Network Identification with Latent Nodes via Auto-Regressive Models

Erfan Nozari Yingbo Zhao Jorge Cortés

***Abstract*—We consider linear time-invariant networks with unknown topology where only a *manifest* subset of the nodes can be directly actuated and measured while the state of the remaining *latent* nodes and their number are unknown. Our goal is to identify the transfer function of the manifest subnetwork and determine whether interactions between manifest nodes are direct or mediated by latent nodes. We show that, if there are no inputs to the latent nodes, the manifest transfer function can be approximated arbitrarily well in the H_∞-norm sense by the transfer function of an auto-regressive model and present a least-squares estimation method to construct the auto-regressive model from measured data. We show that the least-squares auto-regressive method guarantees an arbitrarily small H_∞-norm error in the approximation of the manifest transfer function, exponentially decaying once the model order exceeds a certain threshold. Finally, we show that when the latent subnetwork is acyclic, the proposed method achieves perfect identification of the manifest transfer function above a specific model order as the length of the data increases. Various examples illustrate our results.**

I. Introduction

Network reconstruction problems are widespread in many areas of science and engineering. In systems biology, for instance, genetic network identification uses data from RNA micro-array experiments to identify the interaction pattern between genes in a regulatory network [2], [3]. In neuroscience, researchers seek to understand how different regions of the brain cooperate with each other by having subjects perform certain goal-directed tasks while measuring their brain activity via multi-channel recordings such as electroencephalograms (EEG) [4]–[8]. Similar examples exist in other areas including finance, social networks, and physics. Roughly speaking, the objective in network identification is to determine causal relationships among the nodes in the network that model the direction and strength of the interactions between them. While network control and coordination has made much progress on problems where the interaction topology is either given or the design objective itself, not so much attention has been devoted to develop techniques to address the identification of unknown topologies from measured data. The need for the latter is especially apparent in the context of complex, large-scale networks, where it is often not possible to measure or actuate all nodes, or even know their number. In this paper, we seek to contribute to this body of work by studying the effect that the presence of unmeasured nodes has on the identification of networked linear systems with arbitrary topology.

Literature review: An increasing number of works study topology identification problems to better understand the interactions in large-scale networks and their role in determining the network behavior. A complex network is commonly represented as a directed graph, and the interactions among neighboring nodes are represented by directed edges whose weights reflect the interaction strength. In this sense, topology identification aims at identifying the adjacency matrix of the network graph [9] or its Boolean structure [10]. The work [11] studies the complete characterization of the interaction topology of consensus-type networks using a series of node-knockout experiments, where nodes are sequentially forced to broadcast a zero state without being removed from the network. The work [12] also uses node-knockout experiments to identify the topology of directed linear time-invariant networks relying on the cross-power spectral densities of the network response to wide-sense stationary noise. The work [13] presents a method to infer the topology of a network of coupled phase oscillators from its stable response dynamics, assuming that one can manipulate every individual node and perform large number of experiments. In general, without such assumption, it is difficult or impossible, depending on the additional structural information available, to accurately identify the topology of a general network. As a result, a main focus has been on particular network realizations that explain the measured data, such as the sparsest realization, sometimes with a design parameter to manage the trade-off between model accuracy and sparsity, see e.g., [3], [14]. Along these lines, the work [15] considers the identification of networked linear systems with tree topologies. The above-referenced works rely on knowledge of the number of nodes in the network. However, it is often impossible to sample the state of all nodes, or even know the existence of some of them. The work [16] studies the problem of learning latent tree graphical models where samples are available only from a subset of the nodes, and proposes computationally efficient algorithms for learning trees without any redundant hidden nodes. The work [17] proposes a method to identify the latent nodes and consistently reconstruct the topology under the assumptions that the network is a polytree and the degree of each latent node is at least three, with out-degree of at least two. Unlike the topology identification algorithms proposed in [15], [17], our approach here allows for the possibility of cycles in the network topology. Using the notion of the dynamical structure function of a network with latent nodes [18], the work [19] proposes a convex

A preliminary version appeared as [1] at the American Control Conference.

E. Nozari and J. Cortés are with the Department of Mechanical and Aerospace Engineering, University of California at San Diego, La Jolla, CA 92093, {enozari,cortes}@ucsd.edu. Y. Zhao is with Cymer Corporation, San Diego, CA 92127, yingbo.zhao@asml.com

optimization-based approach to find the best Boolean structure among manifest nodes which consists of computing and comparing the distance between an estimated transfer matrix or data to all possible Boolean structures. The problem of minimal state-space realization of a given dynamical structure function was further studied in [20]. In the present work, however, we use a least-square autoregressive identification approach to identify not only whether a pair of manifest nodes are dynamically connected, but also whether this connection is direct or indirect (latent-mediated) and, in the latter case, the length of the shortest path between the two. Recent work has employed sparse plus low-rank (S+L) decomposition to identify general graphical models (with the possibility of cycles) with latent variables for static [21] and dynamic [22] models. The present paper has two main differences with respect to this work. First, the S+L decomposition assumes that the subnetwork among manifest nodes is sparse and the number of latent nodes is (considerably) smaller than the number of manifest ones, while our method is applicable to arbitrary networks. Second, although the identification procedure of [22] also leads to an auto-regressive (AR) model, it is based on the so-called maximum-entropy covariance extension. This method, with origins in seismic vibrations and human voice analysis, seeks to *maximize* the prediction error [23] (while our approach seeks to *minimize* it), leading to very different models. Finally, our work is inspired by the wide use in neuroscience of AR models to analyze brain data via Granger causality and its variants and the study of effective connectivity among different areas of the brain, see e.g., [5], [6], [24]. The Granger causality measure is a mainly descriptive tool that captures influence and interconnection among time series. A popular variant of Granger causality, direct directed transfer function (dDTF) [8], [25] distinguishes between direct and indirect interconnections between two nodes by multiplying the directed transfer function (DTF, the normalized transfer function between the two nodes) by the partial coherence between them in the frequency domain. We are motivated here by understanding to what extent the reconstruction results obtained via methods that build on Granger causality are sensitive to the presence of latent nodes. Furthermore, we propose a method using (multivariate) AR models for networks with latent nodes that distinguishes between direct and indirect (i.e., latent-mediated) interconnections between two nodes in the time domain based on the order of the interconnection between them.

Statement of contributions: We consider a scenario where one can only directly actuate and measure a subset of the nodes, termed manifest, of a large linear time-invariant network whose total number of nodes and interaction topology are unknown. The objective is to identify the manifest transfer function, which is the submatrix corresponding to the manifest nodes of the transfer function matrix of the entire network. To achieve this, we study the transfer functions provided by linear AR models. Our discussion shows how AR models can be used to effectively distinguish direct interactions between manifest nodes from indirect interactions mediated by latent nodes. Our first contribution shows that, if no inputs act on the latent nodes, then there exists a class of AR models whose transfer functions converge exponentially in the H_∞ norm to the manifest transfer function as the model order increases. We also show that, if the latent subnetwork is acyclic, then this approximation is exact above a specific model order. Our second contribution characterizes the properties of using least-squares auto-regressive estimation to construct the AR model from measured data. We establish that the least-squares matrix estimate converges in probability to the optimal matrix sequence identified in our first contribution, enabling us to determine whether two manifest nodes interact directly or indirectly through latent nodes. We also show that the least-squares auto-regressive method guarantees an arbitrarily small H_∞-norm error as the length of data and the model order grow. In fact, once the order of the AR model candidates exceeds a certain threshold, the H_∞-norm error decays exponentially. Finally, we show that, when the latent subnetwork is acyclic, the method achieves perfect identification of the manifest transfer function. Throughout a series of remarks in the paper, we also discuss how our results can be extended to the identification of linear network models of arbitrary order. Simulations on a directed ring network, Erdős–Rényi random graphs, and real EEG data illustrate our results.

Notation: For a vector $x \in \mathbb{R}^n$, we use x_i to denote its i-th element. Given a sequence $\{x(k)\}_{k=0}^{\infty} \subset \mathbb{R}^n$ and $j_1 \leq j_2 \in \mathbb{Z}_{\geq 0}$, we use $\{x\}_{j_1}^{j_2}$ to denote the finite sequence $\{x(j_1), x(j_1+1), \ldots, x(j_2)\}$. We omit j_1 if $j_1 = 0$. We denote $\|\{x\}_{j_1}^{j_2}\| \triangleq \left(\sum_{k=j_1}^{j_2} x^T(k)x(k)\right)^{\frac{1}{2}}$. A sequence of random variables $\{x\}$ converges in probability to a random variable X, denoted $\operatorname{plim}_{k\to\infty} x(k) = X$, if $\lim_{k\to\infty} \Pr(|x(k) - X| \geq \varepsilon) = 0$ for all $\varepsilon > 0$. Accordingly, a sequence of random matrices $\{A\}$ converges to a random matrix A_∞ in probability if $\operatorname{plim}_{k\to\infty} A_{ij}(k) = A_{\infty,ij}$ for all i, j. For a real matrix $M \in \mathbb{R}^{m\times n}$, we denote its singular values in decreasing order as $\sigma_1(M) \geq \sigma_2(M) \geq \cdots \geq \sigma_{\min(m,n)}(M) \geq 0$ and its spectral norm by $\|M\| = \sigma_1(M)$. The max norm of M is $\|M\|_{\max} = \max_{i,j} |M_{ij}|$. We denote by $\rho(M)$ the spectral radius of a square matrix M. The matrix M is Schur stable if and only if $\rho(M) < 1$. We let $\mathbf{0}_{m\times n}$ denote the $m \times n$ matrix with all zero elements and by I_n the identity matrix of dimension $n \times n$. The H_∞-norm of a discrete transfer function T is $\|T\|_\infty \triangleq \sup_{-\pi\leq\omega\leq\pi} \|T(\omega)\|$.

II. Problem formulation

We consider a discrete-time, linear time-invariant (LTI) network dynamics with state-space representation

$$\begin{aligned} x(k+1) &= Ax(k) + u(k), \\ y(k) &= Cx(k), \end{aligned} \tag{1}$$

where $k \in \mathbb{Z}_{\geq 0}$ is the time index, $x(k) \in \mathbb{R}^n$ is the network state (with $x_i(k)$ representing the state of node $i \in \{1, \ldots, n\}$), $u(k) \in \mathbb{R}^n$ is the control input (with $u_i(k)$ acting on node i), and $y(k) \in \mathbb{R}^m$ is the network output. Here, $A \in \mathbb{R}^{n\times n}$ is the adjacency matrix of the network, characterizing the interactions among neighboring nodes, and $C \in \mathbb{R}^{m \times n}$ is the output matrix. Since natural system are usually driven by noise, the input, state, and output sequences are in general stochastic processes over the sample space of noise realizations. For simplicity, the dynamical description (1) assumes that all nodes are of order 1, that is, $x(k+1)$ depends directly only on $x(k)$ and is conditionally independent of $\{x\}^{k-1}$ given $x(k)$. Nevertheless, as we discuss later (see e.g., Remark 3.4), all of the subsequent results are generalizable to systems whose dynamics (in the original "physical" variables) are described by difference equations of order higher than 1.

Even though there is a control input at every node in the network dynamics (1), we do not assume that all the control inputs are user-specified. In fact, in a large-scale network, it is common that one can actuate only a small subset of the nodes due to computational constraints, physical limitations, or cost. A similar observation can be made regarding the number of nodes whose state can be directly measured. For these reasons, here we assume that the nodes of the network are divided into $n_m \leq n$ *manifest* nodes, which can be directly actuated and measured by the user, and $n - n_m$ *latent* nodes, which can neither be directly actuated nor measured by the user. With this distinction, and using a permutation of the indices in $(1, 2, \dots, n)$ if necessary, we can decompose the network and input state as $x = [x_m, x_l]$ and $u = [u_m, u_l]$, respectively, where the subindex 'm' corresponds to manifest nodes and the subindex 'l' corresponds to latent nodes. With this convention, the output matrix takes the form $C = [I_{n_m \times n_m}, \mathbf{0}_{n_m \times (n-n_m)}]$. With the decomposition of the nodes into manifest and latent, the state-space representation (1) becomes

$$\begin{bmatrix} x_m(k+1) \\ x_l(k+1) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_m(k) \\ x_l(k) \end{bmatrix} + \begin{bmatrix} u_m(k) \\ u_l(k) \end{bmatrix}, \\ y(k) = x_m(k). \tag{2}$$

In the remainder of this paper, we consider the network in the relabeled form (2). Fig. 1 illustrates this relabeling procedure (corresponding to a linear transformation) in a ring.

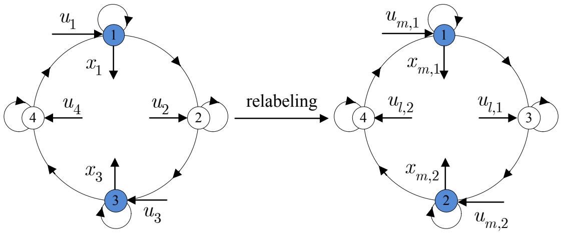


Fig. 1. Node relabeling in a directed ring with 4 nodes. Nodes 1 and 3 are manifest, nodes 2 and 4 are latent. The permutation $(1, 2, 3, 4) \rightarrow (1, 3, 2, 4)$ makes manifest and latent nodes have consecutive indices, as in (2).

Since the focus of this work is on network identification and not stabilization, we make the following standard assumption.

Assumption 2.1: The adjacency matrix of the complete network as well as the latent subnetwork are Schur stable, i.e., $\rho(A) < 1$ and $\rho(A_{22}) < 1$.

Remark 2.2: ***(Direct versus latent interactions).*** The interaction graph of the manifest subnetwork is characterized by A_{11}. In particular, the state of node p affects the state of node q *directly* if and only if the entry on the q-th row and the p-th column, denoted by $A_{11}(q, p)$, is nonzero. However, even if $A_{11}(q, p) = 0$, it is still possible that node p affects node q *indirectly* through some latent nodes. The distinction between direct and indirect connections is an important point to which we come back later in our discussion. □

We refer to a latent node as *passive* if its corresponding input is zero. Throughout the paper, we only deal with passive latent nodes, so that $\{u_l\} \equiv 0$. We make the following assumption on the input to the manifest nodes.

Assumption 2.3: The input $\{u_m\}$ to the manifest subnetwork is a zero-mean stochastic process with independent and identically distributed (i.i.d.) absolutely continuous[1] random vectors $u_m(k)$, with covariance I_{n_m}.

Assumption 2.3 guarantees that $\{u_m\}$ is persistently exciting of arbitrary order and its power spectral density does not vanish at any frequency. Similar assumptions are common in system identification, see e.g., [12], [26]. The zero-mean assumption can be relaxed by assuming a nonzero but known $\mathbb{E}[u_m(k)]$ corresponding to the scenario where the designer injects a deterministic stimulating signal into every manifest node, which itself is subject to the disturbance of a zero-mean white noise. Without loss of generality and for simplicity, we assume $\mathbb{E}[u_m(k)] \equiv \mathbf{0}_{n_m}$.

Given the setup above, our objective is to identify the transfer function $T_{x_m u_m}(\omega)$ of the manifest subnetwork, that is, the transfer function from u_m to x_m, absent any knowledge of the latent nodes.

Problem 2.4: ***(Identification of the manifest transfer function).*** Given the measured data $\{y\}_1^N$, find a linear auto-regressive model of order τ, with $N \gg \tau$, of the form

$$\tilde{x}_m(k+1) = \sum_{i=0}^{\tau-1} \tilde{A}_i \tilde{x}_m(k-i) + u_m(k), \tag{3}$$

such that the associated transfer function $T_{\tilde{x}_m u_m}$ from u_m to $\tilde{x}_m$ and the transfer function $T_{x_m u_m}$ from u_m to x_m in (1) are close in the H_∞-norm, i.e., $\|T_{\tilde{x}_m u_m} - T_{x_m u_m}\|_\infty$ is small.

There are alternative methods to identify the transfer function matrix $T_{x_m u_m}$ besides the AR method in (3). Our adoption here of AR model candidates is motivated by their widespread use in neuroscience to determine causality and interconnections in human brain connectivity models, see e.g., [5]–[7][2]. Equipped with time series data obtained during the performance of a cognitive task, the conventional procedure consists of first estimating an AR model, then computing its associated transfer function matrix, and finally evaluating the Granger causality connectivity measure, or generalizations of it, in the frequency domain. We are particularly motivated by the prospect of understanding the sensitivity of these approaches to the presence of latent nodes corresponding to brain regions that are active during the cognitive task but are not directly measured.

[1] Recall that an absolutely continuous random variable/vector is one that has a probability density function (e.g., Gaussian).

[2] In general, the main advantage of AR models over more general models such as ARMA or BJ is their simplicity, only capturing the internal dynamics and assuming negligible input *noise correlation* (though putting no restriction on input *signal correlation*, which is significant in brain dynamics). As a result, prediction error minimization has a closed-form solution for an AR model while it is non-convex in the ARMA or BJ cases.

III. Asymptotically exact identification of the manifest transfer function

In this section we establish that, given an arbitrary precision, there exists an AR model solving Problem 2.4. More precisely, we show that there exists a sequence of AR models of the form (3) with increasing order whose transfer functions converge to $T_{x_m u_m}$ exponentially in the H_∞ sense. We later show that, if the latent subnetwork is acyclic, then this approximation can be made exact.

We start our discussion with a useful auxiliary result.

Lemma 3.1: ***(Upper bound on*** $\|A_{22}^i\|$***).*** For any Schur stable $A_{22} \in \mathbb{R}^{n_l \times n_l}$ and any $\bar{\rho} \in (\rho(A_{22}), 1)$, there exists $\kappa \in \mathbb{R}_{>0}$ such that $\|A_{22}^i\| \leq \kappa \cdot \bar{\rho}^i$, for all $i \in \mathbb{Z}_{\geq 0}$.

Proof. The result is an immediate consequence of the spectral radius formula $\lim_{i\to\infty} \|A_{22}^i\|^{1/i} = \rho(A_{22})$. ■

We are now ready to state the main result of this section.

Theorem 3.2: ***(AR model whose transfer function converges to the manifest transfer function).*** Consider the LTI network described by (2) where all the latent nodes are passive. For any $\bar{\rho} \in (\rho(A_{22}), 1)$, there exists $\bar{\gamma} \in \mathbb{R}_{>0}$ such that for all $\tau \in \mathbb{Z}_{\geq 0}$, the AR model (3) with

$$\tilde{A}_0^* = A_{11}, \quad \tilde{A}_i^* = A_{12}A_{22}^{i-1}A_{21}, \ i \in \{1, \dots, \tau-1\}, \quad (4)$$

guarantees

$$\|T_{\tilde{x}_m u_m}(\cdot, \tau) - T_{x_m u_m}\|_\infty \leq \bar{\gamma} \cdot \bar{\rho}^\tau. \quad (5)$$

Proof. We obtain from (2) that

$$\begin{aligned} T_{x_m u_m}(\omega) &= (zI_{n_m} - A_{11} - A_{12}(zI_{n_l} - A_{22})^{-1}A_{21})^{-1} \\ &\overset{(a)}{=} (zI_{n_m} - A_{11} - \sum_{i=1}^{\infty} z^{-i}A_{12}A_{22}^{i-1}A_{21})^{-1}, \quad (6) \end{aligned}$$

where $z = e^{j\omega}$ and (a) follows by using the relation $(zI_{n_l} - A_{22})^{-1} = \sum_{i=1}^{\infty} z^{-i}A_{22}^{i-1}$. Similarly, from (3) we obtain

$$T_{\tilde{x}_m u_m}(\omega, \tau) = (zI_{n_m} - \sum_{i=0}^{\tau-1} z^{-i}\tilde{A}_i^*)^{-1}. \quad (7)$$

Here we write the transfer function as $T_{\tilde{x}_m u_m}(\omega, \tau)$ to emphasize its dependence on τ. It then follows directly that

$$\begin{aligned} &\|T_{\tilde{x}_m u_m}(\cdot, \tau) - T_{x_m u_m}\|_\infty \\ &= \|T_{x_m u_m}(T_{x_m u_m}^{-1} - T_{\tilde{x}_m u_m}^{-1}(\cdot, \tau))T_{\tilde{x}_m u_m}(\cdot, \tau)\|_\infty \\ &\overset{(a)}{\leq} \|T_{x_m u_m}\|_\infty \|T_{\tilde{x}_m u_m}(\cdot, \tau)\|_\infty \|T_{x_m u_m}^{-1} - T_{\tilde{x}_m u_m}^{-1}(\cdot, \tau)\|_\infty \\ &\overset{(b)}{\leq} \|T_{x_m u_m}\|_\infty \|T_{\tilde{x}_m u_m}(\cdot, \tau)\|_\infty \sum_{i=\tau}^{\infty} \|z^{-i}A_{12}A_{22}^{i-1}A_{21}\|_\infty \\ &\overset{(c)}{\leq} \|T_{x_m u_m}\|_\infty \|T_{\tilde{x}_m u_m}(\cdot, \tau)\|_\infty \|A_{12}\|\|A_{21}\| \sum_{i=\tau}^{\infty} \|A_{22}^{i-1}\| \\ &\overset{(d)}{\leq} \gamma(\tau) \cdot \bar{\rho}^\tau, \end{aligned}$$

where

$$\gamma(\tau) \triangleq \frac{\kappa \|T_{x_m u_m}\|_\infty \|A_{12}\|\|A_{21}\|}{\bar{\rho} - \bar{\rho}^2} \|T_{\tilde{x}_m u_m}(\cdot, \tau)\|_\infty. \quad (8)$$

Here, (a) follows from the sub-multiplicativity of induced norms, (b) follows by the sub-additivity of norms, (c) follows by the definition of the H_∞-norm and also the sub-multiplicativity of induced norms, and (d) follows from Lemma 3.1. The remainder of the proof is devoted to showing the existence of a uniform upper bound $\bar{\gamma}$ for $\gamma(\tau)$. By the definition of the H_∞-norm,

$$\begin{aligned} \|T_{\tilde{x}_m u_m}(\cdot, \tau)\|_\infty &= \sup_{-\pi \leq \omega \leq \pi} \sigma_{\max}\big(T_{\tilde{x}_m u_m}(\omega, \tau)\big) \quad (9) \\ &\overset{(a)}{=} \Big(\inf_{-\pi \leq \omega \leq \pi} \sigma_{\min}\big(T_{\tilde{x}_m u_m}^{-1}(\omega, \tau)\big)\Big)^{-1}, \end{aligned}$$

where (a) holds due to the fact that $\sigma_{\max}(M) = \sigma_{\min}^{-1}(M^{-1})$ for any invertible matrix M. To complete the proof, we only need to show that

$$\vartheta \triangleq \inf_{\tau \in \mathbb{Z}_{\geq 0}} \inf_{-\pi \leq \omega \leq \pi} \sigma_{\min}\big(T_{\tilde{x}_m u_m}^{-1}(\omega, \tau)\big) > 0. \quad (10)$$

We show this in two steps.

(i) It follows from (6) and (7) that

$$\lim_{\tau \to \infty} T_{\tilde{x}_m u_m}^{-1}(\omega, \tau) = T_{x_m u_m}^{-1}(\omega), \quad \forall \omega \in [-\pi, \pi].$$

It is straightforward to show, using the exponential decay of A_{22}^τ and definition of uniform convergence, that each entry of $T_{\tilde{x}_m u_m}^{-1}(\cdot, \tau)$ converges uniformly to the corresponding entry of $T_{x_m u_m}^{-1}$. Hence, given the uniform continuity of matrix eigenvalues as a function of matrix entries [27, Thm 7.8c], $\sigma_{\min}\big(T_{\tilde{x}_m u_m}^{-1}(\cdot, \tau)\big)$ converges uniformly to $\sigma_{\min}\big(T_{x_m u_m}^{-1}\big)$. Thus, since $\inf_{-\pi \leq \omega \leq \pi} \sigma_{\min}\big(T_{x_m u_m}^{-1}(\omega)\big) = \|T_{x_m u_m}\|_\infty > 0$ (which itself holds by Assumption 2.1), there exists $\tau_0 \in \mathbb{Z}_{\geq 0}$ such that

$$\inf_{\tau \geq \tau_0} \inf_{-\pi \leq \omega \leq \pi} \sigma_{\min}\big(T_{\tilde{x}_m u_m}^{-1}(\omega, \tau)\big) > 0.$$

(ii) For any finite τ, we show that $T_{\tilde{x}_m u_m}(\cdot, \tau)$ is BIBO stable and thus has no poles on the unit circle (which in turn guarantees $\inf_{-\pi \leq \omega \leq \pi} \sigma_{\min}\big(T_{\tilde{x}_m u_m}^{-1}(\omega, \tau)\big) > 0$). For any bounded input u_m, let the corresponding outputs of $T_{\tilde{x}_m u_m}(\cdot, \tau)$ and $T_{x_m u_m}$ be denoted by $\tilde{x}_m$ and x_m, resp. (with initial states set to zero). We then have

$$x_m(k) - \tilde{x}_m(k) = A_{12}A_{22}^{\tau-1}x_l(k-\tau),$$

where x_l is the (internal) state of the latent nodes in $T_{x_m u_m}$. By Assumption 2.1, both $x_m(k)$ and $A_{12}A_{22}^{\tau-1}x_l(k-\tau)$ are bounded, proving the BIBO stability of $T_{\tilde{x}_m u_m}(\cdot, \tau)$.

Hence, (10) follows by combining (i) and (ii) and the fact that the decomposition $\mathbb{Z}_{\geq 0} = \{0\} \cup \{1\} \cup \cdots \cup \{\tau_0 - 1\} \cup \{\tau_0, \tau_0 + 1, \dots\}$ is finite. Equivalently, there exists $U > 0$ such that $\|T_{\tilde{x}_m u_m}(\cdot, \tau)\|_\infty < U$ for all $\tau \in \mathbb{Z}_{\geq 0}$, so (5) holds with $\bar{\gamma} = \kappa U \|T_{x_m u_m}\|_\infty \|A_{12}\|\|A_{21}\|/(\bar{\rho} - \bar{\rho}^2)$. ■

Theorem 3.2 shows that the presence of latent nodes in the network, as long as they do not receive any external input, does not affect the achievable accuracy of the identification via AR modeling of the manifest transfer function.

Remark 3.3: ***(Direct versus latent interactions – cont'd).*** It follows from the network dynamics (2) that

$$x_m(k+1) = \sum_{i=0}^{k} \tilde{A}_i^* x_m(k-i) + A_{12}A_{22}^k x_l(0) + u_m(k). \quad (11)$$

By virtue of (11), we can distinguish whether two manifest nodes interact directly or indirectly through latent nodes by looking at the matrix sequence $\{\tilde{A}_i^*\}$. First, the state of manifest node p affects the state of manifest node q directly if and only if $\tilde{A}_0^*(q,p) = A_{11}(q,p) \neq 0$. Similarly, the state of manifest node p affects the state of manifest node q indirectly through latent nodes if and only if $\tilde{A}_i^*(q,p) \neq 0$ for some $i \geq 1$. In particular, from the relation $\tilde{A}_i^* = -A_{12}A_{22}^{i-1}A_{21}$, one can see that the state of p first affects some latent nodes (that correspond to the nonzero entries in the p-th column of A_{21}) through A_{21}, then propagates through the latent subnetwork, reflected by A_{22}^{i-1}, and finally affects q through A_{12}. Furthermore, if the latent subnetwork is acyclic, then $\tilde{A}_i^*(q,p) \neq 0$ implies that there are exactly i latent nodes in a path connecting p to q. □

Remark 3.4: ***(Systems described by higher-order difference equations).*** Unlike the system description in (1), the dynamic behavior of many real-world complex systems such as the brain cortical networks is described by difference equations of orders significantly greater than 1, i.e.,

$$\begin{aligned} x(k+1) = A^{(0)}x(k) + A^{(1)}x(k-1) + \cdots \\ + A^{(\nu-1)}x(k-\nu+1) + u(k), \qquad \nu \gg 1 \end{aligned} \tag{12}$$

where $x_1, \ldots, x_{n_m}$ still denote the manifest (sensed and actuated) nodes and $x_{n_m+1}, \ldots, x_n$ are the latent ones. In this description, the vector x corresponds to some relevant physical variables. Defining the state vector $\xi(k) = [x(k)^T \; x(k-1)^T \; \cdots \; x(k-\nu+1)^T]^T$, one can rewrite (12) in order-1 form as

$$\begin{bmatrix} \xi_m(k+1) \\ \xi_l(k+1) \end{bmatrix} = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix} \begin{bmatrix} \xi_m(k) \\ \xi_l(k) \end{bmatrix} + \begin{bmatrix} u_m(k) \\ 0 \end{bmatrix}, \tag{13}$$

where $\xi_m(k) = x_m(k)$, $\xi_l(k) = [x_l(k)^T \; x_m(k-1)^T \; x_l(k-1)^T \; \cdots \; x_m(k-\nu+1)^T \; x_l(k-\nu+1)^T]^T$, $\mathcal{A}_{11} = A_{11}^{(0)}$, and

$$\mathcal{A}_{12} = \begin{bmatrix} A_{12}^{(0)} & A_{11}^{(1)} & A_{12}^{(1)} & \cdots & A_{11}^{(\tau-1)} & A_{12}^{(\tau-1)} \end{bmatrix},$$

$$\mathcal{A}_{21} = \begin{bmatrix} (A_{21}^{(0)})^T & I_{n_m} & 0 & \cdots & 0 & 0 \end{bmatrix}^T,$$

$$\mathcal{A}_{22} = \begin{bmatrix} A_{22}^{(0)} & A_{21}^{(1)} & A_{22}^{(1)} & \cdots & A_{21}^{(\tau-2)} & A_{22}^{(\tau-2)} & A_{21}^{(\tau-1)} & A_{22}^{(\tau-1)} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ I_{n_l} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & I_{n_m} & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & I_{n_l} & 0 & 0 \end{bmatrix}.$$

In this description, we view ξ_m as the actual "manifest state" of the system while the whole vector ξ_l is the "latent state". The reason for this interpretation is that, at any time k, only $x_m(k)$ is directly sensed/actuated while $x(k-1), \ldots, x(k-\nu+1)$ are quantities stored in the system. Interestingly, for the order-1 description (1), this observation brings up the possibility of some of the latent variables x_l simply being a relayed version of manifest variables. Note that, under this interpretation, the matrices $A_{11}^{(1)}, \ldots, A_{11}^{(\nu-1)}$ represent manifest-latent (rather than manifest-manifest) interactions. From (13), it is clear that all the treatment for (1) is readily applicable. Nevertheless, as ν increases, larger τ is necessary in order for (3) to represent the system accurately. This is both intuitive and clear from (5) and (8), where increasing ν results in larger $\|A_{12}\|$ and $\|A_{21}\|$ as well as (usually) $\|T_{x_m u_m}\|$ and $\rho(A_{22})$. This, in turn, may result in numerical difficulties when one constructs the AR model from recorded input-output data (which is the subject of the next section). □

Next, we show that there exists an AR model (3) whose transfer function coincides with the manifest transfer function if the latent subnetwork is acyclic.

Corollary 3.5: ***(Exact manifest transfer function identification for acyclic latent subnetworks).*** Under the assumptions of Theorem 3.2, further assume that the latent subnetwork is acyclic, i.e., there exists $\tau_{22} \in \mathbb{Z}_{\geq 1}$ such that $A_{22}^{\tau_{22}} = \mathbf{0}_{n_l \times n_l}$. Then, the matrix sequence $\tilde{A}_0^*, \cdots, \tilde{A}_{\tau_{22}}^*$ in (4) ensures $T_{\tilde{x}_m u_m} = T_{x_m u_m}$.

The proof of the result follows by comparing (6) and (7), and using the assumption that the latent subnetwork is acyclic. Theorem 3.2 and Corollary 3.5 show that it is possible to identify the transfer function of the manifest subnetwork without any knowledge of the passive latent nodes. However, (4) cannot be directly applied to determine the auto-regressive model because its evaluation requires knowledge of the adjacency matrix A of the whole network, which is unknown. This problem can be circumvented by employing the measured data sequence $\{y\}_1^N \subset \mathbb{R}^{n_m}$, as explained in the next section.

IV. Identification via least-squares estimation

In this section we employ least-squares estimation to compute from data the sequence of matrices defining the auto-regressive model. We show that the estimates resulting from this method asymptotically converge in probability, as the data length N and model order τ increase, to the optimal matrix sequence identified in Theorem 3.2. Finally, we particularize our discussion to the case of acyclic latent subnetworks.

A. Least-squares auto-regressive estimation

Given a vector sequence $\{y\}_1^N \subset \mathbb{R}^{n_m}$, the problem of least-squares auto-regressive (LSAR) model estimation with order $\tau \in \mathbb{Z}_{\geq 1}$ is to find a matrix sequence $\{\hat{A}\}_0^{\tau-1} \subset \mathbb{R}^{n_m \times n_m}$ that minimizes the 2-norm of the residual sequence $\{e\}_\tau^{N-1} \subset \mathbb{R}^{n_m}$ defined by

$$e(k) = y(k+1) - \sum_{i=0}^{\tau-1} \hat{A}_i y(k-i), \tag{14}$$

for $k \in \{\tau, \ldots, N-1\}$. Equation (14) can be written in compact vector form as

$$\vec{y}_N = \hat{\mathbf{A}}_\tau \Phi_N + \vec{e}_N, \tag{15}$$

where

$$
\begin{aligned}
\vec{y}_N &= \begin{bmatrix} y(\tau+1) & y(\tau+2) & \cdots & y(N) \end{bmatrix} \in \mathbb{R}^{n_m\times(N-\tau)},\\
\vec{e}_N &= \begin{bmatrix} e(\tau) & e(\tau+1) & \cdots & e(N-1) \end{bmatrix} \in \mathbb{R}^{n_m\times(N-\tau)},\\
\hat{\mathbf{A}}_\tau &= \begin{bmatrix} \hat{A}_0 & \hat{A}_1 & \cdots & \hat{A}_{\tau-1} \end{bmatrix} \in \mathbb{R}^{n_m\times n_m\tau},\\
\Phi_N &= \begin{bmatrix} y(\tau) & y(\tau+1) & \cdots & y(N-1)\\ y(\tau-1) & y(\tau) & \cdots & y(N-2)\\ \vdots & \vdots & \ddots & \vdots\\ y(1) & y(2) & \cdots & y(N-\tau) \end{bmatrix}.
\end{aligned}
$$

Using the square of the prediction error [26]

$$\operatorname{tr}(\vec{e}_N\vec{e}_N^T) = \operatorname{tr}\left((\vec{y}_N - \hat{\mathbf{A}}_\tau\Phi_N)(\vec{y}_N - \hat{\mathbf{A}}_\tau\Phi_N)^T\right)$$

as the cost function, we compute its gradient

$$\frac{\partial \operatorname{tr}(\vec{e}_N\vec{e}_N^T)}{\partial\hat{\mathbf{A}}_\tau} = (\vec{y}_N - \hat{\mathbf{A}}_\tau\Phi_N)(-\Phi_N^T) = \hat{\mathbf{A}}_\tau\Phi_N\Phi_N^T - \vec{y}_N\Phi_N^T.$$

Setting this to zero, we get a system of linear equations for which a solution is guaranteed to exist (since the rows of $\vec{y}_N\Phi_N^T$ belong to the row space of $\Phi_N\Phi_N^T$, which is the same as the row space of Φ_N^T). By Assumption 2.3, $\det(\Phi_N\Phi_N^T) \neq 0$ and this solution is unique with probability one.[3] If $\det(\Phi_N\Phi_N^T) = 0$, the minimum-norm solution can be found as

$$\hat{\mathbf{A}}_\tau = \vec{y}_N\Phi_N^T(\Phi_N\Phi_N^T)^{-1} = \vec{y}_N\Phi_N^+, \qquad (16)$$

where $(\cdot)^+$ denotes the Moore-Penrose pseudo-inverse. Since (16) is also valid for the nonsingular case, it is taken as the solution to the LSAR estimation problem. In order to indicate the dependency of the solution upon the measured data sequence, we sometimes use the notation $\hat{\mathbf{A}}_\tau(\{y\}_1^N)$.

B. Convergence in probability to manifest transfer function

Here we study the transfer function resulting from the LSAR estimation method and characterize its convergence properties, as the data length and the model order increase, with respect to the transfer function of the manifest subnetwork. Our first result establishes that the LSAR matrix estimate (16) converges in probability to the optimal matrix sequence identified in Theorem 3.2.

Proposition 4.1: ***(The LSAR estimate converges in probability to optimal matrix sequence).*** Consider the LTI network described by (2) where all latent nodes are passive. Given the measured data sequence $\{y\}_1^N$ generated from the dynamics (2) stimulated by the white noise input $\{u_m\}$ according to Assumption 2.3 and any $\bar{\rho} \in (\rho(A_{22}), 1)$, there exists $\beta \in \mathbb{R}_{>0}$ (depending only on the adjacency matrix A) such that the LSAR estimate $\hat{\mathbf{A}}_\tau(\{y\}_1^N)$ in (16) satisfies

$$\|\operatorname*{plim}_{N\to\infty} \hat{\mathbf{A}}_\tau(\{y\}_1^N) - \tilde{\mathbf{A}}_\tau^*\|_{\max} \leq \beta\tau\bar{\rho}^\tau, \qquad (17)$$

where $\tilde{\mathbf{A}}_\tau^* = \begin{bmatrix} \tilde{A}_0^* & \tilde{A}_1^* & \cdots & \tilde{A}_{\tau-1}^* \end{bmatrix} \in \mathbb{R}^{n_m\times n_m\tau}$ is the optimal matrix sequence given by (4).

Proof. For any quasi-stationary signal[4] $\{s\}$, let

$$R_s(j) \triangleq \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \mathbb{E}[s(i)s(i-j)^T].$$

Using the Birkhoff's Ergodic Theorem [28, Thm 7.2.1] (see also [28, Thm 7.1.3]) and the fact that $\{y\}$ is the output of a stable system (and thus the effects of initial conditions asymptotically vanish), we can show that

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N y(i)y(i-j)^T = R_y(j).$$

As a result, $\frac{1}{N}\Phi_N\Phi_N^T \in \mathbb{R}^{n_m\tau\times n_m\tau}$ also converges in probability and

$$
\begin{aligned}
R_\Phi &\triangleq \operatorname*{plim}_{N\to\infty}\frac{1}{N}\Phi_N\Phi_N^T\\
&= \begin{bmatrix} R_y(0) & R_y(1) & \cdots & R_y(\tau-1)\\ R_y^T(1) & R_y(0) & \cdots & R_y(\tau-2)\\ \vdots & \vdots & \ddots & \vdots\\ R_y^T(\tau-1) & R_y^T(\tau-2) & \cdots & R_y(0) \end{bmatrix}.
\end{aligned}
$$

Define

$$\nu(k) \triangleq y(k+1) - \sum_{i=0}^{\tau-1}\tilde{A}_i^* y(k-i), \qquad (18)$$

and note that the transfer function from u_m to ν is $T^{-1}_{\tilde{x}_m u_m}T_{x_m u_m}$, where $T_{x_m u_m}$ and $T_{\tilde{x}_m u_m}$ are given by (6) and (7), respectively. Equation (18) can be written in compact vector form as

$$\vec{y}_N = \tilde{\mathbf{A}}_\tau^*\Phi_N + \vec{\nu}_N, \qquad (19)$$

with $\vec{\nu}_N \triangleq [\nu(\tau)\ \nu(\tau+1)\ \cdots\ \nu(N-1)] \in \mathbb{R}^{n_m\times(N-\tau)}$. From (16) and (19), it follows that

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty}\hat{\mathbf{A}}_\tau(\{y\}_1^N) &= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\vec{y}_N\Phi_N^T\Big(\frac{1}{N}\Phi_N\Phi_N^T\Big)^{-1}\\
&= \tilde{\mathbf{A}}_\tau^* + \operatorname*{plim}_{N\to\infty}\frac{1}{N}\vec{\nu}_N\Phi_N^T\Big(\frac{1}{N}\Phi_N\Phi_N^T\Big)^{-1}. \qquad (20)
\end{aligned}
$$

Moreover, Assumption 2.3 renders $u_m(k)$ independent of $\{y\}_1^k$, which further implies that $\operatorname{plim}_{N\to\infty}\frac{1}{N}\vec{u}_{m,N}\Phi_N^T = \mathbf{0}_{n_m\times n_m\tau}$, where $\vec{u}_{m,N} \triangleq [u_m(\tau)\ u_m(\tau+1)\ \cdots\ u_m(N-1) \in \mathbb{R}^{n_m\times(N-\tau)}$. Therefore,

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\vec{\nu}_N\Phi_N^T = \operatorname*{plim}_{N\to\infty}\frac{1}{N}(\vec{\nu}_N - \vec{u}_{m,N})\Phi_N^T = \Psi, \qquad (21)$$

where $\Psi \triangleq \begin{bmatrix}\Psi_1 & \Psi_2 & \cdots & \Psi_\tau\end{bmatrix} \in \mathbb{R}^{n_m\times n_m\tau}$, with

$$\Psi_j \triangleq \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=\tau}^{N-1}(\nu(i) - u_m(i))y^T(i-j+1) \in \mathbb{R}^{n_m\times n_m}.$$

Thus, using $\operatorname{plim}_{N\to\infty}(\frac{1}{N}\Phi_N\Phi_N^T)^{-1} = R_\Phi^{-1}$, we have

$$\operatorname*{plim}_{N\to\infty}\hat{\mathbf{A}}_\tau(\{y\}_1^N) - \tilde{\mathbf{A}}_\tau^* = \Psi R_\Phi^{-1}.$$

[3] This is because (each element of) $\{y\}_1^{N-1}$ is an affine function of $\{u\}_0^{N-2}$, and $\det(\Phi_N\Phi_N^T)$ is a polynomial function of $\{y\}_1^{N-1}$, so $\det(\Phi_N\Phi_N^T)$ is a polynomial function of $\{u\}_0^{N-2}$. Therefore, the level set $\mathcal{N} = \{\{u\}_0^{N-2} \mid \det(\Phi_N\Phi_N^T) = 0\}$ has Lebesgue measure zero. Thus, by Assumption 2.3, $\Pr(\mathcal{N}) = 0$.

[4] Basically, a signal is quasi-stationary if it has a well-defined covariance function. See [26, Def 2.1] for a formal definition.

By the sub-additivity of the max norm, it holds for any $j \in \{1, \dots, \tau\}$ that

$$\begin{aligned}\|\Psi_j\|_{\max} &\leq \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=\tau}^{N-1} \|(\nu(i) - u_m(i)) y^T(i-j+1)\|_{\max} \\ &\overset{(a)}{\leq} \operatorname*{plim}_{N\to\infty} \frac{\bar{\rho}^{-\tau}}{N} \sum_{i=\tau}^{N-1} (\nu(i)-u_m(i))^T(\nu(i)-u_m(i)) \\ &\quad + \operatorname*{plim}_{N\to\infty} \frac{\bar{\rho}^{\tau}}{N} \sum_{i=\tau}^{N-1} y^T(i-j+1) y(i-j+1) \\ &= \bar{\rho}^{-\tau} \operatorname{tr}(R_{\nu-u_m}(0)) + \bar{\rho}^{\tau} \operatorname{tr}(R_y(0)),\end{aligned} \tag{22}$$

where (a) follows from Lemma A.1 in the appendix with the positive scalar M chosen as $\bar{\rho}^{\tau}$. Using the fact that the transfer function from u_m to $\nu - u_m$ is $T^{-1}_{\tilde{x}_m u_m} T_{x_m u_m} - I_{n_m}$, we obtain

$$\begin{aligned}R_{\nu-u_m}(0) &\triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \mathbb{E}[(\nu - u_m)(i)(\nu - u_m)^T(i)] \\ &\overset{(a)}{=} \frac{1}{2\pi} \int_{-\pi}^{\pi} (T^{-1}_{\tilde{x}_m u_m} T_{x_m u_m}(\omega) - I_{n_m}) \\ &\quad \cdot (T^{-1}_{\tilde{x}_m u_m} T_{x_m u_m}(\omega) - I_{n_m})^* d\omega \\ &\overset{(b)}{\leq} \|T^{-1}_{\tilde{x}_m u_m} T_{x_m u_m} - I_{n_m}\|^2_{\infty} I_{n_m} \\ &\overset{(c)}{\leq} \|T_{x_m u_m} - T_{\tilde{x}_m u_m}\|^2_{\infty} \|T^{-1}_{\tilde{x}_m u_m}\|^2_{\infty} I_{n_m} \\ &\overset{(d)}{\leq} \hat{\gamma} \bar{\rho}^{2\tau} I_{n_m},\end{aligned} \tag{23}$$

where $\hat{\gamma} \triangleq \bar{\gamma}^2 \big(1 + \|A_{11}\| + \|A_{12}\| \|A_{21}\| \kappa (1 - \bar{\rho})^{-1}\big)^2$ is constant, (a) follows from [29, eq. (9-193)], (b) follows by the definition of H_∞-norm, (c) follows by the sub-multiplicativity of induced norms, and (d) holds because of Theorem 3.2 and the observation that, by Lemma 3.1,

$$\|T^{-1}_{\tilde{x}_m u_m}\|_{\infty} \leq 1 + \|A_{11}\| + \|A_{12}\| \|A_{21}\| \kappa (1 - \bar{\rho})^{-1}.$$

We obtain from (22) and (23),

$$\|\Psi_j\|_{\max} \leq \bar{\rho}^{\tau} (\hat{\gamma} n_m + \operatorname{tr}(R_y(0))),$$

and from (20) and (21),

$$\begin{aligned}\| \operatorname*{plim}_{N\to\infty} \hat{\mathbf{A}}_\tau(\{y\}_1^N) - \tilde{\mathbf{A}}^*_\tau \|_{\max} &= \|\Psi R_\Phi^{-1}\|_{\max} \\ &\leq n_m \tau \|R_\Phi^{-1}\|_{\max} \|\Psi\|_{\max} \\ &= n_m \tau \|R_\Phi^{-1}\|_{\max} \max_j \|\Psi_j\|_{\max} \leq \beta \tau \bar{\rho}^{\tau},\end{aligned}$$

where $\beta = (\hat{\gamma} n_m^2 + \operatorname{tr}(R_y(0)) n_m) \|R_\Phi^{-1}\|_{\max}$, as claimed. ■

When it is clear from context, we refer to $\operatorname{plim}_{N\to\infty} \hat{A}_i(\{y\}_1^N)$ simply as $\hat{A}_i$.

Remark 4.2: ***(Direct versus latent interactions – cont'd).*** Proposition 4.1 shows that $\hat{A}_i$ converges in probability to $\tilde{A}^*_i$ exponentially as the model order τ increases. Therefore, within a margin of error that can be tuned as desired, we deduce from the discussion in Remark 3.3 that the LSAR estimate $\hat{A}_0$ allows us to determine whether two manifest nodes interact directly and the LSAR estimates $\{\hat{A}_i\}_{i\geq 1}$ allow us to determine whether two manifest nodes interact indirectly through latent nodes with high probability as the length of measurement data grows. □

Given the result in Proposition 4.1, we next turn our attention to the transfer function from e to y resulting from the LSAR estimation (14), which we denote by $T_{ye}(\{y\}_1^N, \tau)$. The next result shows that the H_∞-norm of this transfer function is uniformly upper bounded with respect to the model order τ.

Lemma 4.3: ***(H_∞-norm of T_{ye} is uniformly upper bounded).*** Under the assumptions of Proposition 4.1, there exist positive scalars τ_0 and $U^{\infty}_{T_{ye}}$ such that, for $\tau \geq \tau_0$,

$$\| \operatorname*{plim}_{N\to\infty} T_{ye}(\{y\}_1^N, \tau) \|_{\infty} \leq U^{\infty}_{T_{ye}}. \tag{24}$$

Proof. By definition of H_∞-norm, we have

$$\begin{aligned}\| \operatorname*{plim}_{N\to\infty} T_{ye}(\{y\}_1^N, \tau) \|_{\infty} &= \sup_{-\pi\leq\omega\leq\pi} \sigma_{\max}\big(\operatorname*{plim}_{N\to\infty} T_{ye}(\omega, \tau) \big) \\ &= \Big(\inf_{-\pi\leq\omega\leq\pi} \sigma_{\min}\big(\operatorname*{plim}_{N\to\infty} T^{-1}_{ye}(\omega, \tau) \big) \Big)^{-1}.\end{aligned} \tag{25}$$

Note that, for every $\omega \in [-\pi, \pi]$ and $\tau \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned}\operatorname*{plim}_{N\to\infty} T^{-1}_{ye}(\omega, \tau) &= z I_{n_m} - \sum_{i=0}^{\tau-1} z^{-i} \hat{A}_i \\ &= T^{-1}_{\tilde{x}_m u_m}(\omega, \tau) - \sum_{i=0}^{\tau-1} z^{-i} (\hat{A}_i - \tilde{A}^*_i),\end{aligned} \tag{26}$$

where $z = e^{j\omega}$. However, for every $\omega \in [-\pi, \pi]$ and $\tau \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned}\Big\| \sum_{i=0}^{\tau-1} z^{-i} (\hat{A}_i - \tilde{A}^*_i) \Big\| &\leq \sum_{i=0}^{\tau-1} \|\hat{A}_i - \tilde{A}^*_i\| \overset{(a)}{\leq} n_m \sum_{i=0}^{\tau-1} \|\hat{A}_i - \tilde{A}^*_i\|_{\max} \\ &\overset{(b)}{\leq} n_m \tau \max_i \|\hat{A}_i - \tilde{A}^*_i\|_{\max} \leq n_m \beta \tau^2 \bar{\rho}^{\tau},\end{aligned}$$

where (a) follows from the fact that $\|A\| \leq n_m \|A\|_{\max}$ for any matrix $A \in \mathbb{R}^{n_m \times n_m}$ and (b) follows from Proposition 4.1. Therefore, using Weyl's theorem for the perturbation of singular values [30] in (26) and taking $\inf_{-\pi\leq\omega\leq\pi}$ of both sides, we get

$$\begin{aligned}&\inf_{-\pi\leq\omega\leq\pi} \sigma_{\min}\big(\operatorname*{plim}_{N\to\infty} T^{-1}_{ye}(\omega, \tau) \big) \\ &\geq \inf_{-\pi\leq\omega\leq\pi} \sigma_{\min}\big(T^{-1}_{\tilde{x}_m u_m}(\omega, \tau) \big) - \Big\| \sum_{i=0}^{\tau-1} z^{-i} (\hat{A}_i - \tilde{A}^*_i) \Big\| \\ &\geq \inf_{-\pi\leq\omega\leq\pi} \sigma_{\min}\big(T^{-1}_{\tilde{x}_m u_m}(\omega, \tau) \big) - n_m \beta \tau^2 \bar{\rho}^{\tau}.\end{aligned}$$

In view of (10), let τ_0 be such that

$$n_m \beta \tau^2 \bar{\rho}^{\tau} \leq \frac{\vartheta}{2}, \qquad \forall \tau \geq \tau_0. \tag{27}$$

Then, the result follows from (25) with $U^{\infty}_{T_{ye}} = \frac{2}{\vartheta}$. ■

We are finally ready to show that the transfer function T_{ye} obtained from the LSAR method converges in probability to the transfer function $T_{x_m u_m}$ of the manifest subnetwork.

Theorem 4.4: ***(The LSAR method consistently estimates the manifest transfer function).*** Under the assumptions of Proposition 4.1, for any $\bar{\rho} \in (\rho(A_{22}), 1)$, there exist positive scalars $\bar{\beta}$, $\bar{\gamma}$ and τ_0 such that, for $\tau \geq \tau_0$,

$$\| \operatorname*{plim}_{N\to\infty} T_{ye}(\{y\}_1^N, \tau) - T_{x_m u_m} \|_{\infty} \leq (\bar{\beta} \tau^2 + \bar{\gamma}) \bar{\rho}^{\tau}. \tag{28}$$

Consequently, $\operatorname{plim}_{N\to\infty,\tau\to\infty} T_{ye}(\{y\}_1^N,\tau) = T_{x_m u_m}$.

Proof. We only need to prove (28) as it directly implies the last equation in the statement. By the sub-additivity and sub-multiplicity of induced norms,

$$\begin{aligned}&\|T_{ye}(\cdot,\tau) - T_{x_m u_m}\|_\infty \\ &\le \|T_{ye}(\cdot,\tau) - T_{\tilde{x}_m u_m}(\cdot,\tau)\|_\infty + \|T_{\tilde{x}_m u_m}(\cdot,\tau) - T_{x_m u_m}\|_\infty \\ &\le \|T_{ye}(\cdot,\tau)\|_\infty \|T_{\tilde{x}_m u_m}(\cdot,\tau)\|_\infty \|T_{ye}^{-1}(\cdot,\tau) - T_{\tilde{x}_m u_m}^{-1}(\cdot,\tau)\|_\infty \\ &\quad + \|T_{\tilde{x}_m u_m}(\cdot,\tau) - T_{x_m u_m}\|_\infty. \qquad (29)\end{aligned}$$

Next, by (9), Lemma 4.3, and Theorem 3.2, there exist positive scalars τ_0, $U_{T_{ye}}^\infty$ and ϑ such that for $\tau \ge \tau_0$,

$$\begin{aligned}&\|\operatorname*{plim}_{N\to\infty} T_{ye}(\cdot,\tau) - T_{x_m u_m}\|_\infty \qquad (30)\\ &\quad \le U_{T_{ye}}^\infty \vartheta^{-1} \|\operatorname*{plim}_{N\to\infty} T_{ye}^{-1}(\cdot,\tau) - T_{\tilde{x}_m u_m}^{-1}(\cdot,\tau)\|_\infty + \bar{\gamma}\bar{\rho}^\tau.\end{aligned}$$

Finally, according to the definition of $T_{ye}(\cdot,\tau)$ in (14) and $T_{\tilde{x}_m u_m}(\cdot,\tau)$ in (7), it follows that

$$\begin{aligned}&\|\operatorname*{plim}_{N\to\infty} T_{ye}^{-1}(\cdot,\tau) - T_{\tilde{x}_m u_m}^{-1}(\cdot,\tau)\|_\infty = \|\sum_{i=0}^{\tau-1} z^{-i}(\operatorname*{plim}_{N\to\infty} \hat{A}_i - \tilde{A}_i^*)\|_\infty \\ &\overset{(a)}{\le} \sum_{i=0}^{\tau-1} \|\operatorname*{plim}_{N\to\infty} \hat{A}_i - \tilde{A}_i^*\| \overset{(b)}{\le} n_m \beta \tau^2 \bar{\rho}^\tau, \qquad (31)\end{aligned}$$

where (a) holds by the sub-additivity and sub-multiplicity of $\|\cdot\|$ and (b) follows by Proposition 4.1 and the fact that $\|A\| \le n_m \|A\|_{\max}$ for any matrix $A \in \mathbb{R}^{n_m \times n_m}$. Thus, we obtain (28) for $\tau \ge \tau_0$, where $\bar{\beta} \triangleq U_{T_{ye}}^\infty \vartheta^{-1} n_m \beta$ is a constant. ■

According to Theorem 4.4, when the length N of the measurement data is sufficiently large and the model order τ exceeds a certain threshold, the error $\|T_{ye}(\tau) - T_{x_m u_m}\|_\infty$ obtained by the LSAR method decreases exponentially with τ.

Remark 4.5: ***(Identification of manifest transfer function requires higher-order models as stability margin of latent subnetwork decreases).*** Even though an explicit expression of the threshold τ_0 in Theorem 4.4 as a function of the network is difficult to obtain, we can still make some useful observations. From inequality (27) in the proof of Lemma 4.3, one can see that τ_0 is an increasing function of $\bar{\rho}$. Hence, as the latent subnetwork becomes less stable ($\rho(A_{22})$ gets closer to 1), the corresponding τ_0 becomes larger, requiring the order of the AR model to be higher to ensure exponential convergence. □

Remark 4.6: ***(Systems described by higher-order difference equations – cont'd).*** As explained in Remark 3.4, the AR representation of systems with order $\nu > 1$ is identical to the $\nu = 1$ case, although they require larger AR order τ. For large-scale systems ($n \gg 1$), increasing τ rapidly raises the number of parameters in (15), which leads to over-parametrization of the LSAR identification. Our simulations in Section V show how this can be overcome both by increasing N (which is computationally costly) and exponential regularization. Also, note that when $\nu > 1$, the only member of the sequence of matrices $A_{11}^{(0)}, \dots, A_{11}^{(\nu-1)}$ (denoting all current and past interactions among manifest *nodes*) that is identifiable by the LSAR method is $A_{11}^{(0)}$ (representing direct interactions among manifest *states*) while the others are only identifiable in the aggregate form (5). □

C. Exact identification for acyclic latent subnetworks

Here we show that the transfer function of the manifest subnetwork can be perfectly identified using the LSAR method with a finite model order if the latent subnetwork is acyclic. We start by refining the result in Proposition 4.1 and showing how, in this case, the convergence of the LSAR matrix estimate (16) to the optimal matrix sequence identified in Theorem 3.2 holds in the mean-square sense.

Proposition 4.7: ***(The LSAR estimate converges in mean square to optimal matrix sequence for acyclic latent subnetworks).*** Consider the LTI network described by (2) where all latent nodes are passive. Further assume that the latent subnetwork is acyclic, i.e., there exists $\tau_{22} \in \mathbb{Z}_{\ge 1}$ such that $A_{22}^{\tau_{22}} = \mathbf{0}_{n_l \times n_l}$. Given the measured data sequence $\{y\}_1^N$ generated from the dynamics (2) stimulated by the white noise input $\{u_m\}$ according to Assumption 2.3, the LSAR estimate $\hat{\mathbf{A}}_\tau(\{y\}_1^N)$ in (16) satisfies, for any $\tau \ge \tau_{22} + 1$,

$$\lim_{N\to\infty} \mathbb{E}[(\hat{\mathbf{A}}_\tau(\{y\}_1^N) - \tilde{\mathbf{A}}_\tau^*)^T(\hat{\mathbf{A}}_\tau(\{y\}_1^N) - \tilde{\mathbf{A}}_\tau^*)] = \mathbf{0}_{n_m\tau \times n_m\tau}.$$

Proof. If A_{22} is nilpotent, using Corollary 3.5, we deduce that the transfer function from u_m to ν defined in (18) is $T_{\tilde{x}_m u_m}^{-1} T_{x_m u_m} = I_{n_m}$. Consequently, the random vectors $\nu(k)$'s are i.i.d. with zero mean and finite second moment $\mathbb{E}[\nu(k)\nu^T(k)] = I_{n_m}$. Define

$$\begin{aligned}Z_N &\triangleq \frac{1}{N}(\hat{\mathbf{A}}_\tau - \tilde{\mathbf{A}}_\tau^*)\Phi_N \Phi_N^T \\ &\overset{(a)}{=} \frac{1}{N}(\vec{\nu}_N - \vec{e}_N)\Phi_N^T \overset{(b)}{=} \frac{1}{N}\vec{\nu}_N \Phi_N^T,\end{aligned}$$

where (a) follows from (15) and (19) and (b) follows from the fact that the least-squares estimate $\hat{\mathbf{A}}_\tau$ in (16) renders $\vec{e}_N \Phi_N^T = \mathbf{0}_{n_m \times n_m\tau}$. Combining the fact that the $\nu(k)$'s are i.i.d. and the fact that $\{y\}_1^k$ is a function of $\{\nu\}_1^{k-1}$, we deduce that $\nu(k)$ are independent of $\{y\}_1^k$. This further implies that $\mathbb{E}[Z_N] = \mathbf{0}_{n_m \times n_m\tau}$. Furthermore,

$$\begin{aligned}\lim_{N\to\infty} \mathbb{E}[Z_N^T Z_N] &= \lim_{N\to\infty} \frac{1}{N^2}\mathbb{E}[\Phi_N \vec{\nu}_N^T \vec{\nu}_N \Phi_N^T] \\ &= \lim_{N\to\infty} \frac{1}{N} R_\Phi = \mathbf{0}_{n_m\tau \times n_m\tau}.\end{aligned}$$

Therefore, $\lim_{N\to\infty} \mathbb{E}[\hat{\mathbf{A}}_\tau - \tilde{\mathbf{A}}_\tau^*] = \lim_{N\to\infty} \mathbb{E}[Z_N] R_\Phi^{-1} = \mathbf{0}_{n_m \times n_m\tau}$ and $\lim_{N\to\infty} \mathbb{E}[(\hat{\mathbf{A}}_\tau - \tilde{\mathbf{A}}_\tau^*)^T(\hat{\mathbf{A}}_\tau - \tilde{\mathbf{A}}_\tau^*)] = R_\Phi^{-1} \lim_{N\to\infty} \mathbb{E}[Z_N^T Z_N] R_\Phi^{-1} = \mathbf{0}_{n_m\tau \times n_m\tau}$, as claimed. ■

We build on this result to show that the manifest transfer function can be perfectly identified using the LSAR method with a finite model order if the latent subnetwork is acyclic.

Theorem 4.8: ***(Exact manifest transfer function identification for acyclic latent subnetworks).*** Under the assumptions of Proposition 4.7, for any $\tau \ge \tau_{22} + 1$,

$$\operatorname*{plim}_{N\to\infty} T_{ye}(\{y\}_1^N, \tau) = T_{x_m u_m}.$$

Proof. We have $\operatorname{plim}_{N\to\infty} \hat{\mathbf{A}}_\tau(\{y\}_1^N) = \tilde{\mathbf{A}}_\tau^*$ from Proposition 4.7, which combined with (31) implies

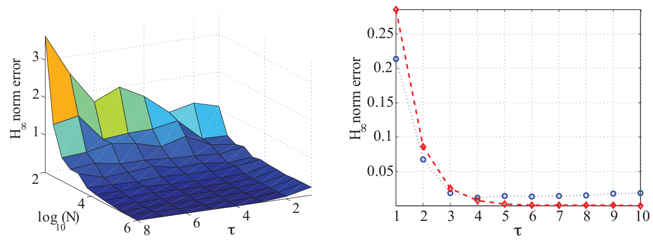
$$\operatorname*{plim}_{N\to\infty} T_{ye}^{-1}(\tau) = T_{\tilde{x}_m u_m}^{-1}(\tau).$$

Moreover, from Corollary 3.5, we have $T_{\tilde{x}_m u_m}(\tau) = T_{x_m u_m}$. The statement then follows from (29) and Lemma 4.3. ■

V. Simulations

In this section, we illustrate the performance of least-squares auto-regressive estimation in identifying the manifest transfer function in two examples, a deterministic directed ring network and a group of Erdős–Rényi random networks. We pay particular attention to the behavior displayed as the length of measured data and the model order change. In both examples, the input signal is a white Gaussian process with unit variance.

Example 5.1: ***(Directed ring network).*** Consider a directed ring network of 40 nodes with self-loops and all edge weights equal to $\alpha = 0.25$. The nodes with indices $\{5, 23, 33, 34, 36\}$ are manifest and the remaining 35 nodes are passive latent. Fig. 2.(a) shows a 3D plot of the identification error $\|T_{ye} - T_{x_m u_m}\|_\infty$ of the LSAR method, with axes corresponding to length of measured data and model order, respectively. We note that, when the measured data length N is small, increasing the AR model order τ does not provide better estimation of the manifest transfer function. Similarly, when the model order τ is too low, increasing the data length N is not helpful either. Instead, when N and τ increase simultaneously, the LSAR method provides good estimation of the manifest transfer function without any knowledge of the latent nodes, as predicted by Theorem 4.4. In Fig. 2.(b), we fix $N = 10^6$ and show that the error of the model obtained by the LSAR method is quite similar to the error $\|T_{\tilde{x}_m u_m} - T_{x_m u_m}\|_\infty$ of the ideal AR model from Theorem 3.2. Note that the latter requires knowledge of the true adjacency matrix A, and we use it here merely for comparison purposes. □



(a) (b)

Fig. 2. H_∞-norm errors for the directed ring network of Example 5.1. (a) The H_∞-norm error of the LSAR method as a function of data length N and model order τ. Performance improves as N and τ increase. (b) Comparison of the H_∞-norm errors of the LSAR method (blue dotted lines) and the optimal AR model from Theorem 3.2 (red dashed lines) for $N = 10^6$.

Example 5.2: ***(Erdős–Rényi random network).*** Here we consider a group of 10 Erdős–Rényi random networks [31]. Each network in the group is of type $G(10, 0.35)$, with 5 manifest nodes chosen randomly and the remaining 5 nodes are latent. Each pair of edges $(i,j), (j,i)$, $1 \le i < j \le 10$ has nonzero weights with probability 0.35 (we choose edges in pairs so that, when plotting the graph, the edge direction can be omitted). The weight of each edge has a uniform distribution in $\{x \in \mathbb{R} \mid 0.1 < x < 0.35\}$ (note that (i,j) and (j,i) can have different weights). Because of rounding errors in the numerical computation, the estimated coefficient matrices (16) of the AR model are usually full matrices. The lower bound on the edge weights allows us to discard entries in $\hat{\mathbf{A}}_0$ that are smaller than 0.1. We consider a fixed length $N = 10^6$ of measured data and analyze the effect of varying model order. Fig. 3 shows a 3D plot of the error in the identification of the manifest transfer function by the LSAR estimation, with axes corresponding to network index and model order, respectively. One can see the improvement in performance as the model order increases for all 10 networks. Fig. 4 compares the identification error of the LSAR method for the networks with indices 1, 6, 8, 10 in Fig. 3 against the error of the optimal AR model from Theorem 3.2. The latent subnetwork of network 6 is acyclic (with $A_{22} = \mathbf{0}_{5\times5}$), and the estimation error goes to 0 when the AR model has order higher than $\tau_{22} = 1$, as predicted by Theorem 4.8. To illustrate our observations in Remark 4.2 regarding the identification of manifest and latent interactions, Fig. 5 shows on the left the networks with indices 1, 6, 8, 10 of Fig. 3 and on the right the corresponding reconstructions obtained with the LSAR method. The indirect interactions represented by dashed edges in these plots imply the presence of latent nodes. For comparison, we have also used the brain connectivity estimator technique called direct directed transfer function

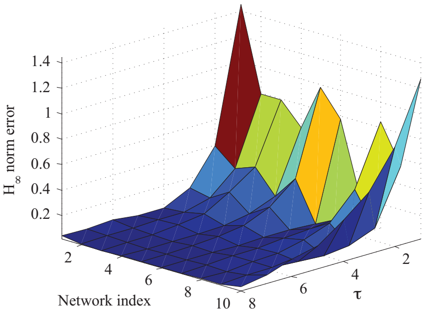


Fig. 3. Illustration of the H_∞-norm error of the LSAR with respect to the model order τ for the group of $G(10, 0.35)$ Erdős–Rényi random networks of Example 5.2. Performance improves as the model order τ increases for all 10 networks. The length of measured data is $N = 10^6$.

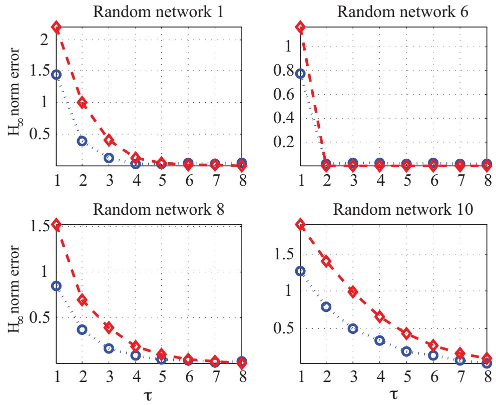
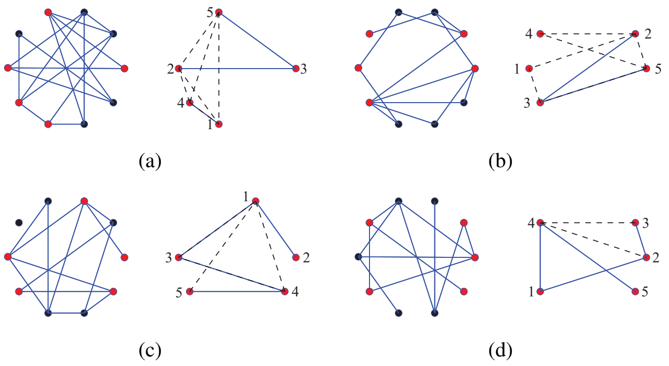


Fig. 4. Comparison of the H_∞-norm errors of the LSAR method (red dashed lines) and the optimal AR model from Theorem 3.2 (blue dotted lines) for the Erdős–Rényi random networks with indices 1, 6, 8, 10 in Fig. 3. The estimation error for network 6 becomes 0 when the AR model has order higher than 1 because the latent subnetwork is acyclic with $\tau_{22} = 1$. The length of measured data is $N = 10^6$.

(dDTF) measure [8], [25] from neuroscience to identify direct connections between nodes. This technique is a refinement of the directed transfer function (DTF) approach, which instead cannot distinguish between direct and indirect connections. We have employed the dynamical modeling method within the Source Information Flow Toolbox (SIFT) [32], [33] in EEGLAB [34], which is a widely used open-source toolbox for EEG analysis. Fig. 6 shows the interaction topology among the 5 manifest nodes in network 10 identified by SIFT using the dDTF measure. The dDTF measure is in the frequency domain and can also be a function of time (e.g., for time-varying networks). Since our networks are time-invariant, the time axis can be ignored. The plot shows that the dDTF identifies roughly equally strong connections for $(2, 4)$ (which is in reality mediated by latent nodes) and $(4, 5)$ (which is a true direct connection). This is in contrast with the identification made with the LSAR method presented in Fig. 5(d). □



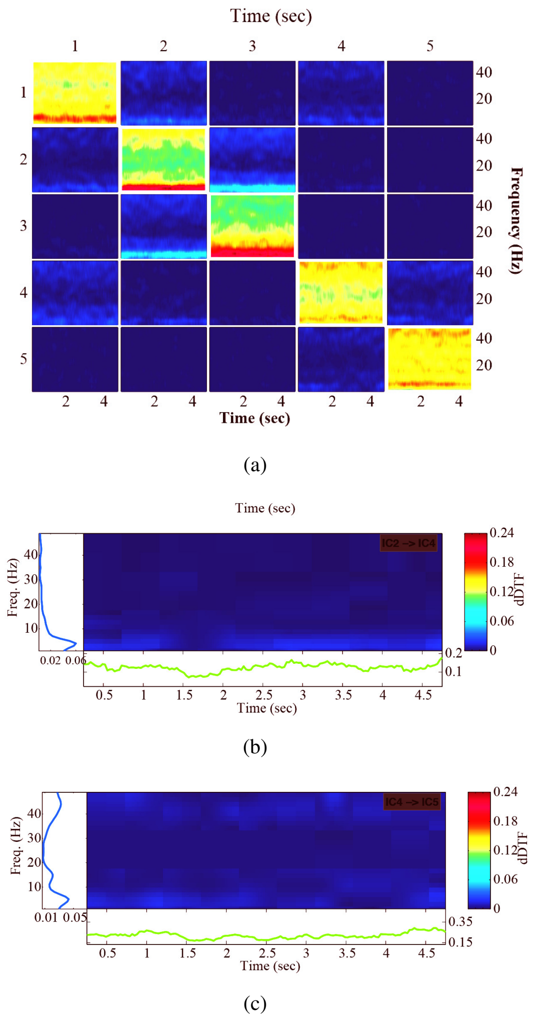
(a)

(b)

(c)

(d)

Fig. 5. Left: Erdős–Rényi random networks corresponding to the networks with indices 1 (a), 6 (b), 8 (c), 10 (d) in Fig. 3, where red circles represent manifest nodes and black circles represent latent nodes. Right: reconstructed interaction graphs of the manifest subnetworks using the LSAR method. The numbers next to these nodes indicate their indices. A blue solid edge represents direct interaction and a black dashed edge represents indirect interaction through latent nodes. Note that the latent subnetwork of network 6 is acyclic.



(a)

(b)

(c)

Fig. 6. (a) The interaction topology identified by the dDTF method for the Erdős–Rényi network with index 10. (b and c) A zoom-in of the (indirect) connection $(2, 4)$ and the (direct) connection $(4, 5)$, resp.

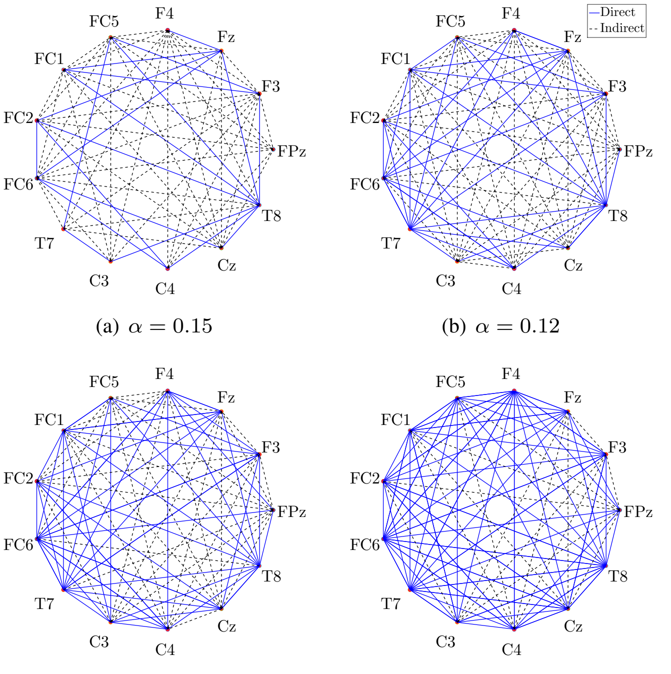
Example 5.3: ***(Cortical brain network identification from EEG data).*** In this example, we apply our method to a multi-channel electroencephalogram (EEG) time-series recorded from a human scalp during a selective visual attention experiment in order to identify the manifest and latent-mediated connections among the channels. The EEG data is taken from the sample dataset available in the EEGLAB MATLAB toolbox [34]. This dataset contains recordings from 32 channels for more than 3 seconds with $T_s = 7.8$ ms sampling time (128 Hz sampling frequency). Channel locations are shown in Fig. 10(a) on a top (axial) view of the skull. During the experiment, the subject is asked to perform specific motor actions in response to certain visual stimuli, requiring coordination among several cortices. We take the first 13 EEG channels corresponding to the fronto-temporal cortical areas (shown as blue squares in Fig. 10(a)) as the manifest nodes and the remaining channels as well as the truly hidden brain regions (the ones not probed in the test) as the latent nodes. In the following, we present the results of identifying the direct and indirect connections among the manifest nodes using the LSAR method as well as the dDTF algorithm [8], [25] and the S+L algorithm of [22]. For each method, we only keep the edges whose identified weights are above a certain threshold θ (which we choose as a proportional constant $\alpha \in (0, 1)$ times the largest edge weight in the network).

In neuroscience, the brain dynamics generating the EEG data are usually approximated by a high-order AR model of the form (12) ($\nu \gtrsim 10$). As mentioned in Remark 4.6, larger τ and thus larger number of parameters are then required, which may lead to over-parametrization. To prevent this, we use an exponentially-regularized version of (16) by minimizing

$$\mathrm{tr}(\vec{e}_N \vec{e}_N^T + \gamma \hat{\mathbf{A}}_\tau P P^T \hat{\mathbf{A}}_\tau^T), \qquad (32)$$

where $P = \mathrm{diag}(1, \rho_0^{-1}, \dots, \rho_0^{-(\tau-1)}) \otimes I_{n_m}$ and, ideally, $\rho_0 = \rho(A_{22})$ (in practice, it is found by trial and error). The role of the exponential regularizer is to encourage the higher-order AR terms to decay exponentially, as $\tilde{A}_i^*$ do. In the simulations that follow, we have used $\gamma = 10$ and $\rho_0 = 0.9$.

Fig. 7 shows the reconstructed manifest subnetwork with direct and indirect connections using the LSAR method for $\tau = 15$ and different values of α. One can observe that the sensitivity of the network structure to the threshold ratio α is significant, showing that the majority of network links are relatively weak with respect to the largest link (which is usually a self-loop). This sensitivity, however, is smaller for the indirect connections. Note that increasing α is a way of enforcing sparsity among the manifest nodes similar (but not equivalent) to [22]. Also, note that unlike [22], the manifest subnetwork estimated by our method is directed (though directions are not shown in Fig. 7 for simplicity).



(a) $\alpha = 0.15$

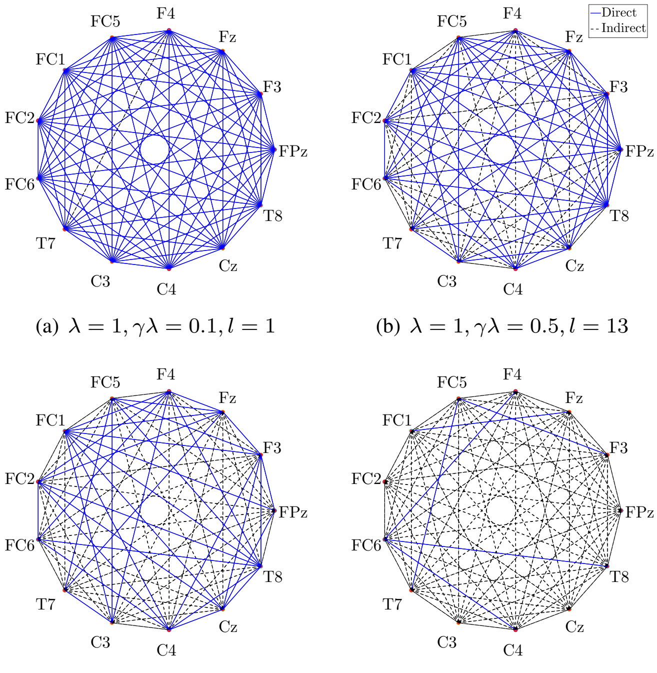
(b) $\alpha = 0.12$

(c) $\alpha = 0.09$

(d) $\alpha = 0.06$

Fig. 7. Reconstructed manifest subnetwork for the EEG data in Example 5.3 using our proposed method with the exponentially-regularized objective function (32) and $\gamma = 10$, $\rho_0 = 0.9$, and $\tau = 15$. The direct (solid blue) and indirect (dashed black) connections are depicted for different values of threshold ratio α. For each value of α, the connections whose weights are smaller than α times the largest network weight are removed.

For comparison, Fig. 8 shows the reconstructed manifest subnetwork with direct and indirect connections using the S+L method of [22] for $n = 5$[5]. Although the use of a threshold value is not prescribed in [22], we have used a fixed value of $\alpha = 0.01$ for all values of (λ, γ), since the absence of a threshold ($\alpha = 0$) results in all nodes being estimated to be (both directly and indirectly) connected. This lack of sparsity occurs for all values of (λ, γ) (no matter how large they are chosen), unless extremely large values are employed, which results in a fully disconnected network. From various plots, we see that even with the use of a threshold value all the nodes are estimated to be indirectly connected, with the sparsity of direct connections and the estimated number of latent nodes being determined by (λ, γ). This abundance of indirect connections and parameter-based tuning of direct connectivity is similar to our results in Fig. 7, even though the details of the reconstructed networks do not exactly match.



(a) $\lambda = 1, \gamma\lambda = 0.1, l = 1$

(b) $\lambda = 1, \gamma\lambda = 0.5, l = 13$

(c) $\lambda = 1, \gamma\lambda = 1, l = 18$

(d) $\lambda = 0.1, \gamma\lambda = 0.5, l = 50$

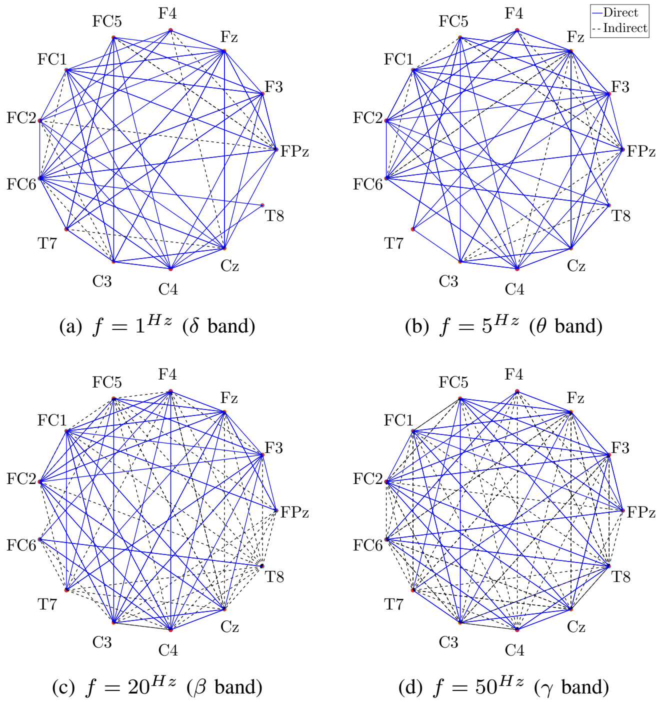
Fig. 8. Reconstructed manifest subnetwork for Example 5.3 using the S+L method in [22]. The direct (solid blue) and indirect (dashed black) connections are depicted for different values of weight parameters (λ, γ) and fixed threshold ratio $\alpha = 0.01$. l represents the estimated number of latent nodes.

Fig. 9 shows the result of applying both the Directed Transfer Function (DTF) [35] and direct Directed Transfer Function (dDTF) methods to the EEG channel data to estimate the indirect and direct connections between the manifest nodes, respectively, for different frequency bands. Both methods are applied to the data using the EEGLAB SIFT plugin for $\tau = 15$ (selected based on SIFT Model Order Selection). In all cases, a constant threshold ratio $\alpha = 0.1$ is used and the value of the threshold is computed with respect to the largest *off-diagonal* link weight in the same frequency. As can be seen, the connectivity pattern is considerably different between lower and higher frequencies, where several pairs are not even indirectly connected over the δ-θ band. This is in contrast to the reconstructed networks of Fig. 7 in which most pairs are at least indirectly connected, even for threshold values as large as $\alpha = 0.15$. Nevertheless, a common feature of all the reconstructed networks in Figs. 7-9 is that the density of direct connections is higher in the fronto-central (FC) areas and lower in central (C) areas and midline frontal pole (FPz). The independence of this sparsity pattern from the employed reconstruction method and parameter value suggests that it is a robust feature of the actual brain connectivity among these areas.

Since the *true* network structure is unknown for this example (and hence the methods are not directly comparable), we validate our LSAR estimated connectivity based on its ability to predict *future* (i.e., unseen) channel activity. Thus, we used the first 80% of data for LSAR estimation and the last 20% for evaluation, which is based on

$$R^2 = 1 - \frac{\sum_{k=N+1}^{N'} \|e(k)\|^2}{\sum_{k=N+1}^{N'} \|y(k)\|^2}, \tag{33}$$

[5] n represents the model order in [22]. While the role of the model order is not discussed in the reference, the use of higher-order models significantly increases the computational cost of the algorithm. Also, note that there is no one-to-one correspondence between the subfigures of Figs. 7-9.



(a) $f = 1^{Hz}$ (δ band)

(b) $f = 5^{Hz}$ (θ band)

(c) $f = 20^{Hz}$ (β band)

(d) $f = 50^{Hz}$ (γ band)

Fig. 9. Reconstructed manifest subnetwork for Example 5.3 using the combination of DTF and dDTF estimation methods. The direct (solid blue) and indirect (dashed black) connections are illustrated for different frequency values and fixed threshold ratio $\alpha = 0.1$.

denoting the percentage of the future channel activity that is correctly predicted by the model [26, §16.4], where $\{y(k)\}_{k=N+1}^{N'}$ is the latter data sequence not used for estimation. The blue curve in Fig. 10(c) shows the value of $R \times 100\%$ for the LSAR method as a function of model order for the same selection of nodes as above (i.e., anterior)[6]. This shows that the method is capable of predicting more than 96.5% of unseen data with model orders $\tau = 15 \sim 20$ (which is relatively low given the large number of latent nodes and the high order of the underlying brain dynamics). It should be noted that the R-value is not a suitable measure for comparison among the networks obtained by the LSAR, S+L, and dDTF methods. On the one hand, the AR model underlying the dDTF method is almost identical to the LSAR model used here, resulting in almost identical R values, while the reconstructed networks are considerably different (c.f. Figs. 7 and 9) due to different interpretations of on the model implications for network connectivity. On the other hand, the R value is not well-defined for the S+L method since the right-hand side of (33) is negative, i.e., the reconstructed AR model has extremely poor *prediction* performance. This is not surprising as the S+L method is aimed at maximizing the entropy (and thus minimizing predictability).

Next, we analyzed the effect of the choice of manifest nodes on the reconstructed network. In addition to selecting the 13 most anterior cortical nodes as above, we performed other runs where we selected the 13 most posterior nodes and 13 random nodes to reconstruct the manifest network using the LSAR method. We show in Fig. 10 these node choices (a), the reconstructed network for the posterior (b) and random (d) selections ($\alpha = 0.12$), and (c) the R values for all three cases. Interestingly, the density of direct connections is significantly higher among the posterior nodes. Also, the LSAR prediction performance is significantly lower in this case, suggesting less conformity of the occipito-parietal cortex to the simplifying assumptions of our AR model (linearity and passivity of latent nodes). Consistently, the network density and R value of the random case interpolates between the anterior and posterior cases, as expected.

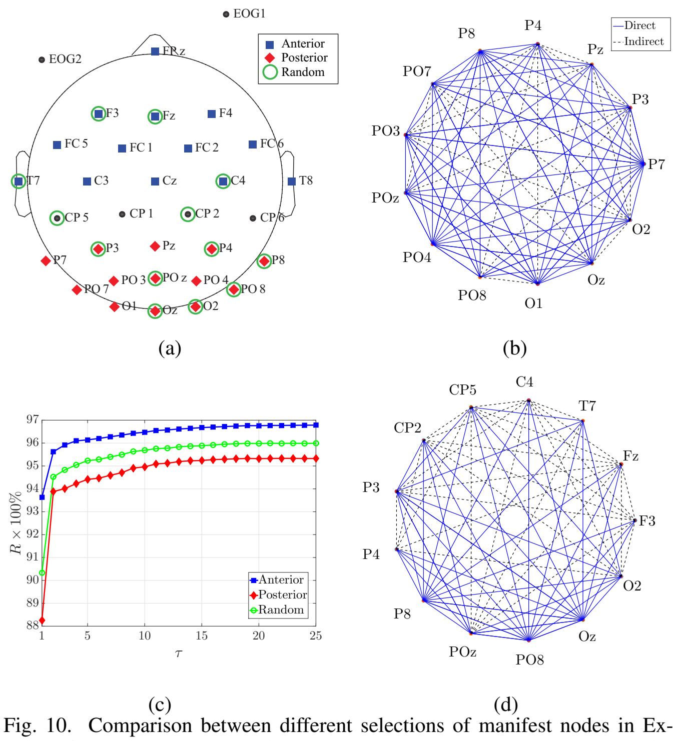
Finally, an interesting observation in Fig. 10(c) is that, even an AR model with $\tau = 2$ can predict about 95% of unseen data in all cases. This, at first glance, questions the need for any higher-order models as far as prediction is concerned. Nevertheless, notice that even an AR model with $\tau = 1$, corresponding to an *isolated* manifest subnetwork, can predict 90% of unseen data, while the visual discrimination task performed by the subject heavily relies on coordination between posterior (visual) and anterior (motor planning and execution) areas. The reason why this model can predict unseen data so well is in the strong dominance of first-order local dynamics of every area (the diagonal of $\tilde{A}_0$) over the rest of network dynamics.[7] Thus, the prediction performance of a first-order model serves as a *baseline* for higher orders, capturing the contribution of local interactions to the overall brain dynamics. This enlightens why the $\sim 1\%$ improvement in prediction performance as we go from $\tau = 2$ to $\tau = 15 \sim 20$ is significant.

VI. Conclusions

We have considered the problem of identifying the interaction structure among a group of nodes, termed manifest, that can be directly actuated and measured, and are part of a larger linear-time invariant network containing an unknown number of latent nodes. We have shown that, if there are no inputs to the latent nodes, then the transfer function of the manifest subnetwork can be approximated to any degree of accuracy by means of an auto-regressive model. We have proposed a least-squares estimation method that uses measured data to generate estimates that converge in probability to this AR model exponentially fast as the length of data and the model order increase. The estimation method does not require any knowledge of the number or the states of the latent nodes. We have illustrated our results in a directed ring network, a group of Erdős–Rényi random graphs, and on a time-series of EEG data recorded from the human brain. Future work will investigate the sensitivity of the estimation's performance to latent nodes, the characterization of particular network structures which are easier or more difficult to identify, the application of our results to the analysis of brain data, and the extension of the results to network models where, in addition to manifest and latent, there are nodes that can be actuated but not measured, and nodes that can be measured but not actuated.

[6] Edge values are not thresholded ($\alpha = 0$) for computing R values.

[7] This can be easily seen by inspecting the AR coefficients $\tilde{A}_i$ estimated from data, and is physiologically justified as each area is composed of millions of neurons that are locally densely connected and serve specific purposes but only (relatively) sparsely connected with remote areas.



(a) (b) (c) (d)

Fig. 10. Comparison between different selections of manifest nodes in Example 5.3: (a) Electrode locations. (b and d) The reconstructed network for the 13 posterior nodes and 13 random nodes, resp. ($\alpha = 0.12$). (c) Prediction performance R for the three different choices of manifest nodes (reconstructed network for anterior selection is given in Fig. 7(b)).

Acknowledgments

The authors would like to thank Dr. John Iversen for numerous discussions, helpful feedback on the contents of the paper, and introducing us to the SIFT Matlab toolbox. This work was supported by NSF Award CNS-1329619.

References

[1] Y. Zhao and J. Cortés, "Identification of linear networks with latent nodes," in *American Cont. Conf.*, (Boston, MA), pp. 173–178, July 2016.

[2] Y. X. R. Wang and H. Huang, "Review on statistical methods for gene network reconstruction using expression data," *Journal of Theoretical Biology*, vol. 362, pp. 53–61, 2014.

[3] A. Julius, M. Zavlanos, S. Boyd, and G. J. Pappas, "Genetic network identification using convex programming," *IET Systems Biology*, vol. 3, no. 3, pp. 155–166, 2009.

[4] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg," *Computers in Biology and Medicine*, vol. 41, no. 12, pp. 1110–1117, 2011.

[5] S. L. Bressler and A. K. Seth, "Wiener–Granger causality: a well established methodology," *Neuroimage*, vol. 58, no. 2, pp. 323–329, 2011.

[6] J. R. Iversen, A. Ojeda, T. Mullen, M. Plank, J. Snider, G. Cauwenberghs, and H. Poizner, "Causal analysis of cortical networks involved in reaching to spatial targets," in *Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, (Chicago, IL), pp. 4399–4402, 2014.

[7] A. Korzeniewska, C. M. Crainiceanu, R. Kuś, P. J. Franaszczuk, and N. E. Crone, "Dynamics of event-related causality in brain electrical activity," *Human Brain Mapping*, vol. 29, no. 10, pp. 1170–1192, 2008.

[8] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biological Cybernetics*, vol. 85, no. 2, pp. 145–157, 2001.

[9] C. D. Godsil and G. F. Royle, *Algebraic Graph Theory*, vol. 207 of *Graduate Texts in Mathematics*. Springer, 2001.

[10] J. Sun, D. Taylor, and E. M. Bollt, "Causal network inference by optimal causation entropy," *SIAM Journal on Applied Dynamical Systems*, vol. 14, no. 1, pp. 73–106, 2015.

[11] M. Nabi-Abdolyousefi and M. Mesbahi, "Network identification via node knockout," *IEEE Transactions on Automatic Control*, vol. 57, no. 12, pp. 3214–3219, 2012.

[12] S. Shahrampour and V. 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.

[13] M. Timme, "Revealing network connectivity from response dynamics," *Physical Review Letters*, vol. 98, no. 22, p. 224101, 2007.

[14] D. Materassi, G. Innocenti, L. Giarré, and M. V. Salapaka, "Model identification of a network as compressing sensing," *Systems & Control Letters*, vol. 62, no. 8, pp. 664–672, 2013.

[15] D. Materassi and G. Innocenti, "Topological identification in networks of dynamical systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1860–1871, 2010.

[16] M. J. Choi, V. Y. Tan, A. Anandkumar, and A. S. Willsky, "Learning latent tree graphical models," *Journal of Machine Learning Research*, vol. 12, pp. 1771–1812, 2011.

[17] D. Materassi and M. V. Salapaka, "Network reconstruction of dynamical polytrees with unobserved nodes," in *IEEE Conf. on Decision and Control*, (Maui, Hawaii, USA), pp. 4629–4634, 2012.

[18] J. Gonçalves and S. Warnick, "Necessary and sufficient conditions for dynamical structure reconstruction of LTI networks," *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1670–1674, 2008.

[19] Y. Yuan, G. B. Stan, S. Warnick, and J. Goncalves, "Robust dynamical network structure reconstruction," *Automatica*, vol. 47, no. 6, pp. 1230–1235, 2011. Special Issue on Systems Biology.

[20] Y. Yuan, K. Glover, and J. Goncalves, "On minimal realisations of dynamical structure functions," *Automatica*, vol. 55, pp. 159–164, 2015.

[21] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," *The Annals of Statistics*, vol. 40, no. 4, pp. 2005–2013, 2012.

[22] M. Zorzi and R. Sepulchre, "AR identification of latent-variable graphical models," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2327–2340, 2016.

[23] C. I. Byrnes, S. V. Gusev, and A. Lindquist, "From finite covariance windows to modeling filters: A convex optimization approach," *SIAM Review*, vol. 43, no. 4, pp. 645–675, 2001.

[24] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[25] A. Korzeniewska, M. Mańczak, M. Kamiński, K. J. Blinowska, and S. Kasicki, "Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method," *Journal of Neuroscience Methods*, vol. 125, no. 1, pp. 195–207, 2003.

[26] L. Ljung, *System Identification: Theory for the User*. Prentice Hall information and system sciences series, Prentice Hall, 1999.

[27] P. Henrici, *Applied and Comp. Complex Analysis, Volume 1: Power Series Integration Conformal Mapping Location of Zero*. Wiley, 1988.

[28] R. Durrett, *Probability: Theory and Examples*. Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 4th ed., 2010.

[29] A. Papoulis and S. U. Pillai, eds., *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.

[30] H. Weyl, "Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung)," *Math. Annalen*, vol. 71, pp. 441–479, 1912.

[31] B. Bollobás, *Random Graphs*. Cambridge University Press, 2 ed., 2001.

[32] T. Mullen, A. Delorme, C. Kothe, and S. Makeig, "An electrophysiological information flow toolbox for EEGLAB," *Biological Cybernetics*, vol. 83, pp. 35–45, 2010.

[33] A. Delorme, T. Mullen, C. Kothe, Z. A. Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig, "EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing," *Computational Intelligence and Neuroscience*, vol. 2011, p. 10, 2011.

[34] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[35] M. J. Kaminski and K. J. Blinowska, "A new method of the description of the information flow in the brain structures," *Biological Cybernetics*, vol. 65, no. 3, pp. 203–210, 1991.

Appendix

Lemma A.1: Given two vectors $a, b \in \mathbb{R}^n$, it holds for any $M \in \mathbb{R}_{>0}$ that $\|ab^T\|_{\max} \leq M^{-1}a^Ta + Mb^Tb$.

Proof. By definition of the max norm,

$$\begin{aligned}\|ab^T\|_{\max} &= \max_{1 \leq i,j \leq n} |a_i b_j| \leq \sum_{i=1}^{n} (M^{-1} |a_i|^2 + M |b_i|^2) \\ &= M^{-1}a^Ta + Mb^Tb.\end{aligned}$$

■