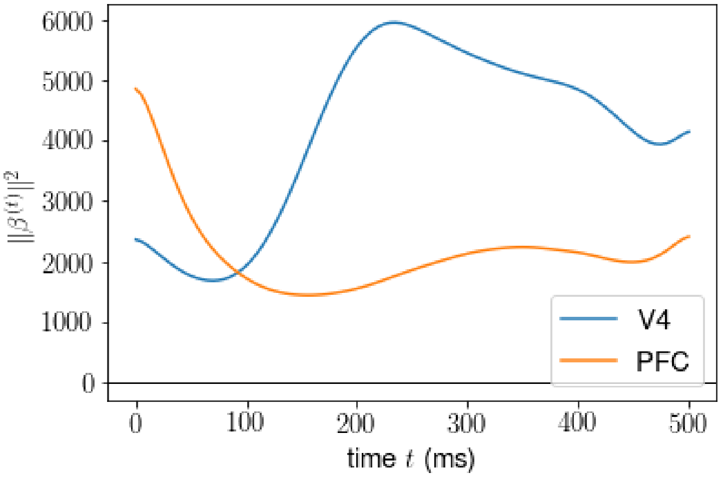


Figure 3.10: **Estimated variance of electrophysiological activity in V4 and PFC driven by latent factors.** *The* ℓ_2 *norm of the factor loading vector* $\beta_k^{(t)}$ *(representing the total variance of the data* $X_k^{(t)}$ *attributable to the latent time series at time* t*, see Eq.* (3.7)*) is plotted across time* t *for V4 (blue) and PFC (orange).*

areas are participating in bidirectionally linked beta oscillations.

It is also possible to get information from the normalized factor loadings, across the electrode arrays, which are displayed in Fig. 3.11 for times $t = 0.15$ and $t = 0.40$ sec. (The loadings have been divided by their maximal value across the array.) Most noticeably, the spatial pattern of activity in PFC at 400 ms, compared with that at 150 ms, becomes more concentrated in the upper left corner of the electrode array. Bootstrap confidence intervals suggest this change in spatial pattern within PFC is not due solely to noise in estimation, see Fig. C.1b in the Supplementary Material. Thus, apparently, the early and late periods of the task may be distinguished not only by very different prediction effects but, in addition, by spatial shifts in lag-correlated activity. (An animation over the complete timeline is available at `github.com/HeejongBong/ladyns`.)

3.5 Discussion

Motivated by the problem of describing interactions across a pair of brain regions, based on high-dimensional neural recordings, we have provided a time-series extension of probabilistic CCA together with a novel sparse estimation methodology. According to our Equation (3.7), each of the two multivariate time series is driven by a single latent time series, with the cross-dependence of these two latent time series representing cross-region interaction. According to Equation (3.8), the latent bivariate time series is a discrete Gaussian process but its correlation matrix is unrestricted, allowing for non-stationarity. The repeated trial structure enabled us to estimate the resulting high-dimensional covariance matrix, by applying sparse estimation and inference methods. We found, and displayed in Fig. 3.9, interesting directional relationships between PFC and V4, in beta power, that appeared during late delay period, where the relationship was bidirectional. These were based on partial R^2 values, computed from the estimated covariance matrices, corresponding

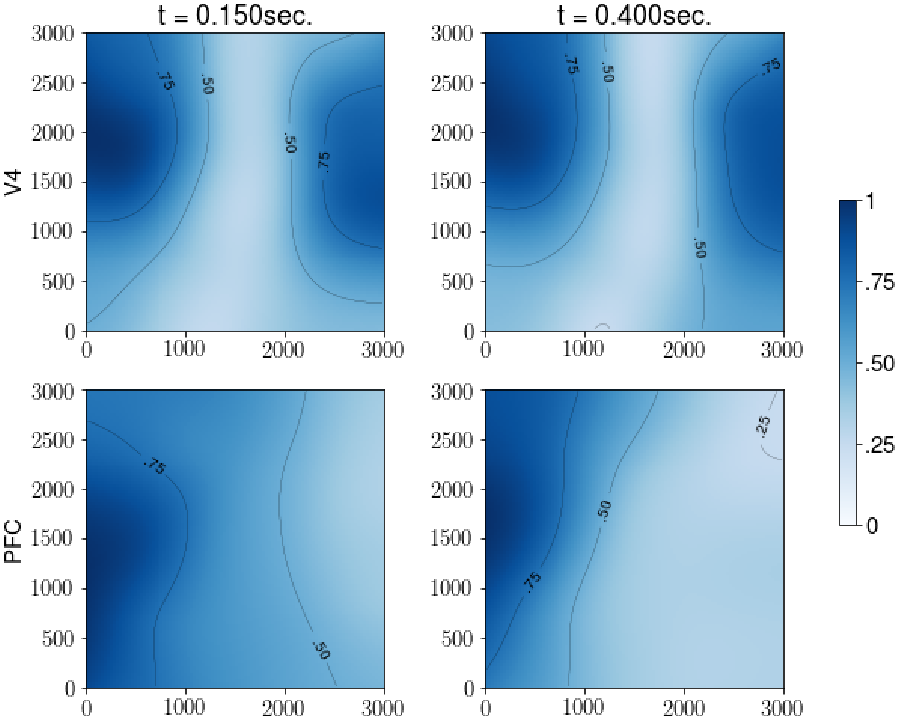


Figure 3.11: *Factor loadings of V4 and PFC, spatially smoothed, normalized, and color coded over the electrode coordinates (μm) at* **(left)** *the first bump of Fig. 3.9b (0.15 sec) and* **(right)** *the second bump of Fig. 3.9b or the first bump of Fig. 3.9a (0.40 sec). Contours at fractions .25, .5, and .75 of maximal power have been added.*

to lagged regressions of one latent time series on the other. The analysis in Fig. 3.9 is in the spirit of Granger causality, but differs from it by allowing for non-stationarity, so that we could obtain the time-varying results.

In addition to making the analysis possible, the repeated trial structure suggests substantive interpretation based on trial-to-trial variability. Neurophysiological experiments like the one generating the data analyzed here take pains to make the experimental setting nearly the same on each trial. The inevitable, relatively small fluctuations in the way the subject interacts with the environment, together with changes in the subject's underlying state (involving fluctuations in motivational drive, for example), lead to observable fluctuations in behavior and in the recorded neural activity. Although the network sources of trial-to-trial variability in the PFC and V4 data are unknown, they produce the kind of correlated activity revealed in Fig. 3.9. To interpret it, we acknowledge there could be some task-irrelevant, trial-varying pulses of inputs that drive beta power in V4 and PFC, having just the right differential time lags to produce the correlated activity picked up by the partial R^2 values. Could such task-irrelevant pulses of activity change across time, within repetitions of the task, in such a way as to produce, the peaks in Figs. 3.9a and 3.9b? It is possible, but it would be surprising, especially when we consider contemporary ideas about beta oscillations during working memory tasks (Miller et al., 2018) along with the well-identified distinction between early and late visual processing, e.g., Yang et al. (2019). The alternative we mentioned, that PFC and V4 are involved, together, in goal-directed visual processing and memory, with PFC influencing V4 early and the two areas acting bidirectionally later, seems a credible possibility.

There are many ways to extend the ideas developed here. For band-pass filtered data, such as those analyzed in Section 3.4, phase analysis (Klein et al., 2020) could be combined with amplitude analysis. Multiple frequencies could be considered (along the lines of Gallagher et al. (2017)), as well. In Bong et al. (2020a) we have described an extension of LaDynS, where the within-region noise vectors ϵ_k were allowed to have general time series structure and the latent time series driving each brain region were multidimensional. That brief report, however, does not provide the details given here, nor does it include inferential procedures. Thus, an important future step will be to show how inference can be carried out for the general models in Bong et al. (2020a). A different direction for additional research would be to simplify the version of LaDynS we have used here by imposing suitable spatiotemporal structure on the latent time series. While we hope such approaches will be fruitful, we believe the general framework of LaDynS can be of use whenever interest focuses on non-stationary interactions among groups of repeatedly-observed multivariate time series.

Chapter 4

Latent Dynamic Factor Analysis of High-Dimensional Neural Recordings

This chapter is a collaboration with Zongge Liu, Zhao Ren, Eric Yttri, Matthew A. Smith, Valérie Ventura, and Robert E. kass. This work is published at NeurIPS2020 (Bong et al., 2020b).

4.1 Introduction

New electrode arrays for recording electrical activity generated by large networks of neurons have created great opportunities, but also great challenges for statistical machine learning (e.g., Steinmetz et al., 2018). For example, Local Field Potentials (LFPs) are signals that represent the bulk activity in relatively small volumes of tissue (Buzsáki et al., 2012; Einevoll et al., 2013), and they have been shown to correlate substantially with the BOLD fMRI brain imaging signal (Logothetis et al., 2001; Magri et al., 2012b). Typical LFP data sets may have dozens to hundreds of time series in each of two or more brain regions, recorded simultaneously across many experimental trials. A motivating example in this paper is LFP recordings from a prefrontal cortex (PFC) and visual area V4 during a visual working memory task. V4 has been reported to retain higher-order information (e.g., color and shape) and attention in visual processing (Orban, 2008; Fries et al., 2001), while PFC is considered to exert cognitive control in working memory (Miller and Cohen, 2001). Despite their spatial distance and functional difference, these regions have been presumed to cooperate during visual working memory tasks. Various approaches have been used to track the interaction among brain regions Adhikari et al. (2010); Jiang et al. (2015); Hultman et al. (2018); Gallagher et al. (2017); Buesing et al. (2014). In particular, delay-specific theta synchrony led by PFC has been discovered during visual memory tasks (Sarnthein et al., 1998; Liebe et al., 2012).

We report here an extension of Gaussian process factor analysis (GPFA, Yu et al., 2009) to two or more groups of time series, where the main interest is non-stationary cross-group interaction; furthermore, the multivariate noise within groups can have both spatial covariation and non-stationary temporal covariation. Here, spatial covariation refers to dependence among the time series and, in the neural context, this results from the spatial arrangement of the electrodes, each of which records one of the time series. Our approach uses probabilistic CCA, but the framework allows rich spatiotemporal dependencies. These generalizations come at a cost: we now have a high-dimensional time series problem within each brain region together with a high-dimensional covariance structure. We solve these high-dimensional problems by imposing sparsity of the dominant effects, building on Bong et al. (2020c), which treats the high-dimensional covariance structure in the context of observational white noise, and by incorporating banded covariance structure as in Bickel and Levina (2008). We thus call our method Latent Dynamic Factor Analysis of High-dimensional time series, LDFA-H.

In a simulation study, based on realistic synthetic time series, we verify the recovery of cross-region structure even when some of our assumptions are violated, and even in the presence of high noise. We then apply the method to 192 LFP time series recorded simultaneously from both Prefrontal Cortex (PFC) and visual area V4, during a memory task, and find time-varying cross-region dependencies.

4.2 Latent Dynamic Factor Analysis of High-dimensional time series

We treat the case of two groups of time series observed, repeatedly, N times. Let $X^1_{:,t} \in \mathbb{R}^{p_1}$ and $X^2_{:,t} \in \mathbb{R}^{p_2}$ be p_1 and p_2 recordings at time t in each of the two groups, for $t = 1, \ldots, T$. As in Yu et al. (2009), we assume that a q-dimensional latent factor $Z^k_{:,t} \in \mathbb{R}^q$ drives each group, here, each brain region, according to the linear relationship

$$X^k_{:,t} \mid Z^k_{:,t} = \mu^k_{:,t} + \beta^k \cdot Z^k_{:,t} + \epsilon^k_{:,t}, \tag{4.1}$$

for brain region $k = 1, 2$, where $\mu^k_{:,t} \in \mathbb{R}^{p_k}$ are mean vectors, $\beta^k \in \mathbb{R}^{p_k \times q}$ are matrices of constant factor loadings, and $\epsilon^k_{:,t} \in \mathbb{R}^{p_k}$ are errors centered at zero (independently of the latent vectors Z). We are interested in the pairwise cross-group dependencies of the latent vectors $Z^1_{f,:}$ and $Z^2_{f,:}$, for $f = 1, \ldots, q$. As in (Bong et al., 2020c), we assume that the time series of these latent vectors follows a multivariate normal distribution

$$\begin{pmatrix} Z^1_{f,:} \\ Z^2_{f,:} \end{pmatrix} \sim \mathrm{MVN}(0, \Sigma_f), \quad f = 1, \ldots, q, \tag{4.2}$$

where Σ_f describes all of their simultaneous and lagged dependencies, both within and between the two vectors. We assume the N sets of random vectors (ϵ, Z) are independent and identically distributed. Fig. 4.1a illustrates the dependence structure of this model. We let P_f be the correlation matrix corresponding to Σ_f, and write its inverse as

$$\Pi_f = \mathrm{P}_f^{-1} = \left(\begin{array}{c|c} \Pi^{11}_f & \Pi^{12}_f \\ \hline \Pi^{12\top}_f & \Pi^{22}_f \end{array}\right) \tag{4.3}$$

where Π^{11}_f and Π^{22}_f are the scaled auto-precision matrices and Π^{12}_f is the scaled cross-precision matrix. We now assume finite-range partial auto-correlation and cross-correlation for $(Z^1_{f,t}, Z^2_{f,t})$,

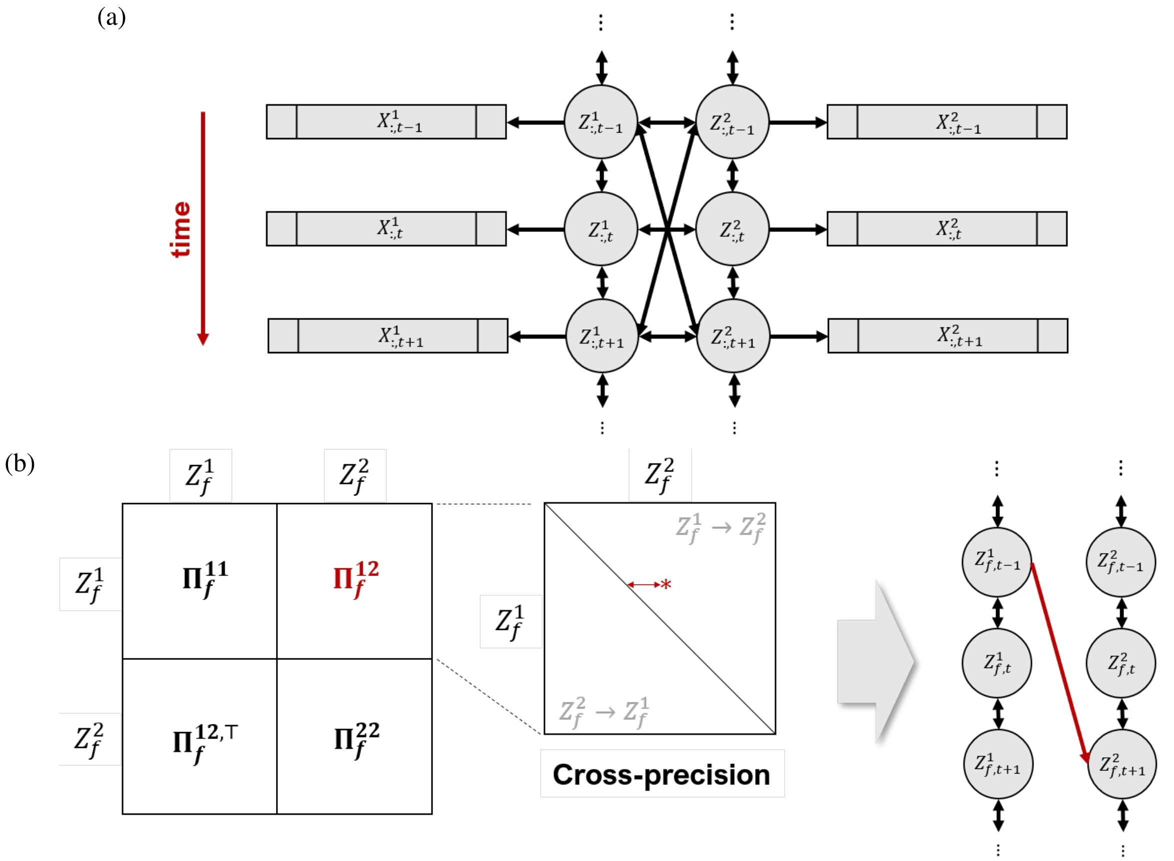


Figure 4.1: **LDFA-H model. (a)** *Dynamic associations between vectors* $X^1_{:,t}$ *and* $X^2_{:,s}$ *are summarized by the dynamic associations between their associated 1D latent variables* $Z^1_{:,t}$ *and* $Z^2_{:,s}$. **(b)** *When a significant cross-precision entry is identified, e.g., the red star in the expanded view of* Π^{12}_f, *its coordinates and distance from the diagonal indicate at what time in the experiment connectivity between two brain areas occurs, and at what lead or lag. Here the red star is in the upper diagonal of* Π^{12}_f, *which means that, at this particular time, region 1 leads region 2, or* $Z^1_f \to Z^2_f$ *in short (a non-zero entry in the lower diagonal would mean* $Z^2_f \to Z^1_f$*). We represent this association by the red arrow on the right-most plot, with a lag of two units of time for illustration.*

so that Π^{11}_f, Π^{22}_f and Π^{12}_f in Equation (4.3) have a banded structure. Specifically, for $k, l = 1, 2$, we assume there is a value h^{kl}_f such that Π^{kl}_f is a $(2h^{kl}_f + 1)$-diagonal matrix. Because our goal is to address the cross-region connectivity and lead-lag relationship, we are particularly interested in the estimation of Π^{12}_f for each latent factor $f = 1, \ldots, q$. Note that the non-zero elements $\Pi^{12}_{f,(t,s)}$, depicted as the red star in the expanded display within Fig. 4.1b, determine associations between the latent pair $Z^1_{f,:}$ and $Z^2_{f,:}$, which are simultaneous when $t = s$ and lagged when $t \neq s$.

Finally, we model the noise in Eq. (4.1) as a Gaussian random vector

$$\mathrm{Vec}(\epsilon^k) = (\epsilon^k_{:,1}; \epsilon^k_{:,2}; \ldots; \epsilon^k_{:,T}) \sim \mathrm{MVN}(0, \Phi^k), \quad k = 1, 2, \tag{4.4}$$

where we allow Φ^k to have non-zero off-diagonal elements to account for within-group spatiotemporal dependence. We assume Φ^k can be written in Kronecker product form

$$\Phi^k = \Phi^k_{\mathcal{T}} \otimes \Phi^k_{\mathcal{S}},\ k = 1, 2, \tag{4.5}$$

where $\Phi^k_{\mathcal{T}}$ and $\Phi^k_{\mathcal{S}}$ are the temporal and spatial components of Φ^k, as is often assumed for spatiotemporal matrix-normal distributions, e.g., (Dawid, 1981). Although this is a strong assumption, implying, for instance, that the auto-correlation of every $X^k_{i,:}$ is proportional to $\Phi^k_{\mathcal{T}}$, we regard Φ_k as a nuisance parameter: our primary interest is Σ_f in Eq. (4.2). We also assume an auto-regressive order at most h^k_ϵ, so that $\Gamma^k_{\mathcal{T}} = \left(\Phi^k_{\mathcal{T}}\right)^{-1}$ is a $(2h^k_\epsilon + 1)$-diagonal matrix. In our simulation we show that we can recover Σ_f accurately even when the Kronecker product and bandedness assumptions fail to hold.

The model in Equations (1)-(5) generalizes other known models. First, when $q = 1$, and $Z^1 = Z^2$ remains constant over time, in the noiseless case ($\epsilon_k = 0$), it reduces to the probabilistic CCA model of Bach and Jordan (2005); see Theorem 2.2 of Bong et al. (2020c) Thus, model (1)-(5) can be viewed as a denoising, multi-level and dynamic version of probabilistic CCA. Second, when $k = 1$, the Gaussian processes are stationary, and the ϵ vectors are white noise, (1)-(5) reduces to GPFA (Yu et al., 2009). Thus, (1)-(5) is a two-group, nonstationary extension of GPFA that allows for within-group spatio-temporal dependence.

Identifiability and sparsity constraints Despite the structure imposed on Φ_k in Eq. (4.5), parameter identifiability issues remain. Our model in Eqs. (4.1), (4.2) and (4.4) induces the marginal distribution of the observed data (X^1, X^2):

$$\left(X^1_{:,1}; X^1_{:,2}; \ldots; X^2_{:,T}\right) \sim \mathrm{N}\left((\mu^1_{:,1}; \mu^1_{:,2}; \ldots; \mu^2_{:,T}), S\right) \tag{4.6}$$

where S is the marginal covariance matrix given by:

$$S = \begin{bmatrix} \Phi^1_{\mathcal{T}} \otimes \Phi^1_{\mathcal{S}} & 0 \\ 0 & \Phi^2_{\mathcal{T}} \otimes \Phi^2_{\mathcal{S}} \end{bmatrix} + \sum_{f=1}^{q} \begin{bmatrix} \Sigma^{11}_f \otimes (\beta^1_f \beta^{1\top}_f) & \Sigma^{12}_f \otimes (\beta^1_f \beta^{2\top}_f) \\ \Sigma^{12\top}_f \otimes (\beta^2_f \beta^{1\top}_f) & \Sigma^{22}_f \otimes (\beta^2_f \beta^{2\top}_f) \end{bmatrix}. \tag{4.7}$$

The family of parameters

$$\theta^{\{\alpha^1,\alpha^2\}} = \left\{ \begin{matrix} \Sigma^{\{\alpha^1_1,\alpha^2_1\}}_1, \ldots, \Sigma^{\{\alpha^1_q,\alpha^2_q\}}_q,\ \Phi^1_{\mathcal{S}} - \sum_{f=1}^{q} \alpha^1_f \beta^1_f \beta^{1\top}_f,\ \Phi^2_{\mathcal{S}} - \sum_{f=1}^{q} \alpha^2_f \beta^2_f \beta^{2\top}_f, \\ \Phi^1_{\mathcal{T}},\ \Phi^2_{\mathcal{T}},\ \beta^1,\ \beta^2,\ \mu^1,\ \mu^2 \end{matrix} \right\}, \tag{4.8}$$

where $\Sigma^{\{\alpha^1_f,\alpha^2_f\}}_f = \left\{ \Sigma_f + \begin{bmatrix} \alpha^1_f \Phi^1_{\mathcal{T}} & 0 \\ 0 & \alpha^2_f \Phi^2_{\mathcal{T}} \end{bmatrix} \right\}$, induce the same marginal distribution in Eq. (4.6), for all $\alpha^1, \alpha^2 \in \mathbb{R}^q$ (notice that $\theta = \theta^{\{\mathbf{0},\mathbf{0}\}} = \{\Sigma_1, \ldots, \Sigma_q,\ \Phi^1_{\mathcal{S}},\ \Phi^2_{\mathcal{S}},\ \Phi^1_{\mathcal{T}},\ \Phi^2_{\mathcal{T}},\ \beta^1,\ \beta^2,\ \mu^1,\ \mu^2\}$ is the original parameter). Preliminary analysis of LFP data indicated that strong cross-region dependence occurs relatively rarely. We therefore resolve this lack of identifiability by choosing the solution given by maximizing the likelihood with an L1 penalty, under the assumption that the inverse cross-correlation matrix Π^{12}_f is a sparse $(2h^{12}_f + 1)$-diagonal matrix.

Latent Dynamic Factor Analysis of High-dimensional time series (LDFA-H) Given N simultaneously recorded pairs of neural time series $\{X^1[n], X^2[n]\}_{n=1,\ldots,N}$, the maximum penalized likelihood estimator (MPLE) of the inverse correlation matrix of the latent variables solves

$$\begin{aligned}\left(\widehat{\Pi}_1, \ldots, \widehat{\Pi}_q\right) = \operatorname{argmin} \; & -\frac{1}{N}\sum_{n=1}^{N} l\left(\theta; X^1[n], X^2[n]\right) + \sum_{f=1}^{q}\sum_{k,l=1}^{2} \left\|\Lambda_f^{kl} \odot \Pi_f^{kl}\right\|_1 \\ \text{s.t. } & \Gamma_{\mathcal{T}}^k \text{ is } (2h_\epsilon^k + 1)\text{-diagonal},\end{aligned} \tag{4.9}$$

where the log-likelihood is

$$l\left(\theta; X^1, X^2\right) = -\log\det S - (X^1_{:,1} - \mu^1_{:,1}; \ldots; X^2_{:,T} - \mu^2_{:,T})^\top S^{-1}(X^1_{:,1} - \mu^1_{:,1}; \ldots; X^2_{:,T} - \mu^2_{:,T}), \tag{4.10}$$

with S defined in Eq. (4.7), and the constraints are

$$\Lambda^{kl}_{f,(t,s)} = \begin{cases} \infty, & (t,s) : |t-s| > h_f^{kl}, \\ \lambda_f, & (t,s) : 0 < |t-s| \leq h_f^{kl}, \;\; k \neq l, \\ 0, & \text{otherwise.} \end{cases} \tag{4.11}$$

for factor $f = 1, \ldots, q$ and brain region $k = 1, 2$. The first constraint forces the corresponding $\Pi^{kl}_{f,(t,s)}$ to zero and thus imposes a banded structure for Π_f^{kl}, and the second assigns the same sparsity constraint λ_f on the off-diagonal elements of Π_f^{12}. Finally, to make calibration of tuning parameters computationally feasible, we set the bandwidth for the latent precisions and the noise precisions within each region to a single value h_{auto}, we set the bandwidth for the latent precisions across regions to a value h_{cross}, and we set the sparsity parameters to a value λ_{cross}, i.e.,

$$h_f^{kk} = h_\epsilon^k = h_{\text{auto}}, \;\; h_f^{12} = h_{\text{cross}} \;\text{ and }\; \lambda_f = \lambda_{\text{cross}},$$

for each factor $f = 1, \ldots, q$ and region $k = 1, 2$. The bandwidths are chosen using domain knowledge and preliminary data analyses. We determine the remaining parameters λ_{cross} and q by 5-fold cross-validation (CV).

Solving Eq. (4.9) requires S^{-1}. Because it is not available analytically and a numerical approximation is computationally prohibitive, we solve Eq. (4.9) using an EM algorithm (Dempster et al., 1977). Let $\theta^{(r)}$ be the parameter estimate at the r-th iteration. We consider the data $\{X^1[n], X^2[n]\}_{n=1,\ldots,N}$ to be incomplete observations of $\{X^1[n], Z^1[n], X^2[n], Z^2[n]\}_{n=1,\ldots,N}$. In the E-step, we estimate the conditional mean and covariance matrix of each $\{Z^1[n], Z^2[n]\}$ with respect to $\{X^1[n], X^2[n]\}$ and $\theta^{(r)}$. Given these sufficient statistics, the problem of computing the MPLE decomposes into two separate minimizations of

1. the negative log-likelihood of Σ_f, w.r.t. the latent factor model (Eq. (4.2)) and
2. the negative log-likelihood of $\Phi^1_{\mathcal{S}}$, $\Phi^2_{\mathcal{S}}$, $\Phi^1_{\mathcal{T}}$, $\Phi^2_{\mathcal{T}}$, β^1, β^2, μ^1, μ^2 w.r.t. the observation model (Eqs. (4.1) and (4.4)).

With the noise correlation and latent factor correlation disentangled, the M-step reduces to easy sub-problems. For example, the minimization with respect to Σ_f is a graphical Lasso problem (Friedman et al., 2008) and the minimization with respect to $\Phi^k_{\mathcal{S}}$ and $\Phi^k_{\mathcal{T}}$ is a maximum likelihood estimation of a matrix-variate distribution (Dawid, 1981). We thus obtain an affordable M-step, and alternating E and M-steps produces a solution to the MPLE problem.

We derive the full formulations in Appendix D.1. Its computational cost is inexpensive: a single iteration of E and M-steps on our cluster server (with 11 Intel(R) Xeon(R) CPU 2.90GHz processors) took in average less than 45 seconds, applied to the experimental data in Section 4.3.2. A single fit on the same data took 42 iterations for around 30 minutes until P and $\{\beta^1, \beta^2\}$ converged under threshold 1e-3 and 1e-5, respectively. The code is provided at `https://github.com/HeejongBong/ldfa`.

4.3 Results

One major novelty of our method is its accounting for auto-correlated noise in neural time series to better estimate cross-region associations in CCA type analysis. This is illustrated in Section 4.3.1 based on simulated data. In Section 4.3.2, we apply LDFA-H to experimental data to examine the lead-lag relationships across two brain areas and the spatial distribution of factor loadings.

4.3.1 LDFA-H retrieves cross-correlations even when noise auto-correlations dominate

We simulated $N = 1000$ i.i.d. neural time series X^k of duration $T = 50$ from Eq. (4.1) for brain regions $k = 1, 2$. The latent time series Z^k were generated from Eq. (4.2) with $q = 1$ pair of factors and correlation matrix P_1 depicted in Fig. 4.2a. The noise ϵ^k was taken to be the $N = 1000$ trials of the experimental data analyzed in Section 4.3.2, first permuted to remove cross-region correlations, then contaminated with white noise to modulate the strength of noise correlation relative to cross-region correlations. The resulting temporal noise correlation matrices, found by averaging correlations over all pairs of simulated time series, are shown in Fig. 4.2b, for four levels of white noise contamination. The magnitudes of cross-region correlation and within-region noise auto-correlation are quantified by the determinant of each matrix, known as the generalized variance (Wilks, 1932); their logarithms are provided atop the panels in Fig. 4.2a and Fig. 4.2b. Generalized variance ranges from 0 (identical signals) to 1 (independent signals). Thus, larger negative values indicate stronger within-region noise correlation (see Appendix D.2). Other simulation details are in Appendix D.2.

We note that the simulation does not satisfy some of the model assumptions in Section 4.2. The noise vectors ϵ^k are not matrix-variate distributed as in Eqs. (4.4) and (4.5) and the derived Γ_T^k does not satisfy a banded structure as in Eq. (4.9). Also, the latent partial auto-correlations (Fig. 4.2) are not banded as assumed in Eq. (4.9).

We applied LDFA-H with $q = 1$ factor, $h_{\text{cross}} = 10$, h_{auto} equal to the maximum order of the auto-correlations in the 2000 observed simulated time series, and λ_{cross} determined by 5-fold CV. Fig. 4.3 shows LDFA-H cross-precision matrix estimates corresponding to the four level of noise correlation shown in Fig. 4.2b. They closely match the true Π_1^{12} shown in the right most panel of Fig. 4.2a.

We also applied five other methods to estimate cross-region connections in the simulated data. They include the popular averaged pairwise correlation (APC); correlation of averaged signals

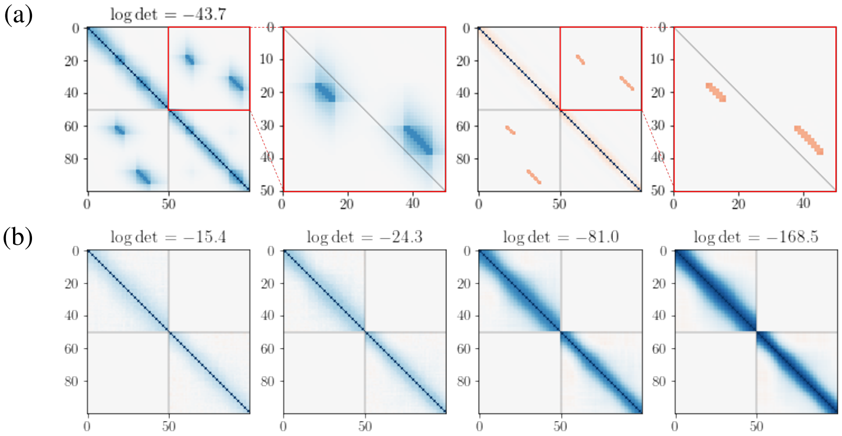


Figure 4.2: **Simulation settings.** **(a)** *(Left to right panels) True correlation matrix* P_1 *for latent factors* $Z^1_{1,:}$ *and* $Z^2_{1,:}$ *from model in Eq.* (4.2)*; close-up of the cross-correlation matrix; corresponding precision matrix* $\Pi_1 = \mathrm{P}_1^{-1}$*; and close-up of cross-precision matrix* Π_1^{12} *(Eq.* (4.3)*). Matrix axes represent the duration,* $T = 50$ *ms, of the time series. Factors* Z^1 *and* Z^2 *are associated in two epochs:* Z^2 *precedes* Z^1 *by* 7*ms from* $t = 13$ *to* 19*ms, and* Z^1 *precedes* Z^2 *by* 7*ms from* $t = 33$ *to* 42*ms.* **(b)** *Noise auto-correlation matrices (Eq.* (4.5)*) for pairs of simulated time series at four strength levels.* $\log\det$ *in (a) and (b) measure correlation strengths.*

(CAS); and CCA (Hotelling, 1936), applied to the NT observed pairs of multivariate random vectors $\{X^1_{:,t}, X^2_{:,t}\}_{n,t\in[N]\times[T]}$ to estimate the cross-correlation matrix between the canonical variables; as well as DKCCA (Rodu et al., 2018) and LaDynS (Bong et al. (2020c)). The first four methods do not explicitly provide cross-precision matrix estimates, so we display their cross-correlation matrix estimates in Fig. 4.4, along with LDFA-H cross-correlation estimates in the last row. It is clear that only LDFA-H successfully recovered the true cross-correlations shown in the second panel of Fig. 4.2a, at all auto-correlated noise levels.

4.3.2 Experimental Data Analysis from Memory-Guided Saccade Task

We now report the analysis of LFP data in areas PFC and V4 of a monkey during a saccade task, provided by Khanna et al. (2020). One trial of the experiment consisted of four stages: (i) fixation: the animal fixated at the center of the screen; (ii) cue: a cue appeared on the screen randomly at one of eight locations; (iii) delay: the animal had to remember the cue location while maintaining eye fixation; (iv) choice: the monkey made a saccade to the remembered cue location. We focused our analysis on the 500 ms delay period, when the animal both processed cue information and prepared a saccade. LFP data were recorded for $N = 1000$ trials by two 96-electrode Utah arrays implanted in PFC and V4, β band-passed filtered, down-sampled from 1 kHz to 100 Hz.

We applied LDFA-H using $h_{\text{auto}} = h_{\text{cross}} = 10$, corresponding to 100 ms (at 100 Hz); the LFP β-power envelopes have frequencies between $12.5Hz$ to $30Hz$, and $h_{\text{auto}} = 10$ enables the slowest filtered signal to complete one full oscillation period. The other tuning parameters were determined by 5-fold CV over $\lambda_{\text{cross}} \in \{0.0002, 0.002, 0.02, 0.2\}$ and $q \in \{5, 10, 15, 20, 25, 30\}$, yielding optimal values $\lambda_{\text{cross}} = 0.02$ and $q = 10$. We also regularize the diagonal elements, due

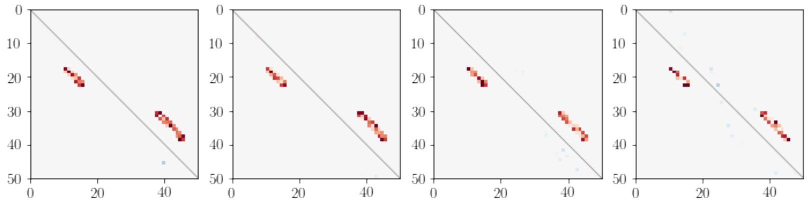


Figure 4.3: **Simulation results: LDFA-H cross-precision matrix estimates.** *Estimates of* Π_1^{12}, *shown in the right-most panel of Fig. 4.2a, using LDFA-H, for the four noise auto-correlation strengths shown in Fig. 4.2b. LDFA-H identified the true cross-area connections at all noise strengths.*

to the otherwise excessively smooth β-power envelopes (see our code or Bong et al. (2020c) for details). The fitted factors were ranked based on the Frobenius norms of their covariance matrices $\|\Sigma_f\|_F^2$; norms are plotted versus f in decreasing order in Fig. D.1, and $\log_{10} \|\Sigma_f\|_F^2$ of the top three factors are provided above each panel in Fig. 4.5a. The estimated cross-precison matrices between two brain regions corresponding to the top three factors are shown in Fig. 4.5a. Note that a positive entry in the precision matrix represents negative association between two regions. We also summarized, for each factor f, the temporal information flow at time t from V4 to PFC and to V4 from PFC with $I_{f,PFC\to V4}(t) = \sum_{t'>t} \left|\widehat{\Pi}^{12}_{f,(t,t')}\right|$ and $I_{f,V4\to PFC}(t) = -\sum_{t'<t} \left|\widehat{\Pi}^{12}_{f,(t,t')}\right|$, respectively, where $\widehat{\Pi}_f$ is the inverse correlation matrix estimate in Eq. (4.9). Fig. 4.5d displays smoothed $I_{f,PFC\to V4}(t)$ and $I_{f,V4\to PFC}(t)$ as functions of t for the top three factors. Lead-lag relationships between V4 and PFC change dynamically over time, and the information flow tends to peak either early in the delay period, when the animal must remember the cue, or later, when it must make a saccade decision. The dominant first factor captures a flow from V4 to PFC centered around 200 milliseconds into the task and a flow from PFC to V4 centered around 320 milliseconds. Factor loadings (subsampled over space) for the 96 V4 and PFC electrodes are shown in Fig. 4.5b and Fig. 4.5c, respectively, for the top three factors (first three columns of the estimate of β^k in Eq. (4.9), with area $k = 1$ being V4 and $k = 2$ being PFC), arranged spatially according to electrode positions on the Utah array. The factors have different spatial modes over the physical space of the Utah array. Confirmation of these patterns would require additional data and analyses.

4.4 Conclusion

To identify dynamic interactions across brain regions we have developed LDFA-H, a nonstationary, multi-group extension of GPFA that allows for within-group spatio-temporal dependence among high-dimensional neural recordings. We applied the method to data during a memory task and found interesting, intuitive results. Although we treated the two-group case, and applied it to interactions across two brain regions, several groups can be handled with straightforward modifications. The approach could, in principle, be applied to many different types of time series, but it has some special features: first, like all methods based on sparsity, it assumes a small number of large effects are of primary interest; second, it uses repetitions, here, repeated trials, to identify time-varying dependence; third, because the within-group spatio-temporal structure is not of interest, the method can remain useful even with some modest within-group model misspecification.

Several restrictive assumptions of LDFA-H, as defined, were helpful here but could be modified for other applications. One is the Kronecker-product form of the noise process. In our simulation study, using a realistic scenario, we showed that LDFA-H can be effective even when the Kronecker-product assumption is violated, but in other cases it may be problematic. In some problems, space and/or time can be decomposed into windows within which the assumption is more reasonable (see Leng and Tang, 2012; Zhou, 2014). Another potentially bothersome assumption is independence between latent factors. It would be possible to include covariance matrix parameters between the factors, but then the model will get computationally prohibitive even with a moderate factor size. State-space models (Buesing et al., 2014; Linderman et al., 2019; Yang et al., 2016) have potential but, to be comparable to LDFA-H, they would have to accommodate nonstationary lead-lag behavior. Computationally efficient methods for identifying time-varying relationships is a vital goal in the analysis of neural data from multiple brain regions.

We applied LDFA-H to LFP data. In contrast, GPFA has been applied mainly to neural spike count data, and it is of course possible to apply LDFA-H to spike counts, as well. However, we have been struck by the strong attenuation of effects due to Poisson-like noise, as discussed in Vinci et al. (2018) and references therein. A version of LDFA-H built for Poisson-like counts, or for point processes, could be the subject of additional research. It may also be advantageous to model spatial dependence explicitly, perhaps based on physical distance between electrodes, analogously to what was done in Vinci et al. (2018), and there may be, in addition, important simplifications available in the temporal structure. It would also be helpful to have additional statistical inference procedures for assessing effects. In the future, we hope to pursue these possible directions, and refine the application of this promising approach to the analysis of high-dimensional neural data.

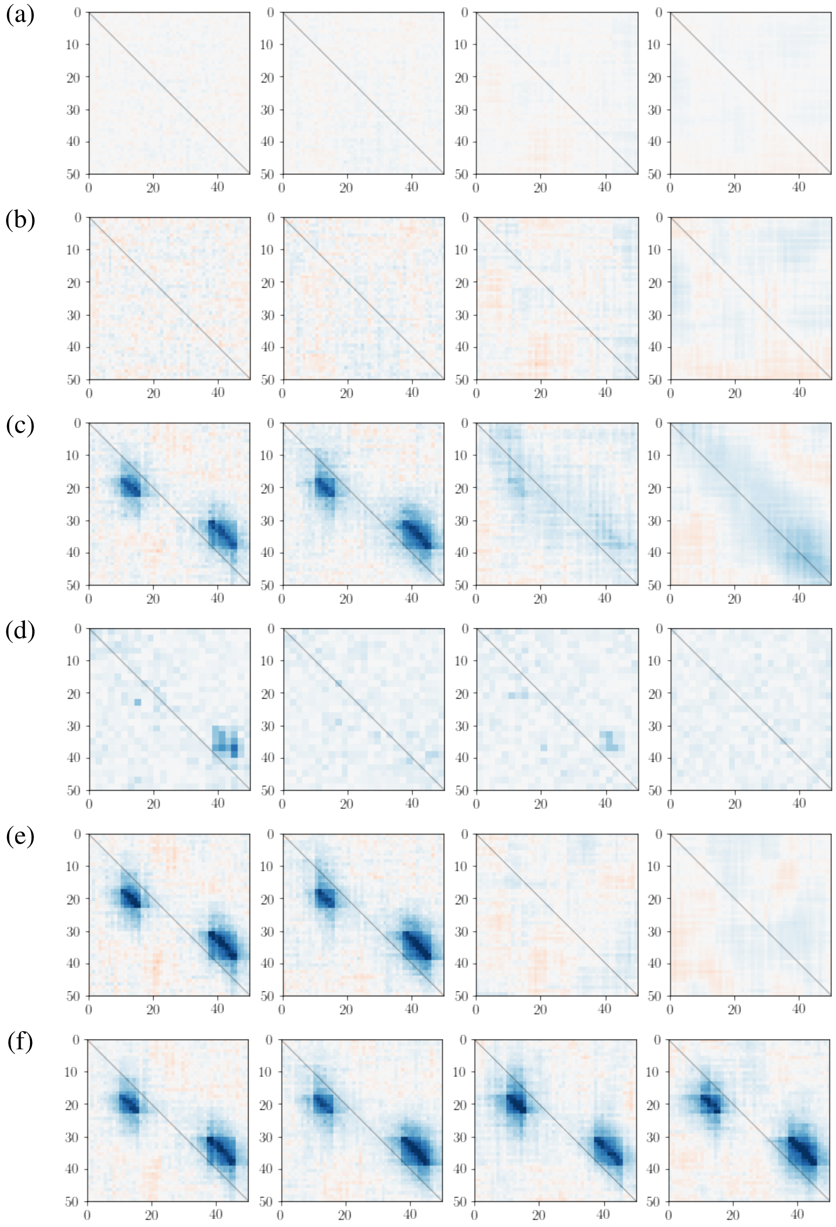


Figure 4.4: **Simulation results: cross-correlation matrix estimates.** *Estimates of* Σ_1^{12} *under four noise correlation levels using* **(a)** *averaged pairwise correlation (APC),* **(b)** *correlation of averaged signal (CAS),* **(c)** *canonical correlation analysis (CCA, Hotelling (1936)),* **(d)** *dynamic kernel CCA (DKCCA, Rodu et al. (2018)),* **(e)** *LaDynS (Bong et al. (2020c)), and* **(f)** *LDFA-H. Only LDFA-H successfully recovered the true cross-correlation at all noise auto-correlation strengths.*

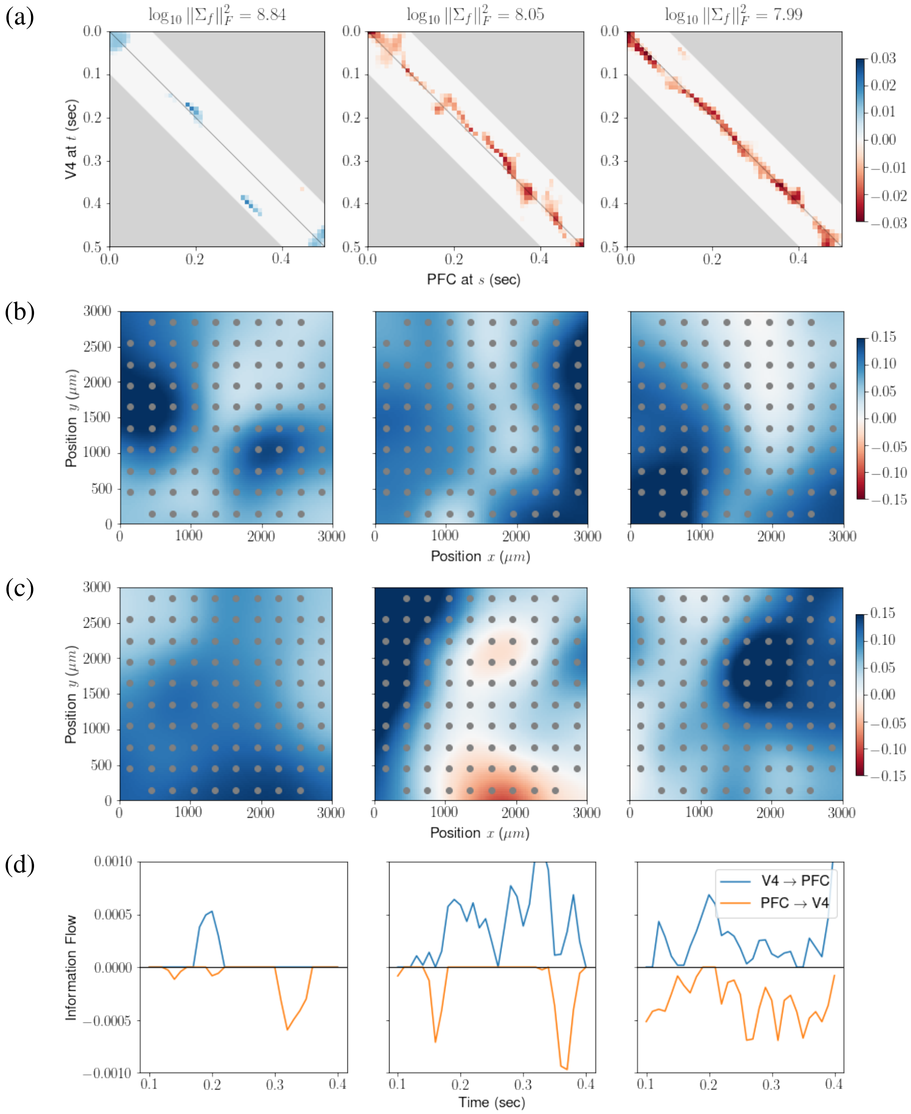


Figure 4.5: **Experimental data results for the top 3 factors.** *(a) Cross-precision matrices. Blue represents positive precision matrix entries, corresponding to negative association. Factors have different connectivity patterns over the experimental trials.* $\log_{10} \|\Sigma_f\|_F^2$*, written atop the panels, measures the strength of each factor. The first factor is more than 6 times larger than the second and third, and displays activity in V4 leading PFC centered around 200 milliseconds and activity in PFC leading V4 centered around 320 milliseconds post cue disappearance. This is also shown in panel (d). (b,c) Factor loadings, smoothed and color coded, plotted on the electrode coordinates* (μm)*. Here, positivity is arbitrary, due to identifiability. Panels (b) and (c) display loadings for the V4 and PFC arrays, respectively. The first factor has activity in V4 centered in two distinct subregions of the array, while activity in PFC is more broadly distributed. (d) Dynamic information flow in the directions* $V4 \rightarrow PFC$ *(blue) and* $PFC \rightarrow V4$ *(orange).*

Chapter 5

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

This chapter is a collaboration with Konrad Urban, Josue Orellana and Robert E. Kass. This work is submitted to Canadian Journal of Statistics and currently under review.

5.1 Introduction

Oscillations in neural circuits have been observed under variety of circumstances, provoking much speculation about their physiological function (Buzsaki and 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 the incorporation of 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 (Schmidt et al., 2019; Miller et al., 2018), and this has led to the idea that alterations of circuit oscillations could indicate brain dysfunction (Mathalon and Sohal, 2015).

From a statistical perspective, regardless of their mechanistic function, neural oscillations can be considered useful indicators of coordinated activity across brain regions. The data we have analyzed, such as that shown in Figure 5.1, come from experiments during which animals are shown a visual stimulus, or they perform a task, repeatedly, across many trials, while neural activity is recorded from electrodes inserted into the brain. If on each trial a pair of oscillatory time series is observed, from two distinct brain regions, and the two oscillation phase measurements rotate forward or backward together, exhibiting correlation, it is a strong indication that the two brain regions are involved in an oscillating neural circuit. Because the phase measurements are noisy angles, an immediate statistical question is, How can such a correlation be established? A natural second question is, How can a graph be constructed to represent interactions across multiple parts of the brain? Neurophysiological investigations often use many electrodes in each

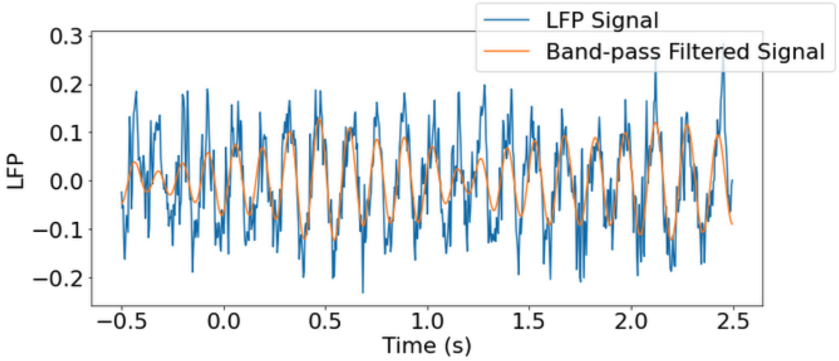


Figure 5.1: Three seconds of data from one electrode (a local field potential or LFP in blue) and a band-pass filtered version of the data (orange).

of two or more specific brain regions, resulting in vectors of oscillating time series in each region, which raises the further question, How can concise and interpretable summaries be used to assess the covariation across groups of oscillating series? The purpose of this paper is to answer these questions under the special circumstances relevant to the multielectrode neural context. We draw results from the literature, including our own work, and also supply several novel observations, re-formulations, and theoretical results, along with one new method. Although we have been motivated by the analysis of neural data, we believe the paper 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.

For stationary time series, coherence is in some important respects a natural, frequency-based analogue to correlation (Brillinger, 2001). Furthermore, graphical models can be constructed from coherence in a manner analogous to Gaussian graphical models (Bach and Jordan, 2004), and their applicability is extended by weakening the assumption from stationarity to local stationarity (Ombao and Van Bellegem, 2008). In the context of task performance, however, many neural signals exhibit striking non-stationarity (e.g. see the introduction to Kass et al. (2018)), and the simplest way to avoid non-stationary effects is to pick a single time point of interest (or several time points) during the task and analyze the covariation of phases across repeated trials. This may be carried out using a circular analogue of Pearson correlation known as Phase-Locking Value (PLV; see Section 5.5.2).

For more than two phases, a recently-developed class of methods uses models called torus graphs (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. Due to their construction as exponential families, torus graphs are probabilistic graphical models. In Section 2 we summarize a few important results about torus graphs. One important consideration is that torus graphs, in general, allow two kinds of association, called rotational, corresponding to positive association, and reflectional, corresponding to negative association, and both are complex-valued. Thus, unlike Pearson correlation, which involves a single real scalar, the general circular analogue of correlation involves two complex numbers. As shown in Section 5.5.2, PLV becomes an analogue of Pearson correlation when reflectional association is absent and the individual circular variables are uniformly distributed. The appearance of a complex measure for circular variables has its origin in the complex representation of the unit circle, as the set of all points of the form $e^{i\theta}$.

Both coherence and PLV are used to measure association of phases in analyzing neural oscillations. The power of an oscillatory signal is its squared amplitude, and as the amplitude evolves across time it is usually called the power envelope for the oscillation. The association of time-varying power in two or more brain regions is of interest, as evidence of association would again (as with phase coupling) indicate joint participation in an oscillatory neural circuit. At a single time point, the ordinary Pearson correlation serves as a measure of association of amplitudes across repeated trials, but it is also possible to assess the association of the power envelopes across time. In Section 5.3 we briefly summarize work we've done to address two challenges: the large numbers of electrodes, which results in large numbers of times series, in each brain region, together with non-stationarity. The method is based on a latent non-stationary time series model.

Given our interest in associations for both phase and amplitude of neural oscillations, it may seem natural to combine them by introducing a complex random vector model, because a complex random variable can be written in polar coordinate form in terms of phase and amplitude (using the general representation for a complex number $z = Re^{i\theta}$). We investigated the use of the complex normal in this context, and provide its definition in Section 5.4. The covariance of two complex random variables X and Y is $\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}\, X)(\overline{Y - \mathrm{E}\, Y})]$, where $\overline{z}$ is the conjugate of a complex number z. This covariance is itself a complex number. The variance then becomes a real number and the complex correlation is defined as in Pearson correlation, by dividing the covariance by the product of the two standard deviations. It is, thus, in general, complex-valued. The second moments that appear in the complex normal include not only the complex covariances but also what are called pseudo-covariances. For two complex random variables X, Y, the pseudo-covariance is defined as $\mathrm{PCov}(X, Y) = \mathrm{E}[(X - \mathrm{E}\, X)(Y - \mathrm{E}\, Y)]$. The covariance and pseudo-covariance for a complex random vector X are then defined analogously to the case of complex random variables, using the Hermitian and transpose operators, $\mathrm{Cov}(X) = E[(X - \mathrm{E}\, X)(X - \mathrm{E}\, X)^H]$ and $\mathrm{PCov}(X) = E[(X - \mathrm{E}\, X)(X - \mathrm{E}\, X)^T]$. When its pseudo-covariance is zero, a complex normal distribution is called *proper*. We show that, under an intuitive restriction on the covariance structure, when we write a complex normal vector in polar coordinate form, the conditional distribution of the vector of phases given the amplitudes forms a torus graph; when the complex normal is proper the torus graph exhibits only rotational dependence; and when, in addition, the mean of the complex normal is zero, the torus graph components are all uniformly distributed. We also note that each of these restrictions on a complex normal family induces a full exponential subfamily. These results are of interest as characterizations of torus graphs, and they shed some light on potential implicit assumptions.

In practice, to extract a phase at a particular frequency either wavelet or Fourier methods are used. This is usually called band-pass filtering, the name adopted in Fourier analysis (see Figure 5.1). Here, though, the filtered signal must be complex-valued. Thus, if real-valued Fourier-based band-pass filtering is applied, it must be followed by the additional step of conversion to the corresponding complex-valued version, which is usually accomplished with the Hilbert transform (see Appendix A).

Coherence is often considered to be the magnitude of the complex correlation at a particular frequency. This is a verbal description of its definition using the Cramér (or Cramér-Khinchin) representation for a bivariate complex stationary process, as displayed in Section 5.5.1. One way to understand this interpretation is to consider the result of filtering the signals with a complex band-pass filter in a band $(\omega_0 - \delta, \omega_0 + \delta)$, for small δ. Ombao and Van Bellegem (2008) note that the resulting narrow-band coherence is the correlation of the filtered signals. We can effectively pass to the limit as $\delta \to 0$ to get the infinitesimal version that appears in the Cramér-Khinchin representation. In a similar vein, we also note that a single-frequency bivariate process with stochastic amplitudes and phases yields coherence as magnitude of complex correlation.

PLV, defined in Section 5.5.2, summarizes a sample of phase pairs by measuring the concentration of the phase differences: the two phases are thereby considered highly associated when their difference is nearly constant across trials. Unlike PLV, coherence is not solely a measure of association of phases; because coherence is based on complex-valued correlation, it also involves amplitudes. In Section 5.5.3, returning to the single-frequency bivariate process, under the assumption that the amplitudes are independent of the phases and the phases are uniformly distributed, we show that as long as the standard deviations of the amplitudes are small relative to their means, coherence and PLV are essentially equivalent. On the other hand, this simplified representation makes it easy to find situations when the two measures can disagree. We also summarize the small literature on their comparison in the analysis of neural oscillations.

Although we had developed methods for analyzing both phase and amplitude associations across brain regions, the use of complex correlation in coherence suggested the possibility of analyzing neural oscillations by combining amplitude and phase through the complex normal distribution. Because we often face large numbers of electrodes in each region we considered a latent variable model. We specify the model in Section 5.6 and apply it to data in Section 5.7. 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 1 KHz, so that each second of data has 1,000 observations. LFPs represent bulk activity near the electrode (roughly within 150-200 microns), involving large numbers of neurons (Buzsáki et al. (2012); Einevoll et al. (2013); Pesaran et al. (2018)).

Appendices A, B, and C contain, respectively, a note on the Hilbert transform, proofs of the new theorems, and details of the fitting procedure for the complex latent variable model.

5.2 Torus Graphs

Figure 5.2 shows simulated, repeated observations from a pair of circular random variables X_1 and X_2. (The simulation displays features of real data, but mildly exaggerated to make our point more vivid.) Most points lie along the diagonal, as might be expected of correlated data, but some appear in the upper left corner of the figure. In the figure, X_1 tends to be slightly ahead of X_2, so as X_1 rotates past π and jumps to $-\pi$, X_2 will not yet have gotten to π. This produces points in the upper left corner of the plot, which is an indication that rectangular coordinates do not do a good job of representing such data.

Instead of thinking about circular random variables on the segment $[-\pi, \pi)$, we can consider their representation as unit-length complex random variables $Y_1 = e^{iX_1}$ and $Y_2 = e^{iX_2}$. The set of such pairs is the product of two circles, a 2-dimensioal torus. In this representation the problem

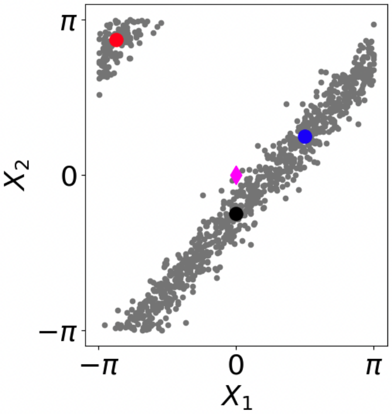


Figure 5.2: Scatter plot showing rotational (positive) dependence. The magenta-diamond marker is at (0,0), black and blue markers are positioned along the diagonal band of points which looks similar to bivariate normal data, except this pattern is not elliptical and the margins are pretty uniform. Also the points near the red marker, on the upper-left corner, look surprising. They appear because, in Cartesian (x, y) coordinates, the angles must jump from π to $-\pi$ as they wrap around the circular domain. Figure has been adapted from a figure in Klein et al. (2020).

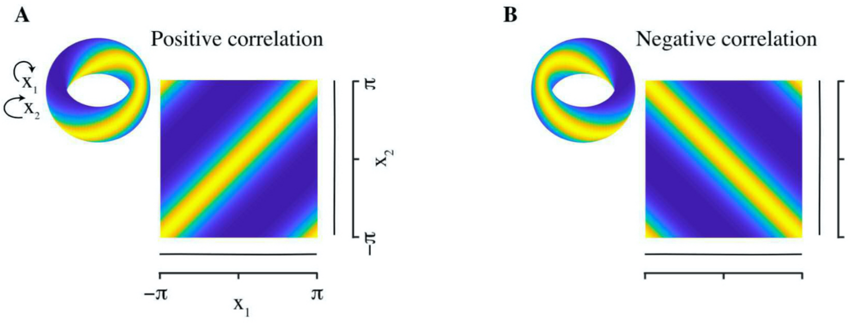


Figure 5.3: The torus is the natural domain for a pair of circular random variables. Illustration of both type of circular covariance in a torus graph bivariate density, with uniform marginal densities, plotted side-by-side on the 2d-torus and cartesian plane. A. Positive rotational dependence. B. Negative reflectional dependence. Figure has been adapted from a figure in Klein et al. (2020).

of weird points due to angular wrapping goes away: Figure 5.3 displays bivariate torus graph densities plotted side-by-side on the torus and on the Cartesian plane.

Using the notations $\overline{C} = E(\cos(X))$ and $\overline{S} = E(\sin(X))$ the first moment for Y_1 can be obtained as

$$\begin{aligned}
E[Y_1] &= E[e^{iX_1}] \\
&= E[\cos(X_1) + i\sin(X_1)] \\
&= E[\cos(X_1)] + i\,E[\sin(X_1)] \\
&= \overline{C_1} + i\overline{S_1} \\
&= R_1 e^{i\mu_1},
\end{aligned}$$

where $R_1 = \left(\overline{C_1}^2 + \overline{S_1}^2\right)^{(1/2)}$ and $\mu_1 = \text{atan2}(\overline{S_1}, \overline{C_1})$. Similarly, $E[Y_2] = R_2 e^{i\mu_2}$. The usual complex covariance is given by

$$\begin{aligned}\text{Cov}(Y_1, Y_2) &= E\left[(Y_1 - E[Y_1])\left(\overline{Y_2 - E[Y_2]}\right)\right]] \\ &= E\left[Y_1\overline{Y_2}\right] - E[Y_1]\,E[\,\overline{Y_2}\,] \\ &= E\left[e^{iX_1}e^{-iX_2}\right] - R_1e^{i\mu_1}\,R_2e^{-i\mu_2} \\ &= E\left[e^{i(X_1-X_2)}\right] - R_1R_2e^{i(\mu_1-\mu_2)}.\end{aligned}$$

However, we also need the pseudo-covariances defined in the introduction, Section 1. In the context of circular data, the covariance and pseudo-covariance are renamed *rotational* and *reflectional*, the latter because it amounts to reflecting an axis:

$$\begin{aligned}\text{Cov}(Y_1, Y_2)_{\text{rotational}} &= \text{Cov}(Y_1, Y_2) \\ \text{Cov}(Y_1, Y_2)_{\text{reflectional}} &= \text{PCov}(Y_1, Y_2)\end{aligned}$$

where

$$\begin{aligned}\text{PCov}(Y_1, Y_2) &= E\left[(Y_1 - E[Y_1])\,(Y_2 - E[Y_2])\right] \\ &= E\left[Y_1Y_2\right] - E[Y_1]\,E[Y_2] \\ &= E\left[e^{iX_1}e^{iX_2}\right] - R_1e^{i\mu_1}\,R_2e^{i\mu_2} \\ &= E\left[e^{i(X_1+X_2)}\right] - R_1R_2e^{i(\mu_1+\mu_2)}.\end{aligned}$$

If rotational covariance measures clockwise-clockwise association, then reflectional covariance measures clockwise-anticlockwise association. When a full exponential family is constructed on the torus (we give the pdf below), both kinds of association must be considered. As a consequence, association of two phases is, in general, represented not by a single scalar as in Pearson correlation, which can represent both positive and negative dependence, but instead by 2 complex numbers (Park and Park, 2018).

As shown in Figure 5.3, the two types of circular covariance measure complementary patterns of linear dependence. The amplitude of the rotational covariance controls the width of the association-band, and the phase of this complex number specifies the shift of this band along the diagonal of the Cartesian plane that runs from bottom-left to top-right (Figure 5.3A). The same notion applies to the reflectional covariance magnitude, but the band is positioned on the other diagonal (Figure 5.3B). The presence of both types of covariation is possible and can lead to concentrated marginals; however, in our experience with phases extracted from LPF neural data, we have almost always observed exclusively rotational dependence. In addition, the marginals are close to uniform. Rotational covariance, with uniform marginals, is the type of covariation measured by PLV, and it also appears in models derived from dynamical systems of coupled oscillators (Cadieu and Koepsell, 2010).

To define an exponential family on a torus with mean and covariance structure, the first and second-order sufficient statistics are needed. Using two-dimensional rectangular coordinates (involving cosines and sines), the first-order sufficient statistics are $U_1 = [\cos(x_1), \sin(x_1)]^T$ and

$U_2 = [\cos(x_2), \sin(x_2)]^T$ and the second-order behavior is summarized by

$$U_1 U_2^T = \begin{bmatrix} \cos x_1 \cos x_2 & \cos x_1 \sin x_2 \\ \sin x_1 \cos x_2 & \sin x_1 \sin x_2 \end{bmatrix}.$$

We can then write a natural exponential family density in the form

$$p(\mathbf{x}; \boldsymbol{\eta}) \propto \exp\left(\boldsymbol{\eta}_1^T \begin{bmatrix} \cos x_1 \\ \sin x_1 \end{bmatrix} + \boldsymbol{\eta}_2^T \begin{bmatrix} \cos x_2 \\ \sin x_2 \end{bmatrix} + \boldsymbol{\eta}_{12}^T \begin{bmatrix} \cos x_1 \cos x_2 \\ \cos x_1 \sin x_2 \\ \sin x_1 \cos x_2 \\ \sin x_1 \sin x_2 \end{bmatrix} . \right) \tag{5.1}$$

Extending Equation (5.1) to $d > 2$, for an exponential family on a d-dimensional torus, yields

$$p(\mathbf{x}; \boldsymbol{\eta}) \propto \exp\left(\sum_{j=1}^{d} \boldsymbol{\eta}_j^T \begin{bmatrix} \cos x_j \\ \sin x_j \end{bmatrix} + \sum_{j<k} \boldsymbol{\eta}_{jk}^T \begin{bmatrix} \cos x_j \cos x_k \\ \cos x_j \sin x_k \\ \sin x_j \cos x_k \\ \sin x_j \sin x_k \end{bmatrix} \right). \tag{5.2}$$

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

$$p(\mathbf{x}; \eta(\boldsymbol{\phi})) \propto \exp\left(\sum_{j=1}^{d} \boldsymbol{\phi}_j^T \begin{bmatrix} \cos x_j \\ \sin x_j \end{bmatrix} + \sum_{j<k} \boldsymbol{\phi}_{jk}^T \begin{bmatrix} \cos(x_j - x_k) \\ \sin(x_j - x_k) \\ \cos(x_j + x_k) \\ \sin(x_j + x_k) \end{bmatrix} \right), \tag{5.3}$$

which uses the same first and second order statistics previously seen in the definitions of complex covariance. Klein *et al.* then defined a *d-dimensional torus graph* to be any member of the family of distributions specified by Equations (5.2) or (5.3) (Klein et al., 2020).

Because a torus graph distribution is an exponential family, it follows that a pair of random variables X_i and X_j will be conditionally independent given all other variables if and only if the four pairwise interaction terms 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 measures such as PLV because for each edge they only consider two random variables in isolation (see Figure 5.4, which displays an example of conditional independence graphs from (Klein et al., 2020). The Torus graph model as seen in Equation (5.3) is composed of three parameter groups: marginal concentrations, rotational covariance, and reflectional covariance. Klein *et al.* showed that the sub-model with only rotational covariance parameters 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.

In unpublished work, 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

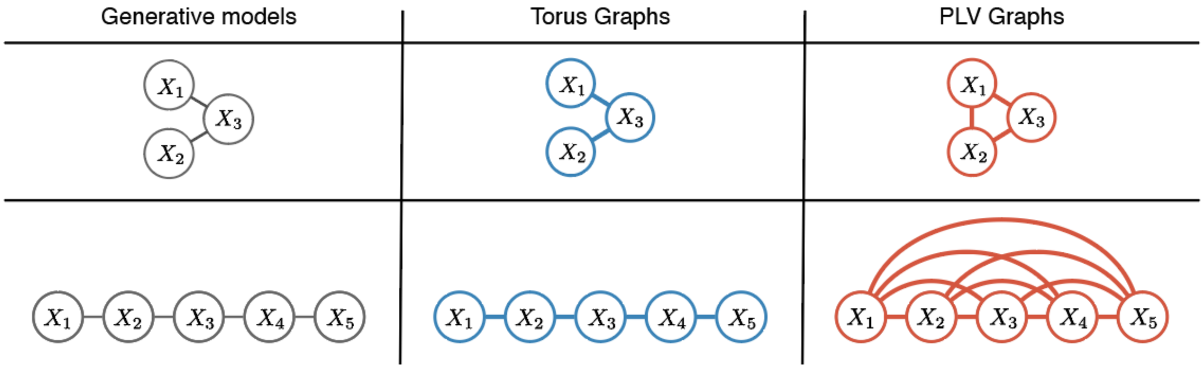


Figure 5.4: Torus Graphs is able to recover conditional dependence graphs but PLV fails. The generative model used simulated data using an independent model; see Klein et al. (2020) for details. Figure has been adapted from a figure in Klein et al. (2020).

structure has the advantage of reducing the number of total parameters, which results in better statistical inference when the amount of data is small in comparison to its dimensionality.

Because the normalization constant is intractable, some methods, including maximum likelihood, are not useful in fitting a torus graph. Klein *et al.* used the relatively recent approach of score matching (Dawid and Musio (2014); Hyvärinen and Dayan (2005); Yu et al. (2016)) to define a convex cost function that avoids the normalization constant. It is interesting to realize that, although torus graphs provide interesting examples of exponential families that could have been defined long ago, they would have been irrelevant to data analysis before the more recent practical developments.

5.3 Latent Dynamical Model

In Figure 5.1, the band-pass filtered signal has a time-varying amplitude. This is a segment from a single trial, and it is of interest to examine trial-to-trial covariation of signals from two brain regions: if the amplitudes tend to get larger or smaller together, then whatever it is that makes one region's signals more powerful, also makes the other region's signals more powerful, providing evidence that the two regions are part of an oscillating circuit. Furthermore, it is possible that an amplitude at one time point in one region is correlated with an amplitude at a slightly later (lagged) time point in the other region. It is important to emphasize the non-stationarity of the covariation, which can only be estimated accurately when the data come from a large number of trials.

To assess time-varying interaction based on multiple such signals recorded from each region, Bong et al. (2021) proposed a method of analyzing these smoothly-varying amplitude time series. They developed a time-series generalization of a factor analysis model, using one latent factor for each region. In their generalization, the two latent factors together formed a non-stationary bivariate time series. In this model, the covariance matrix for the latent bivariate time series summarizes the interactions, including all available time-lagged cross-correlations. They first showed that maximum likelihood in this model (with a suitable identifiability constraint) produced the same solution as a well-known version of multiset canonical correlation analysis (CCA). The method is thus is a time series generalization of probabilistic CCA. Bong *et al.* then considered the problem of estimating the latent time series covariance matrix when there were substantial numbers of electrodes, and thus time series, in each region. In the data they analyzed there were 96 electrodes in each region. Because of the non-stationarity, each combination of time point in one region and time point in the other region could have its own unique correlation across trials, which made the covariance matrix have a large number of free parameters. Bong *et al.* adapted methods from the literature on high-dimensional inference to obtain a reasonable estimate, and they studied its properties. They then showed how it produced interpretable and interesting results when applied to their data.

5.4 The Complex Normal Distribution

A complex random vector $X \in \mathbb{C}^d$ is said to be complex normal if its real and imaginary parts are jointly multivariate normal (Andersen et al., 1995). With $m = \mathrm{E}(X)$, $\Gamma = \mathrm{Cov}(X)$, and $C = \mathrm{PCov}(X)$, we write this as $X \sim CN_d(m, \Gamma, C)$. The distribution of X is said to be *proper* if the pseudo-covariance matrix C is identically 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 α, if and only if X is proper and has mean zero (Adali et al., 2011).

We can view the circularly symmetric, proper, and restriction-less complex normal distributions as full regular exponential families. We provide an abbreviated version of this theorem here; for full details, please see the appendix.

Theorem 5.4.1. *The family of complex normal distributions represented by* $CN(\mu, \Gamma, C)$ *forms a full and regular exponential family. Additionally, the family of proper CN distributions* $CN(\mu, \Gamma, 0)$ *as well as the family of circularly symmetric CN distributions* $CN(0, \Gamma, 0)$ *both form full and regular exponential families.*

We can write the components of a complex normal random vector in polar coordinates and form the conditional distribution of the vector of phases given their amplitudes. Under certain parameter restrictions on a complex normal distribution, this conditional distribution forms a torus graph. These restrictions avoid including second-moment terms, such as $(\cos\Theta_i)^2$, $\cos\Theta_i \sin\Theta_i$ and $(\sin\Theta_i)^2$, that do not contribute to interactions. In the following theorems, when X is a complex random vector, we write the phase and amplitude of its ith component as the pair $(\Theta_i, R_i) = (\arg X_i \mod 2\pi, |X_i|)$.

Theorem 5.4.2. *Suppose* $X \sim CN_d(m, \Gamma, C)$ *with* $(\mathfrak{Re}X, \mathfrak{Im}X) \sim N(\mu, \Sigma)$*. If* $\Sigma^{-1}_{ii} = \Sigma^{-1}_{i+d,i+d}$ *for all* $1 \leq i \leq d$ *and* $\Sigma^{-1}_{i,i+d} = 0$ *for all* $1 \leq i \leq d$*, then* $\Theta | R = r \sim TG(\eta)$ *where* $\eta_i = r_i[(\Sigma^{-1}\mu)_i, (\Sigma^{-1}\mu)_{i+d}]$ *and* $\eta_{ij} = \frac{-1}{2} r_i r_j [\Sigma^{-1}_{i,j}, \Sigma^{-1}_{i,j+d}, \Sigma^{-1}_{i+d,j}, \Sigma^{-1}_{i+d,j+d}]$.

There exists another distribution in the literature, known as the multivariate Generalized von-Mises (mGvM) distribution, which is a more general version of the TG distribution that includes all second moment terms (Navarro et al., 2017). Navarro and colleagues showed that given a $2D$-dimensional normal distribution analogous to the complex normal, $\Theta | R = 1$ is a mGvM distribution without requiring any restrictions on the parameters of the original normal distribution. In the appendix, we generalize this result for any conditional distribution $\Theta | R = r$

and then leverage this result to prove theorem 5.4.2 (see appendix section E.2.2).

We can also study the conditional distributions under the parameter restrictions $C = 0$ and $m = 0$. Doing so allows us to get the following sequence of results under propriety and circular symmetry.

Theorem 5.4.3. *If $X \sim CN_d(m, \Gamma, 0)$, so that the complex normal distribution of X is proper, then $\Theta | R = r \sim TG(\eta(\phi))$, as in Equation (3), where $(\phi_{ij})^3 = (\phi_{ij})^4 = 0$, with $(\phi_{ij})^k$ denoting the kth component of ϕ_{ij}.*

Corollary 5.4.4. *If $X \sim CN_d(m, \Gamma, 0)$, so that the complex normal distribution of X is circularly symmmetric, then $\Theta | R = r \sim T(\eta(\phi))$ with the same restrictions as in theorem 5.4.3 and with the additional restriction that $\phi_i = 0$ for all i.*

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 5.2 of reflectional covariance as pseudo-covariance. Under circular symmetry, because $\phi_i = 0$, each component $\Theta_i | R = r$ 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 bandpass-filtered signals with sufficiently narrow bands are proper; this is a direct consequence of the discussion of narrow-band signals given by Picinbono (1994). In the context of the particular neural data application reported here, we provide empirical evidence for circularity in section 5.7.1.

5.5 Coherence and PLV

5.5.1 Coherence and Complex Correlation

Suppose we have two stationary signals $X_1(t), X_2(t)$ with auto-covariance functions $\Sigma_{kk}(t) = \text{Cov}(X_k(t_0 + t), X_k(t_0))$, for $k = 1, 2$, and a cross-covariance function $\Sigma_{12}(t) = \text{Cov}(X_1(t_0 + t), X_2(t))$ (these definitions do not depend on t_0 due to stationarity). The spectrum and cross-spectrum of the signals are

$$f_{kk}(\omega_0) = \int_0^1 \Sigma_{kk}(t) e^{-i2\pi\omega_0 t} dt$$
$$f_{12}(\omega_0) = \int_0^1 \Sigma_{12}(t) e^{-i2\pi\omega_0 t} dt$$

and the coherency is

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

The coherency is complex-valued. The coherence is the magnitude of the coherency. The Cramér-Khinchin decomposition (see Ch. 3 of Brémaud (2014)) is

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \int_{-0.5}^{0.5} \exp(i2\pi\omega t) d \begin{pmatrix} Z_1(\omega) \\ Z_2(\omega) \end{pmatrix},$$

where $(Z_1(\omega_0), Z_2(\omega_0))^\top$ is a bivariate orthogonal increment random process. In this case, the coherency between X_1 and X_2 at frequency ω_0 is often considered, intuitively, as the complex-valued correlation coefficient between the infinitesimal increments of Z_1 and Z_2 at ω_0,

$$\tau_{12}(\omega_0) = \mathrm{Corr}[dZ_1(\omega_0), dZ_2(\omega_0)].$$

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

$$X_1(t) = A_1 \exp(i(\Theta_1 + 2\pi\omega_0 t)) \text{ and } X_2(t) = A_2 \exp(i(\Theta_2 + 2\pi\omega_0 t)) \tag{5.4}$$

where A_k and Θ_k are random variables, which we think of as representing the trial-specific amplitude and phase of X_k, respectively, for $k = 1, 2$. Then, the auto-covariance and cross-covariance kernels are $\Sigma_{kk}(t) = \mathrm{Var}[A_i e^{i\Theta_k}]e^{i2\pi\omega_0 t}$ and $\Sigma_{12}(t) = \mathrm{Cov}[A_1 e^{i\Theta_1}, A_2 e^{i\Theta_2}]e^{i2\pi\omega_0 t}$. The spectrum and cross-spectrum at frequency ω_0 are

$$f_{kk}(\omega_0) = \int_0^1 \mathrm{Var}[A_k e^{i\Theta_k}]e^{i2\pi\omega_0 t}e^{-i2\pi\omega_0 t}dt = \mathrm{Var}[A_k e^{i\Theta_k}] \text{ for } k = 1, 2, \tag{5.5}$$

and

$$f_{12}(\omega_0) = \int_0^1 \mathrm{Cov}[A_1 e^{i\Theta_1}, A_2 e^{i\Theta_2}]e^{i2\pi\omega_0 t}e^{-i2\pi\omega_0 t}dt = \mathrm{Cov}[A_1 e^{i\Theta_1}, A_2 e^{i\Theta_2}], \tag{5.6}$$

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

$$\tau_{12}(\omega_0) = \frac{f_{12}(\omega_0)}{\sqrt{f_{11}(\omega_0)f_{22}(\omega_0)}} = \mathrm{Corr}[A_1 e^{i\Theta_1}, A_2 e^{i\Theta_2}] = \mathrm{Corr}[X_1(t), X_2(t)]. \tag{5.7}$$

As a generalization of the single-frequency case, it is possible to define "band coherency" over a narrow band of frequencies. If a pair of signals are band-pass filtered, then the band-coherency of the filtered signals is equal to the complex correlation of the filtered signals (Ombao and Van Bellegem, 2008).

5.5.2 PLV as an Analogue of Pearson Correlation

For a sample of pairs of phase measurements $(\theta_{1n}, \theta_{2n})$, with $n = 1, \ldots, N$, phase-locking value is defined as

$$PLV = \left| \frac{1}{N} \sum_{n=1}^{N} e^{i(\theta_{1n} - \theta_{2n})} \right|.$$

In Section 5.2 we noted that the rotational covariance of two random complex unit vectors $Y_1 = e^{iX_i}$ and $Y_2 = e^{iX_2}$ is

$$\mathrm{Cov}(Y_1, Y_2)_{\text{rotational}} = E\left[e^{i(X_1 - X_2)}\right] - R_1 R_2 e^{i(\mu_1 - \mu_2)}.$$

When X_1 and X_2 are uniformly distributed, the means of Y_1 and Y_2 are equal to zero. Thus, in this special case, $R_1 = R_2 = 0$ and we have

$$\mathrm{Cov}(Y_1, Y_2)_{\mathrm{rotational}} = E\left[e^{i(X_1 - X_2)}\right].$$

As a result, when the phases are written as random complex unit vectors, if there is no relectional covariance and the random phases are individually uniformly distributed, PLV becomes a sample-based (moment) estimator of their correlation. Klein *et al.* (2020) reached the same conclusion starting from a general notion of circular correlation (in their supplementary section S7).

5.5.3 Comparing Measures of Covariation

Since PLV was first introduced, people have debated whether to use PLV, coherence, or amplitude correlation to measure connectivity between oscillatory neural signals (e.g. Lachaux et al. (1999); Lepage and Vijayan (2017); Srinath and Ray (2014); Lowet et al. (2016)). Our intention here is to briefly review some of these arguments, and then show a few toy examples illustrating the diverse ways in which various types of dependence are captured or not captured by amplitude covariance, PLV, and complex covariance. As discussed in the previous section, for stationary band-pass oscillatory signals, complex correlation is nearly equivalent to coherence.

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 phases of oscillatory signals. Phase-locking value was introduced in part to overcome the perceived limitations of this dependence (Lachaux et al., 1999). Later work showed that particularly in non-stationary settings, estimators of coherence are not well-behaved and can fail to accurately represent snchrony (Lowet et al., 2016). Other authors have criticized coherence on the grounds that it can be biased by amplitude correlation (Srinath and 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 and Vijayan, 2017).

It is helpful to reconsider the single-frequency signals $X_k(t) = A_k \exp(i(\Theta_k + 2\pi\omega_0 t))$ discussed in Section 5.5.1. Writing the purely angular factor in the rotational phase correlation (which is the factor estimated by PLV) as

$$\mathrm{PLV}_{12}(\omega) = \left|\mathrm{E}[e^{i(\Theta_1 - \Theta_2)}]\right|,$$

in the case that $(A_1, A_2) \perp (\Theta_1, \Theta_2)$, Equations (5.5), (5.6) give:

$$\begin{aligned} f_{kk}(\omega) &= \mathrm{E}[A_k^2] = \mathrm{Var}[A_k] + \mathrm{E}[A_k]^2 \\ |f_{12}(\omega)| &= \mathrm{E}[A_1 A_2]\left|\mathrm{E}[e^{i(\Theta_1 - \Theta_2)}]\right| = (\mathrm{Cov}[A_1, A_2] + \mathrm{E}[A_1]\,\mathrm{E}[A_2])\mathrm{PLV}_{12}(\omega). \end{aligned}$$

Letting $\mathrm{ICV}(A_k) := \frac{\mathrm{E}[A_k]}{\sqrt{\mathrm{Var}[A_k]}}$ be the inverse coefficient of variation of A_k for $k = 1, 2$, the coherence, which is the absolute value of the complex correlation, at frequency ω is

$$\begin{aligned}\tau_{12}(\omega) = |\mathrm{Corr}[X_1(t), X_2(t)]| &= \frac{\mathrm{Cov}[A_1, A_2] + \mathrm{E}[A_1]\,\mathrm{E}[A_2]}{\sqrt{(\mathrm{Var}[A_1] + \mathrm{E}[A_1]^2)(\mathrm{Var}[A_2] + \mathrm{E}[A_2]^2)}} \mathrm{PLV}_{12}(\omega) \\ &= \frac{\mathrm{Corr}[A_1, A_2] + \mathrm{ICV}(A_1)\mathrm{ICV}(A_2)}{\sqrt{(1 + \mathrm{ICV}(A_1)^2)(1 + \mathrm{ICV}(A_2)^2)}} \mathrm{PLV}_{12}(\omega).\end{aligned}$$

Thus, in this independence case, if the amplitude is highly concentrated so that the ICV is large, coherence and PLV will be roughly equal, but if ICV is small or moderate they will differ.

We can observe instances in which the PLV provides stronger evidence for an association than complex correlation. For instance, assume that $\Theta_1 \sim Unif(0, 2\pi)$, $\Theta_2 = \Theta_1 + \epsilon$ (where ϵ is a von Mises random variable which is concentrated around zero) and then let $R_1 = R_2 = |\Theta_1 - \Theta_2|$. In this case, PLV is large, but complex covariance will be relatively small due to the specific form of R_1, R_2.

Finally, there are numerous cases in which complex covariance (and thus coherence in the stationary band-pass filtered case) may reveal information about associations not apparent by examining either PLV of amplitude correlation. We show some of these in Table E.1 in Appendix D. Overall, coherence and PLV can assess phase interaction differently, and in many applications the situations discussed here may usefully inform the choice of one rather than the other.

5.6 A Latent Variable Model for the Proper Complex Normal Distribution

We assume the data take the form of oscillating LFP recordings multiple electrodes embedded in each of several brain regions. The goal is to assess the strength of the associations in the signals across regions. We assume these signals have been band-pass filtered and thus, for each electrode at each moment in time, we have a complex value containing amplitude and phase data. We select a single time point at which we analyze the strength of cross-area interactions.

We develop a latent variable model to estimate a latent covariance matrix which assesses the strength of cross-region interactions between brain areas. The latent variable model is based on a series of proper complex normal distributions.

5.6.1 Model

We assume that there are R brain regions, and each brain region contains D_r electrodes ($1 \leq r \leq R$). We observe vectors of length D_r, denoted $X^{(n),r}$, which represent the signals recorded from each electrode in each brain region. Here n represents the nth trial and so $1 \leq n \leq N$, where N is the total number of trials in the experiment.

We assume that each brain region is represented by a single latent factor $Z_r^{(n)}$ which is a complex random variable. Collectively for each trial a latent random vector $Z^{(n)} = [Z_1^{(n)}, Z_2^{(n)}, ..., Z_R^{(n)}]$ is generated. Each latent factor is associated with a random vector of length D_r which consists of the observed variables; we call this vector $X^{(n),r}$. For the parameters $\Gamma \in \mathbb{C}^{R\times R}$, $\beta^r \in \mathbb{C}^{D_r\times 1}$, and $\eta^r \in \mathbb{C}^{D_r \times D_r}$, we then assume the following generative model:

$$\begin{aligned}X^{(n),r} &= \beta^r Z_r^{(n)} + \epsilon^{(n),r}\\ Z^{(n)} &\sim CN(0, \Gamma, 0)\\ \epsilon^{(n),r} &\sim CN(0, \eta^r, 0)\end{aligned}$$

where the random vectors $\epsilon^{(n),r}$ are noise terms specific to each latent factor. We require that $(\beta^r)^H \beta^r = 1$ to ensure identifiability of the cross-region late covariances. The strength of the cross-region associations is governed by the latent covariance matrix Γ.

Using basic properties of the complex normal distribution, we can write an explicit formula for the marginal distribution of the vector $X^{(n)} = [X^{(n),1}, X^{(n),2}, \ldots, X^{(n),R}]$ which is the concatenation of the observed signals for all electrodes in all regions. In particular, we have:

$$X^{(n)} \sim CN\left(0, \begin{bmatrix} \beta_1\beta_1^H\Gamma_{11} + \eta^1 & \beta_1\beta_2^H\Gamma_{12} & \beta_1\beta_3^H\Gamma_{13} & \ldots \\ \beta_2\beta_1^H\Gamma_{21} & \beta_2\beta_2^H\Gamma_{22} + \eta^2 & \beta_2\beta_3^H\Gamma_{23} & \ldots \\ \beta_3\beta_1^H\Gamma_{31} & \beta_3\beta_2^H\Gamma_{32} & \beta_3\beta_3^H\Gamma_{33} + \eta^3 & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}, 0\right)$$

Fitting the marginal likelihood directly is difficult due to the constraints on the structure of the covariance matrix. Therefore, instead, we can fit the model via expectation maximization (see appendix section E.3). If desired, we expect it would be possible to use the bootstrap to identify significant entries in the latent covariance matrix.

5.6.2 Synthetic Data Example

We evaluate the latent variable model on some synthetic data. Let $Z_n \sim CN(0, \Gamma, 0)$. We then generate the data as $X_n^{(r)} = \mathbb{1}_{D_r} Z_n^r + \gamma_n^{(r)}$ where

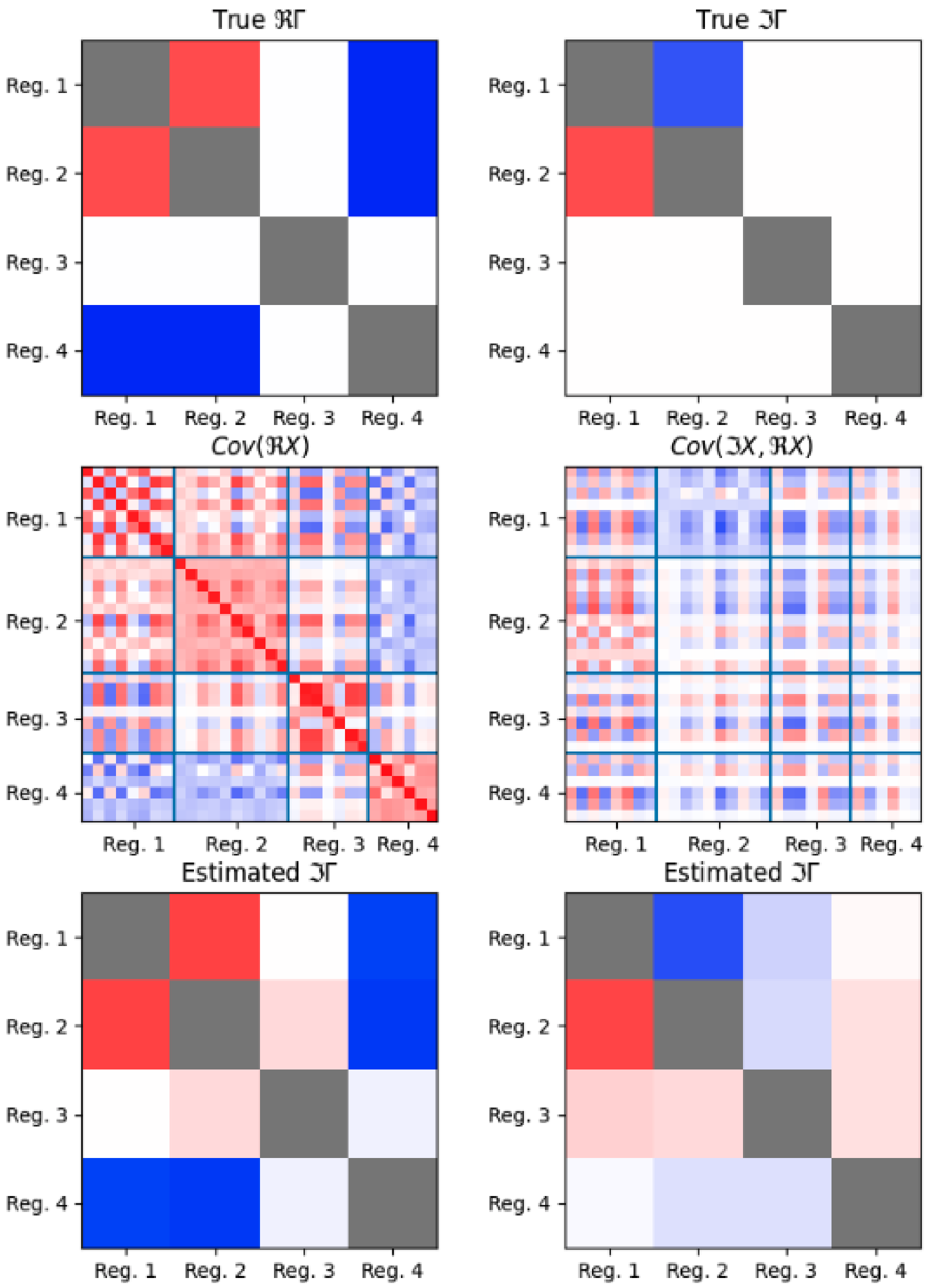
$$\begin{bmatrix}\mathfrak{Re}\gamma_n \\ \mathfrak{Im}\gamma_n\end{bmatrix} \sim N(0, \mathbb{I}_{2D} + UU^T)$$

and U is a 2D-dimensional random vector such that $U_i \sim Unif(-1.0, 1.0)$.

The true latent covariance matrix, the empirical covariance matrix of the observed data, and the estimated latent covariance matrix are shown in Figure 5.5. Since γ allows for interactions between regions in a manner that is not dependent on Z, this example is not consistent with the generative setting for the latent variable model. Nevertheless, the model appears to recover the latent covariance structure.

5.7 Data Analysis

We apply the techniques we've discussed so far to a dataset of local field potential (LFP) data. We use a publicly available dataset of LFP data from the Allen Institute (the dataset was described in Siegle et al. (2021)). In the experiment, six electrode probes were simultaneously inserted into


True ℜΓ

True ℑΓ

Cov(ℜX)

Cov(ℑX, ℜX)

Estimated ℑΓ

Estimated ℑΓ

Figure 5.5: Covariance matrices for synthetic example described in the test. Real and complex parts of the entries in the true Γ matrix used to generate the latent factors are shown in the top row. Diagonal entries are omitted. In the middle row are the empirical covariance matrices of the observed data. Finally, in the bottom row are the real and complex parts of the estimated latent covariance matrix.

the mouse brain, with each probe targeting some area of visual cortex but also recording from other regions. During the experiments, the mice were presented with a variety of visual stimuli;

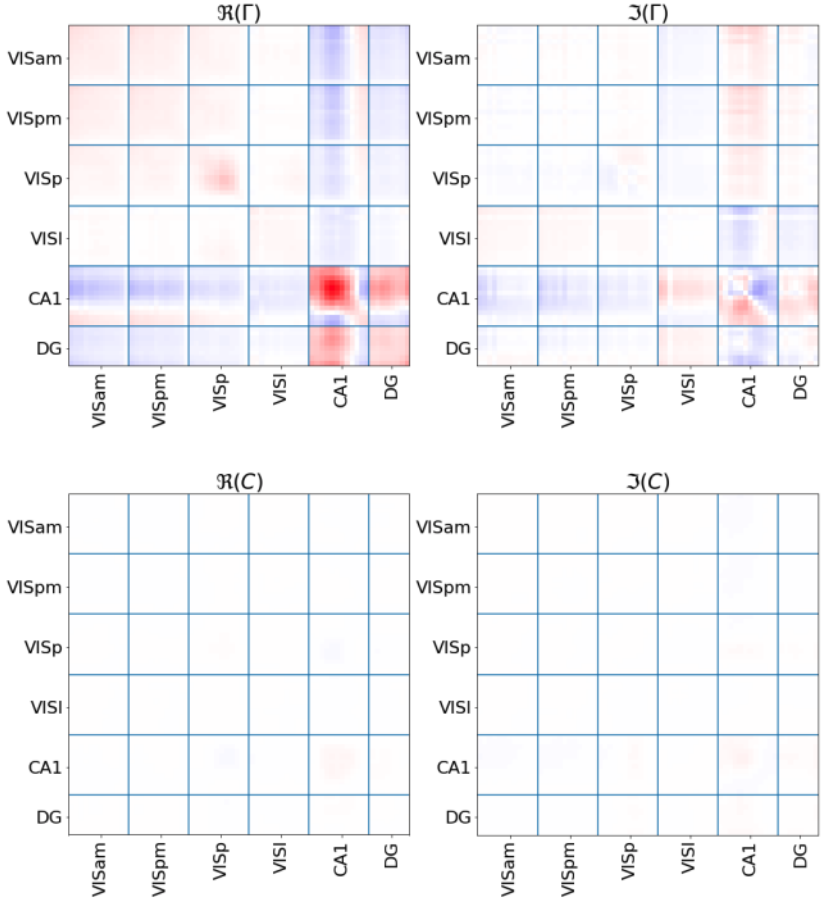


Figure 5.6: Elements of the covariance and pseudo-covariance matrices for LFP data. The covariance matrices (denoted by Γ) were computed, along with the pseudo-covariance matrices (denoted by C). These are plotted above, and appear to show generally larger entries for Γ than for C.

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 of data.

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 a variety 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 well after the stimulus disappears from the screen. We then have a single complex value for each electrode, which is the basis for our analysis.

5.7.1 Analysis of Propriety

We first analyze the dataset to show that our assumption of propriety is reasonable. Recall that the complex normal distribution utilizes covariance (Γ) and pseudo-covariance (C) matrices,

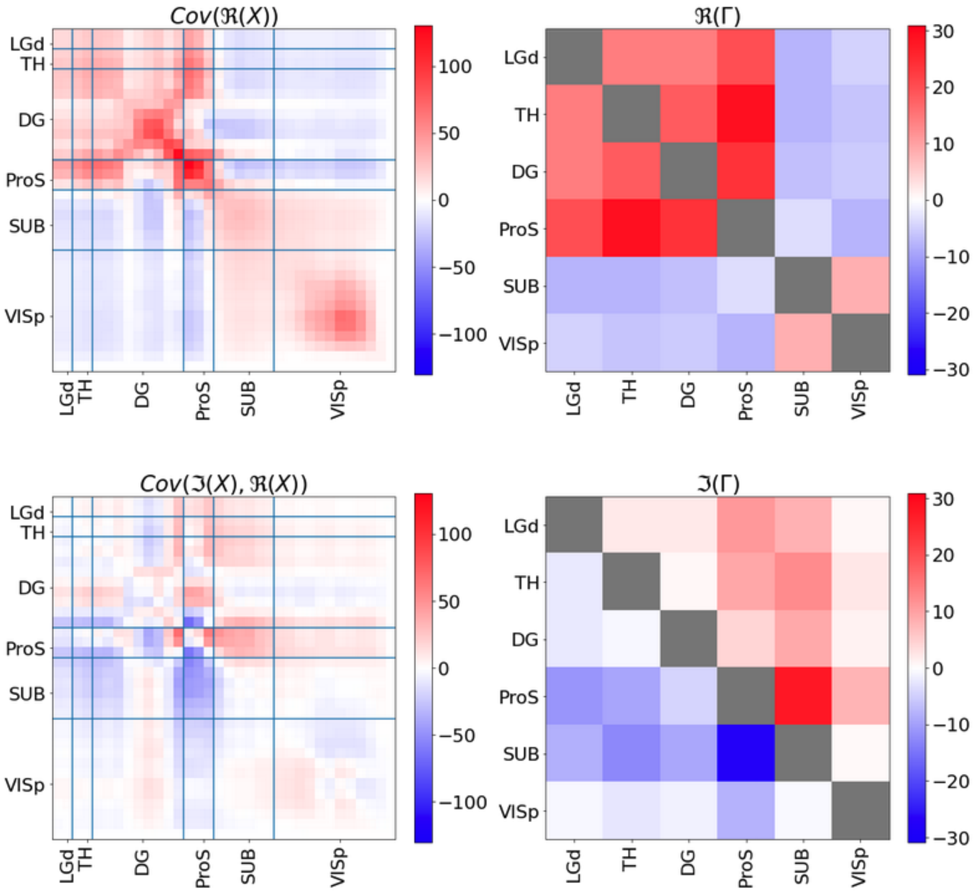


Figure 5.7: An example of the application of the latent variable model to LFP data from the Allen Institute is shown here. The entries in the covariance matrices are arranged consecutively according to their vertical position in the inserted probe. The sample covariance matrices for the real part of the data and the covariance between the real and imaginary parts of the data are shown on the left. Then, the latent variable model was fit to the data, and the estimated latent covariance matrix is denoted by Γ (diagonal entries are omitted). Electrodes are labeled by the anatomical region in which they resided during the experiment.

the second of which is zero for a proper distribution. In Figure 5.6, we plot the covariance and pseudo-covariance matrices of our data. We observe that entries from the pseudo-covariance matrices are generally much smaller than those of the covariance matrices, which is consistent with propriety being a reasonable assumption.

5.7.2 Application of the Latent Variable Model

We then applied the latent variable model to this dataset. For all analyses, we used one latent variable per brain area we analyze. Each brain area has multiple electrodes embedded that record the LFP signals; these signals are the observed variables in our model. In Figures 5.7 and 5.8 we display the empirical covariance matrices of the observed data and the latent covariance matrices

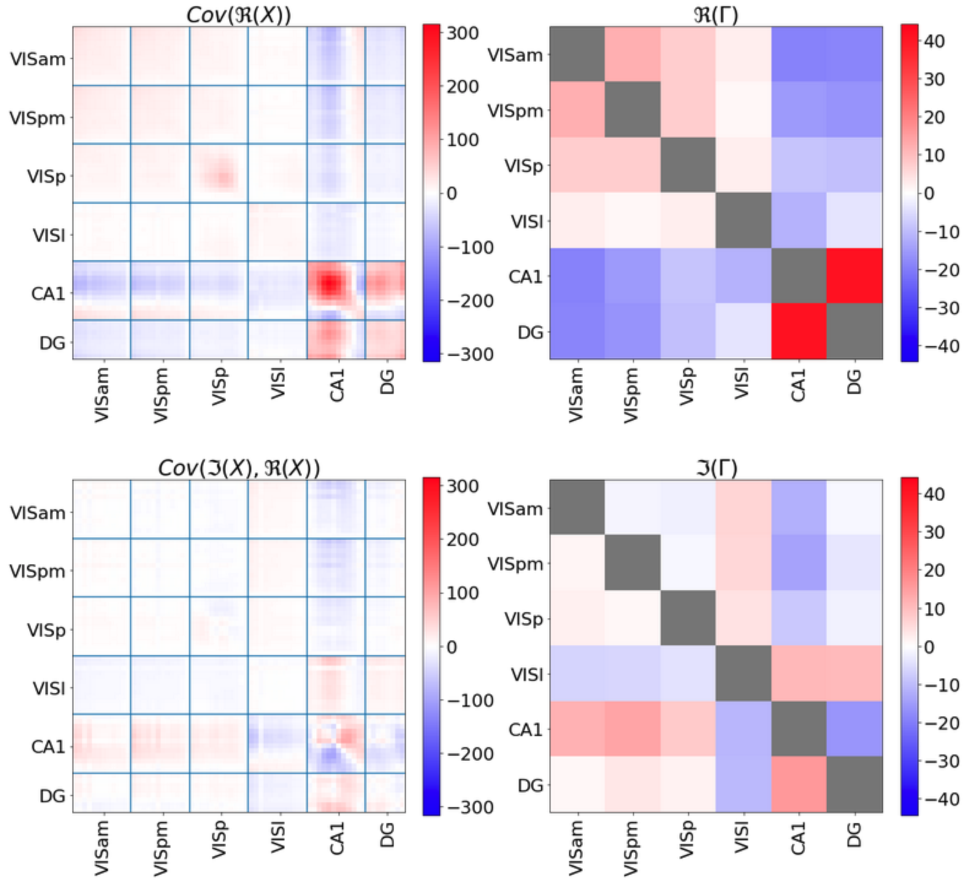


Figure 5.8: This figure was obtained using the same setup as in figure 5.7, except that electrodes were obtained from multiple probes and therefore are only ordered by vertical position within each region.

estimated by our model.

In Figure 5.7, we show data from only a single probe. The elements of the covariance matrices are arranged by their vertical position in the inserted probe. In the real part of the estimated latent covariance matrix, we can observe a relatively strong division of the regions into two groups, based on positive covariances within groups and negative covariances across groups. One group is composed of the subiculum and the primary visual cortex, and a second group is composed of the dentate gyrus in the hippocampus and thalamus, among other regions. The imaginary part of the covariance matrix doesn't appear to show any sharp distinctions between areas.

In Figure 5.8 we show something similar, except that we include electrodes from multiple probes across multiple visual regions, as well as the two hippocampal regions; namely, the dentate gyrus and CA1 region. Once again, in the real part of the latent covariance matrix, we observe the strongest pattern: namely, the four visual areas appear to be positively correlated with each other and negatively correlated with the hippocampal regions, which are themselves highly correlated.

Appendix A

Thesis flow chart with model details.

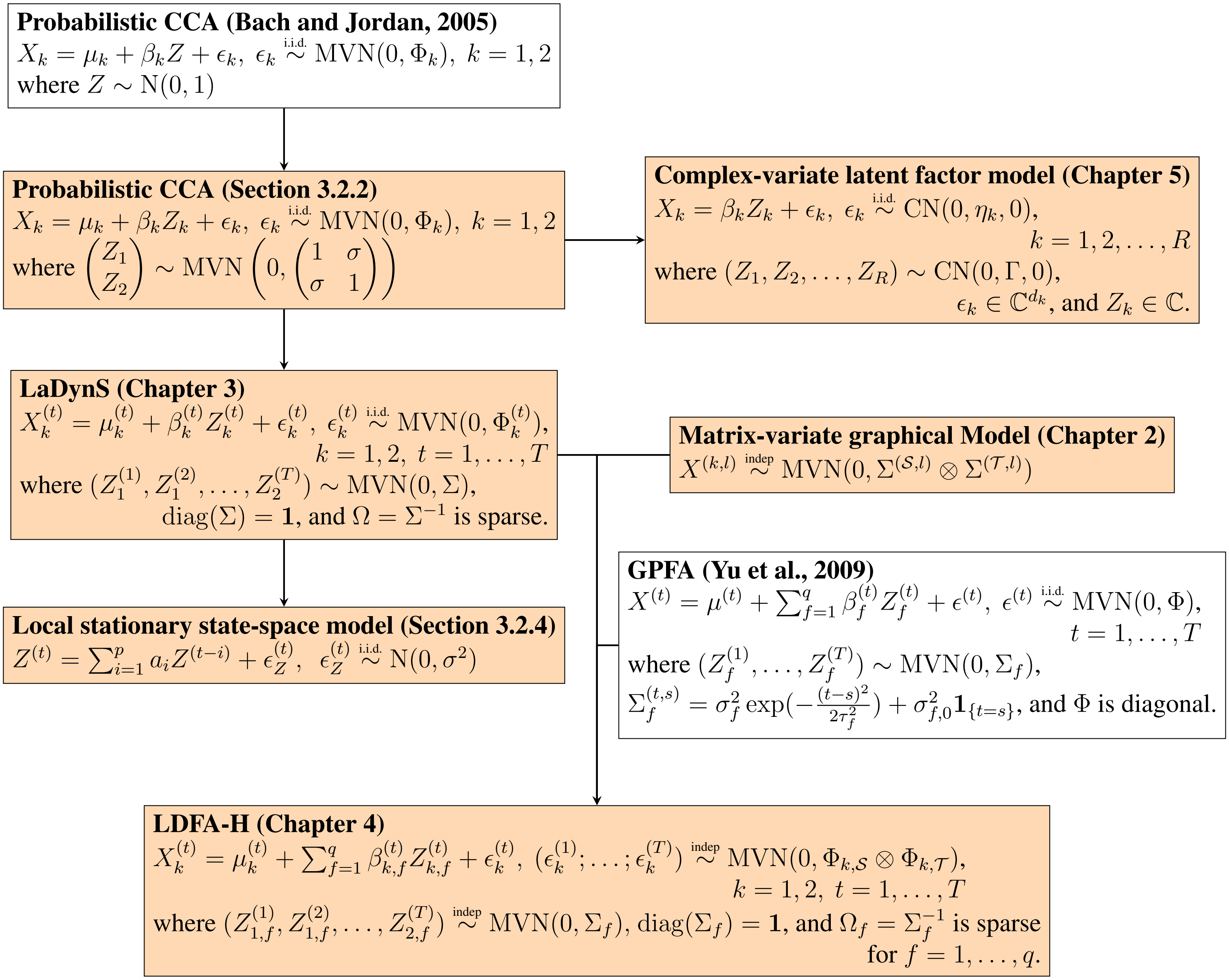

Probabilistic CCA (Bach and Jordan, 2005)
$X_k = \mu_k + \beta_k Z + \epsilon_k,\ \epsilon_k \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, \Phi_k),\ k = 1, 2$
where $Z \sim \text{N}(0, 1)$
Probabilistic CCA (Section 3.2.2)
$X_k = \mu_k + \beta_k Z_k + \epsilon_k,\ \epsilon_k \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, \Phi_k),\ k = 1, 2$
where $\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \text{MVN}\left(0, \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}\right)$
Complex-variate latent factor model (Chapter 5)
$X_k = \beta_k Z_k + \epsilon_k,\ \epsilon_k \overset{\text{i.i.d.}}{\sim} \text{CN}(0, \eta_k, 0),$
$k = 1, 2, \ldots, R$
where $(Z_1, Z_2, \ldots, Z_R) \sim \text{CN}(0, \Gamma, 0),$
$\epsilon_k \in \mathbb{C}^{d_k}$, and $Z_k \in \mathbb{C}$.
LaDynS (Chapter 3)
$X_k^{(t)} = \mu_k^{(t)} + \beta_k^{(t)} Z_k^{(t)} + \epsilon_k^{(t)},\ \epsilon_k^{(t)} \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, \Phi_k^{(t)}),$
$k = 1, 2,\ t = 1, \ldots, T$
where $(Z_1^{(1)}, Z_1^{(2)}, \ldots, Z_2^{(T)}) \sim \text{MVN}(0, \Sigma),$
$\text{diag}(\Sigma) = \mathbf{1}$, and $\Omega = \Sigma^{-1}$ is sparse.
Matrix-variate graphical Model (Chapter 2)
$X^{(k,l)} \overset{\text{indep}}{\sim} \text{MVN}(0, \Sigma^{(\mathcal{S},l)} \otimes \Sigma^{(\mathcal{T},l)})$
GPFA (Yu et al., 2009)
$X^{(t)} = \mu^{(t)} + \sum_{f=1}^{q} \beta_f^{(t)} Z_f^{(t)} + \epsilon^{(t)},\ \epsilon^{(t)} \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, \Phi),$
$t = 1, \ldots, T$
where $(Z_f^{(1)}, \ldots, Z_f^{(T)}) \sim \text{MVN}(0, \Sigma_f),$
$\Sigma_f^{(t,s)} = \sigma_f^2 \exp(-\frac{(t-s)^2}{2\tau_f^2}) + \sigma_{f,0}^2 \mathbf{1}_{\{t=s\}}$, and Φ is diagonal.
Local stationary state-space model (Section 3.2.4)
$Z^{(t)} = \sum_{i=1}^{p} a_i Z^{(t-i)} + \epsilon_Z^{(t)},\ \ \epsilon_Z^{(t)} \overset{\text{i.i.d.}}{\sim} \text{N}(0, \sigma^2)$
LDFA-H (Chapter 4)
$X_k^{(t)} = \mu_k^{(t)} + \sum_{f=1}^{q} \beta_{k,f}^{(t)} Z_{k,f}^{(t)} + \epsilon_k^{(t)},\ (\epsilon_k^{(1)}; \ldots; \epsilon_k^{(T)}) \overset{\text{indep}}{\sim} \text{MVN}(0, \Phi_{k,\mathcal{S}} \otimes \Phi_{k,\mathcal{T}}),$
$k = 1, 2,\ t = 1, \ldots, T$
where $(Z_{1,f}^{(1)}, Z_{1,f}^{(2)}, \ldots, Z_{2,f}^{(T)}) \overset{\text{indep}}{\sim} \text{MVN}(0, \Sigma_f)$, $\text{diag}(\Sigma_f) = \mathbf{1}$, and $\Omega_f = \Sigma_f^{-1}$ is sparse
for $f = 1, \ldots, q$.

Figure A.1: **Flow chart of research projects in this dissertation.** *White boxes code for existing methods by other researchers, orange boxes for completed projects in this thesis. Arrows indicate conceptual and methodological dependence.*

Appendix B

Appendix for Chapter 2

B.1 Additional Figures

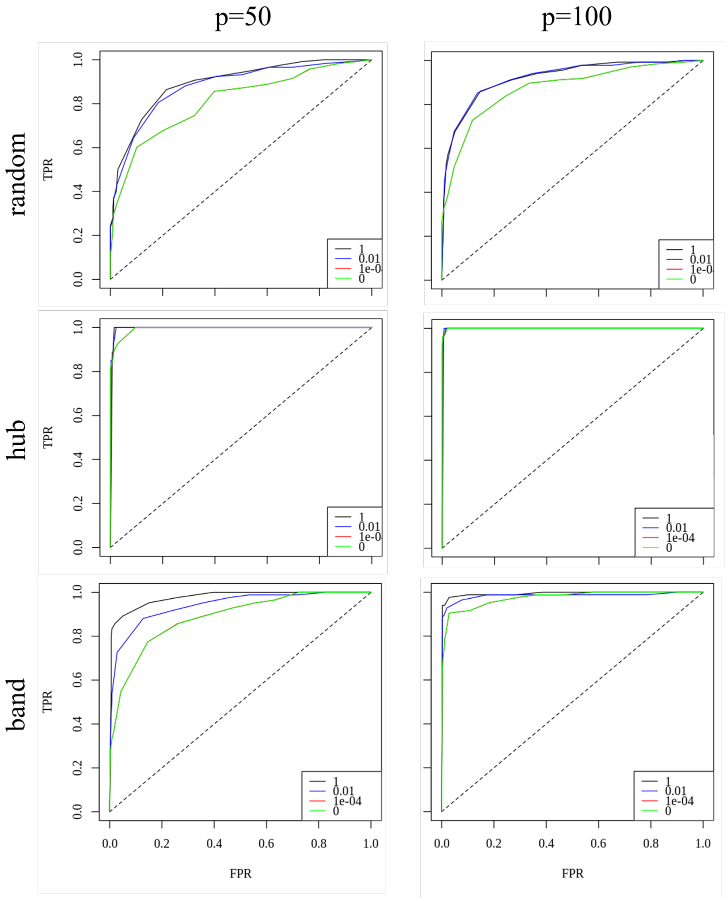


Figure B.1: ROC curve under 3 graph types with $n = 5$, $q = 30$ and $d = 5$ at tuning parameters $\lambda \in (1, 0.01, 1e-4, 1e-6)$. Our method is consistently better regardless of tuning parameters.

B.2 Proofs

B.2.1 Preliminary Lemmas

Here we introduce several lemmas that will help us prove propositions and theorems later. To apply the classical techniques and results in penalized regression analyses, we need to address the dependency among $X_t^{(\mathcal{S},l)}$'s. To do so, we use the fact that, for $k = 1, \dots, n_l$,

$$\sum_{t=1}^{p} X_t^{(k,l)} X_t^{(k,l)\top} = \sum_{t=1}^{p} \lambda_t^{(\mathcal{T},l)} W_t^{(k,l)} W_t^{(k,l)\top} \tag{B.1}$$

where $W_t^{(k,l)} \overset{\text{i.i.d.}}{\sim} N(0, \Sigma^{(\mathcal{S},l)})$ for all $t = 1, \dots, p$. We further let $\xi_{ti}^{(k,l)} = W_{ti}^{(k,l)} - W_{t,\cdot}^{(k,l)} \beta_{\cdot,i}^{(\mathcal{S},l)}$ so that

$$\begin{aligned} \sum_{t=1}^{p} \epsilon_t^{(\mathcal{S},k,l)} X_t^{(k,l)\top} &= \sum_{t=1}^{p} \lambda_t^{(\mathcal{T},l)} \xi_t^{(k,l)} W_t^{(k,l)\top}, \\ \sum_{t=1}^{p} \epsilon_t^{(\mathcal{S},k,l)} \epsilon_t^{(\mathcal{S},k,l)\top} &= \sum_{t=1}^{p} \lambda_t^{(\mathcal{T},l)} \xi_t^{(k,l)} \xi_t^{(k,l)\top}, \end{aligned} \tag{B.2}$$

and $\xi_t^{(k,l)} \overset{\text{i.i.d.}}{\sim} N(0, \Phi^{(\mathcal{S},l)})$ across $t = 1, \dots, p$ and $k = 1, \dots, n_l$.

The subsequent lemmas follow the application to the above equations of sub-exponential concentration inequalities, which could be found in mathematical statistics literature such as Vershynin (2018). The proofs for the lemmas are given in Appendices B.3.1 to B.3.4.

Lemma B.2.1.

$$\mathbb{P}\left[\max_{l=1,\dots,m} \left\| \frac{1}{n_l p} X^{(\mathcal{S},l)\top} X^{(\mathcal{S},l)} - \Sigma^{(\mathcal{S},l)} \right\|_\infty \geq C(\kappa_1, \kappa_3) \sqrt{\frac{\log(qmn_0p)}{n_0 p}}\right] \leq (qmn_0p)^{-1/2}. \tag{B.3}$$

Lemma B.2.2.

$$\begin{aligned} \mathbb{P}\left[\max_{i,j:j\neq i} \frac{1}{n_0 p} \left[\sum_{l=1}^{m} \left(\frac{\sqrt{n_l p}}{\|X_{\cdot,j}^{(\mathcal{S},l)}\|_2} \epsilon_{\cdot,i}^{(\mathcal{S},l)\top} X_{\cdot,j}^{(\mathcal{S},l)} \right)^2 \right]^{1/2} \geq C(\kappa_1, \kappa_3) \sqrt{\frac{m + \log(qmn_0p)}{n_0 p}}\right] \\ \leq (qmn_0p)^{-1/2}. \end{aligned} \tag{B.4}$$

and

$$\begin{aligned} \mathbb{P}\left[\max_{i=1,\dots,q} \frac{1}{n_0 p} \left[\sum_{l=1}^{m} \left(\frac{\sqrt{n_l p}}{\|X_{\cdot,i}^{(\mathcal{S},l)}\|_2} \epsilon_{\cdot,i}^{(\mathcal{S},l)\top} X^{(\mathcal{S},l)} \beta_{\cdot,i}^{(\mathcal{S},l)} \right)^2 \right]^{1/2} \geq C(\kappa_1, \kappa_3) \sqrt{\frac{m + \log(qmn_0p)}{n_0 p}}\right] \\ \leq (qmn_0p)^{-1/2}. \end{aligned} \tag{B.5}$$

Lemma B.2.3.

$$\mathbb{P}\left[\max_{l=1,\ldots,m}\max_{i,j:j\neq i}\left|\frac{1}{n_l p}\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}X_{\cdot,j}^{(\mathcal{S},l)}\right|\geq C(\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_lp}}\right]\leq(qmn_0p)^{-1/2}, \tag{B.6}$$

and

$$\mathbb{P}\left[\max_{l=1,\ldots,m}\max_{i=1,\ldots,q}\left|\frac{1}{n_l p}\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}\beta_{\cdot,i}^{(\mathcal{S},l)}\right|\geq C(\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_0p}}\right]\leq(qmn_0p)^{-1/2}. \tag{B.7}$$

Lemma B.2.4.

$$\mathbb{P}\left[\max_{l=1,\ldots,m}\left\|\frac{1}{n_l p}\epsilon^{(\mathcal{S},l)\top}\epsilon^{(\mathcal{S},l)}-\Phi^{(\mathcal{S},l)}\right\|_\infty\geq C(\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_0p}}\right]\leq(qmn_0p)^{-1/2}. \tag{B.8}$$

B.2.2 Proof of Theorem 2.3.1

Based on the optimality of the group Lasso estimate in Eq. (2.31), we immediately have

$$\begin{aligned}&\frac{1}{2n_0p}\sum_{l=1}^{m}\left(\|X_{\cdot,i}^{(\mathcal{S},l)}-X^{(\mathcal{S},l)}\widehat{\beta}_{\cdot,i}^{(\mathcal{S},l)}\|_2^2-\|X_{\cdot,i}^{(\mathcal{S},l)}-X^{(\mathcal{S},l)}\beta_{\cdot,i}^{(\mathcal{S},l)}\|_2^2\right)\\&\leq\lambda_i\sum_{j:j\neq i}\left(\left\|\beta_{-ji}^{(\mathcal{S},\cdot)}\right\|_2-\left\|\widehat{\beta}_{-ji}^{(\mathcal{S},\cdot)}\right\|_2\right)\end{aligned} \tag{B.9}$$

for every $i=1,\ldots,q$. We notice that the left-hand side satisfies

$$\begin{aligned}&\frac{1}{2n_0p}\sum_{l=1}^{m}\left(\|X_{\cdot,i}^{(\mathcal{S},l)}-X^{(\mathcal{S},l)}\widehat{\beta}_{\cdot,i}^{(\mathcal{S},l)}\|_2^2-\|X_{\cdot,i}^{(\mathcal{S},l)}-X^{(\mathcal{S},l)}\beta_{\cdot,i}^{(\mathcal{S},l)}\|_2^2\right)\\&=\frac{1}{2n_0p}\sum_{l=1}^{m}\left(\|X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)}\|_2^2-2(X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)})^\top\epsilon_{\cdot,i}^{(\mathcal{S},l)}\right).\end{aligned} \tag{B.10}$$

For the first term,

$$\begin{aligned}&\frac{1}{2n_0p}\sum_{l=1}^{m}\|X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)}\|_2^2\\&=\frac{1}{2n_0p}\sum_{l=1}^{m}\Delta_{\cdot,i}^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)}\\&=\sum_{l=1}^{m}\frac{2n_lp}{2n_0p}\left[\Delta_{\cdot,i}^{(\mathcal{S},l)\top}\Sigma^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)}+\Delta_{\cdot,i}^{(\mathcal{S},l)\top}\left\{\frac{X^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}}{2n_lp}-\Sigma^{(\mathcal{S},l)}\right\}\Delta_{\cdot,i}^{(\mathcal{S},l)}\right]\\&\geq\frac{1}{\kappa_3}\sum_{l=1}^{m}\|\Delta_{\cdot,i}^{(\mathcal{S},l)}\|_2^2-\frac{\kappa_1}{\kappa_3}\sum_{l=1}^{m}\Delta_{\cdot,i}^{(\mathcal{S},l)\top}\left\{\frac{X^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}}{2n_lp}-\Sigma^{(\mathcal{S},l)}\right\}\Delta_{\cdot,i}^{(\mathcal{S},l)}\end{aligned} \tag{B.11}$$

where the last inequality resorts to Assumptions 1 and 3. To provide a further lower-bound, we use the fact based on Lemma B.2.1 that

$$\frac{1}{\kappa_3}\sum_{l=1}^{m}\|\Delta_{\cdot,i}^{(\mathcal{S},l)}\|_2^2 \geq C(\kappa_3)\sum_{j:j\neq i}\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\|_2^2 \tag{B.12}$$

and

$$\begin{aligned}
&\sum_{l=1}^{m}\Delta_{\cdot,i}^{(\mathcal{S},l)\top}\left\{\frac{X^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}}{2n_l p}-\Sigma^{(\mathcal{S},l)}\right\}\Delta_{\cdot,i}^{(\mathcal{S},l)}\\
&\leq \sum_{l=1}^{m}\sum_{j_1,j_2}\Delta_{j_1 i}^{(\mathcal{S},l)}\left\{\frac{X_{\cdot,j_1}^{(\mathcal{S},l)\top}X_{\cdot,j_2}^{(\mathcal{S},l)}}{2n_l p}-\Sigma_{j_1 j_2}^{(\mathcal{S},l)}\right\}\Delta_{j_2 i}^{(\mathcal{S},l)}\\
&\leq \sum_{j_1,j_2}\left\|\Delta_{j_1 i}^{(\mathcal{S},\cdot)}\right\|_2\max_{l}\left|\frac{X_{\cdot,j_1}^{(\mathcal{S},l)\top}X_{\cdot,j_2}^{(\mathcal{S},l)}}{2n_l p}-\Sigma_{j_1 j_2}^{(\mathcal{S},l)}\right|\left\|\Delta_{j_2 i}^{(\mathcal{S},\cdot)}\right\|_2 \qquad \text{(B.13)}\\
&\leq \max_{l}\left|\frac{X_{\cdot,j_1}^{(\mathcal{S},l)\top}X_{\cdot,j_2}^{(\mathcal{S},l)}}{2n_l p}-\Sigma_{j_1 j_2}^{(\mathcal{S},l)}\right|\left(\sum_{j}\left\|\Delta_{ji}^{(\mathcal{S},\cdot)}\right\|_2\right)^2\\
&\leq C(\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_l p}}\left(\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\right)^2
\end{aligned}$$

uniformly over $i = 1, \dots, q$ with probability at least $1-(qmn_0p)^{-1/2}$. For the second term,

$$\begin{aligned}
&\frac{1}{n_0 p}\left|\sum_{l=1}^{m}(X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)})^\top\epsilon_{\cdot,i}^{(\mathcal{S},l)}\right| \leq \sum_{j:j\neq i}\left|(\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}X_{\cdot,j}^{(\mathcal{S},l)})\sum_{l=1}^{m}\Delta_{ji}^{(\mathcal{S},l)}\right|\\
&\leq \sum_{j:j\neq i}\left\|\left\{\frac{\sqrt{n_l p}}{\|X_{\cdot,j}^{(\mathcal{S},l)}\|_2}\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}X_{\cdot,j}^{(\mathcal{S},l)}\right\}_{l=1,\dots,m}\right\|_2\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\\
&\leq \max_{j:j\neq i}\left\|\left\{\frac{\sqrt{n_l p}}{\|X_{\cdot,j}^{(\mathcal{S},l)}\|_2}\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}X_{\cdot,j}^{(\mathcal{S},l)}\right\}_{l=1,\dots,m}\right\|_2\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2 \qquad \text{(B.14)}\\
&\leq C(\kappa_1,\kappa_3)\sqrt{\frac{m+\log(qmn_0p)}{n_0 p}}\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2
\end{aligned}$$

uniformly over $i = 1, \dots, q$ with probability at least $1-(qmn_0p)^{-1/2}$ where the last inequality is due Lemma B.2.2. For the right-hand side of Eq. (B.9),

$$\begin{aligned}
&\lambda_i\sum_{j:j\neq i}\left(\left\|\underline{\beta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2-\left\|\underline{\widehat{\beta}}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\right)\\
&\leq \lambda_i\left(\sum_{j:\beta_{ji}^{(\mathcal{S},l)}\neq 0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2-\sum_{j:\beta_{ji}^{(\mathcal{S},l)}=0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\right).
\end{aligned} \tag{B.15}$$

On the one hand, because $\frac{1}{2n_0p}\sum_{l=1}^{m}\|X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)}\|_2^2 \geq 0$ a.s., Eqs. (B.10), (B.14) and (B.15) earn

$$-C(\kappa_1,\kappa_3)\sqrt{\frac{m+\log(qmn_0p)}{n_0p}}\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2 \leq \lambda_i\Bigg(\sum_{j:\beta_{ji}^{(\mathcal{S},l)}\neq 0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2 - \sum_{j:\beta_{ji}^{(\mathcal{S},l)}=0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\Bigg) \tag{B.16}$$

uniformly over $i=1,\ldots,q$ with probability at least $1-C(qmn_0p)^{-1/2}$. Due to Assumption 4, $\sqrt{\frac{m+\log(qmn_0p)}{n_0p}} \to 0$ as $n_0\to\infty$, and therefore

$$\sum_{j:\beta_{ji}^{(\mathcal{S},l)}=0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2 \leq C\sum_{j:\beta_{ji}^{(\mathcal{S},l)}\neq 0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2 \tag{B.17}$$

uniformly over $i=1,\ldots,q$ with probability at least $1-C(qmn_0p)^{-1/2}$ for sufficiently large n_0. On the other hand, Eqs. (B.10) to (B.15) earn

$$\begin{aligned}
&\lambda_i\Bigg(\sum_{j:\beta_{ji}^{(\mathcal{S},l)}\neq 0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2 - \sum_{j:\beta_{ji}^{(\mathcal{S},l)}=0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\Bigg)\\
&\geq C(\kappa_3)\sum_{j:j\neq i}\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\|_2^2 - C(\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_lp}}\Bigg(\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\Bigg)^2\\
&\quad - C(\kappa_1,\kappa_3)\sqrt{\frac{m+\log(qmn_0p)}{n_0p}}\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2
\end{aligned} \tag{B.18}$$

uniformly over $i=1,\ldots,d$ with probability at least $1-2(qmn_0p)^{-1/2}$. Because

$$\Bigg(\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\Bigg)^2 \leq C\Bigg(\sum_{j:\beta_{ji}^{(\mathcal{S},l)}\neq 0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\Bigg)^2 \leq C\,d\sum_{j:\beta_{ji}^{(\mathcal{S},l)}\neq 0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2^2 \tag{B.19}$$

with probability at least $1-C(qmn_0p)^{-1/2}$ for sufficiently large n_0 due to Eq. (B.17), and $d\sqrt{\frac{\log(qmn_0p)}{n_op}}\to 0$ as $n_0\to\infty$ due to Assumption 4, then Eq. (B.18) turns into

$$\begin{aligned}
&\lambda_i\Bigg(\sum_{j:\beta_{ji}^{(\mathcal{S},l)}\neq 0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2 - \sum_{j:\beta_{ji}^{(\mathcal{S},l)}=0}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2\Bigg)\\
&\geq C(\kappa_3)\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2^2 - C(\kappa_1,\kappa_3)\sqrt{\frac{m+\log(qmn_0p)}{n_0p}}\sum_{j:j\neq i}\left\|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\right\|_2
\end{aligned} \tag{B.20}$$

uniformly over $i = 1, \dots, d$ with probability at least $1 - 2(qmn_0p)^{-1/2}$ for a sufficiently large n_0. As a result, with a sufficiently large n_0 and the same probability,

$$\begin{aligned}
&C(\kappa_3) \sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2^2 \\
&\leq \lambda_i \Bigg(\sum_{j:\beta_{ji}^{(\mathcal{S},l)}\neq 0} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 - \sum_{j:\beta_{ji}^{(\mathcal{S},l)}= 0} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 \Bigg) + C(\kappa_1, \kappa_3) \sqrt{\frac{m + \log(qmn_0p)}{n_0p}} \sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 \\
&\leq C(\kappa_1, \kappa_3) \sqrt{\frac{m + \log(qmn_0p)}{n_0p}} \sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 \\
&\leq C(\kappa_1, \kappa_3) \sqrt{d \frac{m + \log(qmn_0p)}{n_0p}} \sqrt{\sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2^2},
\end{aligned} \tag{B.21}$$

uniformly over $i = 1, \dots, d$, and

$$\max_i \sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2^2 \leq C(\kappa_1, \kappa_3)\, d\, \frac{m + \log(qmn_0p)}{n_0p}. \tag{B.22}$$

In addition, due to Eq. (B.17), under the same event,

$$\begin{aligned}
&\max_i \sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 \leq C \max_i \sum_{j:\beta_{ji}\neq 0} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 \\
&\leq C\sqrt{d} \max_i \sqrt{\sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2^2} \leq C(\kappa_1, \kappa_3)\, d \sqrt{\frac{m + \log(qmn_0p)}{n_0p}}.
\end{aligned} \tag{B.23}$$

Last, it follows Eqs. (B.9) and (B.10) that

$$\begin{aligned}
&\max_i \frac{1}{2n_0p} \sum_{l=1}^{m} \left\| X^{(\mathcal{S},l)} \Delta_{\cdot,i}^{(\mathcal{S},l))} \right\|_2^2 \\
&\leq \max_i \left[\lambda_i \sum_{j:j\neq i} \left(\left\| \underline{\beta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 - \left\| \underline{\widehat{\beta}}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 \right) + \frac{1}{n_0p} \sum_{l=1}^{m} (X^{(\mathcal{S},l)} \Delta_{\cdot,i}^{(\mathcal{S},l)})^\top \epsilon_{\cdot,i}^{(\mathcal{S},l)} \right] \\
&\leq \max_i \left[\lambda_i \sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 + C(\kappa_1, \kappa_3) \sqrt{\frac{m + \log(qmn_0p)}{n_0p}} \sum_{j:j\neq i} \left\| \underline{\Delta}_{ji}^{(\mathcal{S},\cdot)} \right\|_2 \right] \\
&\leq C(\kappa_1, \kappa_3)\, d\, \frac{m + \log(qmn_0p)}{n_0p}
\end{aligned} \tag{B.24}$$

under the same event.

B.2.3 Proof of Proposition 2.3.2

For $i, j \in \{1, \dots, q\}$, let

$$\widehat{\phi}_{ij}^{(\mathcal{S},l)} := \frac{1}{n_l p}\left[\widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)\top}\widehat{\epsilon}_{\cdot,j}^{(\mathcal{S},l)} - \epsilon_{\cdot,i}^{(\mathcal{S},l)\top}\epsilon_{\cdot,j}^{(\mathcal{S},l)}\right] + \frac{1}{n_l p}\left(\|\epsilon_{\cdot,i}^{(\mathcal{S},l)}\|_2^2\Delta_{ij}^{(\mathcal{S},l)} + \|\epsilon_{\cdot,j}^{(\mathcal{S},l)}\|_2^2\Delta_{ji}^{(\mathcal{S},l)}\right)\mathbb{I}(i \neq j) \tag{B.25}$$

and

$$\widetilde{\phi}_{ij}^{(\mathcal{S},l)} := \frac{1}{n_l p}\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}\epsilon_{\cdot,j}^{(\mathcal{S},l)} - \Phi_{ij}^{(\mathcal{S},l)}. \tag{B.26}$$

Then, for $i = j$,

$$\widehat{\Phi}_{ii}^{(\mathcal{S},l)} - \Phi_{ii}^{(\mathcal{S},l)} = \frac{1}{n_l p}\widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)\top}\widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)} - \Phi_{ii}^{(\mathcal{S},l)} = \widehat{\phi}_{ii}^{(\mathcal{S},l)} + \widetilde{\phi}_{ii}^{(\mathcal{S},l)}, \tag{B.27}$$

and, for $i \neq j$, because $\beta_{ij}^{(\mathcal{S},l)} = -\Phi_{ij}^{(\mathcal{S},l)}/\Phi_{ii}^{(\mathcal{S},l)}$,

$$\begin{aligned}\Phi_{ij}^{(\mathcal{S},l)} - \widehat{\Phi}_{ij}^{(\mathcal{S},l)} =& -\left(\Phi_{ij}^{(\mathcal{S},l)} + \Phi_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)} + \Phi_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)}\right)\\ &+ \frac{1}{n_l p}\left(\widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)\top}\widehat{\epsilon}_{\cdot,j}^{(\mathcal{S},l)} + \|\widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)}\|_2^2\widehat{\beta}_{ij}^{(\mathcal{S},l)} + \|\widehat{\epsilon}_{\cdot,j}^{(\mathcal{S},l)}\|_2^2\widehat{\beta}_{ji}^{(\mathcal{S},l)}\right)\\ =&\ \widehat{\phi}_{ij}^{(\mathcal{S},l)} + \widehat{\phi}_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)} + \widehat{\phi}_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)}\\ &+ \widetilde{\phi}_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)}\end{aligned} \tag{B.28}$$

These error terms drive $\widehat{\rho}_{ij}^{(\mathcal{S},l)} - \rho_{ij}^{(\mathcal{S},l)}$ for $i, j : i \neq j$ by

$$\begin{aligned}\widehat{\rho}_{ij}^{(\mathcal{S},l)} - \rho_{ij}^{(\mathcal{S},l)} =& -\frac{\widehat{\Phi}_{ij}^{(\mathcal{S},l)}}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}} + \frac{\Phi_{ij}^{(\mathcal{S},l)}}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}}\\ =&\frac{\Phi_{ij}^{(\mathcal{S},l)} - \widehat{\Phi}_{ij}^{(\mathcal{S},l)}}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}} + \Phi_{ij}^{(\mathcal{S},l)}\left(\frac{1}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} - \frac{1}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}}\right)\\ =&\frac{\widehat{\phi}_{ij}^{(\mathcal{S},l)} + \widehat{\phi}_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)} + \widehat{\phi}_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)} + \widetilde{\phi}_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)}}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}}\\ &+ \Phi_{ij}^{(\mathcal{S},l)}\left(\frac{1}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} - \frac{1}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}}\right)\end{aligned} \tag{B.29}$$

We notice that

$$\begin{aligned}\Theta_{ij}^{(\mathcal{S},l)} =&\frac{\widetilde{\phi}_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)}/2 + \widetilde{\phi}_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)}/2}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}}\\ =&\frac{\widetilde{\phi}_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)}}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} + \frac{\Phi_{ij}^{(\mathcal{S},l)}\left(\Phi_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)} + \Phi_{jj}^{(\mathcal{S},l)}\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\right)}{2\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}},\end{aligned} \tag{B.30}$$

because $\beta_{ij}^{(\mathcal{S},l)} = -\Phi_{ij}^{(\mathcal{S},l)}/\Phi_{ii}^{(\mathcal{S},l)}$. Hence, the remainder term turns out to be

$$\begin{aligned}
O_{ij}^{(\mathcal{S},l)} = &\left[\Phi_{ij}^{(\mathcal{S},l)}\left(\frac{1}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} - \frac{1}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}}\right) - \frac{\Phi_{ij}^{(\mathcal{S},l)}\left(\Phi_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)} + \Phi_{jj}^{(\mathcal{S},l)}\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\right)}{2\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}}\right] \\
&+ \left(\widetilde{\phi}_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)} + \widetilde{\phi}_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)}\right)\left(\frac{1}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}} - \frac{1}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}}\right) \\
&+ \frac{\widehat{\phi}_{ij}^{(\mathcal{S},l)} + \widehat{\phi}_{ii}^{(\mathcal{S},l)}\beta_{ij}^{(\mathcal{S},l)} + \widehat{\phi}_{jj}^{(\mathcal{S},l)}\beta_{ji}^{(\mathcal{S},l)}}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}}
\end{aligned} \tag{B.31}$$

For $\widehat{\phi}_{ii}^{(\mathcal{S},l)}$'s,

$$\begin{aligned}
\max_i \sum_{l=1}^{m} |\widehat{\phi}_{ii}^{(\mathcal{S},l)}| \leq &\max_i \sum_{l=1}^{m} \left| \|\widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)}\|_2^2 - \|\epsilon_{\cdot,i}^{(\mathcal{S},l)}\|_2^2 \right| \\
\leq &C(\kappa_3) \max_i \sum_{l=1}^{m} \frac{1}{2n_0 p} \left\| \widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)} - \epsilon_{\cdot,i}^{(\mathcal{S},l)} \right\|_2^2 \\
\leq &C(\kappa_1, \kappa_3)\, d\, \frac{m + \log(qmn_0 p)}{n_0 p}
\end{aligned} \tag{B.32}$$

with probability at least $1 - C(qmn_0p)^{-1/2}$ for a sufficienty large n_0, where the probability inequality at the last line is the result of Theorem 2.3.1. For $\widetilde{\phi}_{ii}^{(\mathcal{S},l)}$'s,

$$\max_i \max_l |\widetilde{\phi}_{ii}^{(\mathcal{S},l)}| \leq C(\kappa_1, \kappa_3) \sqrt{\frac{\log(qmn_0p)}{n_0 p}} \tag{B.33}$$

with probability at least $1 - (qmn_0p)^{-1/2}$ as a result of Lemma B.2.4. Assumption 4 implies that $\max_i \widehat{\Phi}_{ii}^{(\mathcal{S},l)} \leq C(\kappa_1, \kappa_3)$, so both $\widehat{\phi}^{(\mathcal{S},l)}$ and $\widetilde{\phi}^{(\mathcal{S},l)}$ are bounded. As a result, it is easily seen that, with the same probability,

$$|\sum_{l=1}^{m} O_{ij}^{(\mathcal{S},l)}| \leq C(\kappa_1, \kappa_3) \max \left\{ \begin{array}{l} \sum_l |\widehat{\phi}_{ii}^{(\mathcal{S},l)}|, \sum_l |\widehat{\phi}_{ij}^{(\mathcal{S},l)}|, \sum_l |\widehat{\phi}_{jj}^{(\mathcal{S},l)}|, \\ \sum_l |\widetilde{\phi}_{ii}^{(\mathcal{S},l)2}|, \sum_l |\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}|, \sum_l |\widetilde{\phi}_{jj}^{(\mathcal{S},l)2}|, \\ \sum_l |\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{ij}^{(\mathcal{S},l)}|, \sum_l |\widetilde{\phi}_{jj}^{(\mathcal{S},l)}\widetilde{\phi}_{ij}^{(\mathcal{S},l)}|. \end{array} \right\} \tag{B.34}$$

For example, the Taylor's theorem gives

$$\begin{aligned}
&\Phi_{ij}^{(\mathcal{S},l)}\left(\frac{1}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} - \frac{1}{\sqrt{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}}\right) - \frac{\Phi_{ij}^{(\mathcal{S},l)}\left(\Phi_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)} + \Phi_{jj}^{(\mathcal{S},l)}\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\right)}{2\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} \\
&= \frac{\Phi_{ij}^{(\mathcal{S},l)}\left(\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)} - \Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)} - \Phi_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)} - \Phi_{jj}^{(\mathcal{S},l)}\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\right)}{2\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} \\
&\quad + \Phi_{ij}^{(\mathcal{S},l)}\int_{\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}}^{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}} \frac{3}{4t^{5/2}}(\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)} - t)^2 dt \\
&\le C(\kappa_1,\kappa_3,\kappa_4)\max\left\{\begin{array}{r}\widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)} - \Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)} - \Phi_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)} - \Phi_{jj}^{(\mathcal{S},l)}\widetilde{\phi}_{ii}^{(\mathcal{S},l)}, \\ \left(\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)} - \widehat{\Phi}_{ii}^{(\mathcal{S},l)}\widehat{\Phi}_{jj}^{(\mathcal{S},l)}\right)^2\end{array}\right\},
\end{aligned} \tag{B.35}$$

which is easily fit into Eq. (B.34).

For the second order terms of $\widetilde{\phi}^{(\mathcal{S},l)}$ in Eq. (B.34), we recognize that they are sub-Weibull random variables and use the corresponding concentration inequality (Kuchibhotla and Chakrabortty, 2018) formulated in the following lemma. See Appendix B.3.5 for the derivation.

Lemma B.2.5.

$$\mathbb{P}\left[\begin{array}{r}\max_{i,j}\max\left\{\sum_{l=1}^{m}\left|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}\right|, \sum_{l=1}^{m}\left|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{ij}^{(\mathcal{S},l)}\right|\right\} \\ \ge C(\kappa_1,\kappa_3)\,\frac{m+\log^2(qmn_0p)}{n_0p}\end{array}\right] \le (qmn_0p)^{-1/2}. \tag{B.36}$$

For the linear terms of $\widehat{\phi}^{(\mathcal{S},l)}$, we use the following upperbound. See Appendix B.3.6 for the proof.

Lemma B.2.6.

$$\mathbb{P}\left[\max_{i,j}\sum_{l=1}^{m}\left|\widehat{\phi}_{ij}^{(\mathcal{S},l)}\right| \ge C(\kappa_1,\kappa_3)\, d\,\frac{m+\log(qmn_0p)}{n_0p}\right] \le C(qmn_0p)^{-1/2}. \tag{B.37}$$

The desired result is a straightforward result of the two above lemmas.

B.2.4 Proof of Proposition 2.3.4

Let $\Theta_E = \frac{1}{\sqrt{m}}\sum_{l=1}^{m}\sqrt{n_lp}\,\Theta_E^{(\mathcal{S},l)}$, and $O_E = \frac{1}{\sqrt{m}}\sum_{l=1}^{m}\sqrt{n_lp}\,O_E^{(\mathcal{S},l)}$, where $\Theta_E^{(\mathcal{S},l)}$ and $O_E^{(\mathcal{S},l)}$ are the collections of $\Theta_{ij}^{(\mathcal{S},l)}$ and $O_{ij}^{(\mathcal{S},l)}$ for $(i,j)\in E$, respectively. For any $x>0$ and $\delta>0$,

$$\begin{aligned}
&\mathbb{P}[\|\widehat{T}_E - T_E\|_\infty > x] \le \mathbb{P}[\|\Theta_E\|_\infty > x-\delta] + \mathbb{P}[\|O_E\|_\infty > \delta] \\
&\le \mathbb{P}[\|Z\|_\infty > x-\delta] + KS_\infty(\Theta_E, Z) + \mathbb{P}[\|O_E\|_\infty > \delta] \\
&= \mathbb{P}[\|Z\|_\infty > x] + \mathbb{P}[x \ge \|Z\|_\infty > x-\delta] + KS_\infty(\Theta_E, Z) + \mathbb{P}[\|O_E\|_\infty > \delta],
\end{aligned} \tag{B.38}$$

where $KS_\infty(\Theta_E, Z) := \sup_{x>0}|\mathbb{P}[\|\Theta_E\|_\infty > x] - \mathbb{P}[\|Z\|_\infty > x]|$ is the Kolmogorov-Smirnov distance between $\|\Theta_E\|_\infty$ and $\|Z\|_\infty$. Similarly,

$$\begin{aligned} &\mathbb{P}[\|\widehat{T}_E - T_E\|_\infty > x] \geq \mathbb{P}[\|\Theta_E\|_\infty > x + \delta] - \mathbb{P}[\|O_E\|_\infty > \delta] \\ &\geq \mathbb{P}[\|Z\|_\infty > x + \delta] - KS_\infty(\Theta_E, Z) - \mathbb{P}[\|O_E\|_\infty > \delta] \\ &= \mathbb{P}[\|Z\|_\infty > x] - \mathbb{P}[x + \delta \geq \|Z\|_\infty > x] - KS_\infty(\Theta_E, Z) - \mathbb{P}[\|O_E\|_\infty > \delta]. \end{aligned} \tag{B.39}$$

Therefore we conclude that

$$\begin{aligned} &\sup_{x>0}\Big|\mathbb{P}[\|\widehat{T}_E - T_E\|_\infty > x] - \mathbb{P}[\|Z\|_\infty > x]\Big| \\ &\leq KS_\infty(\Theta_E, Z) + \sup_{x>0}\mathbb{P}[x + \delta \geq \|Z\|_\infty \geq x - \delta] + \mathbb{P}[\|O_E\|_\infty > \delta]. \end{aligned} \tag{B.40}$$

We use the anti-concentration inequality of the supremum norm of Gaussian random vectors with mean zero (Corollary 1, Chernozhukov et al., 2015) and the fact that

$$S_{(i,j),(i,j)} = \frac{1}{m}\sum_{l=1}^{m}\frac{\|\Sigma^{(\mathcal{T},l)}\|_F^2}{p}(1 - \rho_{ij}^{(\mathcal{S},l)2})^2 \geq C(\kappa_3) \tag{B.41}$$

to control the second term:

$$\sup_{x>0}\mathbb{P}[x + \delta \geq \|Z\|_\infty \geq x - \delta] \leq C(\kappa_3)\delta\sqrt{\log|E|}, \tag{B.42}$$

for a sufficiently small δ. Setting $\delta = C(\kappa_1, \kappa_3)d\frac{m + \log^2(qmn_0p)}{\sqrt{mn_0p}}$, which converges to 0 as $n \to \infty$ due to Assumption 4, Proposition 2.3.2 provides an upperbound over the size of the last term.

Now it suffices to derive a probabilistic upperbound for $KS_\infty(\Delta_\infty, Z)$. We recall that each element of Θ_E is

$$\begin{aligned} \Theta_{ij} =& \frac{1}{\sqrt{m}}\sum_{l=1}^{m}\sqrt{n_lp}\Theta_{ij}^{(\mathcal{S},l)} \\ =& \frac{1}{\sqrt{m}}\sum_{l=1}^{m}\sqrt{n_lp}\left(\frac{\widetilde{\phi}_{ij}^{(\mathcal{S},l)}}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} - \frac{\Phi_{ij}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}}{2\Phi_{jj}^{(\mathcal{S},l)}\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}} - \frac{\Phi_{ij}^{(\mathcal{S},l)}\widetilde{\phi}_{ii}^{(\mathcal{S},l)}}{2\Phi_{ii}^{(\mathcal{S},l)}\sqrt{\Phi_{ii}^{(\mathcal{S},l)}\Phi_{jj}^{(\mathcal{S},l)}}}\right), \end{aligned} \tag{B.43}$$

where $\widetilde{\phi}_{ij}^{(\mathcal{S},l)} := \frac{1}{n_lp}\sum_{k=1}^{n_l}\sum_{t=1}^{p}\epsilon_{ti}^{(\mathcal{S},k,l)\top}\epsilon_{tj}^{(\mathcal{S},k,l)} - \Phi_{ij}^{(\mathcal{S},l)}$. We notice that $\epsilon_{ti}^{(\mathcal{S},k,l)}\epsilon_{tj}^{(\mathcal{S},k,l)}$ are not independent across $t = 1, \ldots, p$ due to the temporal association, while the state-of-the-art theory of high-dimensional CLT elaborates only over independent random vectors yet. We use Eq. (B.2) to convert $\phi^{(\mathcal{S},l)}$ into a summation of independent random variables:

$$\widetilde{\phi}_{ij}^{(\mathcal{S},l)} = \frac{1}{n_lp}\sum_{k=1}^{n_l}\sum_{t=1}^{p}\lambda_t^{(\mathcal{T},l)}\xi_{ti}^{(k,l)}\xi_{tj}^{(k,l)} - \Phi_{ij}^{(\mathcal{S},l)} = \frac{1}{n_lp}\sum_{k=1}^{n_l}\sum_{t=1}^{p}\lambda_t^{(\mathcal{T},l)}\left(\xi_{ti}^{(k,l)}\xi_{tj}^{(k,l)} - \Phi_{ij}^{(\mathcal{S},l)}\right), \tag{B.44}$$

where $\xi_t^{(k,l)} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \Phi^{(\mathcal{S},l)})$, and the last equation holds based on $\mathrm{tr}(\Sigma^{(\mathcal{T},l)}) = 1$. Then, we rewrite

$$\Theta_{ij} = \frac{1}{\sqrt{Np}}\sum_{l=1}^{m}\sum_{k=1}^{n_l}\sum_{t=1}^{p}\theta_{t,ij}^{(k,l)}, \tag{B.45}$$

where $N = \sum_{l=1}^{m} n_l$, and

$$\theta_{t,ij}^{(k,l)} := \lambda_t^{(\mathcal{T},l)} \sqrt{\frac{N}{n_l m}} \left(\frac{\xi_{t,i}^{(k,l)} \xi_{t,j}^{(k,l)} - \Phi_{ij}^{(\mathcal{S},l)}}{\sqrt{\Phi_{ii}^{(\mathcal{S},l)} \Phi_{jj}^{(\mathcal{S},l)}}} - \frac{\Phi_{ij}^{(\mathcal{S},l)} (\xi_{t,j}^{(k,l)2} - \Phi_{jj}^{(\mathcal{S},l)})}{2\Phi_{jj}^{(\mathcal{S},l)} \sqrt{\Phi_{ii}^{(\mathcal{S},l)} \Phi_{jj}^{(\mathcal{S},l)}}} - \frac{\Phi_{ij}^{(\mathcal{S},l)} (\xi_{t,i}^{(k,l)2} - \Phi_{jj}^{(\mathcal{S},l)})}{2\Phi_{ii}^{(\mathcal{S},l)} \sqrt{\Phi_{ii}^{(\mathcal{S},l)} \Phi_{jj}^{(\mathcal{S},l)}}} \right). \tag{B.46}$$

For fixed i and j, because $\xi_{ti}^{(k,l)}$ and $\xi_{tj}^{(k,l)}$ are Gaussian random variables independent across l, k, and t, $\theta_{t,ij}^{(k,l)}$'s are independent sub-exponential random variable across l, k, and t. It satisfies Condition (E.1) of Corollary 2.1 in Chernozhukov et al. (2012). To obtain a non-asymptotic bound for $\sup_{x>0}|\mathbb{P}[\|T_E\|_\infty > x] - \mathbb{P}[\|Z\|_\infty > x]|$, we apply Theorem 2.2 in Chernozhukov et al. (2012) to obtain

$$KS_\infty(\Theta_E, Z) \leq C(\kappa_1) \max \left\{ \frac{\log^{7/8}(qmn_0p)}{(mn_0p)^{1/8}}, \frac{\log^2(qmn_0p)}{(mn_0p)^{1/2}} \right\}. \tag{B.47}$$

Alongside the results in Proposition 2.3.2 and Eq. (B.42), it obtains the desired result.

B.2.5 Proof of Theorem 2.3.5

We denote the conditional Kolmogorov-Smirnov distance between $\|\widehat{Z}\|_\infty$ and $\|Z\|_\infty$ given observed data $\mathcal{D}$ by $KS_\infty(\widehat{Z}, Z|\mathcal{D})$.

$$KS_\infty(\widehat{Z}, Z|\mathcal{D}) := \sup_{x>0} |\mathbb{P}[\|\widehat{Z}\|_\infty > x|\mathcal{D}] - \mathbb{P}[\|Z\|_\infty > x]|. \tag{B.48}$$

The error of the probability estimate is upperbounded by

$$\begin{aligned} &\sup_{x>0} |\mathbb{P}[\|\widehat{T}_E\|_\infty > x] - \mathbb{P}[\|\widehat{Z}\|_\infty > x|\mathcal{D}]| \\ &\leq \sup_{x>0} |\mathbb{P}[\|\widehat{T}_E\|_\infty > x] - \mathbb{P}[\|Z\|_\infty > x]| + KS_\infty(\widehat{Z}, Z|\mathcal{D}). \end{aligned} \tag{B.49}$$

Proposition 2.3.4 provides an upperbound for the first term. Therefore, it suffices to derive an upperbound for $KS_\infty(\widehat{Z}, Z|\mathcal{D})$. According to Lemma 3.1 in Chernozhukov et al. (2012),

$$KS_\infty(\widehat{Z}, Z|\mathcal{D}) \leq C \|\widehat{S}_{EE} - S_{EE}\|_\infty^{1/3} \{1 \vee \log(|E| / \|\widehat{S}_{EE} - S_{EE}\|_\infty)\}^{2/3}. \tag{B.50}$$

Therefore, it suffices to upperbound the size of $\|\widehat{S}_{EE} - S_{EE}\|_\infty$. In Eq. (2.17), $S_{(i_1,j_1),(i_2,j_2)}$ is given by

$$\begin{aligned} &S_{(i_1,j_1),(i_2,j_2)} \\ &:= \frac{1}{m} \sum_{l=1}^{m} \frac{\|\Sigma^{(\mathcal{T},l)}\|_F^2}{p} \left[\begin{array}{l} \rho_{i_1i_2}^{(\mathcal{S},l)} \rho_{j_1j_2}^{(\mathcal{S},l)} + \rho_{i_1j_2}^{(\mathcal{S},l)} \rho_{i_2j_1}^{(\mathcal{S},l)} + \frac{1}{2} \rho_{i_1j_1}^{(\mathcal{S},l)} \rho_{i_2j_2}^{(\mathcal{S},l)} \left(\rho_{i_1i_2}^{(\mathcal{S},l)2} + \rho_{j_1j_2}^{(\mathcal{S},l)2} + \rho_{i_1j_2}^{(\mathcal{S},l)2} + \rho_{i_2j_1}^{(\mathcal{S},l)2} \right) \\ - \rho_{i_1i_2}^{(\mathcal{S},l)} \rho_{i_2j_2}^{(\mathcal{S},l)} \rho_{i_2j_1}^{(\mathcal{S},l)} - \rho_{i_1i_2}^{(\mathcal{S},l)} \rho_{i_1j_1}^{(\mathcal{S},l)} \rho_{i_1j_2}^{(\mathcal{S},l)} - \rho_{j_1j_2}^{(\mathcal{S},l)} \rho_{i_2j_2}^{(\mathcal{S},l)} \rho_{i_1j_2}^{(\mathcal{S},l)} - \rho_{j_1j_2}^{(\mathcal{S},l)} \rho_{i_2j_1}^{(\mathcal{S},l)} \rho_{i_1j_1}^{(\mathcal{S},l)} \end{array} \right]. \end{aligned} \tag{B.51}$$

Corollary 2 gives a bound on the error in $\|\widehat{\Sigma}^{(\mathcal{T},l)}\|_F^2$:

$$\max_{l=1,\dots,m} \frac{\|\widehat{\Sigma}^{(\mathcal{T},l)}\|_F^2 - \|\Sigma^{(\mathcal{T},l)}\|_F^2}{p} \leq C(\kappa_1, \kappa_3) \frac{\log(qmn_0p)}{(qn_0)^{1-\frac{1}{2(\alpha_0+1)}}} \tag{B.52}$$

with probability at least $1 - (qmn_0p)^{-1/2}$. On the other hand, Proposition 2.3.2 implies that the error in $\widehat{\rho}_{ij}^{(\mathcal{S},l)}$ is mainly driven by $\Theta_{ij}^{(\mathcal{S},l)}$, where the size of $\Theta_{ij}^{(\mathcal{S},l)}$ is bounded by the Berry-Esseen bound for sub-exponential random variables (for example, see Kuchibhotla and Chakrabortty, 2018):

$$\max_{i,j} \sum_{l=1}^{m} |\Theta_{ij}^{(\mathcal{S},l)}| \leq C(\kappa_1, \kappa_3) \sqrt{\frac{m + \log(qmn_0p)}{mn_0p}} \tag{B.53}$$

with probability at least $1 - (qmn_0p)^{-1/2}$. (Notice that, because of Assumption 4, the above probabilistic bound is larger than $d\frac{m+\log^2(qmn_0p)}{n_0p}$, which is given as a probabilistic bound for $\max_{i,j} \sum_{l=1}^m |O_{ij}^{(\mathcal{S},l)}|$ in Proposition 2.3.2.) Because Assumption 3 implies that both $\frac{\|\Sigma^{(\mathcal{T},l)}\|_F^2}{p}$ and $\min_{i,j} \rho_{ij}^{(\mathcal{S},l)}$ are bounded away from both 0 and ∞ by some constants dependent to κ_3,

$$\begin{aligned} \|\widehat{S}_{EE} - S_{EE}\|_\infty &\leq C(\kappa_1, \kappa_3) \max\left\{ \max_{l=1,\dots,m} \frac{\|\widehat{\Sigma}^{(\mathcal{T},l)}\|_F^2 - \|\Sigma^{(\mathcal{T},l)}\|_F^2}{p}, \max_{i,j} \sum_{l=1}^{m} |\Theta_{ij}^{(\mathcal{S},l)}| \right\} \\ &\leq C(\kappa_1, \kappa_3, \kappa_5) \max\left\{ \frac{\log(qmn_0p)}{(n_0q)^{1-\frac{1}{2(\alpha_0+1)}}}, \sqrt{\frac{m + \log(qmn_0p)}{mn_0p}} \right\} \end{aligned} \tag{B.54}$$

with probability at least $1 - C(qmn_0p)^{-1/2}$ for a sufficiently large n_0. Plugging the result into Eq. (B.50), we obtain the conclusion of the theorem.

B.2.6 Proof of Theorem 2.3.6

Based on the definition of $\mathcal{C}_E(1-\alpha)$ in Eq. (2.18),

$$\mathbf{0} \in \mathcal{C}_E(1-\alpha) \iff \|\widehat{T}_E\|_\infty \leq \widehat{q}_{\|\widehat{Z}\|_\infty, 1-\alpha}. \tag{B.55}$$

Under the null hypothesis, $T_E = \mathbf{0}$, and

$$\begin{aligned} \mathbb{P}[\mathbf{0} \in \mathcal{C}_E(1-\alpha)] &= \mathbb{P}[\|\widehat{T}_E\|_\infty \leq \widehat{q}_{\|\widehat{Z}\|_\infty, 1-\alpha}] \\ &\leq \mathbb{P}[\|\widehat{Z}\|_\infty \leq \widehat{q}_{\|\widehat{Z}\|_\infty, 1-\alpha} | \mathcal{D}] + \sup_{x>0} |\mathbb{P}[\|\widehat{T}_E\|_\infty \leq x] - \mathbb{P}[\|\widehat{Z}\|_\infty \leq x | \mathcal{D}]| \\ &= 1 - \alpha + \sup_{x>0} |\mathbb{P}[\|\widehat{T}_E\|_\infty \leq x] - \mathbb{P}[\|\widehat{Z}\|_\infty \leq x | \mathcal{D}]| \end{aligned} \tag{B.56}$$

almost surely. Then, due to Theorem 2.3.5,

$$\mathbb{P}[\mathbf{0} \in \mathcal{C}_E(1-\alpha)] \leq 1 - \alpha + \mathcal{M} \tag{B.57}$$

with probability at least $1 - C(qmn_0p)^{-1/2}$, where $\mathcal{M}$ is the probabilistic upper bound for $\sup_{x>0}|\mathbb{P}[\|\widehat{T}_E\|_\infty \leq x] - \mathbb{P}[\|\widehat{Z}\|_\infty \leq x|\mathcal{D}]|$ in Theorem 2.3.5. Similarly,

$$\mathbb{P}[\mathbf{0} \in \mathcal{C}_E(1-\alpha)] \geq 1 - \alpha - \mathcal{M} \tag{B.58}$$

with the same probability. Because the assumed sample complexity implies that $\mathcal{M} \to 0$, $\mathbb{P}[\mathbf{0} \notin \mathcal{C}_E(1-\alpha)] \overset{p}{\to} \alpha$, which verifies the validity of the testing procedure.

For the second claim, we use existing results in the extreme value theory of Gaussian random vectors (e.g., see Exercise 5.10, Wainwright, 2019) that

$$\mathbb{P}\left[\|\widehat{Z}\|_\infty \geq \mathbb{E}[\|\widehat{Z}\|_\infty|\mathcal{D}] + \nu\middle|\mathcal{D}\right] \leq \exp\left(-\frac{\nu^2}{2\max_{(i,j)\in E}\widehat{S}_{(i,j),(i,j)}}\right). \tag{B.59}$$

and

$$\mathbb{E}[\|\widehat{Z}\|_\infty|\mathcal{D}] \leq C\sqrt{\log q}\max_{(i,j)\in E}\widehat{S}_{(i,j),(i,j)}, \tag{B.60}$$

almost surely. With $\nu = \sqrt{-2\log\alpha\max_{(i,j)\in E}\widehat{S}_{(i,j),(i,j)}}$,

$$\widehat{q}_{\|\widehat{Z}\|_\infty,1-\alpha} \leq (C\sqrt{\log q} + \sqrt{-2\log\alpha})\max_{(i,j)\in E}\widehat{S}_{(i,j),(i,j)}, \tag{B.61}$$

almost surely. As we saw in Appendix B.2.5,

$$\|\widehat{S}_{EE} - S_{EE}\|_\infty \leq C(\kappa_1,\kappa_3,\kappa_5)\max\left\{\frac{\log(qmn_0p)}{(n_0q)^{1-\frac{1}{2(\alpha_0+1)}}}, \sqrt{\frac{m+\log(qmn_0p)}{mn_0p}}\right\} \tag{B.62}$$

with probability at least $1 - C(qmn_0p)^{-1/2}$. Due to Assumptions 4 and 5, the right hand side converges to 0 as $n_0 \to \infty$, and therefore

$$\widehat{q}_{\|\widehat{Z}\|_\infty,1-\alpha} \leq C(\kappa_1,\kappa_3,\kappa_5)(\sqrt{\log q} + \sqrt{-2\log\alpha})\max_{(i,j)\in E}S_{(i,j),(i,j)} \tag{B.63}$$

with the same probability for sufficiently large n_0. Suppose that the assumed alternative hypothesis implies that

$$\|T_E\|_\infty \geq C(\kappa_1,\kappa_3,\kappa_5)\sqrt{\log q}\max_{(i,j)\in E}S_{(i,j),(i,j)} \geq \widehat{q}_{\|\widehat{Z}\|_\infty,1-\alpha} + \widehat{q}_{\|\widehat{Z}\|_\infty,1/q}. \tag{B.64}$$

Then,

$$\begin{aligned}\mathbb{P}[\mathbf{0} \in \mathcal{C}_E(1-\alpha)] &= \mathbb{P}[\|\widehat{T}_E\|_\infty \leq \widehat{q}_{\|\widehat{Z}\|_\infty,1-\alpha}] \\ &\leq \mathbb{P}[\|T_E\|_\infty - \|\widehat{T}_E - T_E\|_\infty \leq \widehat{q}_{\|\widehat{Z}\|_\infty,1-\alpha}] \\ &= \mathbb{P}[\|\widehat{T}_E - T_E\|_\infty \geq \widehat{q}_{\|\widehat{Z}\|_\infty,1/q}] \\ &\leq \mathbb{P}[\|\widehat{Z}\|_\infty \geq \widehat{q}_{\|\widehat{Z}\|_\infty,1/q}|\mathcal{D}] + \sup_{x>0}\left|\mathbb{P}[\|\widehat{T}_E\|_\infty \leq x] - \mathbb{P}[\|\widehat{Z}\|_\infty \leq x|\mathcal{D}]\right| \\ &= 1/q + \sup_{x>0}\left|\mathbb{P}[\|\widehat{T}_E\|_\infty \leq x] - \mathbb{P}[\|\widehat{Z}\|_\infty \leq x|\mathcal{D}]\right|,\end{aligned} \tag{B.65}$$

which converges to 0 as $n_0 \to \infty$, due to the assumed sample complexity. It verifies the second conclusion about the power analysis.

B.2.7 Proof of Proposition 2.3.7

Let $\widetilde{\beta}^{(\mathcal{T},l)}$ be the projected parameter of $\beta^{(\mathcal{T},l)}$ onto the parameter space under the banded Cholesky factor assumption. That is, for $t = 1, \ldots, p$,

$$\begin{aligned} &\widetilde{\beta}^{(\mathcal{T},l)}_{\cdot,t} := \underset{b \in \mathbb{R}^p}{\operatorname{argmin}}\, \mathbb{E}\left[\|X^{(\mathcal{T},l)}_{t,\cdot} - X^{(\mathcal{T},l)\top} b\|_2^2\right] \\ &\text{w.r.t } \; b_s = 0 \; \text{ where } \; s < t - h_l \; \text{ or } \; s \geq t. \end{aligned} \tag{B.66}$$

According to the regression theory, $\widetilde{\beta}^{(\mathcal{T},l)}_{\cdot,t}$ can be rewritten in terms of $\Sigma^{(\mathcal{T},l)}$ only:

$$\widetilde{\beta}^{(\mathcal{T},l)}_{t-h:t-1,t} = -\widetilde{\Omega}^{(\mathcal{T},l)}_{t,h_l+1,1:h_l} / \widetilde{\Omega}^{(\mathcal{T},l)}_{t,h_l+1,h_l+1}, \tag{B.67}$$

where $\widetilde{\Omega}^{(\mathcal{T},l)}_t = \left(\Sigma^{(\mathcal{T},l)}_{t-h_l:t,t-h_l:t}\right)^{-1} \in \mathbb{R}^{h_l+1 \times h_l+1}$. We denote the regression error by

$$\widetilde{\epsilon}^{(\mathcal{T},l)}_{t,i} := X^{(\mathcal{T},l)}_{t,i} - X^{(\mathcal{T},l)\top}_{\cdot,i} \widetilde{\beta}^{(\mathcal{T},l)}_{\cdot,t}, \tag{B.68}$$

and $\widetilde{\Phi}^{(\mathcal{T},l)}_{tt} := \frac{1}{n_l q} \mathbb{E}[\|\widetilde{\epsilon}^{(\mathcal{T},l)}_{t,i}\|_2^2]$. Then,

$$\widetilde{\Phi}^{(\mathcal{T},l)}_{tt} = \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q} \frac{1}{\widetilde{\Omega}^{(\mathcal{T},l)}_{t,h_l+1,h_l+1}}. \tag{B.69}$$

We notice that $\widetilde{\Omega}^{(\mathcal{T},l)}_{t,h_l+1,1:h_l} / \widetilde{\Omega}^{(\mathcal{T},l)}_{t,h_l+1,h_l+1}$ and $1/\widetilde{\Omega}^{(\mathcal{T},l)}_{t,h_l+1,h_l+1}$ are the projected parameter and prediction error of the $AR(h_l)$ regression on vector-variate observations, so the bias analyses in the existing literature (e.g., Liu and Ren (2020)) is useful here. For example, Lemma B.3 in Liu and Ren (2020) implies that under Assumption 5

$$\begin{aligned} \left\|\widetilde{\beta}^{(\mathcal{T},l)}_{\cdot,t} - \beta^{(\mathcal{T},l)}_{\cdot,t}\right\|_2 &\leq C(\kappa_3, \kappa_5)(h_l - 1)^{-\alpha_l - 1/2}, \\ \left|\widetilde{\Phi}^{(\mathcal{T},l)}_{tt} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q} \Phi^{(\mathcal{T},l)}_{tt}\right| &\leq C(\kappa_3, \kappa_5)(h_l - 1)^{-\alpha_l - 1/2}. \end{aligned} \tag{B.70}$$

Now, it suffices to establish the probabilistic error bound of $\widehat{\beta}^{(\mathcal{T},l)}$ and $\widehat{\Phi}^{(\mathcal{T},l)}$ against $\widetilde{\beta}^{(\mathcal{T},l)}$ and $\widetilde{\Phi}^{(\mathcal{T},l)}$, respectively. We take a similar approach with the proof of Theorem 2.3.1. First, we notice that

$$\widehat{\beta}^{(\mathcal{T},l)}_{t-h_l:t-1,t} = \left(X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot} X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot}\right)^{-1} \left(X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot} X^{(\mathcal{T},l)\top}_{t,\cdot}\right). \tag{B.71}$$

Then,

$$\begin{aligned} &\left(\frac{1}{n_l q} X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot} X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot}\right) (\widehat{\beta}^{(\mathcal{T},l)}_{\cdot,t} - \widetilde{\beta}^{(\mathcal{T},l)}_{\cdot,t}) \\ &= \frac{1}{n_l q} X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot} X^{(\mathcal{T},l)\top}_{t,\cdot} - \left(\frac{1}{n_l q} X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot} X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot}\right) \widetilde{\beta}^{(\mathcal{T},l)}_{t-h_l:t-1,t} \\ &= \frac{1}{n_l q} X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot} \widetilde{\epsilon}^{(\mathcal{T},l)\top}_{t,\cdot}. \end{aligned} \tag{B.72}$$

By a similar argument with the proof of Lemma B.2.3,

$$\begin{aligned}
&\mathbb{P}\left[\max_{l=1,\ldots,m}\max_{t=1,\ldots,p}\left\|\left(\frac{1}{n_l q}X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot}X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot}\right)\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\right\|_\infty \geq C(\kappa_1,\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_0q}}\right]\\
&=\mathbb{P}\left[\max_{l=1,\ldots,m}\max_{t=1,\ldots,p}\left\|\frac{1}{n_l q}X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot}\widetilde{\epsilon}^{(\mathcal{T},l)\top}_{t,\cdot}\right\|_\infty \geq C(\kappa_1,\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_0q}}\right]\\
&\leq C(qmn_0p)^{-1/2}
\end{aligned}\tag{B.73}$$

for some positive constants $C(\kappa_3)$ and C, where $\widetilde{\Delta}^{(\mathcal{T},l)} := \widehat{\beta}^{(\mathcal{T},l)} - \widetilde{\beta}^{(\mathcal{T},l)}$. Under the event,

$$\begin{aligned}
\left|\widetilde{\Delta}^{(\mathcal{T},l)\top}_{\cdot,t}\left(\frac{1}{n_l q}X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot}X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot}\right)\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\right| &\leq \left\|\left(\frac{1}{n_l q}X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot}X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot}\right)\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\right\|_\infty\|\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\|_1\\
&\leq C(\kappa_1,\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_0q}}\|\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\|_1\\
&\leq C(\kappa_1,\kappa_3)\sqrt{h_l\frac{\log(qmn_0p)}{n_0q}}\|\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\|_2,
\end{aligned}\tag{B.74}$$

for any $t = 1,\ldots,T$ and $l = 1,\ldots,m$. On the other hand, by a similar argument with the proof of Lemma B.2.1,

$$\begin{aligned}
&\mathbb{P}\left[\max_{l=1,\ldots,m}\max_{s,t}\left|\frac{1}{n_l q}X^{(\mathcal{T},l)}_{s,\cdot}X^{(\mathcal{T},l)\top}_{t,\cdot} - \frac{\mathrm{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Sigma^{(\mathcal{T},l)}_{st}\right| \geq C(\kappa_1,\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_0q}}\right]\\
&\leq C(qmn_0p)^{-1/2}
\end{aligned}\tag{B.75}$$

for some positive constants $C(\kappa_3)$ and C. Under the event,

$$\begin{aligned}
&\left|\widetilde{\Delta}^{(\mathcal{T},l)\top}_{\cdot,t}\left(\frac{1}{n_l q}X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot}X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot} - \frac{\mathrm{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Sigma^{(\mathcal{T},l)}_{t-h_l:t-1,t-h_l:t-1}\right)\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\right|\\
&\leq \left\|\frac{1}{n_l q}X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot}X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot} - \frac{\mathrm{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Sigma^{(\mathcal{T},l)}_{t-h_l:t-1,t-h_l:t-1}\right\|_\infty\|\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\|_1^2\\
&\leq C(\kappa_1,\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_0q}}\|\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\|_1^2 \leq C(\kappa_1,\kappa_3)h_l\sqrt{\frac{\log(qmn_0p)}{n_0q}}\|\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\|_2^2\\
&\leq \frac{1}{2\kappa_3}\|\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\|_2^2,
\end{aligned}\tag{B.76}$$

for any $t = 1,\ldots,T$, $l = 1,\ldots,m$, and a sufficiently large n_0 due to Assumption 5. Moreover with Assumption 3,

$$\left|\widetilde{\Delta}^{(\mathcal{T},l)\top}_{\cdot,t}\left(\frac{1}{n_l q}X^{(\mathcal{T},l)}_{t-h_l:t-1,\cdot}X^{(\mathcal{T},l)\top}_{t-h_l:t-1,\cdot}\right)\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\right| \geq \frac{1}{2\kappa_3}\|\widetilde{\Delta}^{(\mathcal{T},l)}_{\cdot,t}\|_2^2 \tag{B.77}$$

with probability at least $1 - C(qmn_0p)^{-1/2}$. Eqs. (B.74) and (B.77) imply

$$\frac{1}{2\kappa_3}\|\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)}\|_2^2 \le C(\kappa_1,\kappa_3)\sqrt{h_l\frac{\log(qmn_0p)}{n_0q}}\|\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)}\|_2, \tag{B.78}$$

and

$$\|\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)}\|_2 \le C(\kappa_1,\kappa_3)\sqrt{h_l\frac{\log(qmn_0p)}{n_0q}} \text{ and } \|\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)}\|_1 \le C(\kappa_1,\kappa_3)h_l\sqrt{\frac{\log(qmn_0p)}{n_0q}}. \tag{B.79}$$

for any $t = 1, \dots, T$, $l = 1, \dots, m$ and a sufficiently large n_0, with probability at least $1 - C(qmn_0p)^{-1/2}$. Integrating Eqs. (B.70) and (B.79), we obtain the first result that, with the same probability,

$$\begin{aligned}\|\widehat{\beta}_{\cdot,t}^{(\mathcal{T},l)} - \beta_{\cdot,t}^{(\mathcal{T},l)}\|_2 &\le C(\kappa_3,\kappa_5)(h_l-1)^{-\alpha_l-1/2} + C(\kappa_1,\kappa_3)h_l\sqrt{\frac{\log(qmn_0p)}{n_0q}} \\ &\le C(\kappa_1,\kappa_3,\kappa_5)\sqrt{\frac{\log(qmn_0p)}{(n_0q)^{1-1/2(\alpha_0+1)}}},\end{aligned} \tag{B.80}$$

for any $t = 1, \dots, T$, $l = 1, \dots, m$ and a sufficiently large n_0, given that $h_l = \lfloor (n_lq)^{1/2(\alpha_l+1)} \rfloor$.

For the second result, we observe that

$$\left|\widehat{\Phi}_{tt}^{(\mathcal{T},l)} - \widetilde{\Phi}_{tt}^{(\mathcal{T},l)}\right| \le \frac{1}{n_lq}\left|\|\widehat{\epsilon}_{t,\cdot}^{(\mathcal{T},l)}\|_2^2 - \|\widetilde{\epsilon}_{t,\cdot}^{(\mathcal{T},l)}\|_2^2\right| + \left|\frac{1}{n_lq}\|\widetilde{\epsilon}_{t,\cdot}^{(\mathcal{T},l)}\|_2^2 - \widetilde{\Phi}_{tt}^{(\mathcal{T},l)}\right|. \tag{B.81}$$

For the first term,

$$\begin{aligned}&\frac{1}{n_lq}\left|\|\widehat{\epsilon}_{t,\cdot}^{(\mathcal{T},l)}\|_2^2 - \|\widetilde{\epsilon}_{t,\cdot}^{(\mathcal{T},l)}\|_2^2\right| \\ &= \frac{2}{n_lq}\left|\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)\top}X_{t-h:t-1,\cdot}^{(\mathcal{T},l)}\widetilde{\epsilon}_{t,\cdot}^{(\mathcal{T},l)\top}\right| + \frac{1}{n_lq}\left|\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)\top}X_{t-h:t-1}^{(\mathcal{T},l)}X_{t-h:t-1}^{(\mathcal{T},l)\top}\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)}\right| \\ &\le \left\|\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)\top}\right\|_1\left\|\frac{2}{n_lq}X_{t-h:t-1,\cdot}^{(\mathcal{T},l)}\widetilde{\epsilon}_{t,\cdot}^{(\mathcal{T},l)\top}\right\|_\infty + \left\|\widetilde{\Delta}_{\cdot,t}^{(\mathcal{T},l)\top}\right\|_2^2\left\|\frac{1}{n_lq}X_{t-h:t-1}^{(\mathcal{T},l)}X_{t-h:t-1}^{(\mathcal{T},l)\top}\right\|_2 \\ &\le C(\kappa_1,\kappa_3)h\frac{\log p}{n_lq}\end{aligned} \tag{B.82}$$

for any $t = 1, \dots, T$, $l = 1, \dots, m$, and a sufficiently large n_0, with probability at least $1 - C(qmn_0p)^{-1/2}$, based on Eqs. (B.74) and (B.76) and Assumption 3. For the second term, a similar argument with the proof of Lemma B.2.4 implies that

$$\begin{aligned}&\mathbb{P}\left[\max_{l=1,\dots,m}\max_{s,t}\left|\frac{1}{n_lq}\widetilde{\epsilon}_{s,\cdot}^{(\mathcal{T},l)}\widetilde{\epsilon}_{t,\cdot}^{(\mathcal{T},l)\top} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q}\widetilde{\Phi}_{st}^{(\mathcal{T},l)}\right| \ge C(\kappa_1,\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_lq}}\right] \\ &\le C(qmn_0p)^{-1/2}.\end{aligned} \tag{B.83}$$

In sum,

$$\left|\widehat{\Phi}^{(\mathcal{T},l)}_{tt} - \widetilde{\Phi}^{(\mathcal{T},l)}_{tt}\right| \le C(\kappa_1,\kappa_3) h \frac{\log p}{n_0 q} + C(\kappa_1,\kappa_3)\sqrt{\frac{\log(qmn_0p)}{n_l q}} \le C(\kappa_1,\kappa_3)\sqrt{h\frac{\log p}{n_0 q}} \tag{B.84}$$

for any $t = 1, \ldots, T$, $l = 1, \ldots, m$ and a sufficiently large n_0, with probability at least $1 - C(qmn_0p)^{-1/2}$, which implies the second conclusion where $h = \lfloor (n_l q)^{1/2(\alpha_l+1)} \rfloor$.

B.2.8 Proof of Theorem 2.3.8

We recall the Cholesky decomposition of the temporal precision matrix:

$$\begin{aligned}
\Omega^{(\mathcal{T},l)} &= \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q}(I - \beta^{(\mathcal{T},l)\top})\Phi^{(\mathcal{T},l)-1}(I - \beta^{(\mathcal{T},l)}) \\
\Sigma^{(\mathcal{T},l)} &= \frac{q}{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}(I - \beta^{(\mathcal{T},l)\top})^{-1}\Phi^{(\mathcal{T},l)}(I - \beta^{(\mathcal{T},l)})^{-1}
\end{aligned} \tag{B.85}$$

Our temporal covariance matrix estimate is based on the estimated Cholesky factor, $\widehat{\beta}^{(\mathcal{T},l)}$, and noise variance, $\widehat{\Phi}^{(\mathcal{T},l)}$, in Eqs. (2.23) and (2.24). In regard to their estimate errors, we use the following lemma as a corollary of Proposition 2.3.7. See Appendix B.3.7 for the proof.

Lemma B.2.7. *Suppose that $h_l = \lfloor (n_l q)^{1/(1+\alpha)} \rfloor$ and $\eta = C(\kappa_3)$ satisfies $\eta \le \lambda_1(I - \beta^{(\mathcal{T},l)})$ for $l = 1, \ldots, m$. Then, following the procedure defined in Section 2.2.3,*

$$\begin{aligned}
\mathbb{P}\Bigg[& \max_l \frac{1}{p}\left\|\widehat{\Phi}^{(\mathcal{T},l)} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Phi^{(\mathcal{T},l)}\right\|_F^2 \ge C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}}, \\
& \max_l \frac{1}{p}\left\|\widehat{\Phi}^{(\mathcal{T},l)-1} - \frac{q}{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}\Phi^{(\mathcal{T},l)-1}\right\|_F^2 \ge C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}} \Bigg] \\
& \le C(qmn_0p)^{-1/2},
\end{aligned} \tag{B.86}$$

$$\begin{aligned}
\mathbb{P}\Bigg[& \max_l \frac{1}{p}\|P_\eta(I - \widehat{\beta}^{(\mathcal{T},l)}) - (I - \beta^{(\mathcal{T},l)})\|_F^2 \ge C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}}, \\
& \max_l \frac{1}{p}\|P_\eta(I - \widehat{\beta}^{(\mathcal{T},l)})^{-1} - (I - \beta^{(\mathcal{T},l)})^{-1}\|_F^2 \ge C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}} \Bigg] \\
& \le C(qmn_0p)^{-1/2},
\end{aligned} \tag{B.87}$$

for sufficiently large n_0.

We note that Assumption 3 implies that the operator norms $I - \beta^{(\mathcal{T},l)}$, $(I - \beta^{(\mathcal{T},l)})^{-1}$, $\Phi^{(\mathcal{T},l)}$, and $\Phi^{(\mathcal{T},l)-1}$ are bounded; see Lemma B.2 in Liu and Ren (2020). The eigenvalue truncation in Eq. (2.25) implies the bounded operator norm for $P_\eta(I - \widehat{\beta}^{(\mathcal{T},l)})$ and $P_\eta(I - \widehat{\beta}^{(\mathcal{T},l)})^{-1}$. Finally, Theorem 2.3.8 implies the bounded operator norm for $\widehat{\Phi}^{(\mathcal{T},l)}$ and $\widehat{\Phi}^{(\mathcal{T},l)-1}$. Namely,

$$\begin{aligned} \|I-\beta^{(\mathcal{T},l)}\|_{\text{op}}, \|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})\|_{\text{op}} &\leq C(\kappa_3) \\ \|(I-\beta^{(\mathcal{T},l)})^{-1}\|_{\text{op}}, \|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})^{-1}\|_{\text{op}} &\leq C(\kappa_3) \\ \|\Phi^{(\mathcal{T},l)}\|_{\text{op}}, \left\|\frac{\text{tr}(\Sigma^{(\mathcal{S},l)})}{q}\widehat{\Phi}^{(\mathcal{T},l)}\right\|_{\text{op}} &\leq C(\kappa_3) \\ \|\Phi^{(\mathcal{T},l)-1}\|_{\text{op}}, \left\|\frac{q}{\text{tr}(\Sigma^{(\mathcal{S},l)})}\widehat{\Phi}^{(\mathcal{T},l)-1}\right\|_{\text{op}} &\leq C(\kappa_3) \end{aligned} \tag{B.88}$$

for any $l = 1, \ldots, m$ and a sufficiently large n_0, with probability at least $1 - C(qmn_0p)^{-1/2}$.

Now, we move on to connect the error in the temporal covariance matrix estimate with the Frobenius norm bounds in Lemma B.2.7. We use a well known inequality about the Frobenius and operator norms:

$$\|AB\|_F \leq \|A\|_{\text{op}}\|B\|_F \tag{B.89}$$

for any compatible matrices A and B. For the temporal covariance matrix,

$$\begin{aligned} &\left\|\overline{\Sigma}^{(\mathcal{T},l)} - \frac{\text{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Sigma^{(\mathcal{T},l)}\right\|_F \\ &\leq \|(I-\beta^{(\mathcal{T},l)})^{-1}\|_{\text{op}}\left\|\frac{\text{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Phi^{(\mathcal{T},l)}\right\|_{\text{op}}\|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})^{-1} - (I-\beta^{(\mathcal{T},l)})^{-1}\|_F \\ &\quad + \|(I-\beta^{(\mathcal{T},l)})^{-1}\|_{\text{op}}\left\|\widehat{\Phi}^{(\mathcal{T},l)} - \frac{\text{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Phi^{(\mathcal{T},l)}\right\|_F\|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})^{-1}\|_{\text{op}} \\ &\quad + \|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})^{-1} - (I-\beta^{(\mathcal{T},l)})^{-1}\|_F\left\|\frac{\text{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Phi^{(\mathcal{T},l)}\right\|_{\text{op}}\|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})^{-1}\|_{\text{op}} \end{aligned} \tag{B.90}$$

and hence, plugging in the previously obtained bounds,

$$\left\|\overline{\Sigma}^{(\mathcal{T},l)} - \frac{\text{tr}(\Sigma^{(\mathcal{S},l)})}{q}\Sigma^{(\mathcal{T},l)}\right\|_F \leq C(\kappa_1,\kappa_3,\kappa_5)\sqrt{p}\sqrt{\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}}} \tag{B.91}$$

for any $l = 1, \ldots, m$ with probability at least $1 - C(qmn_0p)^{-1/2}$. The second conclusion about $\overline{\Omega}^{(\mathcal{T},l)}$ is similarly obtained.

B.2.9 Proof of Corollary 2.3.9

The error in the Frobenius norm estimate can be upperbounded by a function of the error in the covariance matrix estimate:

$$\begin{aligned} &\frac{1}{p}\left|\|\widehat{\Sigma}^{(\mathcal{T},l)}\|_F^2 - \|\Sigma^{(\mathcal{T},l)}\|_F^2\right| \\ &= \frac{1}{p}\left(\|\widehat{\Sigma}^{(\mathcal{T},l)}\|_F + \|\Sigma^{(\mathcal{T},l)}\|_F\right)\left|\|\widehat{\Sigma}^{(\mathcal{T},l)}\|_F - \|\Sigma^{(\mathcal{T},l)}\|_F\right| \\ &\leq \left(\frac{2}{\sqrt{p}}\|\Sigma^{(\mathcal{T},l)}\|_F + \frac{1}{\sqrt{p}}\|\widehat{\Sigma}^{(\mathcal{T},l)} - \Sigma^{(\mathcal{T},l)}\|_F\right)\frac{1}{\sqrt{p}}\|\widehat{\Sigma}^{(\mathcal{T},l)} - \Sigma^{(\mathcal{T},l)}\|_F. \end{aligned} \tag{B.92}$$

Because $\frac{1}{\sqrt{p}}\|\Sigma^{(\mathcal{T},l)}\|_F$ is bounded away from both 0 and ∞, and Theorem 2.3.8 implies that

$$\frac{1}{\sqrt{p}}\|\widehat{\Sigma}^{(\mathcal{T},l)} - \Sigma^{(\mathcal{T},l)}\|_F \leq C(\kappa_1, \kappa_3, \kappa_5)\sqrt{\frac{\log(qmn_0p)}{p(n_0q)^{1-1/(2\alpha_0+2)}}} \tag{B.93}$$

with probability at least $1 - C(qmn_0p)^{-1/2}$,

$$\begin{aligned} \frac{1}{p}\left|\|\widehat{\Sigma}^{(\mathcal{T},l)}\|_F^2 - \|\Sigma^{(\mathcal{T},l)}\|_F^2\right| &\leq C(\kappa_1, \kappa_3, \kappa_5)\frac{1}{\sqrt{p}}\|\widehat{\Sigma}^{(\mathcal{T},l)} - \Sigma^{(\mathcal{T},l)}\|_F \\ &\leq C(\kappa_1, \kappa_3, \kappa_5)\sqrt{\frac{\log(qmn_0p)}{p(n_0q)^{1-1/(2\alpha_0+2)}}} \end{aligned} \tag{B.94}$$

with probability at least $1 - C(qmn_0p)^{-1/2}$ for sufficiently large n_0.

B.3 Proof of the Lemmas

B.3.1 Proof of Lemma B.2.1

In Eq. (B.1), $W_{ti}^{(\mathcal{S},l)}W_{tj}^{(\mathcal{S},l)}$ is a sub-exponential variable with

$$\mathbb{E}\left[\exp\left(\nu(W_{ti}^{(\mathcal{S},l)}W_{tj}^{(\mathcal{S},l)} - \Sigma_{tt}^{(\mathcal{T},l)}\Sigma_{ij}^{(\mathcal{S},l)})\right)\right] \leq e^{C(\kappa_3)\nu^2}, \tag{B.95}$$

for all $\nu : |\nu| < C(\kappa_3)$ due to Assumption 3, and independent across $t = 1, \ldots, n_lp$. According to Theorem 2.8.2 in Vershynin (2018) and Assumptions 1 and 2, our main claim is a result of maximal ineqauality on sub-exponential random variables.

B.3.2 Proof of Lemma B.2.2

In Eq. (B.2),

$$(\xi_{\cdot,i}^{(\mathcal{S},l)}, W_{\cdot,j}^{(\mathcal{S},l)}) \overset{\text{i.i.d.}}{\sim} N\left(0, I \otimes \text{diag}\left(\frac{1}{\Omega_{ii}^{(\mathcal{S},l)}}, \Sigma_{jj}^{(\mathcal{S},l)}\right)\right). \tag{B.96}$$

Because $W_{\cdot,i}^{(\mathcal{S},l)}$ is independent to $\xi_{\cdot,j}^{(\mathcal{S},l)}$,

$$\left\{\xi_{\cdot,i}^{(\mathcal{S},l)}W_{\cdot,j}^{(\mathcal{S},l)\top}|W_{\cdot,j}^{(\mathcal{S},l)}\right\} \sim N\left(0, \frac{\|W_{\cdot,j}^{(\mathcal{S},l)}\|_2^2}{\Omega_{ii}^{(\mathcal{S},l)}}\right) \quad a.s., \tag{B.97}$$

and therefore

$$\begin{aligned} \left\{\left(\frac{\sqrt{n_lp}}{\|X_{\cdot,j}^{(\mathcal{S},l)}\|_2}\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}X_{\cdot,j}^{(\mathcal{S},l)}\right)^2 \middle| X_{\cdot,j}^{\mathcal{S},l)}\right\} &= \left\{\left(\frac{\sqrt{n_lp}}{\|X_{\cdot,j}^{(\mathcal{S},l)}\|_2}\xi_{\cdot,i}^{(\mathcal{S},l)\top}W_{\cdot,j}^{(\mathcal{S},l)}\right)^2 \middle| W_{\cdot,j}^{\mathcal{S},l)}\right\} \\ &\sim n_lp\frac{\|W_{\cdot,j}^{(\mathcal{S},l)}\|_2^2}{\|X_{\cdot,j}^{(\mathcal{S},l)}\|_2^2}\chi^2(1) \quad a.s. \end{aligned} \tag{B.98}$$

Due to Assumption 3, $\frac{\|W^{(\mathcal{S},l)}_{\cdot,j}\|_2^2}{\|X^{(\mathcal{S},l)}_{\cdot,j}\|_2^2} \leq C(\kappa_3)$. Hence, the sub-exponential inequality in Theorem 2.8.2, Vershynin (2018) implies that, for any $\nu > 0$,

$$\begin{aligned} &\mathbb{P}\left[\sum_{l=1}^{m}\left(\frac{\sqrt{n_l p}}{\|X^{(\mathcal{S},l)}_{\cdot,j}\|_2}\epsilon^{(\mathcal{S},l)\top}_{\cdot,i}X^{(\mathcal{S},l)}_{\cdot,j}\right)^2 - \sum_{l=1}^{m}\mathbb{E}\left(\frac{\sqrt{n_l p}}{\|X^{(\mathcal{S},l)}_{\cdot,j}\|_2}\epsilon^{(\mathcal{S},l)\top}_{\cdot,i}X^{(\mathcal{S},l)}_{\cdot,j}\right)^2 \geq \nu\right] \\ &\leq \exp\left[-C(\kappa_3)\min\left\{\left(\frac{\nu}{n_l p}\right)^2\frac{1}{m},\ \frac{\nu}{n_l p}\right\}\right]. \end{aligned} \tag{B.99}$$

Assumption 3 implies that $\sum_{l=1}^{m}\mathbb{E}\left(\frac{\sqrt{n_l p}}{\|X^{(\mathcal{S},l)}_{\cdot,j}\|_2}\epsilon^{(\mathcal{S},l)\top}_{\cdot,i}X^{(\mathcal{S},l)}_{\cdot,j}\right)^2 \leq C(\kappa_1,\kappa_3)\ mn_0p$. Plugging-in $\nu = C(\kappa_1,\kappa_3)\cdot n_0p\cdot\max\left\{\sqrt{m\log(qmn_0p)},\ \log(qmn_0p)\right\}$,

$$\begin{aligned} &\mathbb{P}\left[\frac{1}{n_0p}\sum_{l=1}^{m}\left(\frac{\sqrt{n_l p}}{\|X^{(\mathcal{S},l)}_{\cdot,j}\|_2}\epsilon^{(\mathcal{S},l)\top}_{\cdot,i}X^{(\mathcal{S},l)}_{\cdot,j}\right)^2 \geq C(\kappa_1,\kappa_3)\left(m+\log(qmn_0p)\right)\right] \\ &\leq \frac{2}{m(m-1)}(qmn_0p)^{-1/2}. \end{aligned} \tag{B.100}$$

The first conclusion follows the maximal inequality applied to the above probability bound. The second conclusion follws a similar argument with

$$(\xi^{(\mathcal{S},l)}_{\cdot,i}, W^{(\mathcal{S},l)}\beta^{(\mathcal{S},l)}_{\cdot,j}) \overset{\text{i.i.d.}}{\sim} N\left(0, I\otimes\operatorname{diag}\left(\frac{1}{\Omega^{(\mathcal{S},l)}_{ii}}, \frac{\Omega^{(\mathcal{S},l)}_{ii}\Sigma^{(\mathcal{S},l)}_{ii}-1}{\Omega^{(\mathcal{S},l)}_{ii}}\right)\right). \tag{B.101}$$

B.3.3 Proof of Lemma B.2.3

In Eq. (B.2),

$$(\xi^{(\mathcal{S},l)}_{\cdot,i}, W^{(\mathcal{S},l)}_{\cdot,j}) \overset{\text{i.i.d.}}{\sim} N\left(0, I\otimes\operatorname{diag}\left(\frac{1}{\Omega^{(\mathcal{S},l)}_{ii}}, \Sigma^{(\mathcal{S},l)}_{jj}\right)\right), \tag{B.102}$$

$$(\xi^{(\mathcal{S},l)}_{\cdot,i}, W^{(\mathcal{S},l)}\beta^{(\mathcal{S},l)}_{\cdot,j}) \overset{\text{i.i.d.}}{\sim} N\left(0, I\otimes\operatorname{diag}\left(\frac{1}{\Omega^{(\mathcal{S},l)}_{ii}}, \frac{\Omega^{(\mathcal{S},l)}_{ii}\Sigma^{(\mathcal{S},l)}_{ii}-1}{\Omega^{(\mathcal{S},l)}_{ii}}\right)\right). \tag{B.103}$$

In other words, $\xi^{(\mathcal{S},l)}_{ti}W^{(\mathcal{S},l)}_{tj}$ and $\xi^{(\mathcal{S},l)}_{ti}W^{(\mathcal{S},l)}_{t,\cdot}\beta^{(\mathcal{S},l)}_{\cdot,i}$ are sub-exponential random variables with mean zero and constant bounded by $C(\kappa_3)$:

$$\mathbb{E}\left[\exp\left(\nu\xi^{(\mathcal{S},l)}_{ti}W^{(\mathcal{S},l)}_{tj}\right)\right] \leq e^{C(\kappa_3)h^2}, \tag{B.104}$$

$$\mathbb{E}\left[\exp\left(\nu\xi^{(\mathcal{S},l)}_{ti}W^{(\mathcal{S},l)}_{t,\cdot}\beta^{(\mathcal{S},l)}_{\cdot,i}\right)\right] \leq e^{C(\kappa_3)h^2}, \tag{B.105}$$

for all $\nu : |\nu| < C(\kappa_3)$ due to Assumption 3, and independent across $t = 1,\ldots,n_lp$. Then, similar arguments as in Lemma B.2.1 induce the desired conclusions.

B.3.4 Proof of Lemma B.2.4

In Eq. (B.2), $\xi_{ti}^{(\mathcal{S},l)}\xi_{tj}^{(\mathcal{S},l)}$ is a sub-exponential variable with

$$\mathbb{E}\left[\exp\left(\nu(\xi_{ti}^{(\mathcal{S},l)}\xi_{tj}^{(\mathcal{S},l)} - \Sigma_{tt}^{(\mathcal{T},l)}\Phi_{ij}^{(\mathcal{S},l)})\right)\right] \le e^{C(\kappa_3)\nu^2}, \tag{B.106}$$

for all $\nu : |\nu| < C(\kappa_3)$ due to Assumption 3, and independent across $t = 1, \ldots, n_l p$. Following the same proof as in Lemma B.2.1, we prove the claim.

B.3.5 Proof of Lemma B.2.5

We first refer to the definitions of Orlicz norms and sub-Weibull random variables in Kuchibhotla and Chakrabortty (2018).

Definition B.3.1. Let g be a non-decreasing non-negative function on $[0, \infty)$ with $g(0) = 0$. For a random variable X, the g-Orlicz norm is defined as

$$\|X\|_g := \inf\{\nu > 0 : \mathbb{E}[g(|X|/\nu)] \le 1\}. \tag{B.107}$$

Definition B.3.2. A random variable X is sub-Weibull of order $\alpha > 0$, denoted by $X \sim$ sub-Weibull(α), if

$$\|X\|_{\psi_\alpha} < \infty, \ \text{ where } \ \psi_\alpha(x) := \exp(x^\alpha) - 1 \ \text{ for } \ x \ge 0. \tag{B.108}$$

Two important examples of sub-Weibull random variables are sub-Gaussian and sub-exponential random variables, which are sub-Weibull(2) and sub-Weibull(1), respectively. Based on the properties of $\|\cdot\|_{\psi_1}$ (See Section 2.7 in Vershynin, 2018), it is easy to show that $\max_{i,j}\|\widetilde{\phi}_{ij}^{(\mathcal{S},l)}\|_{\psi_1} \le C(\kappa_1, \kappa_3)\frac{1}{\sqrt{n_0 p}}$. Following Proposition D.2 in Kuchibhotla and Chakrabortty (2018),

$$\max_{i,j}\max\left\{\left\|\left|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}\right|\right\|_{\psi_{1/2}}, \left\|\left|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{ij}^{(\mathcal{S},l)}\right|\right\|_{\psi_{1/2}}\right\} \le C(\kappa_1, \kappa_3)\frac{1}{n_0 p}. \tag{B.109}$$

Hence, Theorem 3.1 in Kuchibhotla and Chakrabortty (2018) induces, for any given $i, j \in [q]$, and $\nu > 0$,

$$\mathbb{P}\left[\sum_{l=1}^{m}\left(|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}| - \mathbb{E}|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}|\right) \ge C(\kappa_1, \kappa_3)\frac{\sqrt{m\nu} + \nu^2}{n_0 p}\right] \le e^{-\nu}, \tag{B.110}$$

and by setting $\nu = C\log(qmn_0p)$

$$\begin{aligned}&\mathbb{P}\left[\sum_{l=1}^{m}|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}| \ge \sum_{l=1}^{m}\mathbb{E}|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}| + C(\kappa_1, \kappa_3)\frac{\sqrt{m\log(qmn_0p)} + \log^2(qmn_0p)}{n_0 p}\right]\\ &\le \frac{2}{q^2}(qmn_0p)^{-1/2},\end{aligned} \tag{B.111}$$

which similarly applies to $\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{ij}^{(\mathcal{S},l)}$. Thus, by the maximal inequaity,

$$\begin{aligned}&\mathbb{P}\left[\max_{i,j}\max\left\{\sum_{l=1}^{m}|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{jj}^{(\mathcal{S},l)}|,\sum_{l=1}^{m}|\widetilde{\phi}_{ii}^{(\mathcal{S},l)}\widetilde{\phi}_{ij}^{(\mathcal{S},l)}|\right\}\geq C(\kappa_1,\kappa_3)\left(\frac{m+\log^2(qmn_0p)}{n_0p}\right)\right]\\&\quad\leq(qmn_0p)^{-1/2}.\end{aligned}\tag{B.112}$$

B.3.6 Proof of Lemma B.2.6

Because $\widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)}=X_{\cdot,i}^{(\mathcal{S},l)}-X^{(\mathcal{S},l)}\widehat{\beta}_{\cdot,i}^{(\mathcal{S},l)}$ and $\epsilon_{\cdot,i}^{(\mathcal{S},l)}=X_{\cdot,i}^{(\mathcal{S},l)}-X^{(\mathcal{S},l)}\beta_{\cdot,i}^{(\mathcal{S},l)}$,

$$\begin{aligned}&\sum_{l=1}^{m}|\widehat{\phi}_{ij}^{(\mathcal{S},l)}|\\&=\sum_{l=1}^{m}\frac{1}{n_lp}\left|\widehat{\epsilon}_{\cdot,i}^{(\mathcal{S},l)\top}\widehat{\epsilon}_{\cdot,j}^{(\mathcal{S},l)}-\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}\epsilon_{\cdot,j}^{(\mathcal{S},l)}+\left(\|\epsilon_{\cdot,i}^{(\mathcal{S},l)}\|_2^2\Delta_{ij}^{(\mathcal{S},l)}+\|\epsilon_{\cdot,j}^{(\mathcal{S},l)}\|_2^2\Delta_{ji}^{(\mathcal{S},l)}\right)\mathbb{I}(i\neq j)\right| \qquad \text{(B.113)}\\&=\sum_{l=1}^{m}\frac{1}{n_lp}\left|\begin{aligned}\Delta_{\cdot,i}^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}\Delta_{\cdot,j}^{(\mathcal{S},l)}-\epsilon_{\cdot,j}^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)}-\epsilon_{\cdot,i}^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}\Delta_{\cdot,j}^{(\mathcal{S},l)}\\+\left(\|\epsilon_{\cdot,i}^{(\mathcal{S},l)}\|_2^2\Delta_{ij}^{(\mathcal{S},l)}+\|\epsilon_{\cdot,j}^{(\mathcal{S},l)}\|_2^2\Delta_{ji}^{(\mathcal{S},l)}\right)\mathbb{I}(i\neq j)\end{aligned}\right|\end{aligned}$$

For the first term, it follows Theorem 2.3.1 that

$$\sum_{l=1}^{m}\frac{1}{n_lp}\left|\Delta_{\cdot,i}^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}\Delta_{\cdot,j}^{(\mathcal{S},l)}\right|\leq C(\kappa_1,\kappa_3)\;d\;\frac{m+\log(qmn_0p)}{n_0p}\tag{B.114}$$

uniformly over i and j, with probability at least $1-2(qmn_0p)^{-1/2}$. For the rest terms,

$$\begin{aligned}&\sum_{l=1}^{m}\frac{1}{n_lp}\left|\epsilon_{\cdot,j}^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}\Delta_{\cdot,i}^{(\mathcal{S},l)}-\|\epsilon_{\cdot,j}^{(\mathcal{S},l)}\|_2^2\Delta_{ij}^{(\mathcal{S},l)}\mathbb{I}(i\neq j)\right|\\&=\sum_{l=1}^{m}\frac{1}{n_lp}\left|\epsilon_{\cdot,j}^{(\mathcal{S},l)\top}X_{\cdot,j}^{(\mathcal{S},l)}\Delta_{ji}^{(\mathcal{S},l)}\mathbb{I}(i\neq j)-\|\epsilon_{\cdot,j}^{(\mathcal{S},l)}\|_2^2\Delta_{ij}^{(\mathcal{S},l)}\mathbb{I}(i\neq j)+\sum_{k:k\neq i,j}\epsilon_{\cdot,j}^{(\mathcal{S},l)\top}X_{\cdot,k}^{(\mathcal{S},l)}\Delta_{ki}^{(\mathcal{S},l)}\right|\\&\leq\sum_{l=1}^{m}\frac{1}{n_lp}\left(\left|\epsilon_{\cdot,j}^{(\mathcal{S},l)\top}X^{(\mathcal{S},l)}\beta_{\cdot,j}^{(\mathcal{S},l)}\Delta_{ji}^{(\mathcal{S},l)}\mathbb{I}(i\neq j)\right|+\sum_{k:k\neq i,j}\left|\epsilon_{\cdot,j}^{(\mathcal{S},l)\top}X_{\cdot,k}^{(\mathcal{S},l)}\Delta_{ki}^{(\mathcal{S},l)}\right|\right).\end{aligned}\tag{B.115}$$

We notice that Theorem 2.3.1 and Lemma B.2.2 implies that

$$\begin{aligned}
&\sum_{l=1}^{m} \frac{1}{n_l p} \sum_{k:k\neq i,j} \left| \epsilon_{\cdot,j}^{(\mathcal{S},l)\top} X_{\cdot,k}^{(\mathcal{S},l)} \Delta_{ki}^{(\mathcal{S},l)} \right| \\
&\leq \sum_{k:k\neq i,j} \frac{1}{n_0 p} \left[\sum_{l=1}^{m} \left(\frac{\sqrt{n_l p}}{\|X_{\cdot,k}^{(\mathcal{S},l)}\|_2} \epsilon_{\cdot,i}^{(\mathcal{S},l)\top} X_{\cdot,k}^{(\mathcal{S},l)} \right)^2 \right]^{1/2} \|\underline{\Delta}_{ki}^{(\mathcal{S},\cdot)}\|_2 \\
&\leq \left(\max_{k:k\neq i,j} \frac{1}{n_0 p} \left[\sum_{l=1}^{m} \left(\frac{\sqrt{n_l p}}{\|X_{\cdot,k}^{(\mathcal{S},l)}\|_2} \epsilon_{\cdot,i}^{(\mathcal{S},l)\top} X_{\cdot,k}^{(\mathcal{S},l)} \right)^2 \right]^{1/2} \right) \left(\sum_{k:k\neq i,j} \|\underline{\Delta}_{ki}^{(\mathcal{S},\cdot)}\|_2 \right) \\
&\leq C(\kappa_1, \kappa_3)\, d\, \frac{m + \log(qmn_0 p)}{n_0 p}
\end{aligned} \tag{B.116}$$

and

$$\begin{aligned}
&\sum_{l=1}^{m} \frac{1}{n_l p} \epsilon_{\cdot,j}^{(\mathcal{S},l)\top} X^{(\mathcal{S},l)} \beta_{\cdot,j}^{(\mathcal{S},l)} \Delta_{ji}^{(\mathcal{S},l)} \\
&\leq \frac{1}{n_0 p} \left[\sum_{l=1}^{m} \left(\frac{\sqrt{n_l p}}{\|X_{\cdot,j}^{(\mathcal{S},l)}\|_2} \epsilon_{\cdot,i}^{(\mathcal{S},l)\top} X^{(\mathcal{S},l)} \beta_{\cdot,j}^{(\mathcal{S},l)} \right)^2 \right]^{1/2} \|\underline{\Delta}_{ji}^{(\mathcal{S},\cdot)}\|_2 \\
&\leq C(\kappa_1, \kappa_3)\, d\, \frac{m + \log(qmn_0 p)}{n_0 p}
\end{aligned} \tag{B.117}$$

satisfy simultaneously uniformly over i and j with probability at least $1 - C(qmn_0p)^{-1/2}$. We obtain the desired result by plugging the above probability bounds into Eq. (B.113).

B.3.7 Proof of Lemma B.2.7

Because

$$\frac{1}{p} \left\| \widehat{\Phi}^{(\mathcal{T},l)} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q} \Phi^{(\mathcal{T},l)} \right\|_F^2 = \frac{1}{p} \sum_{t=1}^{p} \left| \Phi_{tt}^{(\mathcal{T},l)} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{T},l)})}{q} \Phi_{tt}^{(\mathcal{T},l)} \right|^2, \tag{B.118}$$

Proposition 2.3.7 implies that

$$\max_{l} \frac{1}{p} \left\| \widehat{\Phi}^{(\mathcal{T},l)} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q} \Phi^{(\mathcal{T},l)} \right\|_F^2 \leq C(\kappa_1, \kappa_3) \frac{\log(qmn_0p)}{(n_0 q)^{1-1/(2\alpha_0+2)}} \tag{B.119}$$

with probability at least $1 - C(qmn_0p)^{-1/2}$. For $\widehat{\Phi}^{(\mathcal{T},l)-1}$,

$$\begin{aligned}
&\frac{1}{p} \left\| \widehat{\Phi}^{(\mathcal{T},l)-1} - \frac{q}{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})} \Phi^{(\mathcal{T},l)-1} \right\|_F^2 \\
&\leq \frac{1}{p} \left\| \widehat{\Phi}^{(\mathcal{T},l)} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q} \Phi^{(\mathcal{T},l)} \right\|_F^2 \left\| \frac{q}{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})} \Phi^{(\mathcal{T},l)-1} \right\|_{\mathrm{op}}^2 \left\| \widehat{\Phi}^{(\mathcal{T},l)-1} \right\|_{\mathrm{op}}^2 \\
&\leq C(\kappa_3) \frac{1}{p} \left\| \widehat{\Phi}^{(\mathcal{T},l)} - \frac{\operatorname{tr}(\Sigma^{(\mathcal{S},l)})}{q} \Phi^{(\mathcal{T},l)} \right\|_F^2
\end{aligned} \tag{B.120}$$

with probability at least $1-C(qmn_0p)^{-1/2}$ due to Assumption 3 and the bounded operator norm of $\widehat{\Phi}^{(\mathcal{T},l)-1}$ for sufficiently large n_0 (guaranteed by Proposition 2.3.7). It verifies the first conclusion.

For $\widehat{\beta}$, Proposition 2.3.7 implies that

$$\begin{aligned}\frac{1}{p}\left\|(I-\widehat{\beta}^{(\mathcal{T},l)})-(I-\beta^{(\mathcal{T},l)})\right\|_F^2 &= \frac{1}{p}\sum_{t=1}^{p}\left\|\widehat{\beta}_{\cdot,t}^{(\mathcal{T},l)}-\beta_{\cdot,t}^{(\mathcal{T},l)}\right\|_2^2 \\ &\le C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}}\end{aligned} \tag{B.121}$$

with probability at least $1-C(qmn_0p)^{-1/2}$. Because our estimator uses eigenvalue truncation, we use Lemma B.1 in Liu and Ren (2020) to guarantee the error bound for $P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})$: given that $\eta \le \lambda_1(I-\beta)$,

$$\begin{aligned}\frac{1}{p}\left\|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})-(I-\beta^{(\mathcal{T},l)})\right\|_F^2 &\le \frac{C}{p}\left\|(I-\widehat{\beta}^{(\mathcal{T},l)})-(I-\beta^{(\mathcal{T},l)})\right\|_F^2 \\ &\le C(\kappa_1,\kappa_3,\kappa_5)\frac{\log(qmn_0p)}{(n_0q)^{1-1/(2\alpha_0+2)}}\end{aligned} \tag{B.122}$$

with probability at least $1-C(qmn_0p)^{-1/2}$. Then,

$$\begin{aligned}&\frac{1}{p}\left\|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})^{-1}-(I-\beta^{(\mathcal{T},l)})^{-1}\right\|_F^2 \\ &\le \frac{1}{p}\left\|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})-(I-\beta^{(\mathcal{T},l)})\right\|_F^2\left\|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})\right\|_{\mathrm{op}}^2\left\|(I-\beta^{(\mathcal{T},l)})^{-1}\right\|_{\mathrm{op}}^2 \\ &\le C(\kappa_3)\frac{1}{p}\left\|P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})-(I-\beta^{(\mathcal{T},l)})\right\|_F^2\end{aligned} \tag{B.123}$$

due to Lemma B.2 in Liu and Ren (2020) and the bounded operator norm of $P_\eta(I-\widehat{\beta}^{(\mathcal{T},l)})$. It verifies the second conclusion.

B.4 Derivation of the asymptotic covariance between the edgewise test statistics

In this section, we derive the asymptotic covariance of $\widehat{T}_{i_1j_1}$ and $\widehat{T}_{i_2j_2}$ in Eq. (2.17):

$$\begin{aligned}&S_{(i_1,j_1),(i_2,j_2)} \\ &:= \frac{1}{m}\sum_{l=1}^{m}\frac{\|\Sigma^{(\mathcal{T},l)}\|_F^2}{p}\left[\begin{array}{l}\rho_{i_1i_2}^{(\mathcal{S},l)}\rho_{j_1j_2}^{(\mathcal{S},l)}+\rho_{i_1j_2}^{(\mathcal{S},l)}\rho_{i_2j_1}^{(\mathcal{S},l)}+\frac{1}{2}\rho_{i_1j_1}^{(\mathcal{S},l)}\rho_{i_2j_2}^{(\mathcal{S},l)}\left(\rho_{i_1i_2}^{(\mathcal{S},l)2}+\rho_{j_1j_2}^{(\mathcal{S},l)2}+\rho_{i_1j_2}^{(\mathcal{S},l)2}+\rho_{i_2j_1}^{(\mathcal{S},l)2}\right) \\ -\rho_{i_1i_2}^{(\mathcal{S},l)}\rho_{i_2j_2}^{(\mathcal{S},l)}\rho_{i_2j_1}^{(\mathcal{S},l)}-\rho_{i_1i_2}^{(\mathcal{S},l)}\rho_{i_1j_1}^{(\mathcal{S},l)}\rho_{i_1j_2}^{(\mathcal{S},l)}-\rho_{j_1j_2}^{(\mathcal{S},l)}\rho_{i_2j_2}^{(\mathcal{S},l)}\rho_{i_1j_2}^{(\mathcal{S},l)}-\rho_{j_1j_2}^{(\mathcal{S},l)}\rho_{i_2j_1}^{(\mathcal{S},l)}\rho_{i_1j_1}^{(\mathcal{S},l)}\end{array}\right].\end{aligned} \tag{B.124}$$

By Proposition 2.3.2, $\widehat{\rho}_{ij}^{(\mathcal{S},l)} - \rho_{ij}^{(\mathcal{S},l)}$ has the leading term $\Theta_{ij}^{(\mathcal{S},l)}$. Hence, $\widehat{T}_{i_1j_1}$ and $\widehat{T}_{i_2j_2}$ have asymptotic covariance,

$$\begin{aligned} S_{(i_1,j_1),(i_2,j_2)} &= \mathbb{E}\left[\left(\frac{1}{\sqrt{m}}\sum_{l=1}^{m}\sqrt{n_l p}\,\Theta_{i_1j_1}^{(\mathcal{S},l)}\right)\left(\frac{1}{\sqrt{m}}\sum_{l=1}^{m}\sqrt{n_l p}\,\Theta_{i_2j_2}^{(\mathcal{S},l)}\right)\right] \\ &= \frac{1}{m}\sum_{l=1}^{m} n_l p\,\mathbb{E}\left[\Theta_{i_1j_1}^{(\mathcal{S},l)}\Theta_{i_2j_2}^{(\mathcal{S},l)}\right]. \end{aligned} \tag{B.125}$$

Based on Eq. (2.37), for each $l = 1, \ldots, m$,

$$\begin{aligned} &\mathbb{E}\left[\Theta_{i_1j_1}^{(\mathcal{S},l)}\Theta_{i_2j_2}^{(\mathcal{S},l)}\right] \\ &= \mathbb{E}\left[\begin{aligned} &\left(\frac{\widetilde{\delta}_{i_1j_1}^{(\mathcal{S},l)}}{\sqrt{\Phi_{i_1i_1}^{(\mathcal{S},l)}\Phi_{j_1j_1}^{(\mathcal{S},l)}}} - \frac{\Phi_{i_1j_1}^{(\mathcal{S},l)}\widetilde{\delta}_{j_1j_1}^{(\mathcal{S},l)}}{2\Phi_{j_1j_1}^{(\mathcal{S},l)}\sqrt{\Phi_{i_1i_1}^{(\mathcal{S},l)}\Phi_{j_1j_1}^{(\mathcal{S},l)}}} - \frac{\Phi_{i_1j_1}^{(\mathcal{S},l)}\widetilde{\delta}_{i_1i_1}^{(\mathcal{S},l)}}{2\Phi_{i_1i_1}^{(\mathcal{S},l)}\sqrt{\Phi_{i_1i_1}^{(\mathcal{S},l)}\Phi_{j_1j_1}^{(\mathcal{S},l)}}}\right) \\ &\times\left(\frac{\widetilde{\delta}_{i_2j_2}^{(\mathcal{S},l)}}{\sqrt{\Phi_{i_2i_2}^{(\mathcal{S},l)}\Phi_{j_2j_2}^{(\mathcal{S},l)}}} - \frac{\Phi_{i_2j_2}^{(\mathcal{S},l)}\widetilde{\delta}_{j_2j_2}^{(\mathcal{S},l)}}{2\Phi_{j_2j_2}^{(\mathcal{S},l)}\sqrt{\Phi_{i_2i_2}^{(\mathcal{S},l)}\Phi_{j_2j_2}^{(\mathcal{S},l)}}} - \frac{\Phi_{i_2j_2}^{(\mathcal{S},l)}\widetilde{\delta}_{i_2i_2}^{(\mathcal{S},l)}}{2\Phi_{i_2i_2}^{(\mathcal{S},l)}\sqrt{\Phi_{i_2i_2}^{(\mathcal{S},l)}\Phi_{j_2j_2}^{(\mathcal{S},l)}}}\right) \end{aligned}\right], \end{aligned} \tag{B.126}$$

which is a summation of the second moments of $\widetilde{\delta}^{(\mathcal{S},l)}$. For each second moment,

$$\begin{aligned} &\mathbb{E}[\widetilde{\delta}_{i_1j_1}^{(\mathcal{S},l)}\widetilde{\delta}_{i_2j_2}^{(\mathcal{S},l)}] \\ &= \mathbb{E}\left[\frac{1}{n_l^2p^2}\sum_{k=1}^{n_l}\sum_{t=1}^{p}\left(\epsilon_{ti_1}^{(\mathcal{S},k,l)}\epsilon_{tj_1}^{(\mathcal{S},k,l)} - \Sigma_{tt}^{(\mathcal{T},l)}\Phi_{i_1j_1}^{(\mathcal{S},l)}\right)\cdot\sum_{k=1}^{n_l}\sum_{t=1}^{p}\left(\epsilon_{ti_2}^{(\mathcal{S},k,l)}\epsilon_{tj_2}^{(\mathcal{S},k,l)} - \Sigma_{tt}^{(\mathcal{T},l)}\Phi_{i_2j_2}^{(\mathcal{S},l)}\right)\right] \\ &= \frac{1}{n_l^2p^2}\sum_{t=1}^{p}\lambda_t^{(\mathcal{T},l)}\cdot\mathbb{E}\left[\sum_{k=1}^{n_l}\left(\xi_{ti_1}^{(k,l)}\xi_{tj_1}^{(k,l)} - \Phi_{i_1j_1}^{(\mathcal{S},l)}\right)\cdot\sum_{k=1}^{n_l}\left(\xi_{ti_2}^{(k,l)}\xi_{tj_2}^{(k,l)} - \Phi_{i_2j_2}^{(\mathcal{S},l)}\right)\right] \\ &= \frac{\|\Sigma^{(\mathcal{T},l)}\|_F^2}{n_lp^2}\cdot\mathbb{E}\left[\left(\xi_{ti_1}^{(k,l)}\xi_{tj_1}^{(k,l)} - \Phi_{i_1j_1}^{(\mathcal{S},l)}\right)\cdot\left(\xi_{ti_2}^{(k,l)}\xi_{tj_2}^{(k,l)} - \Phi_{i_2j_2}^{(\mathcal{S},l)}\right)\right], \end{aligned} \tag{B.127}$$

where

$$\mathbb{E}\left[\left(\xi_{ti_1}^{(k,l)}\xi_{tj_1}^{(k,l)} - \Phi_{i_1j_1}^{(\mathcal{S},l)}\right)\cdot\left(\xi_{ti_2}^{(k,l)}\xi_{tj_2}^{(k,l)} - \Phi_{i_2j_2}^{(\mathcal{S},l)}\right)\right] = \Phi_{i_1i_2}^{(\mathcal{S},l)}\Phi_{j_1j_2}^{(\mathcal{S},l)} + \Phi_{i_1j_1}^{(\mathcal{S},l)}\Phi_{i_2j_2}^{(\mathcal{S},l)}. \tag{B.128}$$

Plugging it into Eq. (B.126),

$$\begin{aligned}
&\mathbb{E}\left[\Theta^{(\mathcal{S},l)}_{i_1j_1}\Theta^{(\mathcal{S},l)}_{i_2j_2}\right] \\
&= \frac{\|\Sigma^{(\mathcal{T},l)}\|_F^2}{n_l p^2}\frac{1}{\sqrt{\Phi^{(\mathcal{S},l)}_{i_1i_1}\Phi^{(\mathcal{S},l)}_{j_1j_1}\Phi^{(\mathcal{S},l)}_{i_2i_2}\Phi^{(\mathcal{S},l)}_{j_2j_2}}} \\
&\quad\times\left[\begin{array}{l}
\Phi^{(\mathcal{S},l)}_{i_1i_2}\Phi^{(\mathcal{S},l)}_{j_1j_2}+\Phi^{(\mathcal{S},l)}_{i_1j_2}\Phi^{(\mathcal{S},l)}_{j_2i_1}+\dfrac{\Phi^{(\mathcal{S},l)}_{i_1j_1}\Phi^{(\mathcal{S},l)}_{i_2j_2}}{2\Phi^{(\mathcal{S},l)}_{i_1i_1}\Phi^{(\mathcal{S},l)}_{i_2i_2}}(\Phi^{(\mathcal{S},l)2}_{i_1i_2}+\Phi^{(\mathcal{S},l)2}_{i_1j_2}+\Phi^{(\mathcal{S},l)2}_{j_1i_2}+\Phi^{(\mathcal{S},l)2}_{j_1j_2}) \\
-\dfrac{\Phi^{(\mathcal{S},l)}_{i_2j_2}}{\Phi^{(\mathcal{S},l)}_{i_2i_2}}\Phi^{(\mathcal{S},l)}_{i_1i_2}\Phi^{(\mathcal{S},l)}_{j_1i_2}-\dfrac{\Phi^{(\mathcal{S},l)}_{i_2j_2}}{\Phi^{(\mathcal{S},l)}_{j_2j_2}}\Phi^{(\mathcal{S},l)}_{i_1j_2}\Phi^{(\mathcal{S},l)}_{j_1j_2}-\dfrac{\Phi^{(\mathcal{S},l)}_{i_1j_1}}{\Phi^{(\mathcal{S},l)}_{i_1i_1}}\Phi^{(\mathcal{S},l)}_{i_2i_1}\Phi^{(\mathcal{S},l)}_{j_2i_1}-\dfrac{\Phi^{(\mathcal{S},l)}_{i_1j_1}}{\Phi^{(\mathcal{S},l)}_{j_1j_1}}\Phi^{(\mathcal{S},l)}_{i_2j_1}\Phi^{(\mathcal{S},l)}_{j_2j_1}
\end{array}\right] \\
&= \frac{\|\Sigma^{(\mathcal{T},l)}\|_F^2}{n_l p^2}\left[\begin{array}{l}
\rho^{(\mathcal{S},l)}_{i_1i_2}\rho^{(\mathcal{S},l)}_{j_1j_2}+\rho^{(\mathcal{S},l)}_{i_1j_2}\rho^{(\mathcal{S},l)}_{i_2j_1}+\dfrac{1}{2}\rho^{(\mathcal{S},l)}_{i_1j_1}\rho^{(\mathcal{S},l)}_{i_2j_2}\left(\rho^{(\mathcal{S},l)2}_{i_1i_2}+\rho^{(\mathcal{S},l)2}_{j_1j_2}+\rho^{(\mathcal{S},l)2}_{i_1j_2}+\rho^{(\mathcal{S},l)2}_{i_2j_1}\right) \\
-\rho^{(\mathcal{S},l)}_{i_1i_2}\rho^{(\mathcal{S},l)}_{i_2j_2}\rho^{(\mathcal{S},l)}_{i_2j_1}-\rho^{(\mathcal{S},l)}_{i_1i_2}\rho^{(\mathcal{S},l)}_{i_1j_1}\rho^{(\mathcal{S},l)}_{i_1j_2}-\rho^{(\mathcal{S},l)}_{j_1j_2}\rho^{(\mathcal{S},l)}_{i_2j_2}\rho^{(\mathcal{S},l)}_{i_1j_2}-\rho^{(\mathcal{S},l)}_{j_1j_2}\rho^{(\mathcal{S},l)}_{i_2j_1}\rho^{(\mathcal{S},l)}_{i_1j_1}
\end{array}\right].
\end{aligned} \tag{B.129}$$

Plugging it into Eq. (B.125) retrieves the desired result.

Appendix C

Appendix for Chapter 3

C.1 Proof of Theorem 3.2.3

Let $u_k^{(t)} = S_{kk}^{(t,t)-\frac{1}{2}}\beta_k^{(t)}$ and $\Psi_k^{(t)} = S_{kk}^{(t,t)-\frac{1}{2}}\Phi_k^{(t)}S_{kk}^{(t,t)-\frac{1}{2}}$ where $S_{kl}^{(t,s)} = \beta_k^{(t)}\Sigma_{kl}^{(t,s)}\beta_l^{(s)\top} + \Phi_k^{(t)}\delta_{kl}^{(t,s)}$ is each sub-matrix of the marginal covariance matrix S of $(X_1^{(1)}; X_1^{(2)}; \dots; X_2^{(T)})$ and $\delta_{kl}^{(t,s)} = \begin{cases} 1, & k = l \text{ and } t = s, \\ 0, & \text{otherwise,} \end{cases}$ for $t, s \in [T]$ and $k, l = 1, 2$. Then the minimal conditional entropy condition can be rewritten into $u_k^{(t)\top}u_k^{(t)} = \beta_k^{(t)\top}S_{kk}^{(t,t)-1}\beta_k^{(t)} = 1$ and

$$
\begin{aligned}
u_k^{(t)\top}\Psi_{kk}^{(t,t)}u_k^{(t)} &= \beta_k^{(t)\top}S_{kk}^{(t,t)-1}\Phi_k^{(t)}S_{kk}^{(t,t)-1}\beta_k^{(t)} \\
&= \beta_k^{(t)\top}S_{kk}^{(t)-1}(S_{kk}^{(t)} - \beta_k^{(t)}\beta_k^{(t)\top})S_{kk}^{(t)-1}\beta_k^{(t)} \\
&= \beta_k^{(t)\top}S_{kk}^{(t)-1}\beta_k^{(t)} - (\beta_k^{(t)\top}S_{kk}^{(t)-1}\beta_k^{(t)})^2 \\
&= 1 - 1^2 = 0
\end{aligned}
$$

for $t \in [T]$ and $k = 1, 2$. That is, $u_k^{(t)}$ is orthogonal to $\Psi_{kk}^{(t,t)}$.

Denoting the block diagonal matrix of $\{S_{kk}^{(t,t)} : t \in [T],\ k = 1, 2\}$ by V, $R = V^{-\frac{1}{2}}SV^{-\frac{1}{2}}$ consists of sub-matrices

$$
R_{kl}^{(t,s)} = S_{kk}^{(t,t)-\frac{1}{2}}S_{kl}^{(t,s)}S_{ll}^{(s,s)-\frac{1}{2}} = u_k^{(t)}\Sigma_{kl}^{(t,s)}u_l^{(s)\top} + \Psi_k^{(t)}\delta_{kl}^{(t,s)}.
$$

Due to the orthogonality between $u_k^{(t)}$ and $\Psi_k^{(t)}$, the calculation of $\det(R)$ and R^{-1} is straightforward: $\det(R) = \det(\Omega)/\prod_{k,t}\operatorname{pdet}(\Psi_k^{(t)})$ and $Q = R^{-1}$ consists of sub-matrices

$$
Q_{kl}^{(t,s)} = u_k^{(t)}\Omega_{kl}^{(t,s)}u_l^{(s)\top} + \Psi_k^{(t)+}\delta_{kl}^{(t,s)}
$$

where $\Omega = \Sigma^{-1}$ is the precision matrix and $\operatorname{pdet}(A)$ and A^+ are the pseudo-determinant and Moore-Penrose pseudo-inverse of a positive semi-definite matrix A. Notice that $\Psi_k^{(t)} = I - u_k^{(t)}u_k^{(t)\top} = \Psi_k^{(t)+}$ and hence $\operatorname{pdet}(\Psi_k^{(t)}) = 1$. In turn, the negative log-likelihood under the model (Eqs. (3.7) and (3.8)) of a parameter set $\theta = \{\Sigma\} \cup \{\mu_k^{(t)}, \beta_k^{(t)}, \Phi_k^{(t)} : t \in [T] \text{ and } k = 1, 2\}$ wrt.

observed time-series $\{X_{1,[n]}, X_{2,[n]}\}_{n=1,\dots,N}$ is

$$
\begin{aligned}
\mathrm{nll}(\theta; \{X_{1,[n]}, X_{2,[n]}\}_{n=1,\dots,N}) &= -\log\det(\Omega) + \sum_{k,t}\log\mathrm{pdet}(\Psi_k^{(t)}) + \sum_{k,t}\log\det(S_{kk}^{(t,t)}) \\
&\quad + \mathrm{tr}(\Omega\overline{\Sigma}) + \sum_{k,t}\mathrm{tr}(\Psi_k^{(t)+} S_{kk}^{(t,t)-\frac{1}{2}}\overline{S}_{kk}^{(t,t)} S_{kk}^{(t,t)-\frac{1}{2}}) \\
&= -\log\det(\Omega) + \mathrm{tr}(\Omega\overline{\Sigma}) \\
&\quad + \sum_{k,t}\left\{\log\det(S_{kk}^{(t,t)}) + \mathrm{tr}(\Psi_k^{(t)+}S_{kk}^{(t,t)-\frac{1}{2}}\overline{S}_{kk}^{(t,t)}S_{kk}^{(t,t)-\frac{1}{2}})\right\}
\end{aligned} \tag{C.1}
$$

where

$$
\begin{aligned}
\overline{\Sigma} &= \overline{\mathrm{Var}}\left[(u_1^{(t)} S_{11}^{(t,t)-\frac{1}{2}}X_1^{(t)} : t\in[T]), (u_2^{(t)}S_{22}^{(t,t)-\frac{1}{2}}X_2^{(t)} : t \in [T])\right] \\
\overline{S}_{kl}^{(t,s)} &= \overline{\mathrm{Cov}}[X_k^{(t)}, X_l^{(s)}]
\end{aligned}
$$

for $t, s \in [T]$ and $k, l = 1, 2$. The maximum likelihood estimator satisfies the first optimality of minimizing $\mathrm{nll}(\theta; \{X_{1,[n]}, X_{2,[n]}\}_{n=1,\dots,N})$ wrt. parameters $u_k^{(t)}$ and $S_{kk}^{(t,t)-\frac{1}{2}}$

$$
\begin{aligned}
\nabla_{u_k^{(t)}}\mathrm{nll} &= \sum_{l,s} S_{kk}^{(t,t)-\frac{1}{2}}\overline{S}_{kl}^{(t,s)}S_{ll}^{(s,s)-\frac{1}{2}}u_l^{(s)}\Omega_{lk}^{(s,t)} - S_{kk}^{(t,t)-\frac{1}{2}}\overline{S}_{kk}^{(t,t)}S_{kk}^{(t,t)-\frac{1}{2}}u_k^{(t)}\Sigma_{kk}^{(t,t)} \\
&= \lambda_k^{(t)}\, u_k^{(t)},\ \exists\lambda_k^{(t)} \in \mathbb{R}, \\
\nabla_{S_{kk}^{(t,t)-\frac{1}{2}}}\mathrm{nll} &= \sum_{l,s} u_k^{(t)}\Omega_{kl}^{(t,s)}u_l^{(s)\top}S_{ll}^{(s,s)-\frac{1}{2}}\overline{S}_{lk}^{(s,t)} - u_k^{(t)}\Sigma_{kk}^{(t,t)}u_k^{(t)\top}S_{kk}^{(t,t)-\frac{1}{2}}\overline{S}_{kk}^{(t,t)} \\
&\quad -S_{kk}^{(t,t)\frac{1}{2}} + S_{kk}^{(t,t)-\frac{1}{2}}\overline{S}_{kk}^{(t,t)} \\
&= 0
\end{aligned}
$$

for all $t \in [T]$ and $k = 1, 2$. In this case,

$$
\begin{aligned}
\left(\nabla_{S_{kk}^{(t,t)-\frac{1}{2}}}\mathrm{nll}\right) S_{kk}^{(t,t)-\frac{1}{2}} &= \left(\nabla_{u_k^{(t)}}\mathrm{nll}\right) u_k^{(t)\top} - I + S_{kk}^{(t,t)-\frac{1}{2}}\overline{S}_{kk}^{(t,t)}S_{kk}^{(t,t)-\frac{1}{2}} \\
&= \lambda_k^{(t)}u_k^{(t)}u_k^{(t)\top} - I + S_{kk}^{(t,t)-\frac{1}{2}}\overline{S}_{kk}^{(t,t)}S_{kk}^{(t,t)-\frac{1}{2}} = 0,
\end{aligned}
$$

so $S_{kk}^{(t,t)} = \overline{S}_{kk}^{(t,t)} + \lambda_k^{(t)}\beta_k^{(t)}\beta_k^{(t)\top}$. Plugging it into Eq. (C.1), the maximum likelihood estimation reduces to minimizing

$$
\mathrm{nll}(\theta; \{X_{1,[n]}, X_{2,[n]}\}_{n=1,\dots,N}) = -\log\det(\Omega) + \mathrm{tr}(\Omega\overline{\Sigma}) - \sum_{k,t}\log(1 - \lambda_k^{(t)})
$$

wrt. $\overline{\Sigma}_{kk}^{(t,t)} = 1 - \lambda_k^{(t)}$. It is equivalent to

$$
\operatorname*{argmin}_{\Omega', w_k^{(t)}} -\log\det(\Omega') + \mathrm{tr}(\Omega'\overline{\Sigma}')
$$

where $\overline{\Sigma}'^{(t,s)}_{kl} = w_k^{(t)\top}\overline{S}_{kl}^{(t,s)}w_l^{(s)}$, $\mathrm{diag}(\overline{\Sigma}') = 1$, and $\Omega' = D\Omega D$ for a diagonal matrix D with $D_{kk}^{(t,t)} = \sqrt{1 - \lambda_k^{(t)}} = \Omega_{kk}^{(t,t)}$. It is the same with finding $w_k^{(t)}$ to minimize $\log\det(\overline{\Sigma}')$ under the same constraints, which is the `GENVAR` procedure of Kettenring (1971).

C.2 Fitting LaDynS

To update Ω, the P-gLASSO algorithm of Mazumder and Hastie (2012) is more efficient than the original gLASSO algorithm of Friedman et al. (2008), which is attributed to P-gLASSO's flexibility with initial values: gLASSO operates with a strict choice of initial $\widehat{\Omega}$ ($\overline{\Sigma}^{-1}$ in case of Eq. (3.14)), whereas P-gLASSO allows a warm start: in Algorithm 2, the estimate $\widehat{\Omega}$ from the past iteration serves as a warm start for the next iteration, so that we do not have to redo the entire paths from $\overline{\Sigma}^{-1}$ to $\widehat{\Omega}$. The sparse structure of Ω in Fig. 3.3 allows additional efficiency since the number of parameters reduces from $2T^2$ to $(2Td_{\text{cross}} - d_{\text{cross}}^2) + (2Td_{\text{auto}} - d_{\text{auto}}^2)$. Algorithm 3 is a modification of P-gLASSO that reduces the size of the LASSO sub-problem from $2T$ to $2d_{\text{cross}} + 2d_{\text{auto}}$ and the computational cost of one P-gLASSO iteration from $O(T^4 + T^3N)$ to $O(T((d_{\text{cross}} + d_{\text{auto}})^3 + (d_{\text{cross}} + d_{\text{auto}})^2N)$, when LASSO is solved by the LARS algorithm of Efron et al. (2004).

C.3 Inference on the change of the factor loadings

The left column of Fig. C.1 shows the estimated change (the difference) in the factor loadings for the two brain regions from the first bump (150 ms) and the second bump (400 ms) in Fig. 3.9a. To check the significance of the distinction, we obtained 95% confidence intervals based on 200 bootstrap estimates. In the bootstrap, $N = 3000$ trials was sampled from $\{X_{1,[n]}, X_{2,[n]}\}_{n=1,\ldots,N}$ with replacement. Unlike the permutation bootstrap in Section 3.2.4, the resulting dataset maintains not only the autocorrelation within each of X_1 and X_2 but also the correlated activity between them. The 97.5 and 2.5 percentiles of the bootstrap estimated differences are shown at the middle and right column of Fig. C.1, respectively. In PFC, at 400 ms compared with 150 ms, several electrodes in the upper left corner have substantially larger weights, while the electrodes in the lower middle of the array have substantially smaller weights, and the CIs are clearly separated from 0. There appear to be differences in V4, but they are smaller.

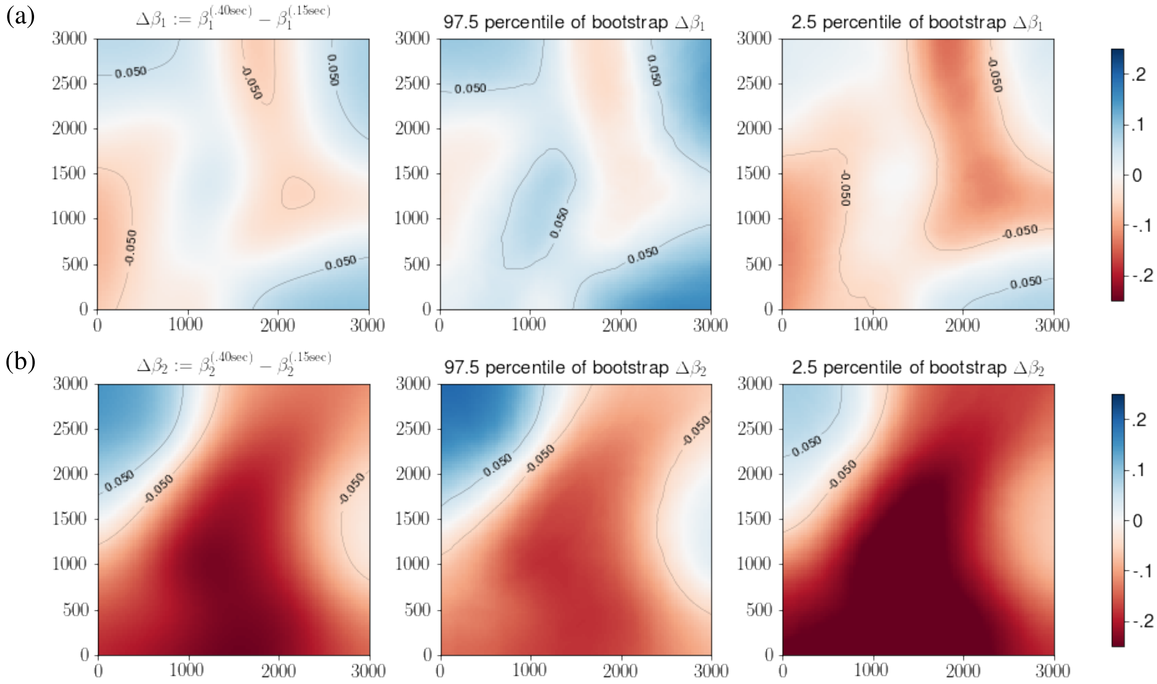


Figure C.1: **Inference on the the factor loading change from 150ms to 400ms.** *The estimated difference in the factor loadings for* **(a)** *V4 and* **(b)** *PFC from 150ms to 400ms alongside the 97.5 and 2.5 percentiles of the bootstrap distribution. The .05 and -.05 contours have been added to the plots to show regions, blue above .05, red below -.05, where the differences may be distinguishable from zero. The differences are more dramatic in PFC: there is a region in the upper left corner of the array where the 2.5 percentile is well above .05, and thus separated from zero, while the differences in the patch of* $(1000, 1500) \times (500, 1500)$ *have 97.5 percentile well below -.05.*

Algorithm 2 Coordinate descent algorithm to fit LaDynS

Input:

$\{X_k : k = 1, \dots, K\}$: input data
$\Lambda \in [0, \infty]^{KT \times KT}$: sparsity penalty matrix
$\text{iter}_{\text{max}} \in \mathbb{N}_+$: maximum iteration
$\text{ths} \in \mathbb{R}_+$: threshold for convergence

Output: Ω and $w_k^{(t)}$'s which solve (Eq. (3.13)) w.r.t. (Fig. 3.3)

Initialization:

```
Initialize $w_k^{(t)}$ so that $w_k^{(t)\top}\overline{\mathrm{Var}}[X_k^{(t)}]w_k^{(t)} = 1$ for all $t \in [T]$ and $k = 1, 2$. e.g.,
    $$w_k^{(t)} \leftarrow \mathbf{1}/\sqrt{\mathbf{1}^\top\overline{\mathrm{Var}}[X_k^{(t)}]\mathbf{1}}. \tag{C.2}$$
and let
    $$\overline{\Sigma} \leftarrow \overline{\mathrm{Var}}[w_1^{(1)\top}X_1^{(1)}, \dots, w_2^{(T)\top}X_2^{(T)}]. \tag{C.3}$$
Initialize $\Sigma$ and $\Omega$ by
    $$\Sigma \leftarrow \overline{\Sigma} + \lambda_{\text{diag}} I_{2T} \text{ and } \Omega \leftarrow \Sigma^{-1}. \tag{C.4}$$
```

Iteration:

```
for iter in 1:$\text{iter}_{\text{max}}$ do
    $\Sigma_{\text{last}} \leftarrow \Sigma$, $\Omega_{\text{last}} \leftarrow \Omega$
    $\Sigma, \Omega \leftarrow \texttt{P-gLASSO}(\Omega_{\text{init}}, \Sigma_{\text{init}}, \overline{\Sigma}, \Lambda, \text{iter}_{\text{max}}, \text{ths})$.

    for $k$ in 1:2 and $t$ in 1:$T$ do
        $A \leftarrow \overline{\mathrm{Cov}}[X_k^{(t)}, (Y_l^{(s)} : (l, s) \neq (k, t))]$.
        $b \leftarrow (\Omega_{kl}^{(t,s)} : (l, s) \neq (k, t))$
        if $Ab \neq \mathbf{0}$ then
            $w_k^{(t)} \leftarrow \overline{\mathrm{Var}}(X_k^{(t)})^{-1}Ab$
            $w_k^{(t)} \leftarrow w_k^{(t)}/\sqrt{w_k^{(t)\top}\overline{\mathrm{Var}}[X_k^{(t)}]w_k^{(t)}}$
        end if
    end for

    $\overline{\Sigma} \leftarrow \overline{\mathrm{Var}}\left[w_1^{(1)\top}X_1^{(1)}, \dots, w_2^{(T)\top}X_2^{(T)}\right]$.
    if $\max(|\Sigma - \Sigma_{\text{last}}|) <$ ths then
        break
    end if
end for
```

Algorithm 3 Modified `P-gLASSO` (Mazumder and Hastie, 2012) algorithm

```
Input:
    $\Omega_{\text{init}}, \Sigma_{\text{init}} \in \mathbb{R}^{P \times P}$ : initial values, $\Sigma_{\text{init}} = (\Omega_{\text{init}})^{-1}$
    $\overline{\Sigma} \in \mathbb{R}^{P \times P}$: sample covariance matrix of a $P$-variate random variable
    $\Lambda \in [0, \infty]^{P \times P}$: sparsity penalty matrix
    $\text{iter}_{\text{max}} \in \mathbb{N}_+$: maximum iteration
    $\text{ths} \in \mathbb{R}_+$: threshold for convergence
Output: $\Omega$ and $\Sigma = \Omega^{-1}$ which solves (Eq. (3.14))
Initialization:
    $\Sigma \leftarrow \Sigma_{\text{init}}, \Omega \leftarrow \Omega_{\text{init}}$
Iteration:
    for iter in 1:max$_{\text{iter}}$ do
        $\Sigma_{\text{last}} \leftarrow \Sigma$, $\Omega_{\text{last}} \leftarrow \Omega$

        for $p$ in 1:$P$ do
            $D_k$ : the collection of $q$'s in $[P]$ s.t. $q \neq p$ and $\Lambda_{p,q} < \infty$
            $I_p$ : the collection of $q$'s in 1:$P$ s.t. $q \neq p$ and $\Lambda_{p,q} = \infty$
            (We notate the submatrix of a matrix $A \in \mathbb{R}^{P \times P}$ of rows in $I \subset [d]$ and columns in
            $J \subset [P]$ by $A_{IJ}$. We moreover use $-p$ as a notation for $[d] \backslash \{p\}$ when it is used as a
            subscript of $A$.)

            $W = (\Omega_{-p,-p})^{-1}$ can be easily calculated by $\Sigma_{-p,-p} - \Sigma_{-p,p}\Sigma_{-p,p}/\Sigma_{p,p}$

            $\Sigma_{p,p} \leftarrow \overline{\Sigma}_{p,p} + \Lambda_{p,p}$
            $\Omega_{p,D_p}, \Omega_{D_p,p} \leftarrow \texttt{LASSO}(\Sigma_{p,p} \cdot W_{D_p,D_p}, -\overline{\Sigma}_{p,D_p}, \Lambda_{p,D_p})$ with an initial value $\Omega_{p,D_p}$.
            $\Omega_{p,I_p}, \Omega_{I_p,p} \leftarrow \mathbf{0}$
            $\Sigma_{p,-p}, \Sigma_{-p,p} \leftarrow -W_{:,D_p}\Omega_{D_p,p}\Sigma_{p,p}$
            $\Omega_{p,p} \leftarrow (1 - \Omega_{p,D_p}\Sigma_{D_p,p})/\Sigma_{p,p}$
            $\Sigma_{-p,-p} = W + \Sigma_{-p,p}\Sigma_{-p,p}/\Sigma_{p,p}$
        end for

        if $\max(|\Sigma - \Sigma_{\text{last}}|) <$ ths then
            break
        end if
    end for
```

Appendix D

Appendix for Chapter 4

D.1 EM-algorithm to fit LDFA-H (Section 2)

Initialization Let $\widehat{\theta}^{(0)} = \{\widehat{\Sigma}_1^{(0)}, \dots, \widehat{\Sigma}_q^{(0)}, \widehat{\Phi}_{\mathcal{S}}^{1,(0)}, \widehat{\Phi}_{\mathcal{S}}^{2,(0)}, \widehat{\Phi}_{\mathcal{T}}^{1,(0)}, \widehat{\Phi}_{\mathcal{T}}^{2,(0)}, \widehat{\beta}^{1,(0)}, \widehat{\beta}^{2,(0)}, \widehat{\mu}^{1,(0)}, \widehat{\mu}^{2,(0)}\}$ be the initial parameter value. Since the MPLE objective function for LDFA-H given in Eq. (4.9) is not guaranteed convex, an EM-algorithm may find a local minimum according to a choice of the initial value. Hence a good initialization is crucial to a successful estimation. Here we suggest an initialization by a canonical correlation analysis (CCA).

Let $\{X^1[n], X^2[n]\}_{n=1,\dots,N}$ be N simultaneously recorded pairs of neural time series. We can view them as NT recorded pairs of multivariate random vectors $\{X^1_{:,t}[n], X^2_{:,t}[n]\}_{(n,t)\in[N]\times[T]}$. We obtain $\widehat{\beta}_1^{1,(0)}$ and $\widehat{\beta}_1^{2,(0)}$ by CCA as follows:

$$\widehat{\beta}_1^{1,(0)}, \widehat{\beta}_1^{2,(0)} = \underset{\beta_1^1\in\mathbb{R}^{p_1}, \beta_1^2\in\mathbb{R}^{p_2}}{\operatorname{argmax}} \frac{\beta_1^{1\top} S^{12} \beta_1^2}{\sqrt{\beta_1^{1\top} S^{11}\beta_1^1}\sqrt{\beta_1^{2\top} S^{22}\beta_1^2}} \tag{D.1}$$

where

$$\begin{aligned} S^{11} &= \frac{1}{NT}\sum_{n,t}(X^1_{:,t}[n] - \frac{1}{NT}\sum_{n,t} X^1_{:,t}[n])(X^1_{:,t}[n] - \frac{1}{NT}\sum_{n,t} X^1_{:,t}[n])^\top \\ S^{22} &= \frac{1}{NT}\sum_{n,t}(X^2_{:,t}[n] - \frac{1}{NT}\sum_{n,t} X^2_{:,t}[n])(X^2_{:,t}[n] - \frac{1}{NT}\sum_{n,t} X^2_{:,t}[n])^\top \\ S^{12} &= \frac{1}{NT}\sum_{n,t}(X^1_{:,t}[n] - \frac{1}{NT}\sum_{n,t} X^1_{:,t}[n])(X^2_{:,t}[n] - \frac{1}{NT}\sum_{n,t} X^2_{:,t}[n])^\top. \end{aligned} \tag{D.2}$$

According to the equivalence between CCA and probablistic CCA shown by A. Anonymous, it gives an estimate of the first latent factors

$$\widehat{Z}_{1,:}^{k,(0)}[n] = \widehat{\beta}_1^{k,(0)} X^k[n] \tag{D.3}$$

for $n = 1, \dots, N$ and $k = 1, 2$. The initial second latent factors $\widehat{Z}_2^{k,(0)}$ and the corresponding factor loading $\widehat{\beta}_2^{k,(0)}$ is similarly set by the second pair of canonical variables, and so on. Then we assign the empirical covariance matrix of $\{\widehat{Z}_f^{1,(0)}[n], \widehat{Z}_f^{2,(0)}[n]\}_{n\in[N]}$ to the initial latent covariance matrix $\widehat{\Sigma}_f^{(0)}$ for $f = 1, \dots, q$ and the matrix-variate normal estimate (Zhou, 2014) on $\{\widehat{\epsilon}^{k,(0)}[n] := X^k[n] - \widehat{\beta}^{k,(0)}\widehat{Z}^{k,(0)}[n]\}_{n\in[N]}$ to $\widehat{\Phi}_{\mathcal{T}}^{k,(0)}$ and $\widehat{\Phi}_{\mathcal{S}}^{k,(0)}$ for $k = 1, 2$. Along $\widehat{\mu}^{k,(0)} := \frac{1}{N}\sum_{n=1}^N X^k[n]$, the above parameters comprises the initial parameter set $\widehat{\theta}^{(0)}$.

However, we cannot run an E-step on the above parameter set because $\widehat{\Phi}^{k,(0)}$ is not invertible. We instead pick one of its unidentifiable parameter sets $\widehat{\theta}^{(0),\{\alpha^1,\alpha^2\}}$, defined in Eq. (4.8), with all $\widehat{\Phi}^{k,(0)}$'s and $\widehat{\Sigma}_f^{(0)}$'s invertible. Specifically, we take

$$\alpha_f^k = \frac{1}{2}\lambda_{\min}\left(\Sigma_f^{1/2}\begin{bmatrix}\Phi_{\mathcal{T}}^1 & 0\\ 0 & \Phi_{\mathcal{T}}^2\end{bmatrix}^{-1}\Sigma_f^{1/2}\right) \tag{D.4}$$

for $f = 1, \dots, q$ and $k = 1, 2$ where $\lambda_{\min}(A)$ is the smallest eigenvalue of symmetric matrix A. Henceforth, we notate $\widehat{\theta}^{(0),\{\alpha^1,\alpha^2\}}$ by $\widehat{\theta}^{(0)}$. For $t = 1, 2, \dots$, we iterate the following E-step and M-step until convergence.

Another promising initialization is by finding time (t, s) on which the canonical correlation between $X_{:,t}^1$ and $X_{:,s}^2$ maximizes. i.e., we initialize $\widehat{\beta}_1^{1,(0)}$ and $\widehat{\beta}_1^{2,(0)}$ by

$$\widehat{\beta}_1^{1,(0)}, \widehat{\beta}_1^{2,(0)} = \underset{\beta_1^1\in\mathbb{R}^{p_1},\beta_1^2\in\mathbb{R}^{p_2}}{\operatorname{argmax}} \frac{\beta_1^{1\top}S_{(t,s)}^{12}\beta_1^2}{\sqrt{\beta_1^{1\top}S_{(t,t)}^{11}\beta_1^1}\sqrt{\beta_1^{2\top}S_{(s,s)}^{22}\beta_1^2}} \quad \text{such that} \quad |t-s| < h_{\text{cross}}. \tag{D.5}$$

where

$$\begin{aligned}
S_{(t,t)}^{11} &= \frac{1}{N}\sum_{n,t}(X_{:,t}^1[n] - \frac{1}{N}\sum_n X_{:,t}^1[n])(X_{:,t}^1[n] - \frac{1}{N}\sum_n X_{:,t}^1[n])^\top \\
S_{(s,s)}^{22} &= \frac{1}{N}\sum_{n,s}(X_{:,s}^2[n] - \frac{1}{N}\sum_n X_{:,t}^2[n])(X_{:,s}^2[n] - \frac{1}{N}\sum_n X_{:,s}^2[n])^\top \\
S_{(t,s)}^{12} &= \frac{1}{N}\sum_{n,t}(X_{:,t}^1[n] - \frac{1}{N}\sum_n X_{:,t}^1[n])(X_{:,s}^2[n] - \frac{1}{N}\sum_n X_{:,s}^2[n])^\top.
\end{aligned} \tag{D.6}$$

for $(t, s) \in [T] \times [T]$. Then the other parameters are initialized as above. We can even take an ensemble approach in which we fit LDFA-H on different initialized values and pick the estimate with the minimum cost function (Eq. (4.9)).

Now, for $r = 1, 2, \dots$, we alternate an E-step and an M-step until the target parameter Π_f convergences.

E-step Given $\widehat{\theta} := \widehat{\theta}^{(r-1)}$ from the previous iteration, the conditional distribution of latent factors $Z^1[n]$ and $Z^2[n]$ with respect to observed data $X^1[n]$ and $X^2[n]$ on trial $n = 1, \dots, N$ follows

$$\left(Z_{1,:}^1[n]; Z_{1,:}^2[n]; \dots; Z_{q,:}^2[n]\right) \mid X^1[n], X^2[n] \sim \text{MVN}\left(m_{\vec{Z}|X}^{(r)}[n], V_{\vec{Z}|X}^{(r)}\right), \tag{D.7}$$

where

$$V^{(r)}_{\vec{Z}|X} = \begin{pmatrix} V^{(r)}_{Z_1,Z_1|X} & \cdots & V^{(r)}_{Z_1,Z_q|X} \\ \vdots & \ddots & \vdots \\ V^{(r)}_{Z_q,Z_1|X} & \cdots & V^{(r)}_{Z_q,Z_q|X} \end{pmatrix} = \begin{pmatrix} W^{(r)}_{Z_1,Z_1|X} & \cdots & W^{(r)}_{Z_1,Z_q|X} \\ \vdots & \ddots & \vdots \\ W^{(r)}_{Z_q,Z_1|X} & \cdots & W^{(r)}_{Z_q,Z_q|X} \end{pmatrix}^{-1} \tag{D.8}$$

and

$$\begin{aligned} m^{(r)}_{\vec{Z}|X}[n] &= \left(m^{(r)}_{Z^1_1|X}; m^{(r)}_{Z^1_2|X}; \ldots; m^{(r)}_{Z^2_q|X}\right) \\ &= V^{(r)}_{\vec{Z}|X}\ \left(\widehat{\beta}^{1\top}_1 \widehat{\Gamma}^1_{\mathcal{S}} X^1[n] \widehat{\Gamma}^1_{\mathcal{T}}\ ;\ \widehat{\beta}^{2\top}_1 \widehat{\Gamma}^2_{\mathcal{S}} X^2[n] \widehat{\Gamma}^2_{\mathcal{T}}\ ;\ \ldots\ ;\ \widehat{\beta}^{2\top}_q \widehat{\Gamma}^2_{\mathcal{S}} X^2[n] \widehat{\Gamma}^2_{\mathcal{T}}\right) \end{aligned} \tag{D.9}$$

given

$$W^{(r)}_{Z_f,Z_g|X} = \begin{pmatrix} (\widehat{\beta}^{1\top}_f \widehat{\Gamma}^1_{\mathcal{S}} \widehat{\beta}^1_g)\ \widehat{\Gamma}^1_{\mathcal{T}} & 0 \\ 0 & (\widehat{\beta}^{2\top}_f \widehat{\Gamma}^2_{\mathcal{S}} \widehat{\beta}^2_g)\ \widehat{\Gamma}^2_{\mathcal{T}} \end{pmatrix} + \mathbb{I}_{\{f=g\}}\ \widehat{\Omega}_f,\ \ \mathbb{I}_{\{f=g\}} = \begin{cases} 1, & f=g \\ 0, & \text{o.w.} \end{cases} \tag{D.10}$$

for $f, g = 1, \ldots, q$.

M-step We find $\widehat{\theta}^{(r)}$ which maximize the conditional expectation of the penalized likelihood under the same constraints in Eq. (4.9), i.e.

$$\begin{aligned} \widehat{\theta}^{(r)} = \operatorname{argmin} &\frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_{Z[n]|X[n],\widehat{\theta}^{(r-1)}} \left[\log p(X^1[n], X^2[n], Z^1[n], Z^2[n]; \widehat{\theta}^{(r-1)})\right] \\ &+ \sum_{f=1}^{q} \sum_{k,l=1}^{2} \left\|\Lambda^{kl}_f \odot \Pi^{kl}_f\right\|_1 \ \text{ s.t. } \widehat{\Gamma}^k_{\mathcal{T}} \text{ is } (2h^k_\epsilon + 1)\text{-diagonal} \end{aligned} \tag{D.11}$$

where p is the probability density function of our model in Eqs. (4.1), (4.4) and (4.5) and the expectation $\mathbb{E}_{Z[n]|X[n],\widehat{\theta}^{(r-1)}}$ follows the conditional distribution in Eq. (D.7). Taking a block coordinate descent approach, we solve the optimization problem by alternating M1 - M4.

M1: With respect to latent precision matrices Ω_f, Eq. (D.11) reduces to a graphical Lasso problem,

$$\widehat{\Omega}^{(r)}_f = \underset{\Omega_f}{\operatorname{argmin}} \left\{ -\log\det(\Omega_f) + \mathbf{tr}\left(\Omega_f \left(V^{(r)}_{Z_f|X} + \widehat{\mathbb{E}}[m^{(r)}_{Z_f|X} m^{(r)\top}_{Z_f|X}]\right)\right) + \sum_{k,l=1}^{2} \left\|\Lambda^{kl}_f \odot \Pi^{kl}_f\right\|_1 \right\} \tag{D.12}$$

for each $f = 1, \ldots, q$ where $\widehat{\mathbb{E}}[m^{(r)}_{Z_f|X} m^{(r)\top}_{Z_f|X}] = \frac{1}{N}\sum_{n=1}^{N} m^{(r)}_{Z_f|X}[n]\ m^{(r)\top}_{Z_f|X}[n]$. The graphical Lasso problem is solved by the P-GLASSO algorithm by Mazumder et al. (2010).

M2: With respect to Γ^k, Eq. (D.11) reduces to an estimation of matrix-variate normal model (Zhou, 2014). The estimation problem can be formulated as

$$\widehat{\Gamma}^{k(r)}_{\mathcal{S}} = \frac{1}{T}\left(\widehat{\mathbb{E}}\left[m^{(r)}_{\epsilon^k|X} m^{(r)\top}_{\epsilon^k|X}\right] + \sum_{f,g=1}^{q} \operatorname{tr}(V^{(r)}_{Z^k_f, Z^k_g|X}) \beta^k_f \beta^{k\top}_g\right) \tag{D.13}$$

and

$$\widehat{\Gamma}_{\mathcal{T}}^{k(r)} = \underset{\Gamma_{\mathcal{T}}^k}{\operatorname{argmin}} \left\{ \begin{aligned} & -\log\det(\Gamma_{\mathcal{T}}^k) \\ & + \frac{1}{p_k} \operatorname{tr}\Big(\Gamma_{\mathcal{T}}^k \Big(\sum_{f,g=1}^{q} (\beta_f^{k\top} \Gamma_{\mathcal{S}}^k \beta_g^k)\, V_{Z_f^k, Z_g^k|X}^{(r)} + \widehat{\mathbb{E}}\Big[m_{\epsilon^k|X}^{(r)\top} \Gamma_{\mathcal{S}}^k m_{\epsilon^k|X}^{(r)} \Big] \Big) \Big) \end{aligned} \right\} \tag{D.14}$$

$$\text{s.t. } \widehat{\Gamma}_{\mathcal{T}}^k \text{ is } (2h_\epsilon^k + 1)\text{-diagonal}$$

for each $k = 1, 2$ where $m_{\epsilon^k|X}^{(r)} = X^k - \beta^k m_{Z^k|X}^{(r)} - \mu^k$ and $\widehat{\mathbb{E}}[A]$ is the empirical mean of a random matrix A. The estimation of $\Gamma_{\mathcal{T}}^k$ under the bandedness constraint is tractable with modified Cholesky factor decomposition approach with bandwidth h_ϵ^k using the procedure by Bickel and Levina (2008).

M3: With respect to β^k, Eq. (D.11) reduces to a quadratic program

$$\widehat{\beta}^{k(r)} = \arg\max_{\beta^k} \left\{ \begin{aligned} & \sum_{t,s} \Gamma_{\mathcal{T},(t,s)}^k \operatorname{tr}\left(\beta^{k\top} \Gamma_{\mathcal{S}}^k \beta_k \, (V_{Z_{:,t}^k, Z_{:,s}^k|X}^{(r)} + \widehat{\operatorname{Cov}}[m_{Z_{:,t}^k|X}^{(r)}, m_{Z_{:,s}^k|X}^{(r)}]) \right) \\ & - 2 \sum_{t,s} \Gamma_{\mathcal{T},(t,s)}^k \operatorname{tr}\left(\Gamma_{\mathcal{S}}^k \beta^k \widehat{\operatorname{Cov}}[X_{:,t}^k, m_{Z_{:,s}^k|X}^{(r)}] \right) \end{aligned} \right\} \tag{D.15}$$

where $\Gamma_{T,(t,s)}^k$ is the (t, s) entry in $\Gamma_{\mathcal{T}}^k$ and $\widehat{\operatorname{Cov}}(A, B)$ is the empirical covariance matrix between random vectors A and B. The analytic form of the solution is given by

$$\beta^k = \left(\sum_{t,s} \Gamma_{\mathcal{T},(t,s)}^k (V_{Z_{:,t}^k, Z_{:,s}^k|X}^{(r)} + \widehat{\operatorname{Cov}}[m_{Z_{:,t}^k|X}^{(r)}, m_{Z_{:,s}^k|X}^{(r)}]) \right)^{-1} \left(\sum_{t,s} \Gamma_{\mathcal{T},(t,s)}^k \widehat{\operatorname{Cov}}[m_{Z_{:,s}^k|X}^{(r)}, X_{:,t}^k] \right) \tag{D.16}$$

M4: With resepct to μ^k, it is straight-forward that Eq. (D.11) yields

$$\widehat{\mu}^{k(r)} = \widehat{\mathbb{E}}\left[X^k - \sum_{f=1}^{q} \beta_f^k m_{Z_f^k|X}^{(r)\top} \right].$$

D.2 Simulation details (Section 3)

We simulated realistic data with known cross-region connectivity as follows. Simulating $q = 1$ pair of latent time-series Z^k from Equation (3.8), we introduced an exact ground-truth for the inverse cross-correlation matrix Π_1^{12} by setting:

$$\Pi_1 = \begin{bmatrix} (\mathrm{P}_{1,0}^{11})^{-1} & 0 \\ 0 & (\mathrm{P}_{1,0}^{22})^{-1} \end{bmatrix} + \begin{bmatrix} D^1 & \Pi_1^{12} \\ \Pi_1^{12\top} & D^2 \end{bmatrix} \tag{D.17}$$

where D^1 and D^2 are diagonal matrices with elements $D_{(t,t)}^1 = \sum_s \Pi_{1,(t,s)}^{12}$ and $D_{(s,s)}^2 = \sum_t \Pi_{1,(t,s)}^{12}$, which ensures that the matrix on the right hand side is positive definite. The matrix on the left hand side contains the auto-precision matrices of the two latent time series, with elements simulated from the squared exponential function:

$$\mathrm{P}_{1,0}^{kk} = \left[\exp\left(-c^k (t-s)^2 \right) \right]_{t,s} + \lambda I_T, \tag{D.18}$$

with $c^1 = 0.105$ and $c^2 = 0.142$, chosen to match the observed LFPs auto-correlations in the experimental dataset (Section 4.3.2). We added the regularizer λI_T, $\lambda = 1$, to render P^{kk} invertible. We designed the true inverse cross-correlation matrix Π^{12} to induce lead-lag relationship between Z^1 and Z^2 in two epochs as depicted in the right-most panel of Fig. 4.2a. Specifically, the elements of Π^{12} were set:

$$\Pi^{12}_{(t,s)} = \begin{cases} -r, & \text{where } Z^1_{1,t} \text{ and } Z^2_{1,s} \text{ partially correlate}, \\ 0, & \text{elsewhere}, \end{cases} \tag{D.19}$$

where the association intensity $r = 0.6$ was chosen to match our cross-correlation estimate in the experimental data (Section 4.3.2). Finally, we rescaled $\mathrm{P}_1 = \Pi_1^{-1}$ to have diagonal elements equal to one. The corresponding factor loading vector β_1^k was randomly generated from standard multivariate normal distribution and then scaled to have $\|\beta_1^k\|_2 = 1$.

We generated the noise ϵ^k from the $N = 1000$ trials of the experimental data analyzed in Section 4.3.2. First, we permuted the trials in one region to remove cross-region correlations. Let $\{Y^1[n], Y^2[n]\}_{n=1,\ldots,N}$ be the permuted dataset. Then we contaminated the dataset with white noise to modulate the strength of noise correlation relative to cross-region correlations. i.e.

$$\epsilon^k_{:,t} = Y^k_{:,t} - \mu^k_{:,t} + \eta^k_{:,t}, \;\; \eta^k_{:,t} \overset{\text{indep}}{\sim} \mathrm{MVN}\left(0, \lambda_\epsilon \widehat{\mathrm{Cov}}[Y^k_{:,t}]\right), \;\text{ and }\; \mu^k_{:,t} = \widehat{\mathbb{E}}[Y^k_{:,t}] \tag{D.20}$$

where $\widehat{\mathbb{E}}[Y^k_{:,t}]$ and $\widehat{\mathrm{Cov}}[Y^k_{:,t}]$ wer the empirical mean and covariance matrix of $Y^k_{:,t}$, respectively, for $k = 1, 2, t = 1, \ldots, T$. The noise auto-correlation level was modulated by $\lambda_\epsilon \in \{2.78, 1.78, 0.44, 0.11\}$. We also obtained Σ_1 by scaling P_1 so that $\Sigma^{kk}_{1,(t,s)} = \beta_1^{k\top} S^k_t \beta^k_1$. Putting all the pieces together, we generated observed time series by Eq. (4.1).

D.3 Experimental data analysis details (Section 3.2)

The strength of each factor, which is characterized by Σ_f, is shown in Fig. D.1.

We also examined an alternative definition of information flow, using non-stationary regresssion in the spirit of Granger causality. For the latent factor f in V4 at time t, we use partial R^2, effectively comparing the full regression model using the full history of latent variables in both area,

$$Z^1_{f,t} \sim Z^1_{f,1:t-1} + Z^2_{f,1:t-1}$$

with the reduced model using history of latent variables in V4 only,

$$Z^1_{f,t} \sim Z^1_{f,1:t-1}.$$

The partial R^2 for $Z^1_{f,t}$ on $Z^2_{f,1:t-1}$ given $Z^1_{f,1:t-1}$ summarizes the contribution of PFC history to V4, after taking account of the autocorrelation in V4, and thus can be viewed as information flow from V4 to PFC at time t. Dynamic information flow from V4 to PFC is defined similarly. The results shown in Fig. D.2 are consistent with those in Fig. 4.5d.

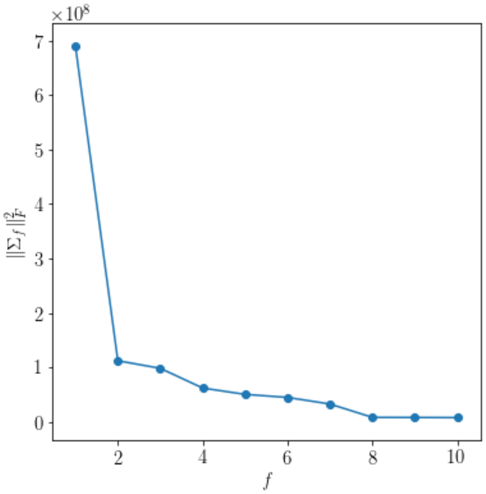


Figure D.1: *Squared Frobenius norms of covariance matrix estimates,* $\widehat{\Sigma}_f$*, for all factors* $f = 1, \ldots, 10$*. Notice that the amplitudes of the top four factors dominate the others.*

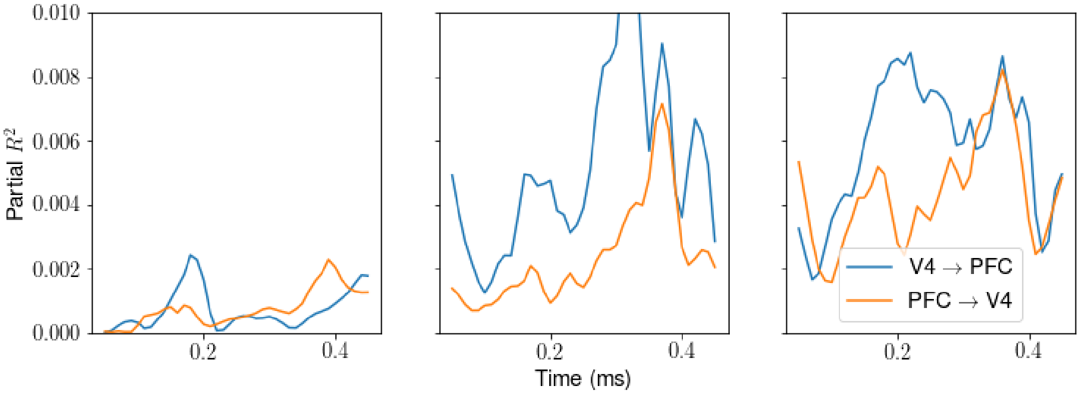


Figure D.2: **Information flow by partial** R^2 **for the top three factors.** *In this figure, we characterize dynamic information flow in terms of partial* R^2*. We show dynamic information flow from* $V4 \rightarrow PFC$ *(blue) and* $PFC \rightarrow V4$ *(orange). The results in the first panel are consistent with those in the first panel of Fig. 4.5d.*

Appendix E

Appendix for Chapter 5

E.1 A Note on the Hilbert Transform

Let $Z_t = X_t + iY_t$ be the output of a complex signal that has been filtered in a band $(\omega_0 - \delta, \omega_0 + \delta)$. Here X_t and Y_t are real. Suppose we observe X_t. The problem is to recover Z_t, which is possible when δ is sufficiently small. The Hilbert transform operates on X_t to produce Y_t according to the following:

$$Y_t = \int_0^1 \frac{X_u}{\pi(t-u)} du.$$

This formula is given in numerous available sources. Its derivation can be made concise, as follows.

Note first that the Fourier transform of Z_t, which we write as $\mathcal{F}(Z_t)$, is concentrated around ω_0 and the transform of its complex conjugate $\mathcal{F}(\overline{Z}_t)$ is concentrated around $-\omega_0$. From the general relations

$$\begin{aligned} X_t &= \frac{1}{2}(Z_t + \overline{Z}_t) \\ Y_t &= \frac{-i}{2}(Z_t - \overline{Z}_t) \end{aligned}$$

we have the Fourier transforms

$$\begin{aligned} \mathcal{F}(X_t) &= \frac{1}{2}(\mathcal{F}(Z) + \mathcal{F}(\overline{Z})) \\ \mathcal{F}(Y_t) &= \frac{-i}{2}(\mathcal{F}(Z) - \mathcal{F}(\overline{Z})). \end{aligned}$$

We write $\mathcal{F}(X_t)$ and $\mathcal{F}(Y_t)$ as functions of a real frequency ω in the form $\mathcal{F}(X_t)[\omega]$ and $\mathcal{F}(Y_t)[\omega]$. Because $\mathcal{F}(X_t)$ and $\mathcal{F}(Y_t)$ are concentrated around ω_0 and $-\omega_0$, when we multiply $\mathcal{F}(X_t)[\omega]$ by $(-i \text{ sgn}(\omega))$ we get

$$(-i \text{ sgn}(\omega))\mathcal{F}(X_t)[\omega] = \mathcal{F}(Y_t)[\omega]$$

where $\operatorname{sgn}(\omega)$ is the sign of ω. As a result, the convolution $X_t * \mathcal{F}^{-1}(-i\text{sgn}(\omega_0))$ satisfies

$$X_t * \mathcal{F}^{-1}(-i \text{ sgn}(\omega)) = Y_t.$$

Plugging the formula for $\mathcal{F}^{-1}(-i\ \mathrm{sgn}(\omega_0))$ (a standard example in texts on Fourier transforms) into the definition of convolution produces the integral appearing above, giving the Hilbert transform of X_t. The resulting Z_t is called the analytic signal.

E.2 Theorems and Proofs

E.2.1 Proof of Theorem 5.4.1

We first provide a fuller statement of the theorem we are trying to prove. To start, we recall some preliminaries - see (Barndorff-Nielsen, 2014) for additional details. An exponential family is a family of distributions $\mathcal{D}$ such that, for every distribution $P \in \mathcal{D}$, we can represent the density of P in the following form:

$$p^\theta(x) = a(\theta)b(x)\exp\left(\sum_{i=1}^{k}\theta_i t_i(x)\right)$$

For some $\theta \in \mathbb{R}^k$ and sufficient statistics $\{t_i(x)\}$. For any exponential family $\mathcal{D}$ and choice of sufficient statistics $\{t_i(x)\}$ (collectively $t(x)$) let $\Theta_{\mathcal{D},t} = \{\theta \in \mathbb{R}^d | \int_x p^\theta(x)dx < \infty\}$. Let P^θ denote the distribution corresponding to the density p^θ. If $\mathcal{D}$ is an exponential family and there exists some choice of statistics $\{t_i(x)\}$ inducing densities $p^\theta(x)$ for which $\mathcal{D} = \{P^\theta | \theta \in \Theta_{\mathcal{D},t}\}$ then we say that $\mathcal{D}$ is full. $\Theta_{\mathcal{D},t}$ is called a parameter space of $\mathcal{D}$. We say that $\Theta_{\mathcal{D},t}$ is linearly dependent if we can express some entry θ_i as a linear combination of other entries in θ for all $\theta \in \Theta_{\mathcal{D},t}$. If $\Theta_{\mathcal{D},t}$ is linearly independent, then the statistics $t_i(x)$ are minimal sufficient statistics (see Barndorff-Nielsen (2014)). Finally, if there exist statistics $\{t_i(x)\}$ for which $\Theta_{\mathcal{D},t}$ is an open set of $\mathbb{R}^d$, then we say that the family $\mathcal{D}$ is regular.

Now, we define the following families. Denote by $\mathcal{D}_1$ the full family of distributions given by densities $\{p^{\mu,\Sigma} | (\mu, \Sigma) \in N_1\}$ where $p^{\mu,\Sigma}$ is the $2d$-dimensional normal density and $N_1 = \{(\mu, \Sigma) | \int_x p^{\mu,\Sigma}(x)dx < \infty, \Sigma \text{ symmetric}\}$ and by $(\mu, \Sigma) \in \mathbb{R}^{4d^2+2d}$ we mean a vector containing all of the entries in μ, Σ. $\mathcal{D}_1$ consists of all distributions which can be written as $CN(m, \Gamma, C)$ by the definition of the complex normal distribution.

Further, by $\mathcal{D}_2$ denote the full family of distributions given by the densities $\{p^{\mu,\Sigma} | (\mu, \Sigma) \in N_2\}$ where $p^{\mu,\Sigma}$ is once again the $2d$-dimensional normal distribution but this time:

$$N_2 = \{(\mu, \Sigma) | \int_x p^{\mu,\Sigma}(x)dx < \infty, \Sigma \in \mathcal{S}\}$$

Where $\mathcal{S}$ is the set of $2d \times 2d$ matrices $\Sigma \in \mathcal{S}$ which can be written as:

$$\Sigma = \begin{bmatrix} A & -B \\ B & A \end{bmatrix}$$

Where $A \in \mathbb{R}^{d\times d}$ is a symmetric matrix and $B \in \mathbb{R}^{d\times d}$ is an anti-symmetric matrix (i.e. $B = -B^T$). From Picinbono (1996), we have that for a random variable $X \sim CN(m, \Gamma, 0)$, we can say that $\Sigma_{11} = \frac{1}{2}\mathfrak{Re}\Gamma, \Sigma_{12} = \frac{-1}{2}\mathfrak{Im}\Gamma, \Sigma_{21} = \frac{1}{2}\mathfrak{Im}\Gamma, \Sigma_{22} = \frac{1}{2}\mathfrak{Re}\Gamma$. Therefore, for $X \sim CN(m, \Gamma, 0)$ it follows that Σ must have the form of the matrices in $\mathcal{S}$. Hence, $\mathcal{D}_2$ is equivalent to the set of distributions expressible as $CN(m, \Gamma, 0)$ - the proper case.

Finally, denote by $\mathcal{D}_3$ the full family of distributions $\mathcal{D}_3 = \{p^{0,\Sigma} | (0, \Sigma) \in N_2\}$. This is the circularly symmetric case.

Before stating the theorem, we first define some statistics which will be necessary in the statement of the theorem. Let $S_i(x) = (\mathfrak{Re}x, \mathfrak{Im}x)_i$ for $1 \leq i \leq 2d$. Let let $T_{ij}(x) = (\mathfrak{Re}x, \mathfrak{Im}x)_i(\mathfrak{Re}x, \mathfrak{Im}x)_j$. Let $S(x) = [S_i(x) | 1 \leq i \leq 2d]$ denote a vector concatenating all of the statistics $S_i(x)$, and likewise define $T(x) = [T_{ij}(x) | 1 \leq i \leq j \leq 2d]$. Let $\phi(x) = \text{Concat}[T(x), S(x)]$ denote the statistic created by concatenating T and S. Further, let $T^1(x)_{ij} = \mathfrak{Re}x_i\mathfrak{Re}x_j + \mathfrak{Im}x_i\mathfrak{Im}x_j$ ($1 \leq i \leq j \leq d$) and $T^2(x)_{ij} = \mathfrak{Re}x_i\mathfrak{Im}x_j - \mathfrak{Im}x_i\mathfrak{Re}x_j$ ($1 \leq i \leq j \leq d$). Define $T^1(x)$ and $T^2(x)$ analogously to above. Let $\phi'(x) = \text{Concat}[T_1, T_2, S]$ and $\phi''(x) = \text{Concat}[T_1, T_2]$ We can now state the theorem:

Theorem E.2.1. *$\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$ are all full, regular exponential exponential families. Then, $\phi(x)$ is a minimal sufficient statistic for $\mathcal{D}_1$, $\phi'(x)$ is a minimal sufficient statistic for $\mathcal{D}_2$ and $\phi''(x)$ is a minimal sufficient statistic for $\mathcal{D}_3$.*

Proof. $\mathcal{D}_1$ is known to be full and regular (Barndorff-Nielsen, 2014). All that remains for $\mathcal{D}_1$ is to show that $\phi(x)$ is a minimal sufficient statistic.

Letting $z = (\mathfrak{Re}x, \mathfrak{Im}x)$ for $X = x$ can write the normal PDF:

$$\begin{aligned} p^{\mu,\Sigma}(x) &= \frac{1}{\sqrt{(2\pi)^{2d}|\Sigma|}} \exp\left(\frac{-1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu)\right) \\ &\propto \exp\left(\frac{-1}{2}z^T\Sigma^{-1}z + \mu^T\Sigma^{-1}z - \frac{1}{2}\mu^T\mu\right) \\ &\propto \exp\left(\frac{-1}{2}\operatorname{Tr}\left(\Sigma^{-1}zz^T\right) + \mu^T\Sigma^{-1}z - \frac{1}{2}\mu^T\mu\right) \\ &\propto \exp\left(\frac{-1}{2}\sum_{i<j} 2\left(\Sigma^{-1}\odot zz^T\right)_{ij} - \frac{1}{2}\sum_i \left(\Sigma^{-1}\odot zz^T\right)_i + \mu^T\Sigma^{-1}z - \frac{1}{2}\mu^T\mu\right) \end{aligned}$$

It is therefore clear that without any constraints on μ, Σ, the statistics $S(X)_i = z_i$ (for $1 \leq i \leq 2d$) and $T(X)_{ij} = (zz^T)_{ij}$ (for $1 \leq i \leq j \leq 2d$) which together form ϕ are minimal sufficient.

Now, consider the case that $C = 0$, corresponding to the family $\mathcal{D}_2$. From Picinbono (1996), we have that $\Sigma_{11} = \frac{1}{2}\mathfrak{Re}\Gamma, \Sigma_{12} = \frac{-1}{2}\mathfrak{Im}\Gamma, \Sigma_{21} = \frac{1}{2}\mathfrak{Im}\Gamma, \Sigma_{22} = \frac{1}{2}\mathfrak{Re}\Gamma$. Therefore, Σ is of the the form specified in lemma E.2.5, and so Σ^{-1} is as well. In particular, we can write Σ^{-1} as:

$$\Sigma^{-1} = \begin{bmatrix} A & B \\ -B & A \end{bmatrix}$$

Let $z_1 = \mathfrak{Re}X$ and $z_2 = \mathfrak{Im}X$. We can then rewrite the density above as follows:

$$\begin{aligned}
p^{\mu,\Sigma}(x) &\propto \exp\left(\frac{-1}{2}\sum_{i<j} 2\left(\begin{bmatrix} A & B \\ -B & A\end{bmatrix} \odot zz^T\right)_{ij} - \frac{1}{2}\sum_{i}\left(\begin{bmatrix} A & B \\ -B & A\end{bmatrix} \odot zz^T\right)_{i} + \mu^T\Sigma^{-1}z\right) \\
&= \exp\left(\frac{-1}{2}\left(\sum_{i<j} 2(A \odot z_1 z_1^T)_{ij} + \sum_{i<j} 2(A \odot z_2 z_2^T)_{ij} + 2\sum_{ij}(B \odot z_1 z_2^T)_{ij}\right)\right. \\
&\qquad\qquad \left. -\frac{1}{2}\sum_{i} A_{ii}(z_i^2 + z_{i+d}^2) + \mu^T\Sigma^{-1}z\right) \\
&= \exp\left(\frac{-1}{2}\left(\sum_{i<j} 2(A \odot z_1 z_1^T)_{ij} + \sum_{i<j} 2(A \odot z_2 z_2^T)_{ij} + \sum_{i} A_{ii}(z_i^2 + z_{i+d}^2)\right)\right. \\
&\qquad\qquad \left. -\sum_{ij}(B \odot z_1 z_2^T)_{ij} + \mu^T\Sigma^{-1}z\right) \\
&= \exp\left(\frac{-1}{2}\left(\sum_{i<j} 2(A \odot (z_1 z_1^T + z_2 z_2^T))_{ij} + \sum_{i} A_{ii}(z_i^2 + z_{i+d}^2)\right)\right. \\
&\qquad\qquad \left. -\sum_{i<j}(B \odot (z_1 z_2^T - z_2 z_1^T))_{ij} + \mu^T\Sigma^{-1}z\right)
\end{aligned}$$

The fourth line follows from observing that $B = -B^T$ (one consequence of which is that $B_{ii} = 0$). We can now see that the statistics $S(X)$ as defined above along with $T^1_{ij}(X) = \mathfrak{Re}X_i\mathfrak{Re}X_j + \mathfrak{Im}X_i\mathfrak{Im}X_j$ (for $i \leq j$) and $T^2_{ij}(X) = \mathfrak{Re}X_i\mathfrak{Im}X_j - \mathfrak{Im}X_i\mathfrak{Re}X_j$ (for $i < j$) which together form ϕ' are minimal sufficient statistics.

Now, we must show that $\mathcal{D}_2$ is full and regular. Let $p^{\theta'}(x) \propto \exp(\phi'(x)^T\theta')$ where $\theta' \in \mathbb{R}^{d^2+2d}$. To show that $\mathcal{D}_2$ is full, we need to show that there doesn't exist any θ' such that $\int_x \exp(\phi'(x)^T\theta')dx < \infty$ but $p^{\theta'}(x)$ does not correspond to any distribution in $\mathcal{D}_2$. Assume towards a contradiction that such a θ' existed. Then, using the derivation above, it is easy to see that we could construct a symmetric A, an antisymmetric B, and vector μ such that $p^{\mu,\Sigma}(x) = p^{\theta'}(x)$. The fact that $\int p^{\theta'}(x)dx < \infty$ implies that $\int p^{\mu,\Sigma}(x)dx < \infty$, so by construction of $\mathcal{D}_2$, we know that $p^{\mu,\Sigma}(x)$ corresponds to a distribution in $\mathcal{D}_2$ and so therefore $p^{\theta'}(x)$ corresponds to a distribution in $\mathcal{D}_1$ which is a contradiction. Hence, all integrable $p^{\theta'}(x)$ correspond to a distribution in $\mathcal{D}_2$ and so $\mathcal{D}_2$ is full.

Now we want to show that $\mathcal{D}_2$ is regular. In particular, we need to show that the set $\Theta' = \{\theta' | \int p^{\theta'}(x)dx < \infty\}$ is open in $\mathbb{R}^{d^2+2d}$. Recall that $\mathcal{D}_1$ is regular. We can see that $\mathcal{D}_1$ is the set of distributions with integrable densities $p^{\theta}(x) \propto \exp(\phi(x)^T\theta)$ where $\theta \in \mathbb{R}^{2d(2d+1)/2+2d}$. Let $\Theta = \{\theta | \int p^{\theta}(x)dx < \infty\}$. Because $\mathcal{D}_1$ is regular, Θ is an open subset of $\mathcal{R}^{2d(2d+1)/2+2d}$.

Since $\mathcal{D}_2 \subset \mathcal{D}_1$, for every $\theta' \in \Theta'$, there exists a $\theta \in \Theta$ such that $p^{\theta'}(x) = p^{\theta}(x)$. In particular, there exists a map from θ' to θ for which $p^{\theta'}(x) = p^{\theta}(x)$. I claim that this map is a linear map; in other words, there exists some matrix $M \in \mathbb{R}^{(d^2+2d)\times(2d(2d+1)/2+2d)}$ such that if $\theta = M^T\theta'$ then $p^{\theta}(x) = p^{\theta'}(x)$. To see that this claim holds true, just note that the sufficient statistics for $\mathcal{D}_2$ are linear combinations of the sufficient statistics in $\mathcal{D}^1$. For instance,

$T^1_{ij}(x) = \mathfrak{Re}x_i\mathfrak{Re}x_j + \mathfrak{Im}x_i\mathfrak{Im}x_j = T_{ij}(x) + T_{i+d,j+d}(x)$; other equalities can be shown for the remaining statistics. In particular, this means that there exists some matrix M such that $\phi'(x) = M\phi(x)$. Then, simply note that $p^{\theta'}(x) \propto \exp(\phi'(x)^T\theta') = \exp((M\phi(x))^T\theta') = \exp((\phi(x))^T M\theta') \propto p^{M\theta'}(x)$. Hence M is a linear map as described.

Assume towards a contradiction that Θ' is not an open set in $\mathbb{R}^{d^2+2d}$. Then, there exists some point $\theta' \in \Theta'$ such that for every ϵ there exists another point $w \in \mathbb{R}^{d^2+2d}$ such that $w \notin \Theta'$ but $||w - \theta'||_2 < \epsilon$. Because $w \notin \Theta'$, this implies that $Mw \notin \Theta$. However, note that $||M^T w - M^T\theta'||_2 = ||M^T(w - \theta')||_2 \leq ||M||_2||w - \theta'||_2 \leq ||M||_2\epsilon$. Since ϵ is arbitrary, our statement would seem to imply that we can find an Mw arbitrarily close to $\theta \in \Theta$ such $Mw \notin \Theta$. However, this is a contradiction of our assumption that Θ is an open set. Therefore, Θ' must be an open set and so $\mathcal{D}_2$ is regular.

Finally, we want to show the same properties about $\mathcal{D}_3$. In this scenario,$\mu = 0$ so we have

$$\begin{aligned}
p_X(x) &\propto \exp\left(\frac{-1}{2}\left(\sum_{i<j} 2(A \odot (z_1 z_1^T + z_2 z_2^T))_{ij} + \sum_i A_{ii}(z_i^2 + z_{i+d}^2)\right.\right. \\
&\qquad \left.\left. - \sum_{i<j}(B \odot (z_1 z_2^T - z_2 z_1^T))_{ij} + \mu^T\Sigma^{-1}z\right)\right) \\
&= \exp\left(\frac{-1}{2}\left(\sum_{i<j} 2(A \odot (z_1 z_1^T + z_2 z_2^T))_{ij} + \sum_i A_{ii}(z_i^2 + z_{i+d}^2)\right.\right. \\
&\qquad \left.\left. - \sum_{i<j}(B \odot (z_1 z_2^T - z_2 z_1^T))_{ij}\right)\right)
\end{aligned}$$

And so the statistics $T^1(X)$ and $T^2(X)$, which together form $\phi''(x)$ are minimal sufficient. The same strategy as applied above can be used to show that $\mathcal{D}_3$ is full and regular. □

E.2.2 The Multivariate Generalized Von Mises Distribution

The multivariate Generalized von Mises (mGvM) distribution generalizes the Torus Graphs distribution by containing second moment terms such as $(\cos\Theta_i)^2$, $(\sin\Theta_i)^2$ and $(\cos\Theta_i)(\sin\Theta_i)$ (Navarro et al., 2017). The definition of the distribution is given below:

Definition E.2.2 (Multivariate Generalized Von-Mises (mGvM) Distribution). We say that a random vector $\Theta_d : \Omega \to [0, 2\pi)^d$ is distributed multivariate generalized von-Mises $\Theta_d \sim mGvM(\nu, \psi)$ for a $2d$-dimensional vector ν and $2d \times 2d$ symmetric matrix ψ if the following holds:

$$p_\Theta(\theta) \propto \exp\left(\nu^T \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix} + \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}^T \psi \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}\right)$$

By constraining the parameters on the mGvM distribution, we can achieve a TG distribution:

Lemma E.2.3 (Relation between mGvM and TG distributions). *If* $\Theta \sim mGvM(\nu, \psi)$ *and* $\psi_{ii} = \psi_{i+d,i+d}$ *for* $1 \leq i \leq d$, *and further* $\psi_{i,i+d} = 0$ *for* $1 \leq i \leq d$, *then* $\Theta \sim TG(\eta)$ *for* $\eta_{ij} = 2[\psi_{ij}, \psi_{i,j+d}, \psi_{i+d,j}, \psi_{i+d,j+d}]^T$ *and* $\eta_{ii} = [\nu_i, \nu_{i+d}]$.

Proof. Assume that $\Theta \sim mGvM(\nu, \psi)$ for the conditions stated in the lemma. Then:

$$\begin{aligned}
p_\Theta(\theta) &\propto \exp\left(\nu^T \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix} + \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}^T \psi \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}\right) \\
&= \exp\Bigg(\sum_i \begin{bmatrix}\nu_i & \nu_{i+d}\end{bmatrix} \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix} + \sum_i ((\cos\theta_i)^2 \psi_{ii} + (\sin\theta_i)^2 \psi_{i+d,i+d}) + \\
&\qquad \sum_{i \neq j} \big(\psi_{i,j} \cos\theta_i \cos\theta_j + \psi_{i,j+d} \cos\theta_i \sin\theta_j + \\
&\qquad\qquad \psi_{i+d,j} \sin\theta_i \cos\theta_j + \psi_{i+d,j+d} \sin\theta_i \sin\theta_j\big)\Bigg) \\
&= \exp\Bigg(\sum_i \begin{bmatrix}\nu_i & \nu_{i+d}\end{bmatrix} \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix} + \sum_i \psi_{ii} \\
&\qquad + \sum_{i<j} 2 \begin{bmatrix}\psi_{ij} & \psi_{i,j+d} & \psi_{i+d,j} & \psi_{i+d,j+d}\end{bmatrix} \begin{bmatrix}\cos\theta_i \cos\theta_j \\ \cos\theta_i \sin\theta_j \\ \sin\theta_i \cos\theta_j \\ \sin\theta_i \sin\theta_j\end{bmatrix}\Bigg) \\
&\propto \exp\left(\sum_i \begin{bmatrix}\nu_i & \nu_{i+d}\end{bmatrix} \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix} + \sum_{i<j} 2 \begin{bmatrix}\psi_{ij} & \psi_{i,j+d} & \psi_{i+d,j} & \psi_{i+d,j+d}\end{bmatrix} \begin{bmatrix}\cos\theta_i \cos\theta_j \\ \cos\theta_i \sin\theta_j \\ \sin\theta_i \cos\theta_j \\ \sin\theta_i \sin\theta_j\end{bmatrix}\right) \\
&= \exp\left(\sum_i \eta_i^T \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix} + \sum_{i<j} \eta_{ij}^T \begin{bmatrix}\cos\theta_i \cos\theta_j \\ \cos\theta_i \sin\theta_j \\ \sin\theta_i \cos\theta_j \\ \sin\theta_i \sin\theta_j\end{bmatrix}\right)
\end{aligned}$$

Where the final line has the form of the TG distribution, with the values of η given in the lemma. $\square$

Further, by conditioning the angles of the CN distribution on the amplitudes, we achieve an mGvM distribution for any parameterization of the normal distribution:

Lemma E.2.4 (Complex Normal and mGvM)**.** *Assume* $X \sim CN(m, \Gamma, C)$*. By definition, there exists* μ, Σ *such that* $(\mathfrak{Re}X, \mathfrak{Im}X) \sim N(\mu, \Sigma)$*. Let* $\Theta_i = \arg X_i$ *denote the angle of* X_i*, and let* $R_i = |X_i|$ *denote the magnitude of* X_i*. Then:* $\Theta | R = r \sim mGvM(\nu, \psi)$ *where* $\nu = \begin{bmatrix}r \\ r\end{bmatrix} \odot (\Sigma^{-1}\mu)$ *and* $\psi = \frac{-1}{2}\Sigma^{-1} \odot \begin{bmatrix}r \\ r\end{bmatrix} \begin{bmatrix}r \\ r\end{bmatrix}^T$.

Proof. Let $Z = (\mathfrak{Re}X, \mathfrak{Im}X)$. By change of variables, note that

$$p_Z(r_1 \cos\theta_1, \ldots, r_d \cos\theta_d, r_1 \sin\theta_1, \ldots, r_d \sin\theta_d) \prod_i r_i = p_{\Theta,R}(\theta, r)$$

Then:

$$\begin{aligned}
p_{\Theta|R}(\theta|r) = \frac{p_{\Theta,R}(\theta,r)}{p_R(r)} &\propto p_{\Theta,R}(\theta,r) \propto p_Z(r_1\cos\theta_1,\dots,r_d\cos\theta_d,r_1\sin\theta_1,\dots,r_d\sin\theta_d) \\
&\propto \exp\left(\frac{-1}{2}\left(\begin{bmatrix} r\odot\cos\theta \\ r\odot\sin\theta\end{bmatrix}-\mu\right)^T\Sigma^{-1}\left(\begin{bmatrix} r\odot\cos\theta \\ r\odot\sin\theta\end{bmatrix}-\mu\right)^T\right) \\
&\propto \exp\left(\begin{bmatrix} r\odot\cos\theta \\ r\odot\sin\theta\end{bmatrix}^T\Sigma^{-1}\mu-\frac{1}{2}\begin{bmatrix} r\odot\cos\theta \\ r\odot\sin\theta\end{bmatrix}\Sigma^{-1}\begin{bmatrix} r\odot\cos\theta \\ r\odot\sin\theta\end{bmatrix}\right) \\
&= \exp\left(\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}^T\left((\Sigma^{-1}\mu)\odot\begin{bmatrix} r \\ r\end{bmatrix}\right)-\frac{1}{2}\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}\left(\Sigma^{-1}\odot\begin{bmatrix} r \\ r\end{bmatrix}\begin{bmatrix} r \\ r\end{bmatrix}^T\right)\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}\right) \\
&= \exp\left(\nu^T\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}+\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}^T\psi\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}\right)
\end{aligned}$$

For the values of ν, ψ given in the lemma. □

E.2.3 Proof of Theorem 5.4.2

Proof. Note that $(\mathfrak{Re}X, \mathfrak{Im}X) \sim N(\mu, \Sigma)$. By theorem E.2.4,

$$\Theta|R = r \sim mGvM(\nu,\psi) = mGvM\left(\begin{bmatrix} r \\ r\end{bmatrix}\odot(\Sigma^{-1}\mu), \frac{-1}{2}\Sigma^{-1}\begin{bmatrix} r \\ r\end{bmatrix}\begin{bmatrix} r \\ r\end{bmatrix}^T\right)$$

If $\psi_{ii} = \psi_{i+d,i+d}$ and $\psi_{i,i+d} = 0$, then lemma E.2.3 applies and the corollary follows. By construction, $\Sigma^{-1}_{ii} = \Sigma^{-1}_{i+d,i+d}$ and $\Sigma^{-1}_{i,i+d} = 0$, and hence

$$\begin{aligned}
\psi_{ii} &= \left(\frac{-1}{2}\Sigma^{-1}\begin{bmatrix} r \\ r\end{bmatrix}\begin{bmatrix} r \\ r\end{bmatrix}^T\right)_{ii} = \frac{-1}{2}r_ir_i\Sigma^{-1}_{ii} \\
&= \frac{-1}{2}r_ir_i\Sigma^{-1}_{i+d,i+d} = \left(\frac{-1}{2}\Sigma^{-1}\begin{bmatrix} r \\ r\end{bmatrix}\begin{bmatrix} r \\ r\end{bmatrix}^T\right)_{i+d,i+d} = \psi_{i+d,i+d} \\
\psi_{i,i+d} &= \left(\frac{-1}{2}\Sigma^{-1}\begin{bmatrix} r \\ r\end{bmatrix}\begin{bmatrix} r \\ r\end{bmatrix}^T\right)_{i,i+d} = \frac{-1}{2}r_{i,i+d}\Sigma^{-1}_{i,i+d} = 0
\end{aligned}$$

which meets the requirement. □

E.2.4 Structure of Proper Matrices

We need the following auxiliary result for the proof of theorem 5.4.3 that propriety of the CN distribution induces constraints on the conditional TG distribution.

Lemma E.2.5. *Assume W is a $2d \times 2d$ symmetric matrix such that, for a partition $W = \begin{bmatrix} W_{11} & W_{12} \\ W_{12}^T & W_{22} \end{bmatrix}$, we have that W_{12} is antisymmetric (i.e. $W_{12} = -W_{12}^T$) and that $W_{11} = W_{22}$. Assume that inverses of each block of W (that is, W_{11}^{-1} and W_{12}^{-1}) exist. Then, W^{-1} has the same form; that is, it is a symmetric matrix for which $(W^{-1})_{12} = -(W^{-1})_{12}^T$ and $(W^{-1})_{11} = (W^{-1})_{22}$*

Proof. Let W be a symmetric matrix that can be written as $W = \begin{bmatrix} A & B \\ -B & A \end{bmatrix}$. Using the block-matrix inversion formula which leverages the Schur complement of a matrix (see Zhang (2006)), I can write:

$$\begin{aligned} W^{-1} &= \begin{bmatrix} (A - BA^{-1}(-B))^{-1} & -(A - BA^{-1}(-B))^{-1}BA^{-1} \\ -A^{-1}(-B)(A - BA^{-1}(-B))^{-1} & A^{-1} + A^{-1}(-B)(A - BA^{-1}(-B))^{-1}BA^{-1} \end{bmatrix} \\ &= \begin{bmatrix} (A + BA^{-1}B)^{-1} & -(A + BA^{-1}B)^{-1}BA^{-1} \\ A^{-1}B(A + BA^{-1}B)^{-1} & A^{-1} - A^{-1}B(A + BA^{-1}B)^{-1}BA^{-1} \end{bmatrix} \end{aligned}$$

There are two important aspects of this expression to note. First, the upper right and lower left blocks are antisymmetric (in other words, that $W^{-1}_{i,j+d} = -W^{-1}_{j,i+d}$, for $1 \leq i \leq d, 1 \leq j \leq d$):

$$\begin{aligned} -(A + BA^{-1}B)^{-1}BA^{-1} &= -(A + BA^{-1}B)^{-1}(AB^{-1})^{-1} \\ &= -((AB^{-1})(A + BA^{-1}B))^{-1} = -(AB^{-1}A + B)^{-1} \\ \implies [-(A + BA^{-1}B)^{-1}BA^{-1}]^T &= [-(AB^{-1}A + B)^{-1}]^T \\ &= [-(AB^{-1}A + B)^T]^{-1} = [-(A(B^T)^{-1}A + B^T)]^{-1} \\ &= (AB^{-1}A + B)^{-1} = (A + BA^{-1}B)^{-1}BA^{-1} \end{aligned}$$

Further, I claim that the top left and bottom right blocks of A^{-1} are equal (that $A^{-1}_{ij} = A^{-1}_{i+d,j+d}$ for $1 \leq i \leq d, 1 \leq j \leq d$). To see this, note that:

$$\begin{aligned} &A^{-1} - A^{-1}B(A + BA^{-1}B)^{-1}BA^{-1} \\ &= A^{-1}(A + BA^{-1}B)(A + BA^{-1}B)^{-1} - A^{-1}B(A + BA^{-1}B)^{-1}BA^{-1} \\ &= (I + A^{-1}BA^{-1}B)(A + BA^{-1}B)^{-1} - A^{-1}B(A + BA^{-1}B)^{-1}BA^{-1} \\ &= (A + BA^{-1}B)^{-1} + A^{-1}BA^{-1}B(A + BA^{-1}B)^{-1} - A^{-1}B(A + BA^{-1}B)^{-1}BA^{-1} \end{aligned}$$

Therefore, the desired condition is met if $A^{-1}BA^{-1}B(A + BA^{-1}B)^{-1} = A^{-1}B(A + BA^{-1}B)^{-1}BA^{-1}$. To show that this is the case, note that:

$$\begin{aligned} A^{-1}BA^{-1}B(A + BA^{-1}B)^{-1} &= (B^{-1}AB^{-1}A)^{-1}(A + BA^{-1}B)^{-1} \\ &= ((A + BA^{-1}B)B^{-1}AB^{-1}A)^{-1} \\ &= (AB^{-1}AB^{-1}A + A)^{-1} \\ A^{-1}B(A + BA^{-1}B)^{-1}BA^{-1} &= (B^{-1}A)^{-1}(A + BA^{-1}B)^{-1}(AB^{-1})^{-1} \\ &= (AB^{-1}(A + BA^{-1}B)(B^{-1}A))^{-1} \\ &= (AB^{-1}AB^{-1}A + A)^{-1} \end{aligned}$$

Therefore, W^{-1} has the desired form. □

We also need this auxiliary result relating the parametrization of the TG distribution $TG(\eta)$ to the parametrization $TG(\eta(\phi))$ as specified in equations 5.2 and 5.3.

Lemma E.2.6. *Assume* $\Theta \sim TG(\eta)$. *Then,* $\Theta \sim TG2(\phi)$ *for* $\phi_i = \eta_i$, $\phi_{ij} = \frac{1}{2}[(\eta_{ij})^1 + (\eta_{ij})^4, -(\eta_{ij})^2 + (\eta_{ij})^3, (\eta_{ij})^1 - (\eta_{ij})^4, (\eta_{ij})^2 + (\eta_{ij})^3]$, *where* $(\eta_{ij})^k$ *denotes the kth component of* η_{ij}.

Proof. From basic properties of trig functions:

$$\phi_{ij}^T \begin{bmatrix} \cos(\theta_i - \theta_j) \\ \sin(\theta_i - \theta_j) \\ \cos(\theta_i + \theta_j) \\ \sin(\theta_i + \theta_j) \end{bmatrix} = \eta_{ij}^T \begin{bmatrix} \cos\theta_i \cos\theta_j \\ \cos\theta_i \sin\theta_j \\ \sin\theta_i \cos\theta_j \\ \sin\theta_i \sin\theta_j \end{bmatrix} = \frac{1}{2}\eta_{ij}^T \begin{bmatrix} \cos(\theta_i + \theta_j) + \cos(\theta_i - \theta_j) \\ \sin(\theta_i + \theta_j) - \sin(\theta_i - \theta_j) \\ \sin(\theta_i + \theta_j) + \sin(\theta_i - \theta_j) \\ \cos(\theta_i - \theta_j) - \cos(\theta_i + \theta_j) \end{bmatrix}$$

$$\implies \phi_{ij} = \frac{1}{2}[(\eta_{ij})^1 + (\eta_{ij})^4, -(\eta_{ij})^2 + (\eta_{ij})^3, (\eta_{ij})^1 - (\eta_{ij})^4, (\eta_{ij})^2 + (\eta_{ij})^3]$$

The proof then follows by comparing the PDFs. □

E.2.5 Proof of Theorem 5.4.3

Proof. Let $\Sigma = Cov((\mathfrak{Re}X, \mathfrak{Im}X), (\mathfrak{Re}X, \mathfrak{Im}X))$. From (Picinbono, 1996), we have that $\Sigma_{11} = \frac{1}{2}\mathfrak{Re}\Gamma, \Sigma_{12} = \frac{-1}{2}\mathfrak{Im}\Gamma, \Sigma_{21} = \frac{1}{2}\mathfrak{Im}\Gamma, \Sigma_{22} = \frac{1}{2}\mathfrak{Re}\Gamma$. Therefore, Σ is of the the form specified in lemma E.2.5, and so Σ^{-1} is as well. Note that Σ^{-1} meets the conditions of theorem 5.4.2, since $(\Sigma^{-1})_{ii} = (\Sigma^{-1})_{i+d,i+d}$ and $(\Sigma^{-1})_{i,i+d} = (\Sigma^{-1})_{i+d,i} = 0$ (the second fact follows from the observation that the diagonal of any antisymmetric matrix must be zero to achieve antisymmetry). Therefore, $\Theta | R = r \sim TG(\eta)$ where $\eta_{ij} = \frac{-1}{2} r_i r_j [\Sigma^{-1}_{i,j}, \Sigma^{-1}_{i,j+d}, \Sigma^{-1}_{i+d,j}, \Sigma^{-1}_{i+d,j+d}]$. Therefore:

$$\begin{aligned} \phi_{ij} &= \frac{1}{2}[(\eta_{ij})^1 + (\eta_{ij})^4, -(\eta_{ij})^2 + (\eta_{ij})^3, (\eta_{ij})^1 - (\eta_{ij})^4, (\eta_{ij})^2 + (\eta_{ij})^3] \\ &= \frac{1}{2}\left(\frac{-1}{2} r_i r_j\right) [2(\Sigma^{-1})_{ij}, -2(\Sigma^{-1})_{i,j+d}, 0, 0] \end{aligned}$$

Which is exactly the form desired (i.e. $(\phi_{ij})^3 = (\phi_{ij})^4 = 0$). Note then that all parameters are zero except for those on terms of the form $\cos(\theta_i - \theta_j)$ and $\sin(\theta_i - \theta_j)$. Clearly, these terms are invariant to a circular shift since $(\theta_i + \epsilon) - (\theta_j + \epsilon) = \theta_i - \theta_j$ for all ϵ. Therefore, Θ is circularly symmetric. □

E.2.6 Proof of Corollary 5.4.4

Proof. We have the same setup as in theorem 5.4.3, so ϕ_{ij} has the same form. In addition, from theorem 5.4.2 we have $\eta_i = r_i[(\Sigma^{-1}\mu)_i, (\Sigma^{-1}\mu)_{i+d}]$ and from lemma E.2.6 $\eta_i = \phi_i$ so therefore since $\mu = E[(\mathfrak{Re}X, \mathfrak{Im}X)] = 0$ we have $\phi_i = 0$ in addition. □

E.3 Fitting Procedure for Latent Variable Model

As discussed in the text, the model we aim to fit is the following:

$$\begin{aligned} Z^{(n)} &\sim CN(0,\Gamma) \\ X^{(n),r} &= \beta^r Z_r^{(n)} + \epsilon^{(n),r} \\ \epsilon^{(n),r} &\sim CN(0,\eta^r) \end{aligned}$$

To alleviate identifiability issues in the estimation of Γ, we constrain $(\beta^r)^H\beta^r = 1$ and further that $\mathfrak{Im}(\beta_1^r) = 0$ where β_1^r denotes the first element of β^r. The likelihood function for this model is the following:

$$\begin{aligned} L(\theta;\{x^{(n)}\},\{z^{(n)}\}) = \prod_{n=1}^{N}\Bigg[&\frac{1}{\pi^R\det(\Gamma)}\exp(-(z^n)^H\Gamma^{-1}z^{(n)}) \\ &\prod_{r=1}^{R}\left[\frac{1}{\pi^{R(r)}\det(\eta^r)}\exp(-(x^{(n),r}-\beta^r z_r^{(n)})^H(\eta^r)^{-1}(x^{(n),r}-\beta^r z_r^{(n)}))\right]\Bigg] \end{aligned}$$

We use expectation-maximization to fit the model. First, we define auxiliary variables $L(t) \in \mathbb{C}^{R\times R}, W(t) \in \mathbb{C}^{R\times R}, V^{(n)}(t) \in \mathbb{C}^{R\times 1}$

$$\begin{aligned} &[L(t)]_{ij} = \begin{cases} (\beta^i)^H(\eta^i)^{-1}\beta^r & i = j \\ 0 & i \neq j \end{cases} \\ &W(t) = (\Gamma^{-1}(t) + L(t))^{-1} \\ &[V^{(n)}(t)]_r = (\beta^r)^H(\eta^r)^{-1}x^{(n),r} \\ &\mu^{(n)}(t) = W(t)V^{(n)}(t) \end{aligned}$$

We note that $Z^{(n)}|X^{(n)} \sim CN(W(t)V^{(n)}(t), W(t))$. We then compute the following "Q" function:

$$\begin{aligned} Q(\theta(t+1)) &= E_{Z|X,\theta(t)} \log L(\theta; X, Z) \\ &\propto \sum_n \Bigg[-\ln\det\Gamma + \sum_r(-\ln\det\eta_r) - \mathrm{Tr}(\Gamma^{-1}(W(t) + \mu^{(n)}(\mu^{(n)})^H)) \\ &\qquad + \sum_r \Bigg[-(x^{(n),r})^H(\eta^r)^{-1}x^{(n),r} + \overline{\mu_r^{(n)}}(\beta^r)^H(\eta^r)^{-1}x^{(n),r} \\ &\qquad\qquad + (x^{(n),r})^H(\eta^r)^{-1}\beta^r\mu_r^{(n)} - (\beta^r)^H(\eta^r)^{-1}\beta^r(W_{rr}(t) + \mu_r^{(n)}\overline{\mu_r^{(n)}})\Bigg] \end{aligned}$$

Note that since the model's parameters are complex-valued, we must use Wirtinger calculus to take derivatives (see Adali et al. (2011) for an introduction to Wirtinger calculus). By taking derivatives according to the Wirtinger calculus and setting them equal to zero, we achieve the following update steps:

$$
\begin{aligned}
\Gamma(t+1) &= W(t) + \frac{1}{N}\sum_{n=1}^{N} \mu^{(n)}(t)(\mu^{(n)}(t))^H \\
\beta^r(t+1) &= \frac{\sum_n x^{(n),r}(\overline{\mu_r^{(n)}(t)})}{\sum_n (W_{rr}(t) + \mu_r^{(n)}(t)(\overline{\mu_r^{(n)}(t)}))} \\
\eta^r(t+1) &= \frac{1}{N}\sum_n \Bigg[x^{(n),r}(x^{(n),r})^H - (x^{(n),r})(\beta^r(t+1))^H(\overline{\mu_r^{(n)}(t)})) \\
&\qquad - \mu_r^{(n)}\beta^r(t+1)(x^{(n),r})^H + \beta^r(t+1)(\beta^r(t+1))^H(W_{rr}(t) + \mu_r^{(n)}(\overline{\mu_r^{(n)}(t)})\mu_r(t) \Bigg]
\end{aligned}
$$

To satisfy the constraints $(\beta^r)^H(\beta^r) = 1$ and $\mathfrak{Im}\beta_1^r = 0$, we renormalize β^r at the end of each iteration and multiply it by a complex scalar of norm one such that $\mathfrak{Im}\beta^r = 0$. We then apply the opposite transformation to the appropriate column/row of Γ. This transformation will not change the marginal likelihood $L(\theta; X)$, which is the objective for EM.

E.4 Table of Examples Comparing PLV, Amplitude Correlation, and Complex Correlation

Explanation	Setting	Result
PLV, amplitude covariance and complex means are zero, but complex covariance is nonzero	$\Theta_1, \Theta_2 \sim Unif(0, 2\pi)$ $R_1 = \Theta_2$ $R_2 = \Theta_1$	$PLV(\Theta_1, \Theta_2) = 0$ $Cov(R_1, R_2) = 0$ $EX_1 = EX_2 = 0$ $\lvert Cov(X_1, X_2)\rvert > 0$ $\lvert PCov(X_1, X_2)\rvert > 0$
Another example in which PLV, amplitude covariance and complex means are zero, but complex covariance is nonzero	$\Theta_1, \Theta_2 \sim Unif(0, 2\pi)$ $R_1 = \Theta_1 + \Theta_2 \mod 2\pi$ $R_2 = \Theta_1 - \Theta_2 \mod 2\pi$	$PLV(\Theta_1, \Theta_2) = 0$ $Cov(R_1, R_2) = 0$ $EX_1 = EX_2 = 0$ $\lvert Cov(X_1, X_2)\rvert > 0$
PLV, and amplitude covariance are zero, but complex means and complex covariance are nonzero	$\Theta_1, \Theta_2 \sim Unif(0, 2\pi)$ $R_1 = \Theta_1 + \Theta_2$ $R_2 = \Theta_1 - \Theta_2 + 2\pi$	$PLV(\Theta_1, \Theta_2) = 0$ $Cov(R_1, R_2) = 0$ $EX_1 \neq 0, \quad EX_2 \neq 0$ $\lvert Cov(X_1, X_2)\rvert > 0$
PLV, amplitude covariance are complex covariance are zero but complex means are nonzero	$\Theta_1, \Theta_2 \sim Unif(0, 2\pi)$ $R_1 = \Theta_1$ $R_2 = \Theta_2$	$PLV(\Theta_1, \Theta_2) = 0$ $Cov(R_1, R_2) = 0$ $EX_1 \neq 0, \quad EX_2 \neq 0$ $\lvert Cov(X_1, X_2)\rvert = 0$ $\lvert PCov(X_1, X_2)\rvert = 0$

Table E.1: Examples of various ways in which complex covariance and means can capture associations not captured by PLV or amplitude correlation. In all cases, assume that variables are independent unless otherwise stated. We define $X_1 = R_1 \exp(i\Theta_1)$ and $X_2 = R_2 \exp(i\Theta_2)$.

Bibliography

Adali, T., Schreier, P. J., and Scharf, L. L. (2011). Complex-valued signal processing: The proper way to deal with impropriety. *IEEE Transactions on Signal Processing*, 59(11):5101–5125.

Adhikari, A., Sigurdsson, T., Topiwala, M. A., and Gordon, J. A. (2010). Cross-correlation of instantaneous amplitudes of field potential oscillations: a straightforward method to estimate the directionality and lag between brain areas. *Journal of neuroscience methods*, 191:191–200.

Allen, G. I. and Tibshirani, R. (2010). Transposable regularized covariance models with an application to missing data imputation. *Ann. Appl. Stat.*, 4:764–790.

Andersen, H., Højbjerre, M., Sørensen, D., and Eriksen, P. (1995). The multivariate complex normal distribution. In *Linear and Graphical Models*, pages 15–37. Springer.

Armington, J. C. and Bloom, M. B. (1974). Relations between the amplitudes of spontaneous saccades and visual responses. *Journal of Optical Society of America*, 64(9):1263–1271.

Bach, F. R. and Jordan, M. I. (2004). Learning graphical models for stationary time series. *IEEE transactions on signal processing*, 52(8):2189–2199.

Bach, F. R. and Jordan, M. I. (2005). A probabilistic interpretation of canonical correlation analysis. Technical Report 688, Department of Statistics, University of California, Berkeley, Berkeley, CA.

Barndorff-Nielsen, O. (2014). *Information and exponential families: in statistical theory*. John Wiley & Sons.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.*, 36:199–227.

Bong, H., Liu, Z., Ren, Z., Smith, M. A., Ventura, V., and Kass, R. E. (2020a). Latent dynamic factor analysis of high-dimensional neural recordings. *Advances in Neural Information Processing Systems*, 33.

Bong, H., Liu, Z., Ren, Z., Smith, M. A., Ventura, V., and Kass, R. E. (2020b). Latent dynamic factor analysis of high-dimensional neural recordings. *Advances in Neural Information Processing Systems*, 33.

Bong, H., Ventura, V., Smith, M., and Kass, R. (2020c). Latent cross-population dynamic time-series analysis of high-dimensional neural recordings. *Manuscript submitted for publication.*

Bong, H., Ventura, V., Yttri, E. A., Smith, M. A., and Kass, R. E. (2021). Latent cross-population dynamic time-series analysis of high-dimensional neural recordings. *arXiv preprint arXiv:2105.03508*.

Brémaud, P. (2014). Fourier analysis of stochastic processes. In *Fourier Analysis and Stochastic Processes*, pages 119–179. Springer.

Brillinger, D. R. (2001). *Time series: data analysis and theory*. SIAM.

Brincat, S. L. and Miller, E. K. (2015). Frequency-specific hippocampal-prefrontal interactions during associative learning. *Nature Neuroscience*, 18:576–581.

Brincat, S. L. and Miller, E. K. (2016). Prefrontal cortex networks shift from external to internal modes during learning. *Journal of Neuroscience*, 36(37):9739–9754.

Buesing, L., Machado, T. A., Cunningham, J. P., and Paninski, L. (2014). Clustered factor analysis of multineuronal spike data. In *Advances in Neural Information Processing Systems*, pages 3500–3508.

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6):407–420.

Buzsaki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *science*, 304(5679):1926–1929.

Cadieu, C. F. and Koepsell, K. (2010). Phase coupling estimation from multivariate phase statistics. *Neural computation*, 22(12):3107–3126.

Cai, T. T., Li, H., Liu, W., and Xie, J. (2016a). Joint Estimation of Multiple High-dimensional Precision Matrices. *Statistica Sinica*, 26(2):445–464.

Cai, T. T., Ren, Z., and Zhou, H. H. (2016b). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics*, 10(1):1–59.

Cardin, J. A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.-H., and Moore, C. I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature*, 459(7247):663–667.

Chang, J., Qiu, Y., Yao, Q., and Zou, T. (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics*, 206:57 – 82.

Chen, X. and Liu, W. (2015). Graph Estimation for Matrix-variate Gaussian Data. *arXiv e-prints*, page arXiv:1509.05453.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2012). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *arXiv e-prints*, page arXiv:1212.6906.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162:47–70.

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., Bradley, D. C., Smith, M. A., Kohn, A., Movshon, J. A., Armstrong, K. M., Moore, T., Chang, S. W., Snyder, L. H., Lisberger, S. G., Priebe, N. J., Finn, I. M., Ferster, D., Ryu, S. I., Santhanam, G., Sahani, M., and Shenoy, K. V. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, 13(3):369–378.

Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373–397.

Dandekar, S., Privitera, C., Carney, T., and Klein, S. A. (2012). Neural saccadic response estimation during natural viewing. *Journal of Neurophysiology*, 107(6):1776–1790. PMID: 22170971.

Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, 68:265–274.

Dawid, A. P. and Musio, M. (2014). Theory and applications of proper scoring rules. *Metron*, 72(2):169–183.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Ding, S. and Dennis Cook, R. (2018). Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(2):387–408.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11):770–785.

Fornito, A., Zalesky, A., and Breakspear, M. (2013). Graph analysis of the human connectome: Promise, progress, and pitfalls. *NeuroImage*, 80:426 – 444. Mapping the Connectome.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9(10):474–480.

Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291(5508):1560–1563.

Gallagher, N., Ulrich, K. R., Talbot, A., Dzirasa, K., Carin, L., and Carlson, D. E. (2017). Cross-spectral factor analysis. In *Advances in Neural Information Processing Systems*, pages 6842–6852.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American statistical association*, 77(378):304–313.

Goris, R. L. T., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17(6):858–865.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28:321–377.

Hotelling, H. (1992). Relations between two sets of variates. In *Breakthroughs in statistics*, pages 162–190. Springer.

Hultman, R., Ulrich, K., Sachs, B. D., Blount, C., Carlson, D. E., Ndubuizu, N., Bagot, R. C., Parise, E. M., Vu, M.-A. T., Gallagher, N. M., et al. (2018). Brain-wide electrical spatiotemporal dynamics encode depression vulnerability. *Cell*, 173:166–180.

Hyvärinen, A. and Dayan, P. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4).

Jankova, J. and Van De Geer, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229.

Jiang, H., Bahramisharif, A., van Gerven, M. A., and Jensen, O. (2015). Measuring directionality between neuronal oscillations of different frequencies. *Neuroimage*, 118:359–367.

Johnston, R., Snyder, A. C., Khanna, S. B., Issar, D., and Smith, M. A. (2020). The eyes reflect an internal cognitive state embedded in the population activity of cortical neurons. *bioRxiv*.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236.

Kass, R. E., Amari, S.-I., Arai, K., Brown, E. N., Diekman, C. O., Diesmann, M., Doiron, B., Eden, U. T., Fairhall, A. L., and Fiddyment, G. M. (2018). Computational neuroscience: Mathematical and statistical perspectives. *Annual review of statistics and its application*, 5:183–214.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58:433–451.

Khanna, S. B., Scott, J. A., and Smith, M. A. (2020). Dynamic shifts of visual and saccadic signals in prefrontal cortical regions 8Ar and FEF. *Journal of Neurophysiology*. In press.

Klein, N., Orellana, J., Brincat, S. L., Miller, E. K., and Kass, R. E. (2020). Torus graphs for multivariate phase coupling analysis. *Annals of Applied Statistics*, 14(2):635–660.

Kuchibhotla, A. K. and Chakrabortty, A. (2018). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *arXiv preprint arXiv:1804.02605*.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208.

Leavitt, M. L., Mendoza-Halliday, D., and Martinez-Trujillo, J. C. (2017). Sustained activity encoding working memories: Not fully distributed. *Trends in Neurosciences*, 40(6):328 – 346.

Lee, W. and Liu, Y. (2015). Joint Estimation of Multiple Precision Matrices with Common Structures. *Journal of Machine Learning Research*, 16:1035–1062.

Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *Journal of the American Statistical Association*, 107:1187–1200.

Lepage, K. Q. and Vijayan, S. (2017). The relationship between coherence and the phase-locking value. *Journal of Theoretical Biology*, 435:106–109.

Liebe, S., Hoerzer, G. M., Logothetis, N. K., and Rainer, G. (2012). Theta coupling between v4 and prefrontal cortex predicts visual short-term memory performance. *Nature neuroscience*, 15:456.

Linderman, S., Nichols, A., Blei, D., Zimmer, M., and Paninski, L. (2019). Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in c. elegans. *bioRxiv*, page 621540.

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.*, 41:2948–2978.

Liu, Y. and Ren, Z. (2020). Minimax Estimation of Large Precision Matrices with Bandable Cholesky Factor. *To appear, The Annals of Statistics*, page arXiv:1712.09483.

Liu, Y., Taniguchi, M., and Ombao, H. (2021). Statistical inference for local granger causality. *arXiv preprint arXiv:2103.00209*.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fmri signal. *nature*, 412(6843):150–157.

Lowet, E., Roberts, M. J., Bonizzi, P., Karel, J., and De Weerd, P. (2016). Quantifying neural oscillatory synchronization: a comparison between spectral coherence and phase-locking value approaches. *PloS one*, 11(1):e0146443.

Magri, C., Schridde, U., Murayama, Y., Panzeri, S., and Logothetis, N. K. (2012a). The amplitude and timing of the bold signal reflects the relationship between local field potential power at different frequencies. *Journal of Neuroscience*, 32(4):1395–1407.

Magri, C., Schridde, U., Murayama, Y., Panzeri, S., and Logothetis, N. K. (2012b). The amplitude and timing of the bold signal reflects the relationship between local field potential power at different frequencies. *Journal of Neuroscience*, 32:1395–1407.

Mathalon, D. H. and Sohal, V. S. (2015). Neural oscillations and synchrony in brain dysfunction and neuropsychiatric disorders: it's about time. *JAMA psychiatry*, 72(8):840–844.

Mazumder, R. and Hastie, T. (2012). The graphical lasso: New insights and alternatives. *Electronic journal of statistics*, 6:2125.

Mazumder, R., Hastie, T., and Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11:2287–2322.

McKee, Matthew (2019). A braingate electrode array with a dime for size comparison. [Online; accessed Jan 16, 2020].

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34:1436–1462.

Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24:167–202.

Miller, E. K., Lundqvist, M., and Bastos, A. M. (2018). Working memory 2.0. *Neuron*, 100(2):463–475.

Navarro, A. K., Frellsen, J., and Turner, R. E. (2017). The multivariate generalised von mises distribution: inference and applications. In *Thirty-first AAAI conference on artificial intelligence*.

Ombao, H. and Pinto, M. (2021). Spectral dependence. *arXiv preprint arXiv:2103.17240*.

Ombao, H. and Van Bellegem, S. (2008). Evolutionary coherence of nonstationary signals. *IEEE Transactions on Signal Processing*, 56(6):2259–2266.

Orban, G. A. (2008). Higher order visual processing in macaque extrastriate cortex. *Physiological reviews*, 88(1):59–89.

Ossandón, J. P., Helo, A. V., Montefusco-Siegmund, R., and Maldonado, P. E. (2010). Superposition model predicts eeg occipital activity during free viewing of natural scenes. *Journal of Neuroscience*, 30(13):4787–4795.

Park, K. I. and Park, M. (2018). *Fundamentals of probability and stochastic processes with applications to communications*. Springer.

Pesaran, B., Vinck, M., Einevoll, G. T., Sirota, A., Fries, P., Siegel, M., Truccolo, W., Schroeder, C. E., and Srinivasan, R. (2018). Investigating large-scale brain dynamics using field potential recordings: analysis and interpretation. *Nature neuroscience*, 21(7):903–919.

Peterson, C., Stingo, F. C., and Vannucci, M. (2015). Bayesian inference of multiple gaussian graphical models. *Journal of the American Statistical Association*, 110(509):159–174. PMID: 26078481.

Picinbono, B. (1994). On circularity. *IEEE Transactions on signal processing*, 42(12):3473–3482.

Picinbono, B. (1996). Second-order complex random vectors and normal distributions. *IEEE Transactions on Signal Processing*, 44(10):2637–2640.

Ren, Z., Kang, Y., Fan, Y., and Lv, J. (2019). Tuning-free heterogeneous inference in massive networks. *Journal of the American Statistical Association*, 114:1908–1925.

Ren, Z., Sun, T., Zhang, C.-H., and Zhou, H. H. (2015). Asymptotic normality and optimalities in estimation of large gaussian graphical models. *Ann. Statist.*, 43:991–1026.

Rodu, J., Klein, N., Brincat, S. L., Miller, E. K., and Kass, R. E. (2018). Detecting multivariate cross-correlation between brain regions. *Journal of neurophysiology*, 120(4):1962–1972.

Sapountzis, P. and Gregoriou, G. (2017). Neural signatures of attention: Insights from decoding population activity patterns. *Frontiers In Bioscience*, 22:2069–2090.

Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G., and Von Stein, A. (1998). Synchronization between prefrontal and posterior association cortex during human working memory. *Proceedings of the National Academy of Sciences*, 95:7092–7096.

Schmidt, R., Ruiz, M. H., Kilavik, B. E., Lundqvist, M., Starr, P. A., and Aron, A. R. (2019). Beta oscillations in working memory, executive control of movement and thought, and sensorimotor function. *Journal of Neuroscience*, 39(42):8231–8238.

Seth, A. K., Barrett, A. B., and Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35:3293–3297.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American statistical Association*, 88(422):486–494.

Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852):86–92.

Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., Groblewski, P. A., Ahmed, R., Arkhipov, A., Bernard, A., Billeh, Y. N., Brown, D., Buice, M. A., Cain, N., Caldejon, S., Casal, L., Cho, A., Chvilicek, M., Cox, T. C., Dai, K., Denman, D. J., de Vries, S. E. J., Dietzman, R., Esposito, L., Farrell, C., Feng, D., Galbraith, J., Garrett, M., Gelfand, E. C., Hancock, N., Harris, J. A., Howard, R., Hu, B., Hytnen, R., Iyer, R., Jessett, E., Johnson, K., Kato, I., Kiggins, J., Lambert, S., Lecoq, J., Ledochowitsch, P., Lee, J. H., Leon, A., Li, Y., Liang, E., Long, F., Mace, K., Melchior, J., Millman, D., Mollenkopf, T., Nayan, C., Ng, L., Ngo, K., Nguyen, T., Nicovich, P. R., North, K., Ocker, G. K., Ollerenshaw, D., Oliver, M., Pachitariu, M., Perkins, J., Reding, M., Reid, D., Robertson, M., Ronellenfitch, K., Seid, S., Slaughterbeck, C., Stoecklin, M., Sullivan, D., Sutton, B., Swapp, J., Thompson, C., Turner, K., Wakeman, W., Whitesell, J. D., Williams, D., Williford, A., Young, R., Zeng, H., Naylor, S., Phillips, J. W., Reid, R. C., Mihalas, S., Olsen, S. R., and Koch, C. (2019). A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas. *bioRxiv*.

Smith, M. A. and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience*, 28(48):12591–12603.

Srinath, R. and Ray, S. (2014). Effect of amplitude correlations on coherence in the local field potential. *Journal of neurophysiology*, 112(4):741–751.

Steinmetz, N. A., Koch, C., Harris, K. D., and Carandini, M. (2018). Challenges and opportunities for large-scale electrophysiology with neuropixels probes. *Current opinion in neurobiology*, 50:92–100.

Tibshirani, R. J. and Taylor, J. (2012). Degrees of freedom in lasso problems. *The Annals of Statistics*, 40(2):1198–1232.

Ventura, V. (2004). Testing for and estimating latency effects for poisson and non-poisson spike trains. *Neural Computation*, 16(11):2323–2349.

Ventura, V., Cai, C., and Kass, R. E. (2005). Statistical assessment of time-varying dependency between two neurons. *Journal of Neurophysiology*, 94(4):2940–2947.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Vinci, G., Ventura, V., Smith, M. A., and Kass, R. E. (2018). Adjusted regularization of cortical covariance. *Journal of Computational Neuroscience*, 45:83–101.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, pages 471–494.

Yang, Y., Aminoff, E., Tarr, M., and Kass, R. E. (2016). A state-space model of cross-region dynamic connectivity in meg/eeg. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 1234–1242.

Curran Associates, Inc.

Yang, Y., Tarr, M. J., Kass, R. E., and Aminoff, E. M. (2019). Exploring spatiotemporal neural dynamics of the human visual cortex. *Human brain mapping*, 40(14):4213–4238.

Ye, Y., Xia, Y., and Li, L. (2019). Paired Test of Matrix Graphs and Brain Connectivity Analysis. *arXiv e-prints*, page arXiv:1908.08095.

Yin, J. and Li, H. (2012). Model selection and estimation in the matrix normal graphical model. *Journal of Multivariate Analysis*, 107:119 – 140.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology*, 102(1):614–635. PMID: 19357332.

Yu, M., Kolar, M., and Gupta, V. (2016). Statistical inference for pairwise graphical models using score matching. In *Advances in Neural Information Processing Systems*, pages 2829–2837.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68:49–67.

Zhang, F. (2006). *The Schur complement and its applications*, volume 4. Springer Science & Business Media.

Zhao, Y. and Park, I. M. (2017). Variational latent gaussian process for recovering single-trial dynamics from population spike trains. *Neural computation*, 29(5):1293–1316.

Zhou, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *Ann. Statist.*, 42:532–562.

Zhu, Y. and Cribben, I. (2018). Sparse graphical models for functional connectivity networks: best methods and the autocorrelation issue. *Brain connectivity*, 8(3):139–165.

Zhu, Y. and Li, L. (2018). Multiple matrix gaussian graphs estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80:927–950.

Zou, H., Hastie, T., and Tibshirani, R. (2007). On the “degrees of freedom” of the lasso. *The Annals of Statistics*, 35(5):2173–2192.