IDENTIFICATION OF NETWORKS OF WILSON-COWAN NEURONAL OSCILLATORS BY INVERSE SIGMOIDAL TRANSFORMATION

Cassiano O. Becker[1], *Ankit N. Khambhati*[2], *Danielle S. Bassett*[1,2], *and Victor M. Preciado*[1]

University of Pennsylvania
[1]Department of Electrical and Systems Engineering
[2]Department of Bioengineering
Philadelphia, PA 19104 USA

ABSTRACT

Network neuroscience is an expanding interdisciplinary field at the intersection of engineering, math, physics, and neuroscience dedicated to understanding connectivity in the brain in health and disease. A critical challenge in network neuroscience is inferring brain connectivity from statistical relationships in the functional dynamics of individual brain regions. The ability to map brain connectivity and retain neurophysiologic interpretation of those connections is crucial in the study of cognition and in the diagnosis and treatment of brain network disorders. In this work, we propose a method for estimating structural brain connections from excitatory and inhibitory neuron populations using a neuronal network model of Wilson-Cowan oscillators. Our technique estimates the weights of a neuronal network comprised of Wilson-Cowan oscillators based on the observation of times series data and on the knowledge of individual oscillator parameters. Specifically, we employ a derivative estimation technique and develop an inverse nonlinear transformation, which leads to an estimation problem that is linear on the target network weights. To solve the associated large-scale optimization problem we apply a proximal-type optimization algorithm. Finally, to demonstrate the effectiveness of our method, we perform computational experiments using simulations based on neuroimaging connectivity data, showing that network weights are recovered with high accuracy. Our method contributes to integrating brain connectivity with dynamical models of brain function, and may have an impact in diagnosing and understanding pathophysiology in brain disorders such as Parkinson's, epilepsy, or schizophrenia, in which imbalances in excitation and inhibition affect functional connectivity.

***Index Terms*—** neuronal-mass models, time-series analysis, network identification

I. INTRODUCTION

Complex brain networks describe an intricate wiring system for the brain that constrains communication between populations of neurons and supports behavior and cognition. Anatomical connections –white-matter fiber tracts at the macro-scale and axonal projections at the micro-scale– link the functional roles of distributed populations of neurons, forming a map, or mathematical graph, of the brain. Brain graphs commonly describe functional activity of specific brain regions, or nodes, and the structural connections between them, or edges [1]. In network neuroscience, a long-standing question has been: "How might structure facilitate functional communication between brain regions?" [2, 3]. Or, more specifically, "How similar are structural brain graphs to functional brain graphs?". Studying such structure-function relationships in the brain would help network neuroscientists better understand the anatomical substrates of cognition and the functional consequences of its loss in disease.

To infer how brain regions communicate, network neuroscientists measure functional connectivity by computing the statistical similarity between brain dynamics of different regions using methods grounded in signal processing, information theory, and time-series analysis [4, 5, 6]. There is overwhelming evidence that different features of brain signal dynamics underlie functional interactions between specific neuron types. This guiding principle has motivated the use of different similarity functions, such as the Pearson correlation coefficient, power spectral coherence, wavelet coherence, and mutual information, to uncover linear and non-linear relationships between functional brain dynamics. While these similarity measures have been used to map behavioral and cognitive states to stereotyped patterns of functional interactions, the structural connections underlying inter-regional functional relationships remain largely a mystery.

A critical limitation of existing statistical techniques for measuring functional connectivity is that they are indirect measurements of anatomic and physiologic features at the source of measured brain signals. Compelling evidence that one of the basic processing units in the neocortex is the cortical column, heterogeneously comprised of interconnected excitatory and inhibitory neurons that also communicate with neurons in other cortical columns was first presented in [7]. Intuitively, the collective activity of columnar neurons drives fluctuations in the local electric field potential. The resulting neural field is believed to underlie a robust neuron population-level code, which was modeled by Wilson and Cowan by a coarse-grained mean field model of spatially-localized excitatory and inhibitory currents [8]. Since then, Wilson-Cowan's neural field model has been studied in the context of and linked to spatiotemporal phenomena underlying network interactions in mammalian neocortex. In contrast to time-series models of functional brain dynamics, the Wilson-Cowan model is grounded

C.O.B. is supported by CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil. V.M.P. is supported by the NSF under grants CNS-1302222 and IIS-1447470. D.S.B. would also like to acknowledge support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Army Research Laboratory and the Army Research Office through contract numbers W911NF-10-2-0022 and W911NF-14-1-0679, the National Institute of Health (2-R01-DC-009209-11, 1R01HD086888-01, R01-MH107235, R01-MH107703, and R21-M MH-106799), the Office of Naval Research, and the National Science Foundation (BCS-1441502, CAREER PHY-1554488, and BCS-1631550). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

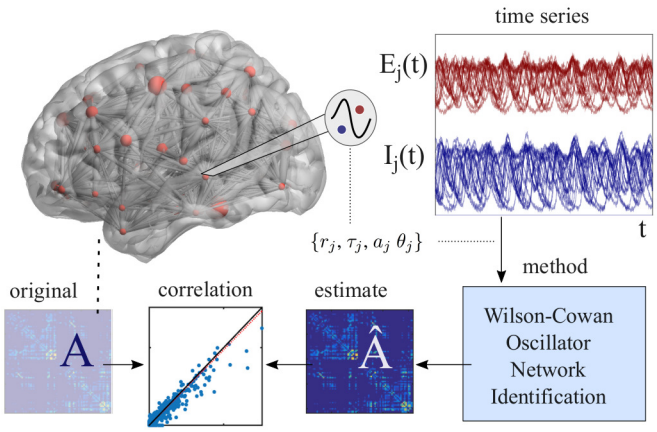

Fig. 1 *Problem overview*. Brain dynamics is modeled as a network of Wilson-Cowan oscillators. Excitatory rates and inhibitory rates of activity $E_j(t)$ and $I_j(t)$ are observed at discrete time instants t_k for each node j in the network. The problem consists of producing estimate $\hat{A}$ and local connectivity $\hat{c}_{1,j}, \ldots, c_{1,j}$ of the original parameters, given the time series observations and knowledge of local oscillator parameters $\{r_j, \tau_j, a_j\ \theta_j\}$.

in basic neurophysiology and incorporates fundamental prior knowledge of the brain's anatomic organization.

For network neuroscientists, fitting functional brain dynamics to neurophysiologically-relevant, dynamical neural mass models would improve interpretation of what it means for brain regions to be functionally connected, beyond mathematical intuition. However, a major technical challenge has been the coupled complexity of the large-scale and nonlinear nature of the models. Our method addresses these problems by applying an inverse transformation on the dynamic equations so that the transformed equations are linearized and become amenable to large-scale estimation methods.

While our method requires observation of excitatory and inhibitory activity in neural signal recordings, made possible by recent advancements in optogenetics [9], integrating this information to infer brain connectivity is especially important for diagnosing and understanding pathophysiology in brain network disorders such as Parkinson's, epilepsy, or schizophrenia, in which imbalances in excitation and inhibition affect functional connectivity [10, 11, 12, 13].

II. Preliminaries and Problem Statement

In this section, we review the dynamic equations of the Wilson-Cowan model and formalize the network identification as an estimation problem. In our description, we adopt standard terminology and notation[1].

Networks of Wilson-Cowan oscillators

The Wilson-Cowan model [8] describes firing rate of two types of neuronal populations in a cortical micro-column, i.e. a cylindrical portion of cortical tissue. The influence that each micro-column exerts in the firing rate of another through synaptic coupling is captured by network edges, which can be modeled as a weighted additive coupling between excitatory populations [14, 15]. More formally, for a network with n nodes, we define $E_j(t)$ (respectively $I_j(t)$) the proportion of excitatory (resp. inhibitory) cells at column $j = 1, \ldots, n$ which are firing per unit time, at the instant t. Likewise, let $P_j(t)$ (resp. $Q_j(t)$) be the external stimulus to column j at time t (e.g. incoming from the thalamus [16]). The coupling weight between node l and node j is denoted A_{jl}, and the connectivity within each column is captured by the coefficients $c_{1,j}, \ldots, c_{4,j}$. The dynamical response of nodes $j = 1, \ldots, n$ is described by a system of $2n$ nonlinear ordinary differential equations, as

$$\tau_{e,j}\frac{dE_j(t)}{dt} = -E_j(t) + (r_{e,j} - E_j(t))\times S_e(c_{1,j}E_j(t) - c_{2,j}I_j(t) + \sum_{l=1,l\neq j}^{n} A_{jl}E_j(t) + P_j(t)), \quad (1)$$

$$\tau_{i,j}\frac{dI_j(t)}{dt} = -I_j(t) + (r_{i,j} - I_j(t))\times S_i\left(c_{3,j}E_j(t) - c_{4,j}I_j(t) + Q_j(t)\right), \quad (2)$$

where $\tau_{e,j}$ (resp. $\tau_{i,j}$) are time constants, and $r_{e,j}$ (resp. $r_{i,j}$) encode the fraction of refractory neurons available to fire associated with the excitatory (inhibitory) populations. Typical values and the biophysical interpretation of these parameters are given in [14]. The last element in the description is the sigmoidal function $S_e(x)$ (resp. $S_i(x)$), which models the effect of saturation in the firing rate of each column due to the neuronal refractory period. It is a monotonically increasing function of x on the interval $(-\infty, \infty)$, attaining the asymptotic values 0 and 1 as x approaches $-\infty$ and ∞. Any function $S(x)$ fulfilling these conditions is considered a valid sigmoidal function [8]. In this work we adopt the following definition:

$$S(x) := \frac{1}{1+e^{-a(x-\theta)}} - \frac{1}{1+e^{a\theta}},$$

where parameter a is associated with the maximum rate of the function, and θ gives the position of the maximum slope. Note that the subtracting term in the sum is chosen such that $S(0) = 0$.

Problem Statement

Addressing our problem from a signal processing standpoint, we assume that samples of the dynamical system variables are collected at regular time instants $t_k := kT$, $k = 1, \ldots, m$, where $T := 1/f_s$ is the sampling interval associated with sampling frequency f_s. We therefore define the data associated with the problem as

- the *time series dataset* $\mathcal{T}$, consisting of a batch of data samples $E_j(t_k) := E_j[k]$, $I_j(t_k) := I_j[k]$ and $P_j(t_k) := P_j[k]$, for samples $k = 1, \ldots, m$, and nodes $j = 1, \ldots, n$.
- the *set of local oscillator parameters* $\mathcal{P}$, consisting of values $r_{e,j}, \tau_{e,j}, a_{e,j}, \theta_{e,j}$, (resp. $r_{i,j}, \tau_{i,j}, a_{i,j}, \theta_{i,j}$), for nodes $j = 1, \ldots, n$.

The identification of the network can now be formalized as an estimation problem, as follows:

[1]**Notation**: We denote by $\mathbf{x}$ a column vector in $\mathbb{R}^n$ and by x_i its i-th entry. Bold capital letters are reserved for matrices, for instance $\mathbf{X} \in \mathbb{R}^{m\times n}$, where $[\mathbf{X}]_{ij}$ denotes the entry in the i-th row and j-th column of $\mathbf{X}$. An indexed vector $\mathbf{x}_i$ (respectively matrix $\mathbf{X}_i$) denotes an element in an indexed set of vectors (respectively, matrices). The $n \times n$ identity matrix is denoted by $\mathbf{I}_n = \text{diag}(\mathbf{1}_n)$, where $\text{diag}(\mathbf{x})$ is a diagonal matrix having the entries of $\mathbf{x}$ in its diagonal , and $\mathbf{1}_n$ is the $n \times 1$ vector of ones. The vectorization operator $\text{vec}(\mathbf{X}) : \mathbb{R}^{m\times n} \to \mathbb{R}^{mn}$ vertically concatenates the columns of $\mathbf{X}$ onto a vector $\mathbf{x} \in \mathbb{R}^{mn}$. The Frobenius norm of a matrix is defined as $\|\mathbf{X}\|_F = (\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2)^{\frac{1}{2}}$.

Problem. *Given the time series dataset $\mathcal{T}$ and the set of oscillator parameters $\mathcal{P}$, provide estimates $\hat{A}_{jl}$ of the ground truth network weights A_{jl}, such that the error $\varepsilon_A(\hat{A}_{jl}, A_{jl}) := \sum_{j,l=1}^{n}(\hat{A}_{jl} - A_{jl})^2$ is minimized. Additionally, provide estimates $\hat{c}_{1,j}, ..., \hat{c}_{4,j}$ of the intra-column connectivity parameters $c_{1,j}, \dots, c_{4,j}$, such that the error $\varepsilon_c(\hat{c}_{lj}, c_{lj}) := \sum_{j=1}^{n}\sum_{l=1}^{4}(\hat{c}_{lj} - c_{lj})^2$ is minimized.* The problem is illustrated in Figure 1.

III. Method

We start with a high-level description of the method, as follows. In the first step, we perform an empirical estimation of the derivatives of the dynamic quantities $E_j(t)$ and $I_j(t)$ based on their respective time samples. Next, given these estimates and the data associated with the problem, we apply an inverse transformation on the dynamic equations, which allows us to obtain a set of equations that is linear in the parameters to be estimated. We then assemble this set of equations in matrix form and propose an optimization problem that seeks to minimize the time series reconstruction error over the set of allowable parameters values, i.e., our optimization variables. Finally, to solve the resulting optimization problem, we resort to a modern algorithm that is applicable to large-scale constrained convex problems. Each of these steps is described in further detail below.

Derivative estimation

We adopt the recent approach proposed in [17] due to its optimal statistical characteristics. It consists of a local polynomial fitting method that achieves a consistent and minimum variance estimation of the derivative by taking a weighted combination of p symmetric differences of the time samples. Formally, we denote the estimates of the derivative of the excitatory and inhibitory rates as $E'_j[k] := dE(t_k)/dt$ (resp. $I'_j[k] := dI_j(t_k)/dt$). For a given signal $x[k]$, its first order derivative estimate $\hat{x}'[k]$ is defined as

$$\hat{x}'[k] = \sum_{h=1}^{p} w_h \cdot \left(\frac{x[k+h] - x[k-h]}{t[k+h] - t[k-h]} \right), \tag{3}$$

whose optimal weights w_h [17, Appendix A] are given by

$$w_h = \frac{6h^2}{p(p+1)(2p+1)}, \quad h = 1, \dots, p. \tag{4}$$

Therefore, applying (3) to the time series dataset $E_j[k]$ and $I_j[k]$ we obtain the derivative estimates $\hat{E}'_j[k]$ and $\hat{I}'_j[k]$ for all $j = 1, \dots, n$.

Inverse sigmoidal transformation

In addressing our estimation formulation, we first note the nonlinear dependency between the dynamic equation (1) and the network weights A_{jl} through the sigmoidal function S_e. This relationship would, in principle, imply a hard, nonconvex, optimization problem to be solved in order to perform the estimation of the parameters. Notwithstanding, as developed next, knowledge of the observations $E_j[k]$ and $P_j[k]$ (resp. $I_j[k]$ and $Q_j[k]$), parameters $\tau_{e,j}$ and $r_{e,j}$ (resp. $\tau_{e,j}$ and $r_{e,j}$), and of the derivative estimates $\hat{E}'[k]$ (resp. $\hat{I}'[k]$), will allow us to perform a functional transformation yielding an estimation problem that will be *linear* in the network weights and local coupling parameters. Considering the dynamic equations of the excitatory nodes (1), we introduce the auxiliary *inverse node dynamics* data $y_j[k]$ for the excitatory population, which we define as:

$$y_j[k] := S_e^{-1}\left(\frac{\tau_{e,j}\hat{E}'_j[k] + E_j[k]}{r_{e,j} - E_j[k]} \right) - P_j[k], \tag{5}$$

by applying the inverse sigmoidal function $S_e^{-1}(S_e(x)) = x$. This inverse function is assured to exist by the monotonicity property of S_e. With $y_j[k]$, the dynamic equations, for all samples $k = 1, \dots, m$ and nodes $j = 1, \dots, n$, can now be written as linear equations in the network weights as

$$y_j[k] = \hat{c}_{1,j}E_j[k] - \hat{c}_{2,j}I_j[k] + \sum_{l=1, l\neq j}^{n} A_{jl}E_j[k] \tag{6}$$

which, when taking the parameters $c_{1,j}$, $c_{2,j}$, A_{jl}, $A_{jj} := 0$, will constitute an overconstrained systems of linear equations when the number of linear independent equations associated with the m measurements is greater than the $n + 1$ parameters that are to be estimated per node $j = 1, \dots, n$. Similarly, an analogous set of linear equations can be obtained by applying an inverse sigmoidal transformation on the set of inhibitory dynamic equations in (2), which relate the time series dataset to the local coupling parameters $c_{3,j}$ and $c_{4,j}$. In the following sections, we focus on the excitatory equations and parameters, since the inhibitory counterpart can be treated as a sub-case of the former.

Estimation of network weights

We now proceed to the main step in the identification of the network, the estimation of network weights. We will use the class of M-estimators [18], i.e., estimators that are obtained as the minima of sums of functions of the data, of which least-squares estimators are a special case. Introducing the estimates $\hat{c}_{1,j}$, $\hat{c}_{2,j}$, $\hat{A}_{jl}$ associated with the parameters $c_{1,j}$, $c_{2,j}$, A_{jl} and accounting for estimation errors $\delta_j[k]$, we rewrite (6) equivalently as

$$y_j[k] = \hat{c}_{1,j}E_j[k] - \hat{c}_{2,j}I_j[k] + \sum_{l=1, l\neq j}^{n} \hat{A}_{jl}E_j[k] + \delta_j[k]. \tag{7}$$

The mn equations defined by (7) can be put in matrix form, as described next. We define the data matrices

$$\mathbf{Y} \in \mathbb{R}^{n\times m}, [\mathbf{Y}]_{il} := y_i[l], \qquad \mathbf{E} \in \mathbb{R}^{n\times m}, [\mathbf{E}]_{il} := E_i[l],$$
$$\mathbf{N} \in \mathbb{R}^{n\times m}, [\mathbf{N}]_{il} := I_i[l], \qquad \mathbf{\Delta} \in \mathbb{R}^{n\times m}, [\mathbf{\Delta}]_{il} := \delta_i[l],$$

as well as the network weights matrix $\hat{\mathbf{A}} \in \mathbb{R}^{n\times n}$ and the vectors of local weights $\hat{\mathbf{c}}_\mathbf{1}, \hat{\mathbf{c}}_\mathbf{2} \in \mathbb{R}^n$. The set of mn equations can now be compactly written as

$$\mathbf{Y} = \mathrm{diag}(\hat{\mathbf{c}}_1)\,\mathbf{E} - \mathrm{diag}(\hat{\mathbf{c}}_2)\,\mathbf{N} + \hat{\mathbf{A}}\,\mathbf{E} + \mathbf{\Delta}. \tag{8}$$

As mentioned previously, the estimation of these parameters can be cast as an optimization problem. In our case, as the main optimization criteria, we minimize the squared Euclidean norm of the error matrix, i.e., $\|\mathbf{\Delta}\|_F^2$. Further, the optimization framework allows us to encode additional assumptions on the network as optimization objectives and constraints. For example, the assumption

of network sparsity can be included by penalizing $\|\text{vec}(\hat{\mathbf{A}})\|_1$, which is the ℓ_1 norm of the vector containing the elements of $\mathbf{A}$. Similarly, one can also choose to penalize larger coefficients of network weights by adding the ℓ_2 norm penalty $\|\text{vec}(\hat{\mathbf{A}})\|_2^2$. Elementwise constraints on the weight values can be also directly encoded, such as positive and upper bounded coefficients $0 \leq \mathbf{A}_{lj} \leq A_{max}$, or network-wide enforcement of symmetry, i.e., $\mathbf{A} = \mathbf{A}^T$. The resulting optimization problem including these factors can be now written in full as

$$\begin{aligned} \underset{\hat{\mathbf{c}}_1,\,\hat{\mathbf{c}}_2,\,\hat{\mathbf{A}}}{\text{minimize}} \quad & \| \mathbf{Y} - \hat{\mathbf{A}}\,\mathbf{E} - \text{diag}(\hat{\mathbf{c}}_1)\,\mathbf{E} - \text{diag}(\hat{\mathbf{c}}_2)\,\mathbf{N} \|_F^2 \\ & + \lambda_1 \frac{m}{n} \|\text{vec}(\hat{\mathbf{A}})\|_1 + \lambda_2 \frac{m}{n} \|\text{vec}(\hat{\mathbf{A}})\|_2^2 \\ \text{subject to} \quad & \mathbf{A} = \mathbf{A}^T; \\ & 0 \leq [\mathbf{A}]_{jl} \leq A_{max}, \qquad j,l = 1,\ldots,m; \\ & [\mathbf{A}]_{jj} = 0, \qquad\qquad\quad j = 1,\ldots,m. \end{aligned} \tag{9}$$

where $(m/n)\lambda_1$ and $(m/n)\lambda_2$ are regularization parameters encoding the relative importance of each penalty, normalized such that their relative importance is independent of specific values of m and n. The optimization problem in (9) is a non-smooth convex quadratic problem with linear constraints, whose solutions yielding a minimum value can be determined efficiently [19].

Joint estimation of network weights and external inputs

Examining equation (1), we note that its dependency on the external stimulus $P_j(t)$ is of the same algebraic nature as the one associated with the network weights A_{jl}, i.e., a linear argument to the sigmoidal function S_e. We can, therefore, extend our method and include $P_j(t)$ as an estimation variable, so as to perform joint inference of network weights and external inputs. As a possible approach, we assume the external inputs to be constant in the period of observation ($k = 1,\ldots,m$), i.e $P_j(t) = p_j$. Therefore, to obtain a linear regression problem, we define our inverse sigmoidal data samples as

$$z_j[k] := S_e^{-1}\left(\frac{\tau_{e,j}\hat{E}'_j[k] + E_j[k]}{r_{e,j} - E_j[k]}\right). \tag{10}$$

By including the optimization data $\mathbf{Z} \in \mathbb{R}^{n\times m}$ and additional variable $\mathbf{p} \in \mathbb{R}^n, [\mathbf{p}]_i = p_i$, the optimization problem can be written in matrix form as:

$$\begin{aligned} \underset{\hat{\mathbf{c}}_1,\,\hat{\mathbf{c}}_2,\,\hat{\mathbf{p}},\,\hat{\mathbf{A}}}{\text{minimize}} \quad & \| \mathbf{Z} - \hat{\mathbf{A}}\,\mathbf{E} - \text{diag}(\hat{\mathbf{c}}_1)\,\mathbf{E} - \text{diag}(\hat{\mathbf{c}}_2)\,\mathbf{N} - \text{diag}(\hat{\mathbf{p}})\,\mathbf{1}_n\mathbf{1}_m^T \|_F^2 \\ & + \lambda_1 \frac{l}{n} \|\text{vec}(\hat{\mathbf{A}})\|_1 + \lambda_2 \frac{l}{n} \|\text{vec}(\hat{\mathbf{A}})\|_2^2 \\ \text{subject to} \quad & \mathbf{A} = \mathbf{A}^T, \\ & 0 \leq [\mathbf{A}]_{jl} \leq A_{max}, \\ & [\mathbf{A}_{jj}] = 0, \quad l,j = 1,\ldots,m. \end{aligned} \tag{11}$$

which is an optimization problem of the same type as the one in (9), allowing for the same type of optimization algorithms.

Large-scale optimization

The solution to optimization problems in (9) and (11) can be readily computed using standard quadratic solvers such as [20]. Typical problem instances, however, involve matrices of size mn, with

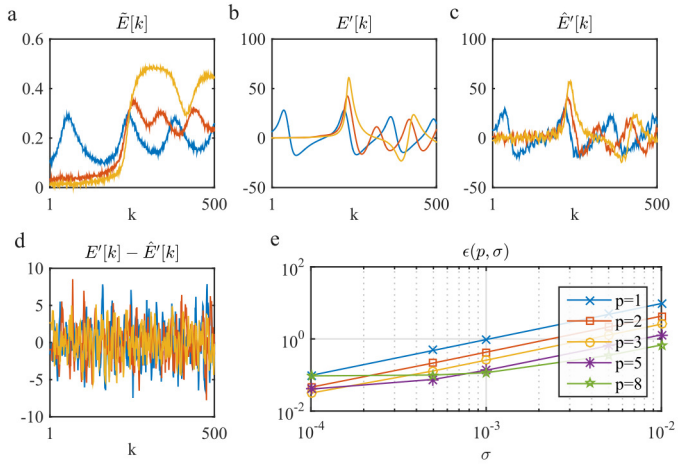

Fig. 2 *Derivative estimation.* In (a) the noisy samples $\tilde{E}[k]$, in (b) the ideal derivatives $E'[k]$, in (c) the estimated derivatives $\hat{E}'[k]$, and in (d) the difference $\hat{E}'[k] - E'[k]$, all from a segment of 500 samples and a subset of 3 representative excitatory nodes, for $p = 5$ and $\sigma = 1e-3$. In (e) we present the behavior of the empirical relative error $\epsilon(p,\sigma)$ for different noise levels σ and derivative orders p.

number of nodes $n = 83$ and samples $m = 10\,000$. Given that one might need to run repeated experiments to determine optimal values for hyper-parameters λ_1 and λ_2 (i.e., compute the regularization path), as well as evaluate multiple datasets, the overall computing time taken might become a relevant factor. An effective solution is to employ proximal splitting optimization algorithms, such as the Alternating Direction Method of Multipliers [21, 22]. These methods provide gains of efficiency by allowing, for example, the use of *factorization caching* and *warm-start*. With factorization caching, the algorithm maintains intermediate computation results in memory across iterations, such as the matrix factorizations involved with the quadratic objective function. With warm-start, intermediate results are re-used across problem instances, being applicable when the optimal regularization hyper-parameters λ_1 and λ_2 are computed. By applying these algorithms in our experiments, the typical type for the solution of one problem instance decreased at least one order of magnitude, from over 10 minutes to a few seconds, when compared to general purpose solvers [20].

IV. Computational Experiments

We now evaluate the performance of the proposed method under relevant situations. We first evaluate the derivative estimation with respect to sensitivity to noise. Next, we apply the full method to estimate the unknown weights of a network of 83 oscillators, whose time series are simulated using edge weights obtained from actual diffusion imaging measurements. Finally, we perform an experiment where we jointly estimate network weights and the magnitude of external input at each node.

For all simulations, the following parameters were applied. The time series was obtained by numerically integrating the system of differential equations for excitatory and inhibitory populations for $n = 83$ using the Runge-Kutta 4^{th} order method [23] and initial conditions $E_j(0) = 0, I_j(0) = 0, j = 1,\ldots,n$. The generated times series was sampled at 5 KHz up to the final time of 2 s, generating 10 000 samples for each node. Different levels of noise σ were applied to the time series following a normal distribution $\mathcal{N}(0, \sigma I_n)$.

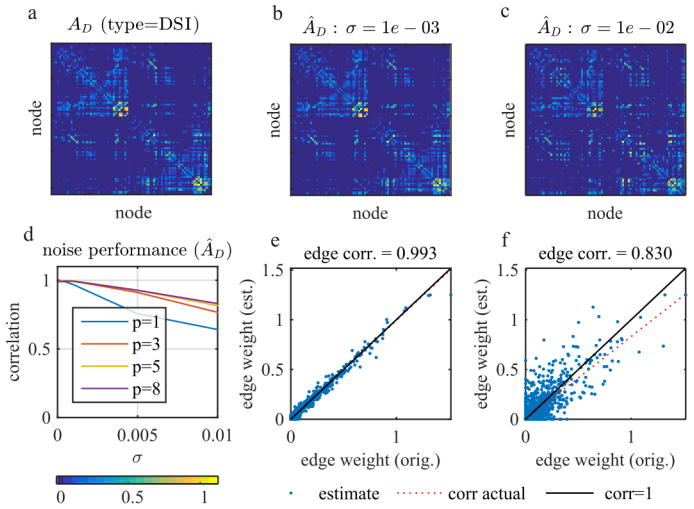

Fig. 3 *Recovered network.* In (a) the network adjacency matrix A_D used as ground truth in our computational experiments. In (b) and (c) we plot the adjacency matrices $\hat{A}_D$ with the estimated network weights for $\sigma = 1e-2$ (b) and $\sigma = 1e-2$ (c). Correspondingly, in (e) and (f) we illustrate the correlation achieved in each case. In (c) we plot the correlation between original and estimated network weights for different noise levels σ and derivative estimation parameter p.

Derivative estimation

In this subsection, we evaluate the performance of the derivative estimation for different levels of the parameter p and noise level σ. The experiment consists of providing a noisy time series $\tilde{E}_j[k] := E_j[k] + w_j[k]$ with $w_j[k] \sim \mathcal{N}(0, \sigma)$ to the derivative estimation method, and evaluating the resulting error in the derivative estimation. This error is defined as the relative Euclidean norm of the difference between derivative estimate $\hat{E}'[k]$ (obtained from observing the noisy time series) and the original derivative value $E'[k]$, which can be directly obtained from the Wilson-Cowan equations at each time sample. The empirical relative error $\epsilon(p, \sigma)$ over all the nodes in the network is defined as

$$\epsilon(p,\sigma) := \sqrt{\frac{\sum_{j=1}^{n}\sum_{k=1}^{m}(\hat{E}'_j[k] - E'_j[k])^2}{\sum_{j=1}^{n}\sum_{k=1}^{m}(E'_j[k])^2}}. \qquad (12)$$

In Figure 2 (a-d) we present, respectively, the noisy samples $\tilde{E}[k]$, the ideal derivatives $E'[k]$, the estimated derivatives $\hat{E}'[k]$, and the difference $\hat{E}'[k] - E'[k]$, from a segment of 500 samples and a subset of 3 representative excitatory nodes, for $p = 5$ and $\sigma = 1e-3$. Qualitatively, by a visual comparison between (b) and (c), one can argue that the major features of the derivative signal are being successfully captured by the derivative estimation algorithm. Quantitatively, we summarize in (e) the behavior of the empirical relative error $\epsilon(p, \sigma)$ for different noise levels σ and derivative orders p. In particular, looking at the result for higher values of p (e.g. $p = 8$), one can observe that higher parameter values are preferred when estimation is performed at higher noise levels. At lower noise levels, however, the performance for high p values is overcome by the choice of lower values, as can be seen for $\sigma = 1e-4$. Such effects can be attributed to the low-pass effect of the averaging operations induced by taking higher values of p, as described in (3).

Recovery at different noise levels

We now proceed to the evaluation of the method with regard to its main aspect, i.e., the inference of network weights given the observation of time series at each node. For that purpose, we generate a simulated time series of excitatory and inhibitory rates in a system with $n = 83$ nodes. The underlying network weights are derived from empirically collected measurements of the density of white matter streamlines connecting different brain regions, which were obtained by processing diffusion spectrum imaging (DSI) data acquired from a healthy individual, as described in [15, Section 'Methods']. The weighted adjacency matrix A_D associated with the network weights is presented in Figure 3 (a), where one can see the occurrence of a few clusters presenting higher weight values. To perform the estimation of network weights, we first obtained the derivative estimates $\hat{E}'_j[k]$, adopting $p = 8$. With the values of $\hat{E}'_j[k]$, $E_j[k]$, $I_j[k]$, $P_j[k]$, as well as of parameters $r_{e,j}$ and $\tau_{e,j}$, we applied the inverse transformation described in equation (3) to obtain $y_j[k]$ for each j and k. Given those, we built the regression matrices $\mathbf{Y}, \mathbf{E}$ and $\mathbf{N}$ to solve the optimization problem in (9) and obtained the estimates $\hat{A}$ for the network weights. In Figure 3 (b) and (c) we plot the adjacency matrices $\hat{A}_D$ with the estimated network weights for $\sigma = 1e-2$ (b) and $\sigma = 1e-2$ (c). Correspondingly, in (d) and (e) we illustrate the correlation achieved in each case, by plotting the original and estimated edge weights for all $\binom{83}{2} = 3403$ values. In particular, the correlation achieved for the lower noise value ($\sigma = 1e-3$) is remarkably high, as can be observed qualitatively in (b) and quantitatively in (d). For the higher noise case, correlation is still significant, and although the scatter plot in (e) presents more widespread values, one can observe in (c) that the most prominent features of the network, such as its overall cluster organization, have been largely preserved. In (c) we summarize the correlation achieved for different noise levels σ and derivative estimation parameter p. It can be seen that higher values of the parameter p offer a better performance at higher noise levels, whereas at low noise values, performance is very high for all values of p considered.

Joint estimation

Lastly, we investigate the performance of the method when both network weights and the external input magnitudes are jointly subject to estimation. We proceeded with steps similar to those described in the previous subsection to obtain the regression matrices $\mathbf{Z}, \mathbf{E}$ and $\mathbf{N}$ with the transformed data $z_j[k]$ being now obtained according to (10). We then solved the optimization problem (11) to obtain estimates $\hat{A}_D$ of network weights and $\hat{p}$ of the external input magnitudes. The results are presented in Figure 4, with (a) displaying the recovered network weights at noise level $\sigma = 1e-3$, and in (c) the corresponding scatter plot, illustrating the correlation level between original and estimated edge weights. In (b) we present the original and estimated input magnitude values for each node, and in (d) we illustrate their correlation. The correlation level achieved in both network weights and input magnitudes is high, confirming the analysis presented in the corresponding Method subsection.

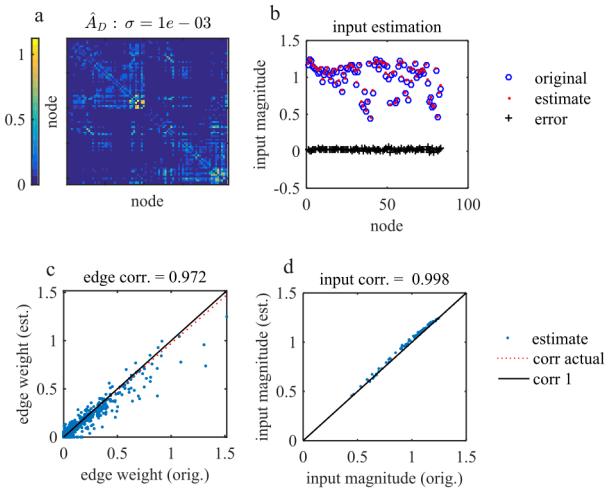

Fig. 4 *Joint estimation with external inputs*. In (a) the recovered network weights at noise level $\sigma = 1e-3$ is presented. In (c) the corresponding scatter plot illustrates the correlation level between original and estimated edge weights. In (b) we present the original and estimated input magnitude values for each node, and in (d), their correlation.

V. Conclusion and Future Work

In this study, we developed a novel method for inferring structural brain connectivity from excitatory and inhibitory neuronal dynamics generated from interconnected Wilson-Cowan oscillators –a well-reputed model of neural activity in cortical columns. In our experiments, we simulated excitatory and inhibitory dynamics on an empirically-derived structural brain network of white-matter connections, and demonstrated that our method could accurately recover structural brain connectivity in the presence of observation noise. Although the method relies on the knowledge of individual oscillator parameters and on the separate observation of excitatory and inhibitory rates, we believe that it presents a principled contribution to the challenge of uncovering information from structural brain networks underlying dynamic function. As future work, we plan to relax the assumptions needed for the method to be applicable to clinical electrophysiological measurements. Our methodological contributions may also have translational impact in the diagnosis of brain network disorders, such as epilepsy, in which dysfunction is driven by imbalances in excitatory and inhibitory neural populations.

VI. References

[1] Edward T. Bullmore and Danielle S. Bassett, "Brain graphs: graphical models of the human brain connectome.," *Annual review of clinical psychology*, vol. 7, no. 1, pp. 113–140, apr 2011.

[2] H J Park and Karl Friston, "Structural and functional brain networks: from connections to cognition," *Science (New York, N.Y.)*, vol. 342, no. 6158, pp. 1238411, 2013.

[3] Bratislav Mišić, Richard F. Betzel, Marcel A. de Reus, Martijn P. van den Heuvel, Marc G. Berman, Anthony R. McIntosh, and Olaf Sporns, "Network-Level Structure-Function Relationships in Human Neocortex," *Cerebral Cortex*, vol. 26, pp. 3285–3296, 2016.

[4] R. Matthew Hutchison, Thilo Womelsdorf, Elena A. Allen, Peter A. Bandettini, Vince D. Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H. Duyn, Gary H. Glover, Javier Gonzalez-Castillo, Daniel A. Handwerker, Shella Keilholz, Vesa Kiviniemi, David A. Leopold, Francesco de Pasquale, Olaf Sporns, Martin Walter, and Catie Chang, "Dynamic functional connectivity: Promise, issues, and interpretations," *NeuroImage*, vol. 80, pp. 360–378, oct 2013.

[5] Mahyar Fazlyab and Victor M Preciado, "Robust topology identification and control of lti networks," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*. IEEE, 2014, pp. 918–922.

[6] Shahin Shahrampour and Victor M Preciado, "Topology identification of directed dynamical networks via power spectral analysis," *IEEE Transactions on Automatic Control*, vol. 60, no. 8, pp. 2260–2265, 2015.

[7] Vernon Mountcastle, "An organizing principle for cerebral function: the unit model and the distributed system," in *The Mindful Brain*, Gerald M. Edelman and Vernon B. Mountcastle, Eds. MIT Press, Cambridge, Mass., 1978.

[8] Hugh R Wilson and Jack D Cowan, "Excitatory and inhibitory interactions in localized populations of model neurons," *Biophysical journal*, vol. 12, no. 1, pp. 1, 1972.

[9] O. Yizhar, L. E. Fenno, M. Prigge, F. Schneider, T. J. Davidson, D. J. O'Shea, V. S. Sohal, I. Goshen, J. Finkelstein, J. T. Paz, K. Stehfest, R. Fudim, C. Ramakrishnan, J. R. Huguenard, P. Hegemann, and K. Deisseroth, "Neocortical excitation/inhibition balance in information processing and social dysfunction," *Nature*, vol. 477, no. 7363, pp. 171–178, Sep 2011.

[10] M. C. Ridding, R. Inzelberg, and J. C. Rothwell, "Changes in excitability of motor cortical circuitry in patients with Parkinson's disease," *Ann. Neurol.*, vol. 37, no. 2, pp. 181–188, Feb 1995.

[11] J. Lisman, "Excitation, inhibition, local oscillations, or large-scale loops: what causes the symptoms of schizophrenia?," *Curr. Opin. Neurobiol.*, vol. 22, no. 3, pp. 537–544, Jun 2012.

[12] Premysl Jiruska, Marco de Curtis, John G. R. Jefferys, Catherine A. Schevon, Steven J. Schiff, and Kaspar Schindler, "Synchronization and desynchronization in epilepsy: controversies and hypotheses.," *The Journal of physiology*, vol. 591, no. Pt 4, pp. 787–97, nov 2013.

[13] Ankit N. Khambhati, Kathryn A. Davis, Brian S. Oommen, Stephanie H. Chen, Timothy H. Lucas, Brian Litt, and Danielle S. Bassett, "Dynamic Network Drivers of Seizure Generation, Propagation and Termination in Human Neocortical Epilepsy," *PLOS Computational Biology*, vol. 11, no. 12, pp. e1004608, dec 2015.

[14] ShiNung Ching, Emery N Brown, and Mark A Kramer, "Distributed control in a mean-field cortical network model: Implications for seizure suppression," *Physical Review E*, vol. 86, no. 2, pp. 021920, 2012.

[15] Sarah Feldt Muldoon, Fabio Pasqualetti, Shi Gu, Matthew Cieslak, Scott T Grafton, Jean M Vettel, and Danielle S Bassett, "Stimulation-based control of dynamic brain networks," *arXiv preprint arXiv:1601.00987*, 2016.

[16] G Bard Ermentrout and David H Terman, *Mathematical foundations of neuroscience*, vol. 35, Springer Science & Business Media, 2010.

[17] Kris De Brabanter, Jos De Brabanter, Bart De Moor, and Irène Gijbels, "Derivative estimation with local polynomial fitting," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 281–301, 2013.

[18] Peter J Huber, *Robust statistics*, Springer, 2011.

[19] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[20] Michael Grant, Stephen Boyd, and Yinyu Ye, "Cvx: Matlab software for disciplined convex programming," 2008.

[21] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[22] Neal Parikh, Stephen P Boyd, et al., "Proximal algorithms.," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[23] John C Butcher, *Numerical methods for ordinary differential equations*, John Wiley & Sons, 2008.