

Sparse Dynamic Network Reconstruction Through L_1 -regularization of a Lyapunov Equation

Ian Xul Belaustegui^a, Marcela Ordorica Arango^a, Roman Rossi-Pool^b,
Naomi Ehrich Leonard^a, Alessio Franci^c

Abstract—An important problem in many areas of science is that of recovering interaction networks from high-dimensional time-series of many interacting dynamical processes. A common approach is to use the elements of the correlation matrix or its inverse as proxies of the interaction strengths, but the reconstructed networks are necessarily undirected. Transfer entropy methods have been proposed to reconstruct directed networks, but the reconstructed network lacks information about interaction strengths. We propose a network reconstruction method that inherits the best of the two approaches by reconstructing a directed weighted network from noisy data under the assumption that the network is sparse and the dynamics are governed by a linear (or weakly nonlinear) stochastic dynamical system. The two steps of our method are *i*) constructing an (infinite) family of candidate networks by solving the covariance matrix Lyapunov equation for the state matrix and *ii*) using L_1 -regularization to select a sparse solution. We further show how to use prior information on the (non)existence of a few directed edges to dramatically improve the quality of the reconstruction.

I. INTRODUCTION

In many areas of science it is common to record the evolution in time of many interacting processes. A natural and important problem is then to develop data-driven methods that can recover, or reconstruct, from the resulting high-dimensional time-series the underlying interaction network. Ideally, the reconstructed network would include information about both the direction and the strength of the interactions. Interaction networks reconstruction has applications, e.g., in neuroscience (between neurons or neuronal populations), in ecology (i.e., between species populations in a community), in molecular biology (i.e., between expression of genes), and in social sciences (i.e., between opinionated individuals).

An established approach, commonly employed in neuroscience applications [1]–[3], is to use correlation measures to approximate interaction networks through the entries of the covariance matrix or its inverse, the precision matrix. Networks reconstructed from the precision matrix were found to better correlate to the underlying anatomical network [1]. However, whether using the correlation or the precision matrix, the reconstructed network will necessarily be undirected because both matrices are symmetric.

^a Department of Mechanical and Aerospace Engineering, Princeton University; {ianxul,m.ordorica,naomi}@princeton.edu

^b Instituto de Fisiología Celular - Neurociencias and Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, México; romanr@ifc.unam.mx

^c Department of Electrical Engineering and Computer Science, University of Liege, and WEL Research Institute, Wavre, Belgium; afranci@uliege.be

In order to reconstruct directed network interactions, transfer entropy (TE) [4] has emerged as a powerful information-theoretical measure of directed information transfer between dynamical processes, with many applications in neuroscience [5]–[9]. However, transfer entropy methods can only reconstruct unweighted directed interaction networks, i.e., they can only provide information about whether a process directly affects (or not) another but not the strength or sign of this directed interaction.

A commonly used, mathematically tractable approximation is to assume that the network dynamics can be modelled by a linear (or weakly nonlinear) stochastic differential equation driven by a white noise vector. When used to describe resting-state brain activity, the resulting model is known as the (linear) Dynamical Causal Model (DCM) [10], [11]. Under this assumption, the covariance matrix of the model is the solution to a Lyapunov equation that links it to the state matrix of the network [12], [13]. If the state matrix is assumed to be symmetric (i.e., the network is undirected) and the noise components are uncorrelated, then solving the Lyapunov equation for the state matrix returns exactly the precision matrix [1]. But if no symmetry assumptions are made, then solving the Lyapunov equation for the state matrix returns a whole affine subspace of candidate solutions. Methods using differential covariance [2] enable singling out asymmetric solutions, but these methods require computing the time-derivative of the measured time-series, which can be problematic or unfeasible when measurement or intrinsic noise are high.

We introduce a novel method to reconstruct weighted, thus directed, dynamical networks, using the idea that under the extra assumption that the interaction network is sparse, L_1 -regularization can effectively select such a sparse solution from among the candidates. We also show how prior knowledge of edge existence can be included in the constraints of the resulting linear programming (LP) problem to enhance the method performance.

The first and main contribution of the paper is to formulate an L_1 -regularization problem for sparse network reconstruction as an LP optimization problem over the affine solution space of all state matrices that solve the covariance Lyapunov equation for a given covariance matrix. Crucially, this requires the preliminary construction of an isomorphism from such a solution space to the solution space of an underdetermined linear system of equations. The second contribution is to show how to introduce prior knowledge on edge existence into the constraints of the LP problem. The

third contribution is to suggest the use of a TE-based network inference algorithm as a way of obtaining some priors on edge existence and show that it significantly improves the quality of network reconstruction.

The paper is structured as follows. Notation is introduced in Section II. The network model and its interpretation are introduced in Section III. Section IV studies the geometry of the solution space of the covariance Lyapunov equation for the state matrix and introduces the isomorphism needed for the LP formulation of an L_1 -regularization problem over this space. Section V explicitly describes the LP formulation of the L_1 -regularization problem. Section VI shows how prior information about edge existence can be incorporated into the constraints of the LP problem. Section VII introduces an algorithm based on TE to provide prior knowledge on edge existence and show how to use it to constrain the LP problem. Finally, Section VIII presents extensive numerical validation results obtained over large families of randomly generated sparse Hurwitz matrices. Conclusions and perspectives are presented in Section IX.

II. NOTATION

We denote $[n] = \{1, \dots, n\}$. Given $n, m \in \mathbb{N}$, we write $\mathbb{R}^{m \times n}$ for the set of m by n matrices with real entries. For $i \in [m]$, and $j \in [n]$, the value at the i -th row and j -th column of $A \in \mathbb{R}^{m \times n}$ is $A_{i,j}$. The column-stack vectorization isomorphism $\text{vec} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{mn}$ is given by $(\text{vec}(A))_{m(j-1)+i} = A_{i,j}$. Vector $\mathbf{1}_n \in \mathbb{R}^n$ consists of n entries equal to 1 (and similarly for $\mathbf{0}_n$). The set of eigenvalues of $B \in \mathbb{R}^{n \times n}$ is $\sigma(B)$. The real part of $z \in \mathbb{C}$ is denoted $\Re(z)$. For a pair of matrices $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times n}$, \otimes is the Kronecker product given by $\forall r, v, s, w \in [n], (A \otimes B)_{n(r-1)+v, n(s-1)+w} = (A_{r,s})(B_{v,w})$, so that $A \otimes B \in \mathbb{R}^{n^2 \times n^2}$.

III. MODEL FORMULATION

We consider a network model described by the linear stochastic differential equation (SDE) [13]

$$\frac{dx}{dt}(t) = Ax(t) + w(t) \quad (1)$$

where w is a white noise vector (with noise covariance $Q = I$) and $A \in \mathbb{R}^{n \times n}$ is a Hurwitz matrix, i.e., all of its eigenvalues have a strictly negative real part. Any solution x for system (1) is a stochastic process with mean 0 and covariance matrix $\Gamma = \mathbf{E}(xx^T)$ (see for example Section 4.3 in [13]).

In a network modeling setting, A is interpreted as a weighted and signed adjacency matrix. A non-zero element $A_{ij} \neq 0$ is interpreted as the existence of a directed edge from node j to node i . The sign of A_{ij} determines the nature of the interaction (excitatory or inhibitory) and $|A_{ij}|$ determines its strength.

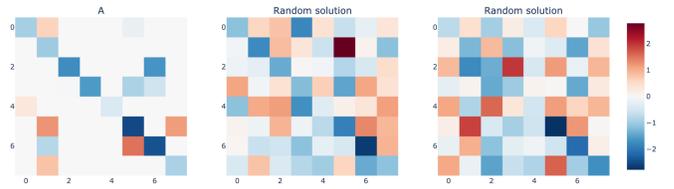


Fig. 1. Heatmaps for three examples of weighted connectivity matrices of dimension 8×8 . The leftmost heatmap is for a sparse matrix $A \in \mathbb{R}^{8 \times 8}$ generated randomly using the algorithm in Section VIII. The other two matrices were selected randomly from the solution space \mathcal{S}_Γ for the matrix A . This means the solution to the system in (1) has the same covariance matrix for all three matrices.

IV. SOLVING THE COVARIANCE LYAPUNOV EQUATION FOR THE STATE MATRIX

It is known [12], [13] that A and Γ satisfy the following Lyapunov equation:

$$\Gamma A^T + A\Gamma = -I. \quad (2)$$

Given A , the stable covariance matrix solution is given by

$$\text{vec}(\Gamma) = -(I \otimes A + A \otimes I)^{-1} \text{vec}(I),$$

which is readily computed. However, because the covariance matrix is symmetric, the inverse problem, i.e., solving for A given Γ , does not have a unique solution [14]. The differential covariance approach [2] uses the fact that the expected value of $\dot{x}x^T$ is equal to $A\Gamma$ to obtain a unique inverse solution A . However, when measurement or intrinsic noise are large, computing time-series derivatives can be problematic or unfeasible. Our goal is to propose a new algorithmic method that, under suitable assumptions, can single out a unique solution to the inverse covariance Lyapunov equation (2). Following [14], given Γ , define its orthonormal spectral decomposition:

$$\Gamma = UCU^T \quad (3)$$

where U is an orthonormal matrix and C is a diagonal matrix of entries c_i , the eigenvalues of Γ . Let $\bar{A} = U^T A U$. Then we can rewrite (2) as

$$U^T (UCU^T A^T + AUCU^T) U = C\bar{A}^T + \bar{A}C = -I$$

which has solutions \bar{A} easily shown to satisfy

$$\bar{A}_{i,j} = \begin{cases} \frac{-1}{2c_i}, & \text{if } i = j \\ -\frac{c_i}{c_j} \bar{A}_{j,i}, & \text{if } i \neq j \end{cases} \quad (4)$$

The set of matrices \bar{A} satisfying (4) is an affine subspace of dimension $m = \frac{n(n-1)}{2}$. Since $A = U\bar{A}U^T$, this means that the set of possible solutions A to the inverse covariance Lyapunov equations is also an m -dimensional affine space $\mathcal{S}_\Gamma \subseteq \mathbb{R}^{n \times n}$. Figure 1 shows a Hurwitz matrix A , with correlation matrix Γ , together with other two matrices taken randomly from the solution space \mathcal{S}_Γ , hence associated to the same covariance matrix.

The following proposition characterizes the matrices in \mathcal{S}_Γ with a linear equality constraint on the vectorized matrices.

It is needed to formulate an L_1 -regularization problem over \mathcal{S}_Γ as a LP problem.

Proposition 4.1: Given a covariance matrix $\Gamma = UCU^T$, define a matrix \mathbf{M} of dimensions $\frac{n(n+1)}{2} \times n^2$ and a vector \mathbf{b} of dimension $\frac{n(n+1)}{2}$ with entries, respectively, given by

$$\mathbf{M}_{\iota_1(i,j),\iota_2(\ell,k)} = \begin{cases} c_j U_{\ell,i} U_{k,j} + c_i U_{\ell,j} U_{k,i} & \text{if } i \neq j \\ U_{\ell,i} U_{k,j} & \text{if } i = j \end{cases}$$

and

$$\mathbf{b}_{\iota_1(i,j)} = \begin{cases} 0 & \text{if } i \neq j \\ \frac{-1}{2c_i} & \text{if } i = j \end{cases}$$

where $\iota_1 : \{(i,j) \mid i,j \in [n], i \leq j\} \rightarrow [\frac{n(n+1)}{2}]$ and $\iota_2 : [n]^2 \rightarrow [n^2]$ are one-to-one functions that give the index of the vectorizations of the upper triangle of \bar{A} (including the diagonal) and of the full matrix respectively. Then for $v \in \mathbb{R}^{n^2}$, we have that $\text{vec}^{-1}(v) \in \mathcal{S}_\Gamma$ if and only if $\mathbf{M}v = \mathbf{b}$.

Proof: Let $v \in \mathbb{R}^{n^2}$ and take $A = \text{vec}^{-1}(v)$. Suppose $A \in \mathcal{S}_\Gamma$. From (4), if $i \neq j$ then $\sum_{\ell=1}^n \sum_{k=1}^n U_{\ell,i} A_{\ell,k} U_{k,j} = (U^T A U)_{i,j} = \bar{A}_{i,j} = -\frac{c_i}{c_j} \bar{A}_{j,i} = -\frac{c_i}{c_j} \sum_{\ell=1}^n \sum_{k=1}^n U_{\ell,j} A_{\ell,k} U_{k,i}$, from which it follows that

$$\sum_{\ell=1}^n \sum_{k=1}^n (c_j U_{\ell,i} U_{k,j} + c_i U_{\ell,j} U_{k,i}) A_{\ell,k} = 0. \quad (5)$$

Similarly, when $i = j$,

$$\bar{A}_{i,j} = \sum_{\ell=1}^n \sum_{k=1}^n U_{\ell,i} U_{k,j} A_{\ell,k} = \frac{-1}{2c_i}. \quad (6)$$

Equations (5),(6) are precisely the $\frac{n(n+1)}{2}$ identities that are encoded in \mathbf{M} and \mathbf{b} , which implies that $\mathbf{M}v = \mathbf{b}$.

Following the same argument, for any $v \in \mathbb{R}^{n^2}$, if $\mathbf{M}v = \mathbf{b}$, then $\text{vec}^{-1}(v) \in \mathcal{S}_\Gamma$. ■

V. L_1 -REGULARIZATION ON \mathcal{S}_Γ AND SPARSE NETWORK RECONSTRUCTION

Assume that the interaction network we are reconstructing is sparse. Then, a reasonable way to recover a particular solution from \mathcal{S}_Γ is to look for matrices with minimum L_1 norm; this is a common approach for promoting sparsity of solutions [15]. We pose the problem of minimizing the L_1 norm of the vectorized matrices as follows:

$$\begin{aligned} & \text{Minimize } \|v\|_1 \\ & \text{Subject to } \mathbf{M}v = \mathbf{b}. \end{aligned} \quad (7)$$

In turn, (7) can be transformed into a Linear Programming (LP) problem by taking a new variable $u \in \mathbb{R}^{n^2}$ and a vector $\mathbf{s} = (\mathbf{1}_{n^2}, \mathbf{0}_{n^2})$, and defining the LP problem:

$$\begin{aligned} & \text{Minimize} && \mathbf{s}^T \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{Subject to} && \begin{bmatrix} \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{b} \\ & \text{and} && \begin{bmatrix} -I & I \\ -I & -I \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \leq \mathbf{0}_{2n^2}. \end{aligned} \quad (8)$$

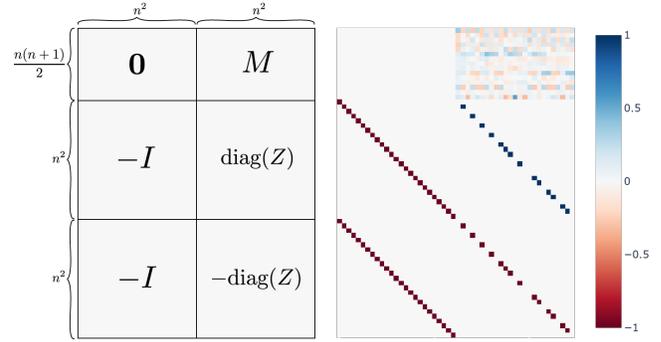


Fig. 2. Diagram illustrating the way the optimization problem (7) can be encoded as a linear programming problem. The top $\frac{n(n+1)}{2} \times 2n^2$ sub-matrix is used as an equality constraint, which restricts the search to matrices on \mathcal{S}_Γ . The $2n^2 \times 2n^2$ sub-matrix below is used as an inequality constraint which ensures that for each ℓ, k we have $|v_{(\ell,k)}| \leq u_{(\ell,k)}$. An example of this encoding for the case $n = 5$ is show on the right.

The LP problem (8) minimizes the objective $\sum_{(i,j)} u_{(i,j)}$ subject to the restriction that $\text{vec}^{-1}(v) \in \mathcal{S}_\Gamma$ and $|v_{(i,j)}| \leq u_{(i,j)}$ for all $i, j \in [n]$. A solution to the LP problem (8) immediately gives a solution to the problem (7) and *vice-versa* – in this sense the two problems are equivalent [15].

Solving (8) will return a sparse solution from \mathcal{S}_Γ . However, because there might be (and usually are) many sparse matrices in \mathcal{S}_Γ , the solution to the L_1 -regularization problem might still be very different from the true network adjacency matrix, including under the sparsity assumption and mostly depending on the initial problem data. In the following section we present a way to constrain the optimization problem by including approximate prior knowledge on edge existence to bias the reconstruction toward solutions that are consistent with the priors.

VI. INCLUDING PRIORS ON EDGE EXISTENCE

The idea for including priors comes from the observation that if we knew that exactly p directed edges (and which ones) existed in the network, then A would lie in $\mathcal{R}_\Gamma \subseteq \mathbb{R}^{n \times n}$ of dimension p , i.e., the subspace parameterized by the p nonzero elements of A . By Proposition 4.1, $A \in \mathcal{S}_\Gamma$. Thus, $A \in \mathcal{S}_\Gamma \cap \mathcal{R}_\Gamma$, that is, A must lie in the intersection of two affine subspaces of dimension $m = \dim(\mathcal{S}_\Gamma) = \frac{n(n-1)}{2}$ and $p = \dim(\mathcal{R}_\Gamma)$, respectively.

It is well known (a consequence of Theorem 2.43 in [16]) that the transversal intersection of two linear subspaces of $\mathbb{R}^{n \times n}$ of dimension p and m , with $m + p \leq n^2$ is either empty or a point. This means that if we have *exact* knowledge of which edges exist in the network and the network has a sufficiently small number of connections (i.e., $m + p \leq n^2$), then solving (8) restricted to \mathcal{R}_Γ would return A *exactly*. However, given only imperfect knowledge of edge existence and if $m + p \leq n^2$, then our noisy approximation $\hat{\mathcal{R}}_\Gamma$ of \mathcal{R}_Γ and \mathcal{S}_Γ will almost surely not intersect at all. In spite of this, looking for the matrix $A \in \mathcal{S}_\Gamma$ with minimal distance to $\hat{\mathcal{R}}_\Gamma$ should bias the solution (8) toward the real A . This is the approach we use here.

Based on this idea we define a vector with entries representing the relative importance assigned to each connection during the optimization, $Z \in [0, 1]^{n^2}$. The objective of the minimization then becomes $\sum_{(i,j)} Z_{(i,j)} |v_{(i,j)}|$. In the simplest case, we can introduce information about the presence of edges in the following way. Let $\mathcal{I} = \{(i, j) \in [n]^2 \mid (i, j) \text{ is a known edge}\}$. Then we can define Z as

$$Z_{(i,j)} = \begin{cases} 1 & \text{if } (i, j) \notin \mathcal{I} \\ 0 & \text{otherwise.} \end{cases}$$

Written as an LP the problem becomes

$$\begin{aligned} & \text{Minimize} && \mathfrak{s}^T \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{Subject to} && \begin{bmatrix} \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{b} \\ & \text{and} && \begin{bmatrix} -I & \text{diag}(Z) \\ -I & -\text{diag}(Z) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \leq \mathbf{0}_{2n^2}. \end{aligned} \quad (9)$$

When there are no edge priors, $Z = \mathbf{1}_{n^2}$, and the problem in (9) reduces to (8). An illustrative example of the constraint matrix in (9) is shown in Figure 2. Note in Figure 2 that this matrix is sparse in most places except the top right corner; thus, optimization methods that take advantage of sparse structure can be employed. Problem (9) leads to an optimization that prioritizes finding a solution where the edges *not in* \mathcal{I} have weight close to zero in absolute value. The values of Z could also be adjusted to correspond with the level of certainty about prior edge existence. For simplicity, in the following we only consider $Z \in \{0, 1\}^{n^2}$.

VII. EDGE PRIORS FROM TRANSFER ENTROPY

There are various ways to obtain prior knowledge on edge existence, e.g., from expert knowledge or from previous analysis. One generally applicable tool for inferring the existence of directed interactions is Transfer Entropy (TE) [4]. Given two discrete-time stationary stochastic processes X and Y , the TE from X to Y is a measure of the predictive power of X about Y at a *discrete* time \mathfrak{t} . In its simplest form TE is defined as [4]

$$\begin{aligned} T_{X \rightarrow Y}(\mathfrak{t}) &= I(X(\mathfrak{t}-1); Y(\mathfrak{t}) \mid Y(\mathfrak{t}-1)) \\ &= H(Y(\mathfrak{t}) \mid Y(\mathfrak{t}-1)) - H(Y(\mathfrak{t}) \mid X(\mathfrak{t}-1), Y(\mathfrak{t}-1)). \end{aligned}$$

So, $T_{X \rightarrow Y}(\mathfrak{t})$ is the mutual information I of $X(\mathfrak{t}-1)$ and $Y(\mathfrak{t})$ conditioned on $Y(\mathfrak{t}-1)$, where

$$I(X; Y \mid Z) = \int_{\mathcal{Z}} \int_{\mathcal{Y}} \int_{\mathcal{X}} p_{X,Y,Z}(x,y,z) \ln \left(\frac{p_Z(z) p_{X,Y,Z}(x,y,z)}{p_{X,Z}(x,z) p_{Y,Z}(y,z)} \right) dx dy dz,$$

p are probability density functions of the joint distributions of variables represented by the subscripts, and $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are the supports of X, Y, Z respectively. $T_{X \rightarrow Y}(\mathfrak{t})$ can equivalently be written in terms of $H(X \mid Y) = -\int_{\mathcal{X}} \int_{\mathcal{Y}} p_{X,Y}(x,y) \ln \left(\frac{p_{X,Y}(x,y)}{p_Y(y)} \right)$. If the processes are assumed to be stationary then TE is the same for all values of \mathfrak{t} and we can simply write $T_{X \rightarrow Y}$.

This measure has been shown to be a good indicator of directed interactions between pairs of noisy dynamical

variables [7], [17], [18], and can be seen as a nonlinear generalization of Granger causality [19]. It can also be extended to take into account factors such as redundancy and synergy in a partial information decomposition approach [20].

A simple way to reduce the effects of redundancy when inferring connectivity is to use *conditional* TE: $T_{X \rightarrow Y \mid Z}$. The definition of $T_{X \rightarrow Y \mid Z}$ is the same as for $T_{X \rightarrow Y}$ but conditioned on the past of Z . Connectivity can then be inferred from a set of n stochastic processes X_i by calculating $T_{X_i \rightarrow X_j \mid X_{(i,j)}}$ for each pair $1 \leq i, j \leq n$, where $X_{(i,j)}$ is the joint distribution of all the variables except X_i and X_j . Then we say there is a significant interaction from X_i to X_j if $T_{X_i \rightarrow X_j \mid X_{(i,j)}}$ is higher than what would be expected by pure chance (using suitable data shuffling and significance testing). However, this approach is usually not feasible for large networks due to the amount of required data.

To bypass this problem in cases where the number of edges in the network is small (i.e., the connectivity matrix is sparse), Novelli et al. [7] proposed a greedy algorithm that finds a set of sources for each target node by *i*) creating a candidate set of source nodes with significant TEs when conditioned only on previously selected nodes and *ii*) removing all nodes for which TE became non-significant due to redundancy with newly added nodes. This algorithm (and others) is implemented in the IDTxl Python package [21].

In order to keep inference fast, we adapted the greedy algorithm in [7] to a simplified version where only Step *i*) of the algorithm is used. This did not seem to drastically impact the performance of the method. However, we observed an increase in false positives, so we introduced a strict upper bound on the number of inferred edges.

VIII. NUMERICAL VALIDATION

We tested our method on randomly generated matrices of the form

$$A = -I + \frac{1 - \varepsilon}{b_{max}} B,$$

where B is an Erdős-Rényi graph with normally distributed edge weights, $\varepsilon \in (0, 1]$ is a parameter that controls the spectrum of A by weighting B , and $b_{max} = \max\{\Re(\lambda) \mid \lambda \in \sigma(B)\}$. We get $a_{max} = \max\{\Re(\lambda) \mid \lambda \in \sigma(A)\} = -1 + \varepsilon$.

A. Implementation

For numerical validation, our algorithm was implemented in standard Python libraries. The code can be found at <https://github.com/ianxul/SDE-net-reconstruction>. For estimating conditional TE interactions, we used the GPU-enabled estimator in IDTxl package [21], and the adapted algorithm described in Section VII. We further simplified the permutation-based significance test suggested in [7] by using a less computationally intensive threshold heuristic similar to [18]. L_1 -optimization was performed using CVXPY package [22] with the splitting conic solver (SCS).

Simulations of SDEs were run with Julia's DifferentialEquations solver. We used the LambdaEM solver for SDEs with a step size of 0.1 for 10,000 time steps (giving 100,000 data

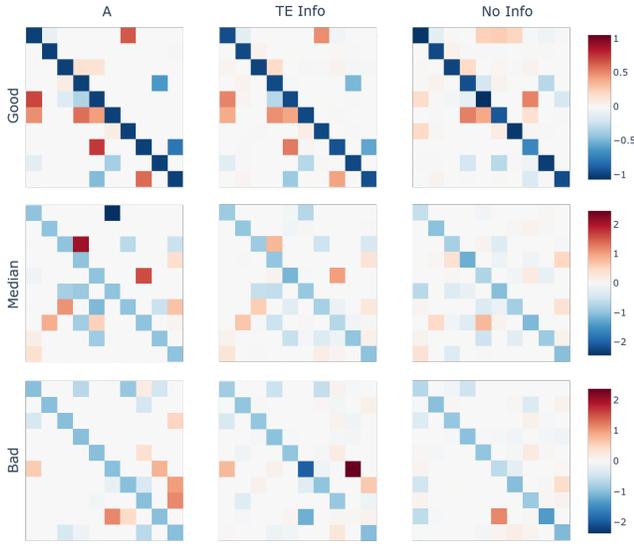


Fig. 3. Three examples of reconstructed connected networks using our method with edge priors inferred using TE (middle), and no info (right), applied to time-series generated from the weakly-nonlinear system in (10). The ground truth connectivity matrix A is shown on the left. The top row shows examples of good reconstruction (high alignment between reconstructed matrix and the ground truth), the middle row shows reconstruction with median alignment, and the bottom row shows a bad reconstruction (low alignment). The alignments for the reconstructions with TE Info were 0.99, 0.90 and 0.31 for the good, median, and bad respectively. For the No Info the alignments were 0.43, 0.42, and 0.53 for the good, median, and bad respectively. Examples were selected from 600 simulated dynamics of matrices with 20 edges and different ε values by sorting according to alignment of TE Info reconstruction and selecting the best, median, and worst reconstructions.

points for each simulation). We applied our method to the data generated from the linear SDE system (1) and from the weakly nonlinear (saturated network interactions) SDE system

$$\frac{dx_i}{dt} = A_{i,i}x_i + \sum_{i \neq j} (\tanh(A_{i,j}x_j)) + w(t). \quad (10)$$

To compare different methods we computed the alignment between the vectorized matrices, excluding the diagonal, as it is not related to network reconstruction:

$$\text{align}(A, B) = \frac{\sum_{i \neq j} A_{i,j} B_{i,j}}{\left(\sum_{i \neq j} A_{i,j}^2\right)^{\frac{1}{2}} \left(\sum_{i \neq j} B_{i,j}^2\right)^{\frac{1}{2}}}. \quad (11)$$

We compared three cases: *i*) “Full Info”, in which full information on edge existence is given and which should lead to perfect reconstruction; *ii*) “TE Info”, in which edge existence is inferred using TE methods; *iii*) “No Info”, in which we simply perform L_1 -regularization over \mathcal{S}_Γ as in (8). Results are exemplified in Figure 3 and summarized in Figure 4.

B. Results and discussion

Our results show that adding edge existence priors obtained through TE [7] to our reconstruction method improves the alignment of reconstructions considerably when compared against doing L_1 -regularization with no prior

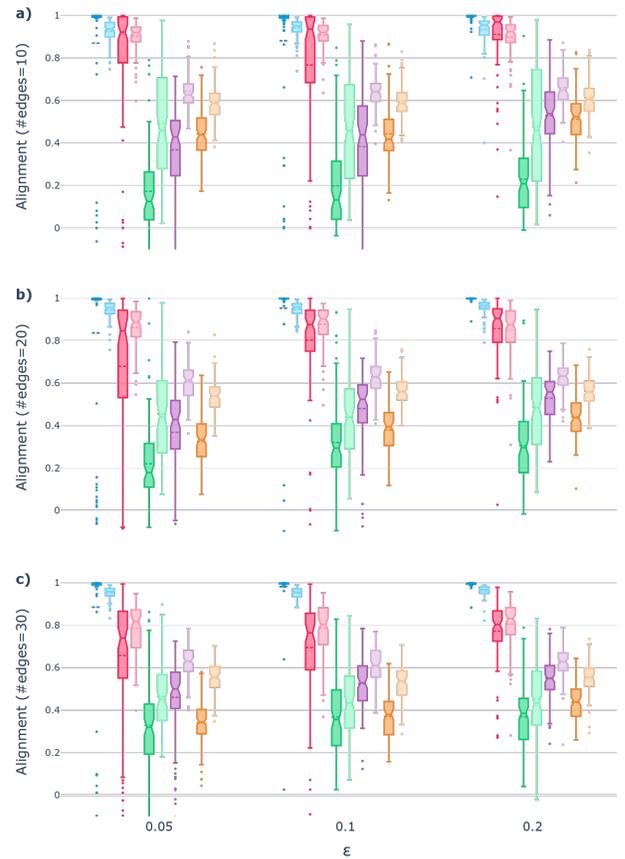


Fig. 4. Plots showing the performance of our method for 100 samples of random Hurwitz matrices of size 10×10 generated with different values of ε , and with **a)** 10, **b)** 20, **c)** 30 edges. The L_1 optimization methods compared are: Full Info (blue), TE Info (red), No Info (green) on edge existence. Performance of precision and correlation matrices were added as reference in purple and orange respectively. Darker colors correspond to reconstructions for data simulated using the linear system in (1), and lighter colors for the weakly-nonlinear system in (10). Notches in boxes represent median values and dotted lines mean values. Boxes encompass 50% of the data points. Isolated points are outliers.

information, and against the precision and correlation matrices, as can be seen in Figures 3 and 4. Interestingly, L_1 -regularization with no edge priors performed worse than the symmetric methods for both the linear and weakly nonlinear cases. A possible reason is that for a given covariance matrix the solution space \mathcal{S}_Γ might contain many matrices with small L_1 norm and, due to noise, one of these (potentially very far from the true solution) can turn into a global minimum. We observed, for instance, cases where edges were flipped (see the top right matrix in Figure 3). By biasing the optimization towards matrices consistent with prior knowledge we increase the likelihood that the correct sparse solution is selected.

All methods performed increasingly worse as the ground truth solution A became less sparse or as ε became smaller (i.e., as the spectrum of A got closer to the imaginary axis). This could simply be a result of not having an optimal algorithm for inferring edges using TE, since the performance of the method with full prior knowledge of existing edges (“Full

Info”) is mostly unaffected by changes in ε and edge number. Applying the full IDTxI method [7] could improve results further due to more rigorous significance tests, a redundant edge cleanup pass and no hard limit on inferred edges. The mean and median performance of the “TE Info” method was always strictly better than all other methods we tested (apart, of course, from “Full Info”) by a significant margin (bootstrap test $p < 10^{-5}$).

Another interesting observation is that introducing a weak nonlinearity in the model usually led to only minor changes in the median performance but often drastically reduced the performance variance, and in some cases increased the median performance as well. In other words, our method performed more consistently (and sometimes better) in the presence of weakly nonlinear network interactions as compared to a purely linear network model. This effect cannot be explained by the use of nonlinear edge inference methods like TE, since fully linear methods like “No Info” and those based on precision and correlation matrices also benefited from the presence of weakly nonlinear interactions. We will investigate this interesting phenomenon in future research.

With respect to computational complexity, our model is fast and high performing since it has the computational complexity of LP optimization (not including the prior edge inference with TE, which is highly parallelizable [21]). This has been shown to be theoretically less than or equal to the current complexity of matrix multiplication [23], which is $\approx O(\#^{2.3})$, where $\#$ is the number of variables in the optimization. In our case, the number of variables is $\# = 2n^2$, and so the (theoretical) computational complexity of the method is $\approx O(n^{4.6})$. For reference, solving the LP optimization for a 100-dimensional system takes about 8 minutes on a single core CPU node with 2.6 GHz Intel Skylake processor and 16 GB of memory.

IX. CONCLUSIONS AND PERSPECTIVES

We introduced a fast and high performing method to reconstruct directed and weighted networks and analyzed its performance using both linear and weakly nonlinear network models. Future works will aim at improving the heuristics used for the TE inference step. Theoretically it will also be important to understand the role of nonlinearities on network inference, which might be key for application to (usually nonlinear) neuronal and other kind of biological networks. Measures of the intrinsic timescale of each node, estimated through auto-correlation analysis [24], can further be used to further constraint the L_1 -optimization step. This will also require consideration of the role of network self-loops, i.e., the diagonal elements of A , more explicitly in the method. All these ideas will soon be implemented and tested on experimental neuroscience (EEG) datasets.

REFERENCES

[1] R. Liégeois, A. Santos, V. Matta, D. V. D. Ville, and A. H. Sayed, “Revisiting correlation-based functional connectivity and its relationship with structural connectivity,” *Network Neuroscience*, vol. 4, no. 4, p. 1235–1251, 2020.

[2] T. W. Lin, A. Das, G. P. Krishnan, M. Bazhenov, and T. J. Sejnowski, “Differential covariance: A new class of methods to estimate sparse connectivity from neural recordings,” *Neural Computation*, vol. 29, no. 10, p. 2581–2632, 2017.

[3] R. Mohanty, W. A. Sethares, V. A. Nair, and V. Prabhakaran, “Re-thinking measures of functional connectivity via feature extraction,” *Scientific Reports*, vol. 10, no. 1, p. 1298, 2020.

[4] T. Schreiber, “Measuring information transfer,” *Physical Review Letters*, vol. 85, no. 2, p. 461–464, Jul. 2000.

[5] M. Lobier, F. Siebenhühner, S. Palva, and J. M. Palva, “Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions,” *NeuroImage*, vol. 85, p. 853–872, 2014.

[6] M. Ursino, G. Ricci, and E. Magosso, “Transfer entropy as a measure of brain connectivity: A critical analysis with the help of neural mass models,” *Frontiers in Computational Neuroscience*, vol. 14, 2020.

[7] L. Novelli, P. Wollstadt, P. Mediano, M. Wibral, and J. T. Lizier, “Large-scale directed network inference with multivariate transfer entropy and hierarchical statistical testing,” *Network Neuroscience*, vol. 3, no. 3, pp. 827–847, 2019.

[8] N. M. Timme and C. Lapish, “A tutorial for information theory in neuroscience,” *eneuro*, vol. 5, no. 3, pp. ENEURO.0052–18.2018, 2018.

[9] P. Sharma, D. J. Bucci, S. K. Brahma, and P. K. Varshney, “Communication network topology inference via transfer entropy,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, p. 562–575, 2017.

[10] K. J. Friston, L. Harrison, and W. Penny, “Dynamic causal modelling,” *Neuroimage*, vol. 19, no. 4, pp. 1273–1302, 2003.

[11] U. Casti, G. Baggio, D. Benozzo, S. Zampieri, A. Bertoldo, and A. Chiuso, “Dynamic brain networks with prescribed functional connectivity,” in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 709–714.

[12] L. Socha, *Linearization methods for stochastic dynamic systems*. Springer Science & Business Media, 2007, vol. 730.

[13] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, 1st ed., ser. Institute of Mathematical Statistics textbooks. Cambridge University Press, 2019, vol. 10.

[14] K. Fernando and H. Nicholson, “Solution of lyapunov equation for the state matrix,” *Electronics Letters*, vol. 17, no. 5, pp. 204–205, 1981.

[15] E. Candes and J. Romberg, “11-magic: Recovery of sparse signals via convex programming,” *URL: www.acm.caltech.edu/11magic/downloads/11magic.pdf*, vol. 4, no. 14, p. 16, 2005.

[16] S. Axler, *Linear algebra done right*. Springer Nature, 2023.

[17] T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, *An Introduction to Transfer Entropy*. Cham: Springer International Publishing, 2016.

[18] A. Rahimzamani and S. Kannan, “Network inference using directed information: The deterministic limit,” in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Monticello, IL, USA: IEEE, 2016, p. 156–163.

[19] L. Barnett, A. B. Barrett, and A. K. Seth, “Granger causality and transfer entropy are equivalent for gaussian variables,” *Physical Review Letters*, vol. 103, no. 23, p. 238701, 2009.

[20] P. L. Williams and R. D. Beer, “Generalized measures of information transfer,” no. arXiv:1102.1507, 2011, arXiv:1102.1507 [physics].

[21] P. Wollstadt, J. T. Lizier, R. Vicente, C. Finn, M. Martínez-Zarzuela, P. Mediano, L. Novelli, and M. Wibral, “Idtxl: The information dynamics toolkit xl: a python package for the efficient analysis of multivariate information dynamics in networks,” *Journal of Open Source Software*, vol. 4, no. 34, p. 1081, 2019, arXiv:1807.10459 [cs, math].

[22] S. Diamond and S. Boyd, “CVXPY: A Python-embedded modeling language for convex optimization,” *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[23] M. B. Cohen, Y. T. Lee, and Z. Song, “Solving linear programs in the current matrix multiplication time,” *Journal of the ACM (JACM)*, vol. 68, no. 1, p. 1–39, 2021.

[24] R. Rossi-Pool, A. Zainos, M. Alvarez, S. Parra, J. Zizumbo, and R. Romo, “Invariant timescale hierarchy across the cortical somatosensory network,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 3, p. e2021843118, 2021.