UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Reconstruction of sparse network dynamics from data

Edmilson Roque dos Santos

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP SÃO CARLOS

Edmilson Roque dos Santos

Reconstruction of sparse network dynamics from data

Thesis submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Doctor in Science. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Tiago Pereira da Silva

USP – São Carlos
March 2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

R786r

Roque dos Santos, Edmilson
Reconstruction of sparse network dynamics from data / Edmilson Roque dos Santos; orientador Tiago Pereira da Silva. -- São Carlos, 2024.
152 p.

Tese (Doutorado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2024.

1. Network dynamics. 2. Sparse networks. 3. Dynamical Systems. 4. Ergodic theory. 5. Sparse Recovery methods. I. Pereira da Silva, Tiago , orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Edmilson Roque dos Santos

Reconstrução de dinâmica de redes esparsas a partir de dados

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Tiago Pereira da Silva

USP – São Carlos
Março de 2024

ACKNOWLEDGEMENTS

First and foremost, I am deeply indebted to my advisors. I would like to thank, Tiago Pereira, for his unwavering support, invaluable guidance, and patience. The way I have grown during the past few years (8 years to be precise) has undoubtedly his influence. Also, I am deeply thankful to Sebastian van Strien for his guidance, time, and patience, always asking the hard questions that until today I remember. Their expertise has been instrumental in shaping the direction of my research and helping me navigate the challenges along the way.

I would like to express my gratitude to Jeroen Lamb for his support, valuable insights, constructive discussions, and guidance. Always allowed me to stay around, fostering my research development.

I want to express my gratitude to all individuals who directly or indirectly contributed throughout the journey of completing this thesis. I thank Tiago Martinelli, Alex Freitas, Thomas Peron, Gabriela Depetri, Michael Hartl, Jaap Eldering, Zeray Hagos, Hans Muller, Ali Tahzibi, Dmitry Turaev, Richard Cubas, Herbert Maquera, Fernando Cordeiro, Robert Ronge, Ivan Silva, Eddie Nijolt, Sajjad Bakrani, Kalle Temperi, Wey Hao Tey, Kevin Kühl, Boumediene Hamzi, Matteo Tanzi, Deniz Eroglu, Asrat Mekonen, Zheng Bian, Marcel Novaes, Douglas Finamore, Krerley Oliveira, Lucas Resende, Juliano Genari, Guilherme Goedert, Ismael Ledoino, Arkady Pikovsky, Lu Yang, Hugo Chu, Matheus Manzatto, Irem Topal, Narcicegi Kiran, and Erik Bollt. Your willingness to share your experiences and insights was crucial to the success of this thesis.

I would like to acknowledge the financial support of CAPES for two months in the early months of my PhD. Also, I would like to thank the support by FAPESP grant 2018/10349-4, FAPESP CEMEAI Grant No. 2013/07375-0 and Serrapilheira Institute (Grant No.Serra-1709-16124) that gave me everything I needed to develop this research. Also, I would like to thank the Newton Advanced Fellow of the Royal Society NAF\R1\180236).

To all my old and new friends who accompanied me throughout this path. Directly or indirectly all of you offered something that I carried during these years. Undoubtedly, I am truly thankful.

To my partner and friend, Larissa Regina Diniz, for her support, understanding, and love, and for providing a listening ear during moments of doubt. From COVID-19 quarantine to a long-distance relationship, we faced together different adventures. You definitely have a part in the conclusion of this thesis.

Lastly, I am grateful to Edmilson dos Santos, Gabriela Roque dos Santos, and Rojanira

Roque dos Santos, my family, for their constant encouragement, understanding, and love. My full gratitude for all their investment of time, patience, and support during not only the last five years but 21 and 29 years (hu3!), respectively. This thesis would not have been possible without their support. Thanks for being a part of this journey.

RESUMO

ROQUE SANTOS, E. **Reconstrução de dinâmica de redes esparsas a partir de dados**. 2024. 152 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

As dinâmicas de redes complexas são comuns em diversos sistemas naturais, abrangendo desde a física até a neurociência. Essas redes apresentam estruturas de interação esparsas, onde apenas uma fração de todas as conexões possíveis existe. Essa estrutura de interação fornece valiosas perspectivas sobre a dinâmica das redes. Por exemplo, interrupções nas redes neuronais frequentemente resultam de problemas relacionados à conectividade. No entanto, em configurações experimentais, geralmente temos acesso a dados de séries temporais multivariadas em vez da própria rede. Nosso objetivo principal é desenvolver métodos para prever e antecipar possíveis novos comportamentos dentro do sistema. Esta tese é dedicada à reconstrução de equações de movimento que descrevem a dinâmica de redes esparsas a partir de dados. Combinamos teoria de sistemas dinâmicos e teoria ergódica com métodos de recuperação esparsa para garantir uma reconstrução exata e única. Para começar, introduzimos um método chamado Ergodic Basis Pursuit (EBP). Este método minimiza os dados de medição necessários, garantindo uma reconstrução precisa, enquanto identifica robustamente a estrutura de interação a partir de dados experimentais, revelando assim a estrutura original da rede. Posteriormente, demonstramos a aplicabilidade deste método redes com clusters. Aproveitando as informações de clusters da rede, o EBP adota uma abordagem de reconstrução dividir-e-conquistar. A reconstrução da rede é dividida em subproblemas, cada um restrito a um cluster específico e resolvido independentemente. As soluções são então combinadas para revelar a estrutura completa da rede. Por fim, empregamos métodos de recuperação esparsa para reconstruir equações de movimento a partir da dinâmica de redes com bursting.

Palavras-chave: Dinâmica de redes, redes esparsas, sistemas dinâmicos, teoria ergódica, métodos de recuperação esparsa.

ABSTRACT

ROQUE SANTOS, E. **Reconstruction of sparse network dynamics from data**. 2024. 152 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Complex network dynamics are prevalent in various natural systems, spanning from physics to neuroscience. These networks feature sparse interaction structures, where only a fraction of all possible connections exist. This interaction structure provides valuable insights into network dynamics. For instance, disruptions in neuronal networks often arise from issues related to connectivity. However, in experimental settings, we typically have access to multivariate time series data rather than the network itself. Our primary goal is to develop methods for predicting and anticipating potential new behaviors within the system. This thesis is dedicated to reconstructing governing equations that describe the dynamics of sparse networks from data. We merge dynamical systems theory and ergodic theory with sparse recovery methods to ensure exact and unique reconstruction. To begin, we introduce a method called Ergodic Basis Pursuit (EBP). This method minimizes the required measurement data, guaranteeing exact reconstruction while robustly identifying the interaction structure from experimental data, thereby revealing the original network structure. Subsequently, we demonstrate the applicability of this method to clustered networks. By leveraging cluster information within the network, EBP adopts a divide-and-conquer reconstruction approach. The network reconstruction is divided into subproblems, each restricted to a specific cluster and solved independently. The solutions are then combined to reveal the complete network structure. Finally, we employ sparse recovery methods to reconstruct governing equations from the dynamics of bursting networks.

Keywords: Network dynamics, sparse networks, dynamical systems, ergodic theory, sparse recovery methods.

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF TABLES

LIST OF SYMBOLS

$\{x(t)\}_{t\geq 0}$ — Multivariate time series

n — Length of time series

N — Number of nodes (vertices)

f — Isolated map

h — Coupling function

G — Graph - network structure

M — Phase space of the isolated dynamics f

F — Network dynamics map

$\mathscr{L}$ — Library - set of basis functions $\mathscr{L} = \{\phi_1, \ldots, \phi_m\}$

$\Phi(X)$ — Library matrix

$\bar{X}$ — Data matrix - one time iterate

m — Number of basis functions - cardinality of $\mathscr{L}$

C — Matrix of coefficients

c_i — Coefficient vector of node i

$[N]$ — The node set $[N] := \{1, \ldots, N\}$

n_0 — Minimum length of time series to a successful reconstruction

μ — Physical measure of the network dynamics F

ν — Product measure that is close to the physical measure

$\mathscr{L}_\nu$ — Orthonormal network library with respect to the product measure ν

$\Phi_\nu(X)$ — Library matrix associated to $\mathscr{L}_\nu$

CONTENTS

CHAPTER

1

INTRODUCTION

Complex networks of coupled dynamical systems play a crucial role in various fields of science, including biology (WINFREE, 2001), chemistry (KURAMOTO, 1984) to physics (STANKOVSKI *et al.*, 2017) and neuroscience (ZAMORA-LÓPEZ; ZHOU; KURTHS, 2011). The network dynamics consists of two key components. The first component is the isolated dynamics, which describe how individual units (nodes) evolve in time. For example, single neurons can exhibit periodic or chaotic behavior (IZHIKEVICH, 2007). The second component is the interaction among nodes within a complex network structure. This interaction results in a high-dimensional and nonlinear dynamical system. Unlike a scenario with an 'all-to-all' interaction structure, these networks typically are sparse (GENIO; GROSS; BASSLER, 2011; BROIDO; CLAUSET, 2019), i.e., only a small subset of all possible connections exists. For instance, empirical observations consistently demonstrate the sparsity of neuronal networks (MASON; NICOLL; STRATFORD, 1991; WATERS; HELMCHEN, 2006; HE; CHEN; EVANS, 2007; GUZMAN *et al.*, 2016). As an example, the *C.elegans*, one of the few nervous systems completely mapped at the cellular scale, exhibits a low connectivity density, approximately around 10% (SCHRÖTER; PAULSEN; BULLMORE, 2017).

The presence of sparse network structure has a profound impact on the dynamics of various systems (SPORNS; TONONI; KÖTTER, 2005; PEREIRA; STRIEN; TANZI, 2020; SCHÄFER *et al.*, 2018; MOLNAR; NISHIKAWA; MOTTER, 2021). Neurological disorders are often linked to disruptions in network structure (BOHLAND *et al.*, 2009), while the stability of power grids depends on the coherent functioning of its units in the face of disturbances (WITTHAUT *et al.*, 2022; MOLNAR; NISHIKAWA; MOTTER, 2021). Consequently, comprehending sparse network dynamics opens the possibility of gaining insights into predicting nonlinear phenomena across diverse systems.

The current state of technology only allows us to access a multivariate time series of nodes' states (PARK; FRISTON, 2013). Typically, due to time resolution or even the cost to acquire the data, we only have access to a limited length of time series in comparison to the

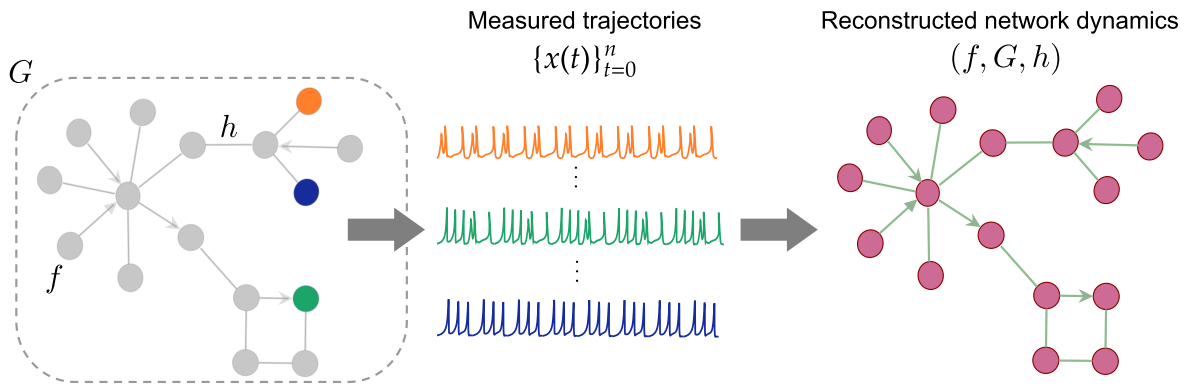

Figure 1 – **Reconstruction of network dynamics from data.** The unknown network dynamics (f, G, h) is accessed through the multivariate time series of all nodes in the network $\{x(t)\}_{t=0}^{n}$. The length of the time series is n and the size of the network is N. The challenge is to reverse engineering and to obtain (f, G, h) back in the regime of the limited amount of time series, i.e., $n \ll N$. This opens the possibility to gain predictive capabilities.

system size (ERNST; BAR-JOSEPH, 2006; WANG *et al.*, 2008; GERVEN *et al.*, 2009; HEMPEL *et al.*, 2011). The challenge is to reverse engineer and extract the dynamics of large networks from a limited amount of data. In other words, to obtain both the nodes' evolution over time when isolated, as well as their interactions with one another within the overall network structure.

Throughout this thesis, the main assumption is that we have access to the time series of *all* nodes in the network, so we obtain $\{x(t)\}_{t\geq 0}$, see Figure 1. This thesis is devoted to solving such a reconstruction problem:

> *to obtain the governing equations of large and sparse network dynamics from a limited amount of multivariate time series* $\{x(t)\}_{t\geq 0}$.

Achieving this allows for predicting the behavior of the system ahead of time, potentially identifying critical transitions. This is challenging mostly due to the limited information. In fact, extracting models of high-dimensional nonlinear systems from limited data is a non-trivial task because there is not enough information from the phase space to rule out all possible models and select one good model for the system.

Reconstruction state of the art: A great deal of effort to reconstruct dynamical systems from multivariate data in the past years has generated different research lines. Despite their success, they fail in the scenario we are interested in. In a nutshell, they can be divided into three different approaches:

- *Informational tools* (BUTTE; KOHANE, ; FUENTE *et al.*, 2004; NAWRATH *et al.*, 2010; BRESSLER; SETH, 2011; SUN; TAYLOR; BOLLT, 2015) focus to identity the presence, strength, and direction of a connection between pair of nodes by measuring the statistical dependence between their time series such as correlations, mutual information or causal

relations. These methods alone do not describe the dynamics on top of the uncovered directed network, which is necessary to predict critical transitions.

- *Reservoir computers* (PATHAK *et al.*, 2018; LU; HUNT; OTT, 2018) (recurrent neural networks, also called echo-state networks) are primarily focused on forecasting dynamical systems, in particular chaotic systems that are inherently difficult to predict. Although reservoirs are only trained through the readout weights and a few hyperparameters, the required length of time series can be as large as the underlying system.

- *Statistical methods for dynamical systems* encompass methods such as dynamic mode decomposition (DMD) and Bayesian inference that use statistical properties to describe the evolution of the dynamical system. Dynamic mode decomposition (BUDIŠIĆ; MOHR; MEZIĆ, 2012; SLIPANTSCHUK; BANDTLOW; JUST, 2020; THIBEAULT *et al.*, 2020) is devoted to identifying a low-dimensional description of the linear but infinite-dimensional evolution of measures. It identifies the current state of the system but fails to extrapolate potential transitions that may occur in the system. Dynamical Bayesian inference (LUCHINSKY *et al.*, 2008; STANKOVSKI *et al.*, 2012) focuses on stochastic dynamical systems that are pumped by noise with known distribution. Although the method can infer low-dimensional systems, it fails to infer large network dynamics because it requires a long length of time series.

Overall we must evoke suitable assumptions to extract the most meaningful model for prediction purposes. Sparse recovery methods (WANG *et al.*, 2011; HAN *et al.*, 2015; WANG; LAI; GREBOGI, 2016; BRUNTON; PROCTOR; KUTZ, 2016; SCHAEFFER; TRAN; WARD, 2018) have evolved as leaders to recover sparse network dynamics. These methods search for sparse representations of the input data, imposing that the network has a low density of connections into a penalized minimization problem, e.g. LASSO (HAN *et al.*, 2015), SINDy (BRUNTON; PROCTOR; KUTZ, 2016) and basis pursuit (NAPOLETANI; SAUER, 2008; WANG *et al.*, 2011). When the network is moderately large, the amount of data required for successful reconstruction is too large, making the network reconstruction challenging. Indeed, in general, the network reconstruction becomes ill-posed and unstable (NAPOLETANI; SAUER, 2008; NOVAES; Roque dos Santos; PEREIRA, 2021). The key idea is that sparsity may promote a decrease in the data length required for the reconstruction (TRAN; WARD, 2017a; SCHAEFFER; TRAN; WARD, 2018).

Main contributions. Together with the sparsity assumption, we consider throughout this thesis mostly chaotic (ergodic) dynamics that exhibit exponential decay of correlations (VIANA, 1997). Evoking dynamical systems and ergodic theory together with sparse recovery methods, we contribute to shifting the reconstruction paradigm. Our main contributions are:

(i) *Minimum length of time series.* Using the statistical properties of the chaotic dynamics, we determine the minimum length of time series for successful reconstruction with high

probability. This minimum length scales quadratically with the node degree (the number of connections the node receives) being probed and logarithmically with the network size.

(ii) *Ergodic Basis Pursuit.* We develop the so-called Ergodic Basis Pursuit (EBP) that uses the inherent ergodicity advantageously to reconstruct the sparse network dynamics from data. In the absence of noise and once the minimum length of time series is attained, EBP exactly recovers the network. We compare EBP against other methods, which it outperforms in all tests. In the presence of measurement noise, EBP is robust, i.e., identifies correctly the network structure, treating the noise level as a tuning parameter. To illustrate the power of EBP, we test EBP in experimental data (HART *et al.*, 2019).

(iii) *Divide-and-conquer reconstruction of clustered networks.* Here, we focus our attention on when there is expert knowledge about the network structure. In particular, the presence of clusters, groups of nodes that share many connections inside themselves but comparatively only a few connections with different groups. Instead of attempting to identify connections among all nodes in the network, the reconstruction is split into subproblems. Each subproblem consists of reconstructing clusters. First, subproblems are solved individually and independently, then combined to yield the global solution, the entire network structure. This approach defines a divide-and-conquer algorithm (LEVITIN, 2007), which is only possible in our case due to the exact reconstruction via EBP. For large networks, using the cluster information speeds up the reconstruction when compared to solving the entire network reconstruction at once.

Reconstruction of sparse network dynamics (f,G,h)**.** The dynamics of N nodes is given by:

$$\begin{aligned} x_i(t+1) &= F_i(x(t)), \\ F_i(x(t)) &= f_i(x_i(t)) + \alpha \sum_{j=1}^{N} A_{ij} h_{ij}(x_i(t), x_j(t)), \quad i = 1,\ldots,N, \end{aligned} \tag{1.1}$$

where $x_i(t)$ represents the state of node i at a time t, $x = (x_1,\ldots,x_N) \in M^N$ is the state of the network, $f_i : M \to M$ corresponds to the isolated map, α is the coupling strength, A_{ij} corresponds to the entry of the adjacency matrix, which encodes the network structure, and $h_{ij} : M \times M \to M$ is the coupling function for the nodes i and j. This model is considered throughout this thesis, and in the case a different model is used, it will be introduced appropriately when needed.

To reconstruct the model of the network dynamics, sparse recovery methods assume that the dynamics F has a sparse representation in an a priori known set of basis functions, a library $\mathscr{L}$. More precisely, the isolated maps f_i and the coupling functions h_{ij} lie in the span of $\mathscr{L} = \{\phi_1,\ldots,\phi_m\}$ with $\phi_l : M^N \to M$. Moreover, for each $i \in [N]$: $F_i = \sum_l (c_i)_l \phi_l$ with $c_i \in \mathbb{R}^m$ is a sparse vector, i.e., a few entries are nonzero. These assumptions convert the reconstruction

problem into solving a linear equation. Consider the library matrix

$$\Phi(X) = \begin{pmatrix} \phi_1(x(0)) & \cdots & \phi_m(x(0)) \\ \phi_1(x(1)) & \cdots & \phi_m(x(1)) \\ \vdots & \ddots & \vdots \\ \phi_1(x(n-1)) & \cdots & \phi_m(x(n-1)) \end{pmatrix} \tag{1.2}$$

and arrange the trajectories into a matrix

$$\bar{X} = \begin{pmatrix} x_1(1) & \cdots & x_N(1) \\ \vdots & \ddots & \vdots \\ x_1(n) & \cdots & x_N(n) \end{pmatrix}. \tag{1.3}$$

Hence, the reconstruction problem becomes in finding the $m \times N$ matrix of coefficients C, which has column vectors given by $\{c_1, c_2, \ldots, c_N\} \subset \mathbb{R}^m$, such that

$$\bar{X} = \Phi(X)C. \tag{1.4}$$

The coefficients encode the network dynamics, in particular the graph structure.

When the amount of data is large in comparison to the network size, the library matrix $\Phi(X)$ might be full column rank, and the unique solution can be found by least square based-methods (LUENBERGER, 1997; BRUNTON; PROCTOR; KUTZ, 2016), i.e., solving the minimization problem $\min_U \|\bar{X} - \Phi(X)U\|_2$. However, for large networks $\Phi(X)$ has more columns than rows, it is a fat matrix. Then, the linear system in Equation (1.4) is underdetermined, see the left panel in Figure 2 for an illustration. So, there are an infinite number of solutions (if there exists at least one) and the reconstruction problem is ill-posed.

For short time series and using the sparsity assumption, one approach is to solve for each node i the basis pursuit (BP) problem

$$\text{(BP)} \quad \min_{u \in \mathbb{R}^m} \|u\|_1 \quad \text{subject to } \Phi(X)u = \bar{x}_i, \tag{1.5}$$

where $\bar{x}_i$ is the i-th column of $\bar{X}$. This minimization problem searches for sparse solutions. In fact, the right panel in Figure 2 illustrates how sparse vectors are optimal solutions.

Coupled logistic maps in a ring network. We will illustrate the reconstruction problem of coupled logistic maps in different network structures. Our contribution (i) and (ii) are valid for general directed networks, as illustrated in Figure 1, but here, we will present it for an undirected ring network. In the end, we look at the special case of undirected networks that contain clusters, which is the setting for our contribution (iii).

The isolated dynamics is given by

$$f_i(x_i) = ax_i(1-x_i), \quad i = 1, \ldots, N,$$

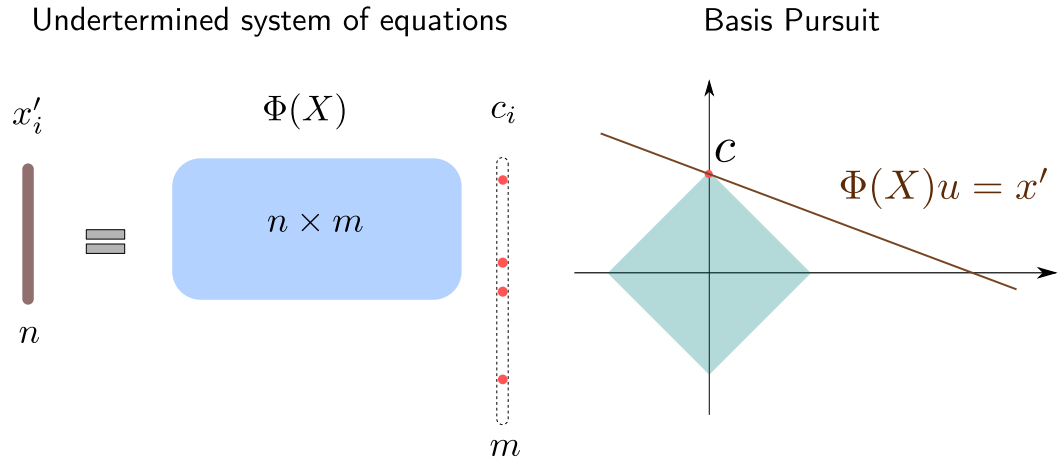

Figure 2 – **Underdetermined system but under sparsity assumptions.** The left panel displays the typical scenario in large networks. A short vector in the left-hand side of the equation, $\bar{x}_i$, corresponds to the column of the matrix $\bar{X}$. On the right-hand side the fat library matrix $\Phi(X)$ and the vector c. This system of equations has an infinite number of solutions. However, the solutions are sparse. In fact, c is sparse because it lies in one of the axes in the plane. So, the vector c_i has only a few non-zero entries, which are illustrated by red dots. The right panel displays the geometric illustration behind the minimization problem in Equation (1.5). The line (or plane in the higher-dimensional scenario) represents the constraint, i.e., the observed trajectory that should be reconstructed. The intersection yields the solution, which is attained by the sparse vector c.

with $a \in [3.9, 4.0)$. The coupling function is

$$h_{ij}(x_i, x_j) = x_i x_j, \quad i, j = 1, \ldots, N.$$

Let us assume the network structure depicted in Figure 3 a), a ring network. Since both f_i and h_{ij} are polynomials, it suffices to consider a set of polynomials of two variables with degree at most $r = 2$ to represent the network dynamics:

$$\mathscr{L} = \{1\} \cup \{\phi_i^p(x_i) = x_i^p\}_{i,p} \cup \{\phi_{ij}^{pq}(x_i, x_j) = x_i^p x_j^q\}_{i,j,p,q}, \tag{1.6}$$

which contains 39 functions, see Section 2.6.1 for details. A node i is only connected to node $i-1$ and $i+1$, so the node i dynamics is expressed as

$$x_i(t+1) = -a\phi_i^1(x(t)) + a\phi_i^2(x(t)) + \alpha\phi_{i-1,i}^{11}(x(t)) + \alpha\phi_{i,i+1}^{11}(x(t)).$$

The same could be done for all nodes in the small network. First, note that solving Equation (1.4) opens the possibility of recovering the network dynamics. Also, the corresponding coefficient vector c_i is sparse since it has only four nonzero terms $\{-a, a, \alpha, \alpha\}$ out of all possible 39 entries. To search solutions of Equation (1.4) while imposing sparsity of the solution is a promising option.

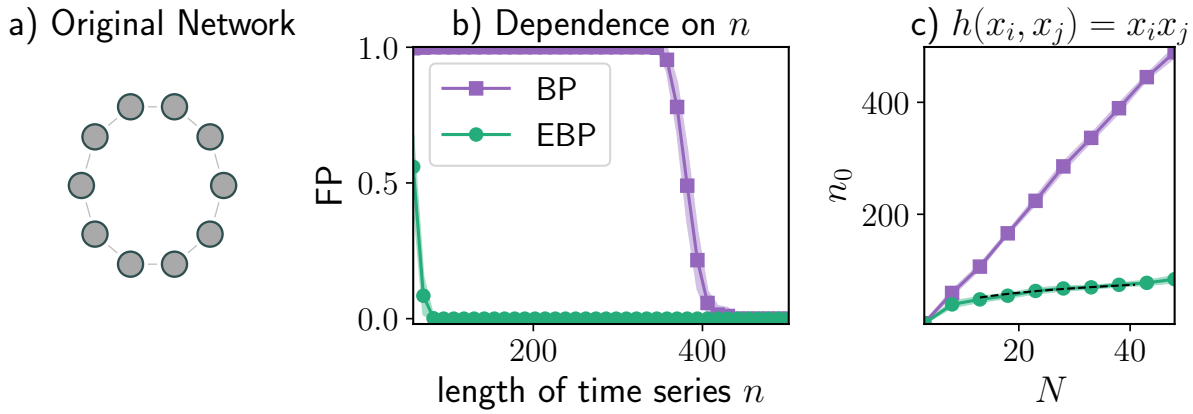

Figure 3 – **Ergodic Basis Pursuit performance requires only short time series.** a) Illustration of a ring graph with $N = 10$. b) False positive (FP) of the reconstructed ring network with respect to the length of time series n for a network size $N = 40$. Both methods predict correctly the true connections, for this reason, we only show the false positive proportion. c) The minimum length of time series n_0 for a successful reconstruction versus system size N. Basis pursuit (BP) and ergodic basis pursuit (EBP) are shown in (purple) squares and (green) circles, respectively. The network dynamics parameters are $a = 3.990$ and coupling strength $\alpha = 5 \times 10^{-4}$. The shaded area corresponds to the standard deviation with respect to 10 distinct initial conditions uniformly drawn in $[0, 1]^N$. The (black) dashed is the scaling $\ln N$ for reference. The Kernel density estimation of ν is used with bandwidth $\chi = 0.05$. The multivariate time series is generated without noise.

BP in Equation (1.5) was used for networks of moderate size (WANG; LAI; GREBOGI, 2016; SCHAEFFER; TRAN; WARD, 2018; SCHAEFFER *et al.*, 2020). However, for large networks, this may lead to spurious linear dependencies among the columns $\Phi(X)$ (NAPOLETANI; SAUER, 2008; NOVAES; Roque dos Santos; PEREIRA, 2021), and (1.5) does not have a unique sparse solution. To illustrate, see Figure 3 b) and c). Ensuring exact network reconstruction, even in the presence of sparse interactions, remains a significant open problem.

Key step: basis adaptation to the dynamics. A priori, the library $\mathscr{L}$, and consequently, the associated library matrix $\Phi(\mathbf{X})$ does not have sufficiently good properties in such a way that Equation (1.5) has unique solutions. Based on compressive sensing theory (CANDES; TAO, 2005; DONOHO; ELAD; TEMLYAKOV, 2006), if the library matrix satisfies certain *good properties*, the corresponding basis pursuit has unique solutions. In an upshot, the good property is: any subset of columns of the library matrix forms a set of nearly orthonormal set of vectors. Hence, our strategy is to use the decay of correlations of the network, so we can introduce a new set of basis functions $\mathscr{L}(\nu)$ whose associated library $\Phi_\nu(\mathbf{X})$ satisfies such good property. The decay of correlations implies that we can control how much time the system needs to evolve such that any two columns of $\Phi_\nu(X)$ become nearly orthogonal to each other.

In practice, we proceed as follows: we estimate a probability distribution ν and orthonormalize the basis functions via the Gram-Schmidt process on the $\mathscr{L}$, see Figure 4. The exponential decay of correlations implies that for a large set of initial conditions, we can find the minimum length of time series n_0 such that $\Phi_\nu(\mathbf{X})$ satisfies the desired conditions. More precisely, as our

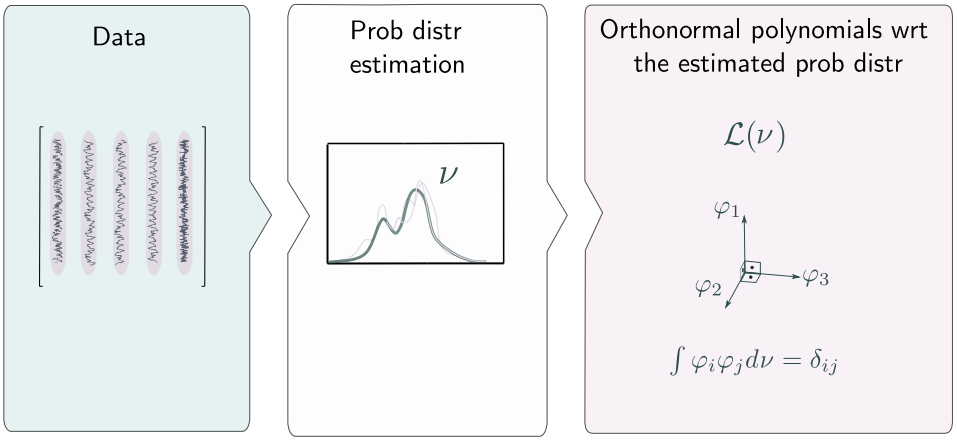

Figure 4 – **Ergodic Basis Pursuit uses dynamical and statistical information.** The input data is generated from a network dynamics that preserves a measure is μ. Using its estimated measure ν, which is a product measure, an orthonormal set of basis functions $L(\nu)$ is constructed, representing the dynamics. Under the assumption that the network dynamics is sparse, the input data and $L(\nu)$ are recast as a minimization problem, whose solution encodes the network.

contribution (i) we show that

$$n_0 \approx k_i^2 \ln(N), \tag{1.7}$$

where k_i is the degree of the probed node. In our contribution (ii) we propose the Ergodic Basis Pursuit (EBP) which consists of solving Equation (1.5) but replacing $\Phi(X)$ by $\Phi_\nu(X)$ for the length of time series $n \geq n_0$, i.e,

$$\text{(EBP)} \quad \min_{u \in \mathbb{R}^m} \|u\|_1 \quad \text{subject to } \Phi_\nu(X)u = \bar{x}. \tag{1.8}$$

See Theorem 2.7.1.1 for the precise statements.

Numerical experiment: coupled logistic maps. Going back to the coupled logistic maps example, we compare the reconstruction performance of the EBP against the classical BP for overall coupling strength $\alpha = 5 \times 10^{-4}$. Figure 3 a) illustrates the ring network with $N = 10$ nodes.

Figure 3 b) shows an evaluation of the reconstruction performance employing the basis pursuit (BP) and the Ergodic Basis Pursuit (EBP) as we increase the length of time series n. The convex minimization problem is solved by employing the `CVXPY` package (DIAMOND; BOYD, 2016; AGRAWAL *et al.*, 2018), in particular, `ECOS` solver (DOMAHIDI; CHU; BOYD, 2013). We also consider the polynomial library as in (1.6) but with the degree at most 3 and allowing product terms $x_i x_j$, so by construction, there exists a sparse representation of the network dynamics in this library.

We observe that the false positive fraction, which calculates the presence of mistakenly found edges regardless of their weights, for BP goes to zero when $n_0 \approx 400$, roughly tenfold the system size. BP (in purple) requires a minimum length of time series n_0 that scales with

the system size to reconstruct a ring network in coupled logistic maps, although the network is sparse. The BP method is inappropriate for large-scale networks.

On the other hand, EBP outperforms the basis pursuit method, reducing the necessary length of time series to reconstruct the network. To evaluate the scaling with respect to the system size, we calculate n_0 as we increase N. In Figure 3 c), we confirm that n_0 scales with the system size for BP instead of $\ln N$ of the EBP method. In Section 2.3.4, we demonstrate that our estimates of n_0 predict the numerical observation when we vary the maximum degree of different network structures.

Prior knowledge: clustered networks. To reconstruct the neighbors of a given node in the system, we solve Equation (1.8), where $\Phi_v(X)$ is constructed using the time series of all nodes in the network. Without any further knowledge about the structure, the reconstruction must be performed assuming that each node is potentially coupled to any other node in the network, i.e., probing the entire network at once.

However, suppose the network has clusters, see Figure 5. Modular networks are typical examples such as the cat connectivity structure. The cat network is a map of the connectivity of the cat cortex (SCANNELL; YOUNG, 1993; SCANNELL; BLAKEMORE; YOUNG, 1995). This network is a so-called rich-club structure, where communities (portions of the network more connected inside than among themselves) interact via a massive community called an integrating cluster.

In this scenario, we can incorporate this information to solve Equation (1.8). Our contribution (iii) is that we can introduce an approach that divides the reconstruction into subproblems that are solved independently. Then, the solutions are combined to yield the entire network structure. The main finding is that for clustered networks such as the ones depicted in Figure 5, this approach speeds up the reconstruction, see Section 4.6 for details.

1.1 Guide to this thesis

Given the importance of establishing a technique for the successful reconstruction of network dynamics from data, this thesis evokes dynamical systems, ergodic theory, and sparse recovery methods to develop an approach that extends the reconstruction to a new level. Our contributions are divided into the following chapters.

Chapter 2 is devoted to detailing our contributions (i) and (ii). We introduce the Ergodic Basis Pursuit (EBP) method that uses the network dynamics' statistical properties to ensure the exact reconstruction of sparse networks when a minimum length of time series is attained. The minimum time series length scales quadratically with the node degree being probed and logarithmically with the network size as in Equation (1.7). Our approach is robust against noise and allows us to treat the noise level as a parameter. We show the reconstruction power of the

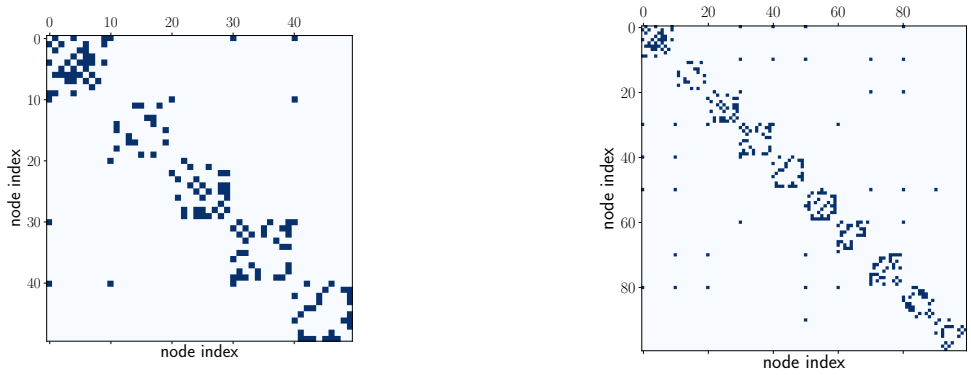

Figure 5 – **Random clustered networks.** A random network model capturing the features of a modular network such as the cat connectivity structure. We construct a network with κ clusters and p nodes in each cluster. Each cluster consists of an Erdos-Renyi (ER) network with with mean degree $\langle k \rangle$ (uniform for all clusters) and the bridge nodes (those nodes sharing inter-cluster connections) forms an integrating cluster, as an ER network as well, with mean degree $\langle k_I \rangle$. The set of bridge nodes are chosen selecting arbitrarily one single node of each cluster. The clusters are identified as block diagonals structures in the plot. The connections among bridge nodes are displayed as dots off the diagonal. Left panel displays a network with $\kappa = 5$ clusters and $p = 5$ node in each cluster. Right panel shows a network contains $\kappa = 10$ and $p = 10$. For both networks the mean degree $\langle k \rangle = 3$ and mean degree in the integrating cluster $\langle k_I \rangle = 4$.

EBP in experimental multivariate time series from optoelectronic networks. Chapter 2 is an adaptation to the thesis of a recently submitted paper entitled *Robust reconstruction of sparse network dynamics* (PEREIRA; SANTOS; STRIEN, 2023).

In Chapter 3 we consider chaotic dynamics in coupled phase oscillators. We reconstruct the network from an initial guess which can include expert knowledge about the system such as main motifs and hubs. When sparsity is taken into account the number of data points needed is drastically reduced when compared to the least-squares recovery. We show that the sparse solution is stable under basis extensions, that is, once the correct network topology is obtained, the result does not change if further motifs are considered. Chapter 3 is adapted from an published paper entitled *Recovering sparse networks: Basis adaptation and stability under extensions* (NOVAES; Roque dos Santos; PEREIRA, 2021).

Chapter 4 details our contribution (iii), merging the above two contributions and pushing forward the following research line. Although the network structure is unknown, there is still a priori knowledge. Hence, we introduce an approach that incorporates expert knowledge of the structure into the reconstruction method. Using this we can break the reconstruction into subproblems and solve each subproblem locally and in parallel. Then, we combine information gathered for each local solution, to form the global solution. That is, if we know which neighbors two distinct nodes have, we can combine this information and obtain a larger uncovered structure. This approach heavily relies on the reconstruction capability of the EBP method.

In Chapter 5 we specialize the reconstruction problem to bursting network dynamics. In particular, we reconstruct governing equations of bursting dynamics, which contain a rational

function, from data. Bursting dynamics are a crucial and common feature in neural networks. Hence, extracting the governing equations can lead to important insights, particularly, control and prediction capabilities. We use implicit-SINDy, which is a sparse recovery method that searches for sparse rational representations of the input data, to reconstruct the dynamics of neurologically relevant motifs. Our main result is that the minimum length of time series to implicit-SINDy recovers the dynamics scales at least quadratically with network size. This bound is useful to characterize the implicit SINDy performance for large network reconstruction.

Finally in Chapter 6 we conclude and list open questions. The appendices are devised to support the associated chapters with further details methods and algorithms used throughout the thesis.

CHAPTER

2

ERGODIC BASIS PURSUIT INDUCES NETWORK RECONSTRUCTION

This chapter is devoted to presenting our work entitled *Robust reconstruction of sparse network dynamics* (PEREIRA; SANTOS; STRIEN, 2023).

Here, we put forward the *Ergodic Basis Pursuit* (EBP) method that reconstructs sparse networks from a limited amount of data. Our method adapts the search for sparse solutions to the statistical properties of the network. We formulate the EBP as a Basis Pursuit problem adapted to the ergodicity of the network. When the network dynamics is ergodic and has decay of correlations, we show that the reconstruction is exact once a minimal length of the time series n_0 is attained. We show that n_0 scales quadratically with the node degree and log of the system size. We also show that the reconstruction is robust against random perturbations. We illustrate the applicability of our method in experimental optoelectronic networks. Our approach enables us to treat the noise level as a tuning parameter to identify the network structure robustly.

2.1 Dynamics on complex networks

We consider dynamics as

$$x_i(t+1) = f_i(x_i(t)) + \alpha \sum_{j=1}^{N} A_{ij} h_{ij}(x_i(t), x_j(t)), \tag{2.1}$$

for each $i \in [N] := \{1,\dots,N\}$, where x_i represents the state of node i, $f_i \colon M_i \to M_i$ corresponds to the isolated map over a bounded set $M_i \subset \mathbb{R}$, α is the coupling strength, A_{ij} equals 1 if node i receives a connection from j and 0 otherwise, and $h_{ij} \colon M_i \times M_j \to M_i$ is the pairwise coupling function. We denote the state of the full network as $x = (x_1,\dots,x_N) \in M^N \equiv \prod_{i\in[N]} M_i$, and $x(t+1) = F(x(t))$. This class of networks is common in applications such as laser dynamics (HART *et al.*, 2019) and can be generalized to higher dimensions.

We consider four assumptions on the network dynamics in (2.1).

(i) Network library. The isolated maps f_i and the coupling functions h_{ij} lie in the span of an ordered library $\mathscr{L} = \{\phi_1, \phi_2, \ldots, \phi_m\}$ where $\phi_l : M^N \to \mathbb{R}$. We consider the polynomials of two variables with degree at most r

$$\mathscr{L} = \{1\} \cup \{x_i^p\}_{i,p} \cup \{x_i^p x_j^q\}_{i,j,p,q}, \tag{2.2}$$

where $i, j \in [N]$ with $i \neq j$ and we remove any redundancy, $p \in [r], q \in [r-1]$, and $p+q \leq r$. The cardinality of $\mathscr{L}$ is given by $m = \binom{N}{2}\binom{r}{2} + Nr + 1$. A directed edge from j to i is given by the presence of a nonzero coefficient in $x_i^p x_j^q$. We discuss the ordering of $\mathscr{L}$ in Section 2.5.1. We say that an ordered library $\mathscr{L}$ is a network library and the functions depend only on pairs of coordinates. Thus, nonzero coefficients in the network representation in $\mathscr{L}$ can be identified with links in the network structure.

(ii) Sparse network. We assume that the network structure is directed and sparse, that is, for each node i, $x_i(t+1) = \sum_{l=1}^m c_i^l \phi_l(x(t))$ for all $t \geq 0$, where $c_i = (c_i^1, \ldots, c_i^m) \in \mathbb{R}^m$ is an s-sparse vector, that is, at most s of its entries are nonzero, see definition 2.5.3. Notice that a node i with degree k_i will have a number of nonzero entries of c_i growing linearly with k_i.

(iii) Exponential mixing. We assume (F, μ) satisfies exponential mixing conditions (HANG; STEINWART, 2017) for the physical measure μ: given a constant $\gamma > 0$ for all $\psi \in \mathscr{C}^1(M^N; M)$ and μ-integrable function φ, there exists $K(\psi, \varphi) > 0$ such that for any $t \geq 0$

$$\left| \int \psi \cdot (\varphi \circ F^t) d\mu - \int \psi d\mu \int \varphi d\mu \right| \leq K(\psi, \varphi) e^{-\gamma t}. \tag{2.3}$$

This assumption is typical for chaotic dynamical systems.

(iv) Near product structure. Since we are dealing with pairwise interactions, given a small $\zeta > 0$ we assume that the network physical measure μ is close to a product measure ν, i.e., $d(\mu, \nu) < \zeta$, where d calculates the maximum difference between integrals with respect to μ and ν over pair of functions in a suitable network library, see Section 2.5.4 for the formal definition. In the weak coupling regime, this assumption is fulfilled (EROGLU *et al.*, 2020). However, this assumption also holds in other scenarios, as we will illustrate later in an experimental application.

First, we consider the network reconstruction problem for the noiseless case and establish the minimum length of time series n_0 so that the EBP exactly reconstructs the network structure. Next, we show that the reconstruction is robust against additive measurement noise, which opens the possibility to apply for the experimental setting.

2.2 Reconstruction problem

To reconstruct the network structure A from the multivariate time series data $\{x(t)\}_{t\geq 0}^{n}$ of (2.1), we consider the library matrix

$$\Phi(X) = \frac{1}{\sqrt{n}} \begin{pmatrix} \phi_1(x(0)) & \cdots & \phi_m(x(0)) \\ \phi_1(x(1)) & \cdots & \phi_m(x(1)) \\ \vdots & \ddots & \vdots \\ \phi_1(x(n-1)) & \cdots & \phi_m(x(n-1)) \end{pmatrix} \tag{2.4}$$

and arrange the trajectories into a matrix

$$\bar{X} = \begin{pmatrix} x_1(1) & \cdots & x_N(1) \\ \vdots & \ddots & \vdots \\ x_1(n) & \cdots & x_N(n) \end{pmatrix}. \tag{2.5}$$

We aim to find the $m \times N$ matrix of coefficients C, which has column vectors given by $\{c_1, c_2, \ldots, c_N\} \subset \mathbb{R}^m$, such that

$$\bar{X} = \Phi(X)C. \tag{2.6}$$

When the amount of data is large in comparison to the network size, the library matrix $\Phi(X)$ might be full column rank, and (sparse) approximations can be found by least square based-methods (BRUNTON; PROCTOR; KUTZ, 2016; MANGAN *et al.*, 2016a; WANG *et al.*, 2018). For short time series, the matrix $\Phi(X)$ has more columns than rows, and one approach is to solve for each node i the basis pursuit (BP) problem

$$(\text{BP}) \quad \min_{u\in\mathbb{R}^m} \|u\|_1 \quad \text{subject to } \Phi(X)u = \bar{x}_i, \tag{2.7}$$

where $\bar{x}_i$ is the i-th column $\bar{X}$. This implementation was used for networks of moderate size (WANG; LAI; GREBOGI, 2016; SCHAEFFER; TRAN; WARD, 2018; SCHAEFFER *et al.*, 2020). For large networks, this may lead to spurious linear dependencies among the columns $\Phi(X)$ (NAPOLETANI; SAUER, 2008; NOVAES; Roque dos Santos; PEREIRA, 2021), and (2.7) does not have a unique sparse solution. In Figure 3 b) and c), we show that the basis pursuit (in purple) requires a minimum length of time series n_0 that scales with the system size to reconstruct a ring network in coupled logistic maps, although the network is sparse. The BP method is inappropriate for large-scale networks.

2.3 Main results: informal statements

To establish conditions for the uniqueness of the reconstruction, successful approaches show that any set of $2s$ columns of $\Phi(X)$ is nearly orthonormal, what is known as the restricted isometry property (RIP) (CANDES; TAO, 2005). By noticing that the inner product of pair of

columns of the original library matrix $\Phi(X)$ can be represented as a Birkhoff sum. We introduce a new library $\mathscr{L}_\nu$ by a Gram-Schmidt process such that the new columns have vanishing Birkhoff sum when the length of the times series diverges. Then, we use a concentration inequality to estimate the minimal length of the time series such that any two distinct column vectors of $\Phi_\nu(X)$ are nearly orthonormal to each other, and RIP for the desired sparsity.

This strategy is implemented with four main results: (I) the introduction of the new basis $\mathscr{L}_\nu$ is a network library which keeps the appropriated sparsity of the original problem; (II) It is possible to obtain a desired RIP constant for $\Phi_\nu(X)$ by exploring the ergodicity of the dynamics; and (III) the reconstruction is unique. Finally, we show (IV) the robustness against measurement noise.

2.3.1 Constructing the adapted network library

First, we notice that using the invariant measure μ directly to obtain a new basis and thus an almost orthonormal structure in the columns of the corresponding $\Phi(X)$ leads to the loss of sparsity in the representation. Indeed, the new orthonormal basis would contain functions that depend on all coordinates, e.g., of the form $\varphi(x_1,\ldots,x_N)$ because μ is not a product measure and mixes all coordinates. Hence, we consider a product measure ν close to μ. More precisely, we perform a Gram-Schmidt (GS) process in the span of $\mathscr{L}$ and obtain a basis $\hat{\mathscr{L}} = \{\hat{\varphi}_1,\ldots,\hat{\varphi}_m\}$. We define $\varphi_i = a_i\hat{\varphi}_i$, where $a_i^2 = 1/\int \hat{\varphi}_i^2 d\nu$, so the new basis $\mathscr{L}_\nu = \{\varphi_i\}_{i=1}^m$ is an orthonormal system with respect to a product measure ν. We assume that each marginal of ν is absolutely continuous with respect to Lebesgue, and the corresponding density is Lipschitz. For our network library $\mathscr{L}$ we have that:

> **Theorem 2.6.1 (Network library is preserved).** GS process maps an s-sparse representation of F in $\mathscr{L}$ to an $\omega_r(s)$-sparse representation in the orthonormal network library $\mathscr{L}_\nu$.

Here, $\omega_r(s) = \left(\left\lfloor \frac{r}{2}\right\rfloor\left(r-\left\lfloor\frac{r}{2}\right\rfloor\right)+r+1\right)s$, where $\lfloor\beta\rfloor$ denotes the largest integer p satisfying $p\leq\beta$. We call $\mathscr{L}_\nu$ the adapted network library and denote the respective library matrix by $\Phi_\nu(X) = \Phi(\mathscr{L}_\nu, X)$. The proof uses that the GS process is a recursive method that involves projections onto preceding functions. Since ν is a product, the projections of the GS are split into products of integrals. Thus, $\mathscr{L}_\nu$ does not have functions that depend on more than two variables and characterize a network library. Also, the representation remains sparse, see Section 2.6.2 for details.

2.3.2 Ergodic basis pursuit

Notice that since library $\mathscr{L}_\nu$ is orthonormal, the set of columns vectors of $\Phi_\nu(X)$ form a set of s nearly orthonormal column vectors. We quantify the orthonormality via the s-th restricted

isometry constant $\delta_s = \delta_s(\Phi_\nu(X))$ as the smallest $\delta \geq 0$ such that

$$(1-\delta)\|u\|_2^2 \leq \|\Phi_\nu(X)u\|_2^2 \leq (1+\delta)\|u\|_2^2 \tag{2.8}$$

for all s-sparse vectors $u \in \mathbb{R}^m$. Next, we determine the minimum length of time series such that $\Phi_\nu(X)$ is RIP with a desired small δ_s. Our second result is

> **Theorem 2.7.1.1 ($\Phi_\nu(X)$ satisfies RIP).** Consider $d(\mu,\nu) < \zeta$ for sufficiently small ζ. For large network sizes and large set of initial conditions if the length of time series n is at least
>
> $$n_0 \approx K_1 \omega_r(s)^2 \ln(Nr), \tag{2.9}$$
>
> for a positive constant K_1, then $\Phi_\nu(X)$ satisfies (2.8) with $\delta_{2\omega_r(s)} \leq \sqrt{2}-1$.

The proof is presented in Section 2.7 and the key steps are as follows. First, we use the coherence (DONOHO; HUO, 2001; DONOHO; ELAD; TEMLYAKOV, 2006) defined as

$$\eta(\Phi_\nu) := \max_{i \neq j} \left|\langle v_i, v_j \rangle\right|$$

over distinct pairs of normalized (Euclidean norm) columns of $\Phi_\nu(X)$. Since we know that $\delta_s \leq \eta(\Phi_\nu)(s-1)$ for any $s \geq 2$ (FOUCART; RAUHUT, 2013), it suffices to introduce a library $\Phi_\nu(X)$ whose coherence is small enough to obtain the desired RIP constant.

Second, let v_i be the i-th column of the matrix $\Phi_\nu(X)$ and notice that the inner product between columns i and j is

$$\langle v_i, v_j \rangle = \frac{1}{n}\sum_{t=0}^{n-1} (\varphi_i \cdot \varphi_j) \circ (F^t(x(0))).$$

Using that μ and ν are close, by triangular inequality and the Bernstein-type inequality, see (HANG; STEINWART, 2017), we control the coherence $\eta(\Phi_\nu)$ by approximating it by $\int \varphi_i \cdot \varphi_j d\mu$. Hence, we can determine a large set of initial conditions such that the RIP of $\Phi_\nu(X)$ is less than $\sqrt{2}-1$, see Section 2.7.1.1. Since $\Phi_\nu(X)$ is RIP, we obtain

> **Theorem 2.7.1.2 (EBP has unique solution).** The convex problem that we call Ergodic Basis Pursuit
>
> $$\text{(EBP)} \quad \min_{u \in \mathbb{R}^m} \|u\|_1 \quad \text{subject to } \Phi_\nu(X)u = \bar{x}, \tag{2.10}$$
>
> has a unique $\omega_r(s)$-sparse solution. That is, c_ν is the only solution of this minimization problem when $\bar{x} = \Phi_\nu(X)c_\nu$.

The proof follows from Theorem 2.7.1.1. EBP can be applied as a network reconstruction method. In terms of the coefficients $\{c_1, \ldots, c_N\}$, we create a weighted edge between node i and j using

$$W_{ij} = \max_{k \in \mathscr{S}_j} c_i^k. \tag{2.11}$$

We reconstruct a weighted subgraph using the node i, its neighbors, and the entry's magnitude of c_i as the edge weight; see details in the Appendix A.1.2.

Remark 2.3.1 (Minimum length of time series for networks)**.** The degree distribution and the condition in (2.9) can be used to estimate the amount of data that ensures the network reconstruction.

- **Erdős-Rényi (ER) networks.** We can apply (2.9) in $\mathscr{O}(1)$ for known random networks. First, note that $\omega_r(s)$ is a linear function with the sparsity level s, and consequently, it is a linear function of the degree $\propto k_i$ of the node i. Also, $\omega_r(s) < (r+1)^2$. The degree distribution is given by a Poisson distribution, so by concentration inequality (CHUNG; LU, 2006), most nodes have their degree close to the mean degree $\langle k \rangle$. Hence, to reconstruct a typical node in ER networks requires (in $\mathscr{O}(1)$) the minimum length of time series given by

$$n_0 = \mathscr{O}\big((r+1)^4 \langle k \rangle^2 \ln(Nr)\big). \tag{2.12}$$

  Note that $\langle k \rangle = pN$, where p is the probability of including an edge in the graph. In the phase where ER network becomes almost sure connected, $p = K \ln N/N$ with $K \geq 1$ (CHUNG; LU, 2006). So,

$$n_0 = \mathscr{O}\big((r+1)^4 \ln(N) \ln(Nr)\big).$$

- **Scale-free networks.** In scale-free networks, the same growth scaling (2.12) is valid for low-degree nodes. However, hubs in Barabási-Albert networks have their degree proportional to $N^{\frac{1}{2}}$, so it requires

$$n_0 = \mathscr{O}\big((r+1)^4 \langle k \rangle^2 N \ln(Nr)\big).$$

- **Regular networks.** All nodes have the same degree. So, the same growth scaling (2.12) is valid for any node in the network.

2.3.3 Robust reconstruction

We now extend the EBP to measurements corrupted by noise

$$y(t) = x(t) + z(t), \tag{2.13}$$

where $(z_n)_{n\geq 0}$ corresponds to independent and identically distributed $[-\xi, \xi]^N$-valued noise process, with probability measure ρ_ξ. The probability measure of the process $(y_n)_{n\geq 0}$ is the convolution $\mu_\xi := \mu * \rho_\xi$ (FOLLAND, 2013), which converges weakly to μ as $\xi \to 0$. We assume that μ_ξ is estimated using a product measure ν. We use that μ_ξ is close to ν to estimate

a new bound for the minimum length of the time series $\tilde{n}_0$ such that $\Phi_\nu(X)$ satisfies RIP with constant $\delta_{2\omega_r(s)} \leq \sqrt{2}-1$.

Since we measure the corrupted data Y instead of X, we use the Mean Value theorem to deduce that

$$\Phi_\nu(Y) = \Phi_\nu(X) + \Lambda(X,\bar{Z}), \tag{2.14}$$

where $\|\Lambda(X,\bar{Z})\|_\infty \leq mNr^2K_1\xi$ and K_1 depends on the density of the marginals of ν. The noisy observation in (2.13) can be recast as a perturbed version of the orthonormal version of (2.6) column-wise

$$\bar{y} = \Phi_\nu(Y)c_\nu + \bar{u}, \tag{2.15}$$

where c_ν is the coefficient vector associated to the network library $\mathscr{L}_\nu$ and $\bar{u}$ is ℓ_2 bounded, see Section 2.8. Thus, we can state our final result:

Theorem 2.8.2 (EBP is robust). If the length of time series $n \geq \tilde{n}_0$, then the family of solutions $\{c^\star_\nu(\varepsilon)\}_{\varepsilon>0}$ to the convex problem (which we call the Quadractically constrained Ergodic Basis Pursuit)

$$\text{(QEBP)} \quad \min_{\tilde{u}\in\mathbb{R}^m} \|\tilde{u}\|_1 \text{ subject to } \|\Phi_\nu(Y)\tilde{u} - \bar{y}\|_2 \leq \varepsilon \tag{2.16}$$

satisfies

$$\|c^\star_\nu(\varepsilon) - c_\nu\|_2 \leq K_2\varepsilon \tag{2.17}$$

for some $K_2 > 0$ as long as $\varepsilon \geq \sqrt{n}\xi\left(1 + mNr^2K_1\|c_\nu\|_\infty\right)$.

2.3.4 Numerical experiment: coupled logistic maps

To compare the reconstruction performance of the EBP against the classical BP, we consider coupled logistic maps, $f(x_i) = ax_i(1-x_i)$ with $a = 3.990$, via the pairwise coupling function $h(x_i,x_j) = x_ix_j$ with overall coupling strength $\alpha = 5\times10^{-4}$, see Introduction 1. Here, we consider a different coupling function given by $h(x_i,x_j) = x_j^2$ and analyze for distinct network structures, see Figure 6. We observe that the EBP method outperforms the basis pursuit on all occasions. If we compare the profile of the curves, all curves look similar to each other. The difference is that in b) and d), EBP requires more data to reconstruct the network structure. This phenomenon was predicted by our estimate in the expression (2.9). Since the maximum degree is larger, the sparsity level s of the target sparse vector is also larger, implying that n_0 grows.

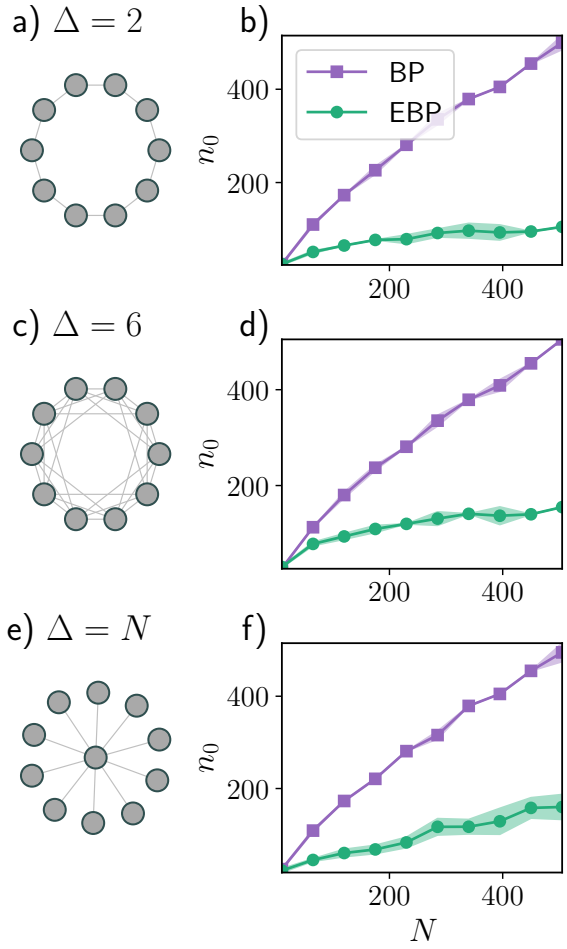

Figure 6 – **Comparison between BP and EBP under different network structures.** a) Ring graph with maximum degree $\Delta = 2$. b) The minimum length of time series n_0 for a successful reconstruction versus system size N, and similarly in d) and f). c) Lattice graph with maximum degree $\Delta = 6$. e) Star graph where the maximum degree grows with the system size. Basis pursuit (BP) and ergodic basis pursuit (EBP) are shown in (purple) squares and (green) circles, respectively. The network dynamics parameters are $a = 3.990$ and coupling strength $\alpha = 1 \times 10^{-3}/\Delta$, so the coupling term in the network dynamics is normalized as we vary N. The shaded area corresponds to the standard deviation with respect to 10 distinct initial conditions uniformly drawn in $[0,1]^N$. The Kernel density estimation of ν is used with bandwidth $\chi = 0.05$. The multivariate time series is generated without noise.

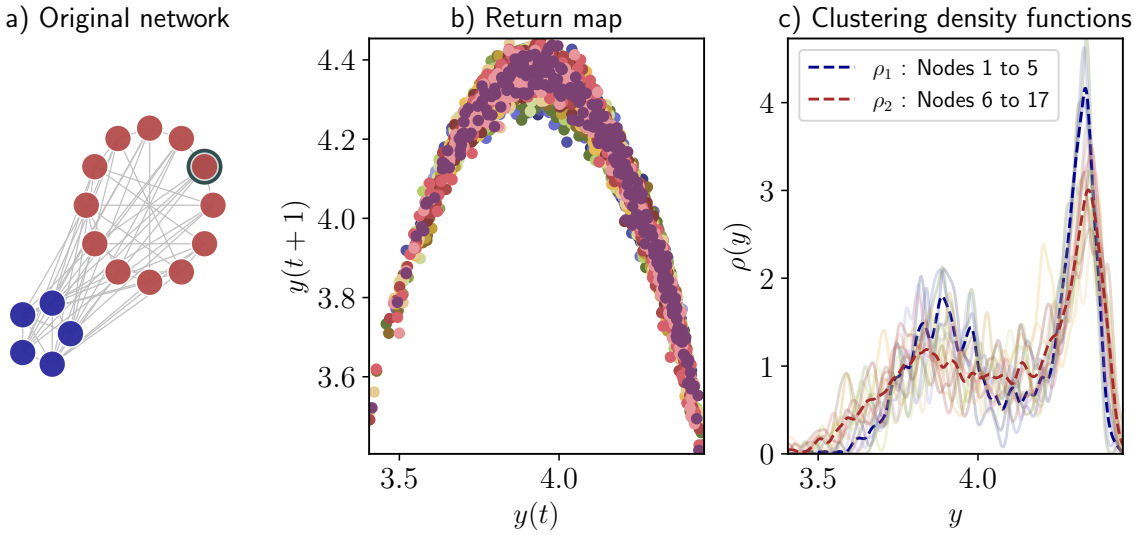

Figure 7 – **Network dynamics of experimental optoelectronic data.** a) Original optoelectronic network with two groups of nodes — dark gray node is marked for future reference. b) Return map for all nodes in the network. c) Densities function ρ_i for each node i (in light color) estimated using each node's time series. Clustering density estimation displays two resulting densities corresponding to two groups of nodes, in blue and red. The density estimation utilizes a Gaussian kernel with bandwidth $\chi = 0.05$.

2.4 Reconstruction of experimental optoelectronic networks

The data is generated from a network of optoelectronic units whose nonlinear component is a Mach-Zehnder modulator (HART *et al.*, 2019). The network is modeled as

$$x(t+1) = \beta I_\theta(x_i(t)) - \alpha \sum_{j=1}^{17} L_{ij} I_\theta(x_j(t)) \bmod 2\pi, \tag{2.18}$$

where the normalized intensity output of the Mach-Zehnder modulator is given by $I_\theta(x) = \sin^2(x+\theta)$, x represents the normalized voltage applied to the modulator, β is the feedback strength, θ is the operating point set to $\frac{\pi}{4}$ and L is the Laplacian matrix — $L_{ij} = \delta_{ij}k_i - A_{ij}$, where δ_{ij} is the Kronecker delta and k_i is the i-th node degree. The experiments were done by varying the coupling strength between the nonlinear elements in an undirected network, depicted in Figure 7 a). We will show results for coupling $\alpha = 0.171875$.

Instead of having access to trajectories from (2.18), we have access to the noisy experimental multivariate time series $\{y_1(t), \ldots, y_{17}(t)\}_{t=1}^{264}$, whose return map is depicted in Figure 7 b). Thus, we are naturally in the setting of (2.16) the randomly perturbed version of the EBP. Typically, for experimental data the noise level ξ is unknown. So, we use the constraint ε in (2.16) as a parameter to tune and search for the correct incoming connections.

The key idea is as follows. For large values of ε we have that $c_\nu^\star(\varepsilon) = 0$ is a solution to (2.16). Next, for moderate values of ε, the coefficients corresponding to the isolated dynamics

appear in $c_\nu^\star(\varepsilon)$. As we decrease ε, we start observing correct connections that are present over multiple values of ε. We aim to identify those robust connections. This can be formulated as an algorithm that we call *relaxing path*, which is described in Appendix A.1. The algorithm consists in solving (2.16) for multiple values of ε while checking which entries of $c_\nu^\star(\varepsilon)$ that correspond to connections persist as ε varies.

To apply these ideas to the experimental data, we first perform a pre-processing. Most of the data are concentrated in a portion of the phase space with scarce excursions to other parts. Thus, we first restrict the data to a portion of the phase space mostly filled, see further details in the Appendix A.2. After this procedure, we obtain a parabolic shape of the return map that corresponds to the restriction of the original optoelectronic network dynamics F onto the interval $\mathscr{A} = [3.4, 4.5]$ over 264-time steps, which we denote $\tilde{F} = F|_{\mathscr{A}}$. Hence, $\tilde{F}$ lies in the span of the quadratic polynomials, and we use $\mathscr{L} = \{\phi_i^p(x_i) = x_i^p : p = 0,1,2\}$. To perform a Gram-Schmidt process, we estimate the ν using all trajectories of a group of nodes through kernel density estimator, improving the estimate accuracy. We assume $d\nu = \rho_1 \times \rho_2 dxdy$ is a product of two densities. Nodes 1 to 5 have the density illustrated in blue, and the remaining nodes have the density in red in the right panel of Figure 7.

The left panel of Figure 8 displays the relaxing path algorithm probing a node (the marked dark gray node in Figure 7) for three distinct ε values. For each ε, we use (2.11) to construct from $c_\nu^\star(\varepsilon)$ the weighted subgraph corresponding to the probed node's neighbors. As we vary ε all edge weights decrease in magnitude (edge thickness), in particular false connections (in orange) that are not robust against variation of ε. In fact, for the smallest ε (in the left) we observe a few false connections whose edge weights are smaller than the true connections (in gray). As we increase ε, a few false connections start to vanish. Further increasing ε only the robust connections are present and the algorithm stops. Since the algorithm is node-dependent, we quantify the overall reconstruction performance in the parameter interval via a weighted false link proportion for each node, expressed in the Appendix A.2.3, and then average over all 17 nodes. The right panel of Figure 8 shows that the algorithm identifies the original network structure successfully within an interval of the parameter ε.

2.5 Mathematical analysis and preliminaries

In the remainder of this paper we prove our main results Theorems 2.6.1, 2.7.1 and 2.8.2. To this end we briefly recall some definitions and established results from compressive sensing (FOUCART; RAUHUT, 2013) and exponentially mixing dynamical systems (HANG; STEINWART, 2017).

We introduce the notation $[m] := \{1,2,\ldots,m\}$. For a matrix $\Phi \in \mathbb{R}^{n\times m}$ and a subset $\mathscr{S} \subseteq [m]$, $\Phi_{\mathscr{S}}$ indicates the column submatrix of Φ consisting of the columns indexed by $\mathscr{S}$.

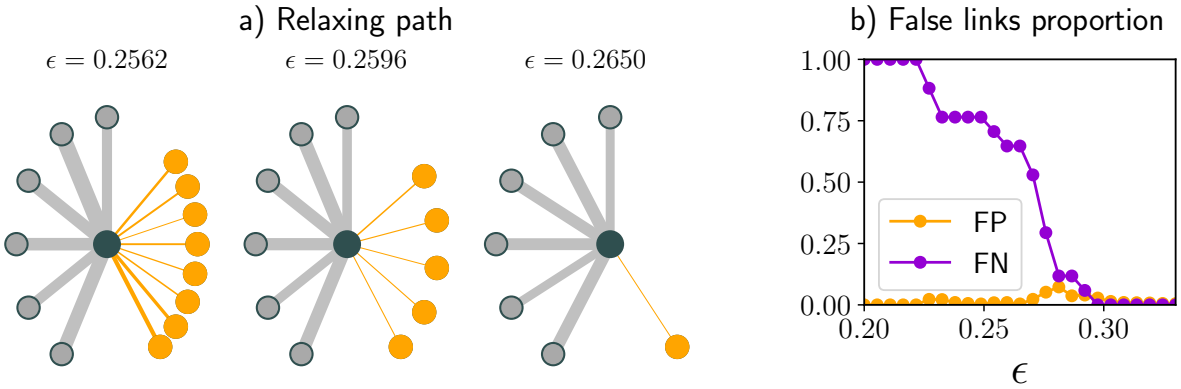

Figure 8 – **Reconstruction of the original network from experimental data.** a) Relaxing path algorithm is performed in the node (in dark gray) from the left panel. There are three different relaxing parameter values, where the edges are colored accordingly: the true edges (in gray) and false positives (in orange) while the thickness is the edge weight, see (2.11). b) False positive (FP) (in orange) and false negative (FN) (in purple) of the reconstructed network versus the parameter ε. We varied the ε parameter through 25 values equally spaced in the interval $\mathscr{E} = [0.20, 0.33]$. We employ ECOS convex optimization solver (DOMAHIDI; CHU; BOYD, 2013) to solve (2.16).

We denote the transpose of Φ as Φ^T. We denote the $L^2(\mu)$ norm of a function $\psi : M^N \to \mathbb{R}$ as

$$\|\psi\|_\mu = \left(\int |\psi(x)|^2 d\mu(x) \right)^{1/2},$$

and denote $\|\psi\|_\infty := \sup_{x \in M^N} |\psi(x)|$. Also, we denote $\lfloor \beta \rfloor$ as is the largest number $p \in \mathbb{N}$ satisfying $p \leq \beta$.

Let $\{M_i\}_{i \in [N]}$ be a collection of subsets of $\mathbb{R}$. For $\mathscr{J} \subset [N]$ denote the canonical projection by

$$\pi_{\mathscr{J}} : M^N \to \prod_{i \in \mathscr{J}} M_i. \tag{2.19}$$

2.5.1 Network library

Consider a network dynamics in (2.1). Suppose that for each $i \in [N]$ there exists $m_i \in \mathbb{N}$ such that the isolated map f_i is in the span of the set $\{\phi_i^p : p \in [m_i]\}$ of functions $\phi_i^p : M_i \to \mathbb{R}$, i.e., $f_i = \sum_{p=1}^{m_i} c_i^p \phi_i^p$. We denote the collection of all these functions as

$$\mathscr{I} = \{\phi_i^p : i \in [N], p \in [m_i]\}.$$

Similarly, for each $i, j \in [N]$ there exist $m_i, m_j \in \mathbb{N}$ such that the pairwise coupling function h_{ij} lies in the span of the set $\{\phi_{ij}^{pq} : p \in [m_i], q \in [m_j]\}$ of functions $\phi_{ij}^{pq} : M_i \times M_j \to \mathbb{R}$, i.e., $h_{ij} = \sum_{q=1}^{m_j} \sum_{p=1}^{m_i} c_{ij}^{pq} \phi_{ij}^{pq}$. We denote the collection of all these functions as

$$\mathscr{P} = \{\phi_{ij}^{pq} : i, j \in [N], p \in [m_i], q \in [m_j]\}.$$

We remove any redundancy in the collections $\mathscr{I}$ and $\mathscr{P}$. In particular, we make explicit the constant function 1 to avoid a trivial redundancy. We define the network library:

Definition 2.5.1 (Network Library)**.** We call *network library* the collection of functions

$$\mathscr{L} = \{1\} \cup \mathscr{I} \cup \mathscr{P} \tag{2.20}$$

that represent the network dynamics map F_α in (2.1).

The network library can capture the network structure because the basis functions correspond to pairwise interactions. For the node i dynamics, a nonzero coefficient of $\phi_{ij}^{pq} \in \mathscr{L}$ are associated with an edge between node i and j in the network. More precisely, the node i of the network is identified by the labeled coordinate on M_i. The following definition identifies the edge:

Definition 2.5.2 (Edge via Network Library)**.** Let $i \in [N]$ and F_i has a representation in $\mathscr{L}$. Let $\mathscr{L}_i \subset \mathscr{L}$ be a subset that contains all necessary basis functions such that $F_i \in \text{span } \mathscr{L}_i$. If $\phi_{ij}^{pq} \in \mathscr{L}_i$ for $j \in [N]$, $p \in [m_i], q \in [m_j]$, then there is an directed edge from j to i.

A priori, the network library has no natural ordering, so that we can introduce an ordered network library. We choose the following ordering: it first disposes of the constant function. Then, it is followed by the functions in $\mathscr{I}$, which are ordered fixing the $i \in [N]$ and letting run the index $p \in [m_i]$. Finally, the set $\mathscr{P}$ is ordered, fixing an element of the index set $\{(i,j) \in [N] \times [N]\}$ (which is organized in lexicographic order) and running through the index set $\{(p,q) \in [m_i] \times [m_j]\}$ (also organized in lexicographic order), i.e.,

$$\begin{aligned}\mathscr{L}^o = \{1, \phi_1^1(x_1), \ldots, \phi_1^{m_1}(x_1), \phi_2^1(x_2), \ldots, \phi_2^{m_2}(x_2), \ldots, \phi_N^1(x_N), \ldots, \phi_N^{m_N}(x_N), \\ \phi_{11}^{11}(x_1,x_1), \ldots, \phi_{NN}^{m_N m_N}(x_N,x_N)\}.\end{aligned} \tag{2.21}$$

We abuse notation and denote the ordered network library simply as $\mathscr{L}$.

We also define an s-sparse representation of the network dynamics F_α in a network library. Let us define an s-sparse vector.

Definition 2.5.3 (Sparse vector)**.** A vector $u \in \mathbb{R}^m$ is said to be s-sparse if it has at most s nonzero entries, i.e.,

$$|\{j \in \{1, \ldots, m\} : u^j \neq 0\}| \leq s.$$

Each node in the network has its sparsity level in the library, but we consider an upper bound in the sparsity level to depend only on one parameter s. To make notation easier in next definition, let $\mathscr{L} = \{\phi_l : M^N \to \mathbb{R} : l \in [m]\}$ be the network library, where m is its cardinality.

Definition 2.5.4 (Sparse Network Dynamics Representation)**.** $F_\alpha : M^N \to M^N$ has an s-sparse representation in $\mathscr{L}$ if there exists a set $\{c_1, \ldots, c_N\} \subset \mathbb{R}^m$ of $s-$sparse vectors such that the coordinate $i \in [N]$ is given by $F_i = \sum_{l=1}^m c_i^l \phi_l$, where $c_i = (c_i^1, \ldots, c_i^m) \in \mathbb{R}^m$.

2.5.2 Sparse recovery

Here we outline the results of sparse recovery employed in the paper. The next proposition states an equivalent expression to the restricted isometry constant and restricted isometry property.

Proposition 2.5.4.1. The $s-$th restricted isometry constant δ_s is given by

$$\delta_s = \max_{\mathscr{S}\subset[m],\text{card}(\mathscr{S})\leq s} \|\Phi_{\mathscr{S}}^T\Phi_{\mathscr{S}} - \mathbf{1}_s\|_2,$$

$\Phi_{\mathscr{S}}$ is the submatrix of Φ composed by the columns supported in $\mathscr{S}\subset[m]$.

Let the coherence of a matrix Φ be given by $\eta(\Phi) := \max_{i\neq j}\left|\langle v_i, v_j\rangle\right|$ defined over distinct pairs of normalized (Euclidean norm) columns of the matrix Φ. The coherence upper bounds the restricted isometry constant, and we use this fact in our proof:

Proposition 2.5.4.2 (Coherence bounds restricted isometry constant)**.** If the matrix $\Phi\in M^{n\times m}$ has ℓ_2-normalized columns $\{v_1,\ldots,v_m\}$, then

$$\delta_1 = 0, \quad \delta_2 = \eta, \quad \delta_s \leq \eta(s-1), s\geq 2.$$

Proof. See proof in (FOUCART; RAUHUT, 2013). □

The uniqueness of solutions of the ergodic basis pursuit is a consequence of the following results.

Theorem 2.5.5 (Uniqueness of noiseless recovery (CANDÈS, 2008; FOUCART; RAUHUT, 2013))**.** Suppose $y=\Phi c$ where $c\in\mathbb{R}^m$ is an $s-$sparse vector. Also, suppose that the $2s$-th restricted isometry constant of the matrix $\Phi\in M^{n\times m}$ satisfies $\delta_{2s}<\sqrt{2}-1$. Then c is the unique minimizer of

$$\min_{u\in\mathbb{R}^m}\|u\|_1 \quad \text{subject to } \Phi u = y.$$

Proof. See proof in (CANDES; TAO, 2005; FOUCART; RAUHUT, 2013). □

In case of measurement corrupted by noise, the following result holds:

Theorem 2.5.6 (Noisy recovery)**.** Suppose $y=\Phi c+z$ with $\|z\|_2\leq\varepsilon$, and denote $c^\star$ the solution to the convex minimization problem

$$\min_{\tilde{u}\in\mathbb{R}^m}\|\tilde{u}\|_1 \text{ subject to } \|y-\Phi\tilde{u}\|_2\leq\varepsilon. \tag{2.22}$$

Assume that $\delta_{2s}<\sqrt{2}-1$. Then the solution to (2.22) obeys

$$\|c^\star - c\|_2 \leq K_0 s^{-1/2}\|c-c_s\|_1 + K_1\varepsilon,$$

for constants $K_0, K_1 > 0$ and c_s denote the vector c with all but the $s-$largest entries set to zero.

Proof. See proof in (CANDÈS, 2008; FOUCART; RAUHUT, 2013). □

2.5.3 Exponential mixing condition

We consider a class of chaotic dynamical systems — exponentially mixing systems — that satisfies a concentration inequality obtained in (HANG; STEINWART, 2017). Here, we state this result applied to network dynamics.

Definition 2.5.7 (Exponential mixing condition)**.** The network dynamics (F,μ) satisfies the exponential mixing condition for some constant $\gamma>0$ if for all $\psi\in\mathscr{C}^1(M^N;\mathbb{R})$ and $\varphi\in L^1(\mu)$ there exists a constant $K(\psi,\varphi)>0$ such that

$$\left|\int_{M^N}\psi\cdot(\varphi\circ F^n)d\mu-\int_{M^N}\psi d\mu\int_{M^N}\varphi d\mu\right|\leq K(\psi,\varphi)e^{-\gamma n},\quad n\geq 0. \tag{2.23}$$

We state an adapted version for network dynamics of the concentration inequality (HANG; STEINWART, 2017) for $\mathscr{C}^1(M^N;\mathbb{R})$ observables.

Theorem 2.5.8 (Bernstein inequality for exponential mixing network dynamics (HANG; STEINWART, 2017).)**.** Let (F,μ) be an exponential mixing network dynamical system on M^N for some constant $\gamma>0$. Moreover, let $\psi\in\mathscr{C}^1(M^N;\mathbb{R})$ be a function such that $\int_{M^N}\psi d\mu=0$ and assume that there exist $\varsigma>0$, $\varkappa>0$ and $\sigma\geq 0$ such that $\|D\psi\|_\infty\leq\varsigma$, $\|\psi\|_\infty\leq\varkappa$, and $\|\psi^2\|_\mu^2\leq\sigma^2$. Let $\mathscr{N}\subset\mathbb{N}$ be defined as

$$\mathscr{N}:=[3,\infty)\bigcap\left\{p\in\mathbb{N}\ :\ p^2\geq\frac{808(3\varsigma+\varkappa)}{\varkappa}\text{ and }\frac{p}{(\ln p)^2}\geq 4\right\}.$$

Then, for all $\varepsilon>0$ and all

$$n\geq n_0:=\max\left\{e^{\frac{3}{\gamma}},\min_{\mathscr{N}}p\right\}, \tag{2.24}$$

we have

$$\mu\left(x_0\in M^N\ :\ \left|\frac{1}{n}\sum_{k=0}^{n-1}\psi\circ F^k(x_0)\right|\geq\varepsilon\right)\leq 4e^{-\theta(n,\varepsilon,\sigma,\varkappa)}, \tag{2.25}$$

where

$$\theta(n,\varepsilon,\sigma,\varkappa):=\frac{n\varepsilon^2}{8(\ln n)^{\frac{2}{\gamma}}(\sigma^2+\varepsilon\varkappa/3)}.$$

2.5.4 Semimetric between probability measures

We consider exponentially mixing systems that have near product structure. To be more precise, we introduce a semimetric between probabilities measures suitable to our results. Let $\mathscr{M}(M^N)$ be the set of probability measures on M^N. We introduce a probability semimetric (RACHEV, 1991) between measures on $\mathscr{M}(M^N)$ over a reference finite set of functions $\mathscr{K}$ that is composed by functions on the given network library $\mathscr{L}$. In other words, elements of $\mathscr{K}$ are of the form $\phi_{ij}^{pq}\circ\pi_{\mathscr{J}}$ with $i,j\in\mathscr{J}\subset[N]$. They are integrated over a lower dimensional space than the ambient space M^N, which motivates to define a semimetric out of it, rather than using other metrics on $\mathscr{M}(M^N)$.

Definition 2.5.9. For any $\mu, \nu \in \mathscr{M}(M^N)$ we define the semimetric over a reference network library $\mathscr{L}$ as

$$d_{\mathscr{K}}(\mu,\nu) = \max_{\psi\in\mathscr{K}} \left| \int_{M^N} \psi d\mu - \int_{M^N} \psi d\nu \right|. \tag{2.26}$$

$d_{\mathscr{K}}(\mu,\nu)$ is a semimetric and not a metric because: it is symmetric, it satisfies the triangular inequality, and when $\mu = \nu$ implies that $d_{\mathscr{K}}(\mu,\nu) = 0$ but not the converse. Indeed, consider the set $\mathscr{K}$ given by

$$\mathscr{K} = \{\psi_i : M_i \to \mathbb{R} : i \in [N], \int \psi_i dx_i = 0, \psi_i(0) = 0\},$$

where we assume that $0 \in M_i$ for any $i \in [N]$. Moreover, let δ_0 be the Dirac measure at 0. Consider the following two product measures

$$\mu = \mathrm{Leb}^N \qquad \nu = \delta_0^N.$$

It follows that $d_{\mathscr{K}}(\mu,\nu) = 0$ but $\mu \neq \nu$.

In what follows in Section 2.7, it is useful to consider the following finite set $\mathscr{K} = (\mathscr{L}\cdot\mathscr{L})$, where $(\mathscr{L}\cdot\mathscr{L}) = \{(\psi_i\cdot\psi_j) : \psi_i, \psi_j \in \mathscr{L}\}$, removing any redundancy.

2.5.5 Orthogonal polynomials

We recall some results for orthonormal polynomials. First, let us state an inequality for orthonormal polynomials in one variable (SZEGÔ, 1939; FTOREK; ORŠANSKY, 2014). Here we consider a system of orthonormal polynomials $\{\varphi_p(x)\}_{p\geq 0}$ with respect to a measure ν that is absolutely continuous to Lebesgue, whose density is ρ. Since we are in the one variable case, the index p corresponds to the degree to which the coefficient x^p is positive.

Theorem 2.5.10 (One variable Korous inequality (SZEGÔ, 1939; FTOREK; ORŠANSKY, 2014))**.** Let $\{\varphi_p(x)\}_{p\geq 0}$ be a generalized system of orthonormal polynomials with respect to (w.r.t.) the density $\lambda(x)$ and $\{\tilde{\varphi}_p(x)\}_{p\geq 0}$ be a system of orthonormal polynomials w.r.t. the density $\tilde{\lambda}(x)$ such that

$$\lambda(x) = \rho(x)\tilde{\lambda}(x),$$

be two weight (density) functions on the segment (a,b), where $\rho(x) \geq \rho_0 > 0$ and ρ is Lipschitz with constant $\mathrm{Lip}(\rho)$. Then the following estimation

$$|\varphi_p(x)| \leq \frac{1}{\rho_0}|\tilde{\varphi}_p(x)| + \frac{K\mathrm{Lip}(\rho)}{\rho_0^{3/2}}(|\tilde{\varphi}_p(x)| + |\tilde{\varphi}_{p-1}(x)|), \tag{2.27}$$

where $\rho_0 = \min_{x\in(a,b)}\rho(x)$, $x \in (a,b)$ and $K = \max\{|a|,|b|\}$.

We also recall a result for the product of orthonormal polynomials (DUNKL; XU, 2014).

Proposition 2.5.10.1 (Proposition 2.2.1 in (DUNKL; XU, 2014)). Let $\rho(x_1,x_2) = \rho_1(x_1)\rho_2(x_2)$, where ρ_1 and ρ_2 are two weight functions of one variable. Let $\{\varphi_1^p(x_1)\}_{p\geq 0}^{\infty}$ and $\{\varphi_2^q(x_2)\}_{q\geq 0}^{\infty}$ with $p,q \in \mathbb{N}$ be sequences of orthogonal polynomials with respect to ρ_1 and ρ_2, respectively. Then a mutually orthogonal basis of the space of orthogonal polynomials of degree r with respect to ρ is given by:

$$\varphi_{12}^{pq}(x_1,x_2) = \varphi_1^p(x_1)\varphi_2^q(x_2), \quad 0 \leq p+q \leq r.$$

Furthermore, if $\{\varphi_1^p(x_1)\}_{p\geq 0}^{\infty}$ and $\{\varphi_2^q(x_2)\}_{q\geq 0}^{\infty}$ are orthonormal with respect to ρ_1 and ρ_2, respectively, then so is $\varphi_{12}^{pq}(x_1,x_2)$ with respect to ρ.

2.5.6 Gram-Schmidt process

Let ν be a measure on M^N that is absolutely continuous with respect to Lebesgue. We address the problem of ortho-normalizing the ordered network library $\mathscr{L}$ with respect to a measure ν. Let us denote the inner product w.r.t. ν as

$$\langle \phi_k, \phi_l \rangle = \int_{M^N} \phi_k \phi_l d\nu \qquad \|\phi_l\|_\nu^2 = \langle \phi_l, \phi_l \rangle. \tag{2.28}$$

We consider the Gram-Schmidt (GS) process, which is a recursive method given as

$$\begin{aligned} \hat{\varphi}_1 &= \phi_1 \\ \hat{\varphi}_{k+1} &= \phi_{k+1} - \sum_{l=1}^{k} \langle \phi_{k+1}, \varphi_l \rangle \varphi_l, \\ \varphi_k &:= \frac{\hat{\varphi}_k}{\|\hat{\varphi}_k\|_\nu}, \quad k \geq 1. \end{aligned} \tag{2.29}$$

From the ordered network library $\mathscr{L}$ the induced library $\mathscr{L}_\nu = \{\varphi_k : M^N \to \mathbb{R} : k \in [m]\}$ is given by each k-th orthonormal function written as a linear combination, whose coefficients are projections on the preceding orthonormal functions.

2.6 Network library is preserved under Gram-Schmidt process

To ensure that the ergodic basis pursuit has a unique solution, the library matrix used in the reconstruction must satisfy the restricted isometry property, as defined in Equation (2.8). However, a priori, the library matrix associated with the network library $\mathscr{L}$, in which F_α has a sparse representation, does not satisfy RIP. Our strategy is to introduce a new library $\mathscr{L}_\nu$ that is orthonormal with respect to a suitable measure ν in $L^2(\nu)$, where the associated library $\Phi_\nu(X)$ satisfies RIP.

2.6.1 The set of pairwise polynomials

We consider a network library given by polynomials in N variables of degree at most r. This also can be applied to trigonometric polynomials in N variables.

Given $r \geq 2$, let us denote the exponent vector set

$$\mathscr{V}_r := \{(p,q) \in [r-1]^2 : p+q \leq r\}, \tag{2.30}$$

which is organized in graded lexicographic order and denoted as $(p',q') \prec (p,q)$. Moreover, denote

$$\begin{aligned}
\mathscr{I}_r &= \{\phi_i^p(x_i) = x_i^p \ : \ i \in [N], p \in [r]\}, \\
\mathscr{P}_r &= \{\phi_{ij}^{pq}(x_i,x_j) = x_i^p x_j^q \ : \ i,j \in [N], i \neq j, (p,q) \in \mathscr{V}_r\},
\end{aligned}$$

where we remove any redundancy. We can unify the notation for both if we denote elements of $\mathscr{I}_r$ as $\phi_{i0}^{p0}(x_i,x_j) = x_i^p$. We define the set of pairwise polynomials in N variables with a degree at most r

$$\begin{aligned}
\mathscr{L} &= \{1\} \cup \mathscr{I}_r \cup \mathscr{P}_r \\
&= \{\phi_{ij}^{pq}(x_i,x_j) = x_i^p x_j^q \ : \ i \in [N], j \in \{0\} \cup [N], i \neq j \\
&\qquad\qquad p = \{0\} \cup [r], q \in \{0\} \cup [r-1], \\
&\qquad\qquad p+q \leq r\},
\end{aligned}$$

whose cardinality is given by $m = \binom{N}{2}\binom{r}{2} + Nr + 1$. In fact, the independent polynomial 1 contributes with one term. The cardinality of $\mathscr{I}_r$ is Nr because for each $i \in [N]$ there are r polynomials in the subset $\{\varphi_{i0}^{p0}\}_{p\in[r]}$. Finally, for $\mathscr{P}_r$ fix a pair $i,j \in [N]$ with $i \neq j$. For each pair, the degree of the pairwise polynomial is $p+q = d \in [r]$. Since they are constrained through their sum, for each degree $d \in [r]$, the first component in the sum $p \in \{1,\ldots,d-1\}$, which also determines the value of q correspondingly. Then, there are total of $\sum_{d=1}^r d - r$ possible combinations. Rewriting it

$$\begin{aligned}
\sum_{d=1}^{r} d - r &= \frac{r(r+1)}{2} - r \\
&= \frac{r(r-1)}{2} \\
&= \binom{r}{2}.
\end{aligned}$$

Running over all possible distinct pairs i,j, we obtain the total cardinality of $\mathscr{P}_r$ equal to $\binom{N}{2}\binom{r}{2}$.

Here we adopt the following ordering: fix $j,q = 0$ and start with $p = 0$. Then, for each $i \in [N]$, we run through $p \in [r]$, covering all monomials that depend on one variable. Subsequently, for each element in $\{(i,j) \in [N]^2 : i \in [N], j = i+1,\ldots,N\}$ (organized in lexicographic order), we run through the exponent vector set $\mathscr{V}_r$.

2.6.2 Network library is preserved

Given a trajectory $\{x(t)\}_{t=0}^{n}$ that is sampled from μ_α, the natural choice would be to orthonormalize with respect to μ_α itself. However, it does not necessarily preserve the sparsity of the representation of F_α in the network library $\mathscr{L}$. The next theorem states that the GS process over $\mathscr{L}$ with respect to a product measure ν introduces a new network library $\mathscr{L}_\nu$, and also, F_α is still sparsely represented in $\mathscr{L}_\nu$. In this new basis, the sparsity level depends on the maximum degree r and the sparsity level of the representation in $\mathscr{L}$.

Denote the product measure as $\nu = \prod_{i=1}^{N} \nu_i$ and denote

$$\mathbb{E}(x_i^p) = \int_M x_i^p \, d\nu_i(x_i), \quad i \in [N]. \tag{2.31}$$

Consider the following

Theorem 2.6.1 (Network library is preserved)**.** Let ν be a product measure on M^N that is absolutely continuous with respect to the Lebesgue measure. Gram-Schmidt process maps an s-sparse representation of F_α in the network library $\mathscr{L}$ to an $\omega_r(s)$-sparse representation in the orthonormal network library $\mathscr{L}_\nu$ in $L^2(\nu)$, with

$$\omega_r(s) = \left(\left\lfloor \frac{r}{2} \right\rfloor \left(r - \left\lfloor \frac{r}{2} \right\rfloor \right) + r + 1 \right) s. \tag{2.32}$$

We divide the proof into two parts: first, we show that the GS process maps the network library $\mathscr{L}$ to another network library $\mathscr{L}_\nu$ that is orthonormal w.r.t. ν. The second part is to calculate the sparsity level of the representation of F_α in $\mathscr{L}_\nu$.

2.6.2.1 Proof of Theorem 2.6.1

When we perform the GS process in $L^2(\nu)$ as in (2.29), to orthonormalize $\mathscr{L}$ with respect to the measure ν, the first element in $\mathscr{L}_\nu$ is evidently 1. Following the order in the network library $\mathscr{L}$ in (2.2), we can show that a general form of all polynomials that depend on only one variable is given by the proposition below.

Proposition 2.6.1.1 (Formula of orthonormal functions in one variable)**.** Let $i \in [N]$ and $p \in [r]$. Then any $\varphi_{i0}^{p0} \in \mathscr{L}_\nu$ is given by

$$\begin{aligned} \hat{\varphi}_{i0}^{p0}(x_i) &= x_i^p - \mathbb{E}(x_i^p) - \sum_{l=1}^{p-1} \langle x_i^p, \varphi_{i0}^{l0} \rangle \varphi_{i0}^{l0}(x_i), \\ \varphi_{i0}^{p0}(x_i) &= \frac{\hat{\varphi}_{i0}^{p0}(x_i)}{\|\hat{\varphi}_{i0}^{p0}\|_\nu}, \end{aligned} \tag{2.33}$$

and

$$\mathbb{E}(\hat{\varphi}_{i0}^{p0}(x_i)) = \mathbb{E}(\varphi_{i0}^{p0}(x_i)) = 0. \tag{2.34}$$

We prove this statement in two parts. First, we continue the Gram-Schmidt process over the ordering of $\mathscr{L}$. So, fix $i = 1$ and run over $p \in [r]$. The next term after the constant function 1 is

$$\begin{aligned}\hat{\varphi}_{10}^{10}(x) &= \phi_{10}^{10}(x_1) - \langle \phi_{10}^{10}, 1\rangle 1 \\ &= x_1 - \mathbb{E}(x_1),\end{aligned}$$

and consequently, $\varphi_{10}^{10}(x_1) = \frac{\hat{\varphi}_{10}^{10}(x_1)}{\|\hat{\varphi}_{10}^{10}\|_\nu}$, which satisfies (2.33) and (2.34). To calculate the next element $\varphi_{10}^{20}(x_1)$, we follow (2.29):

$$\begin{aligned}\hat{\varphi}_{10}^{20}(x_1) &= \phi_{10}^{20}(x_1) - \langle \phi_{10}^{20}, 1\rangle 1 - \langle \phi_{10}^{20}, \varphi_{10}^{10}\rangle \varphi_{10}^{10}(x_1) \\ &= x_1^2 - \mathbb{E}(x_1^2) - \langle x_1^2, \varphi_{10}^{10}\rangle \varphi_{10}^{10}(x_1),\end{aligned}$$

and consequently, $\varphi_{10}^{20}(x_1) = \frac{\hat{\varphi}_{10}^{20}(x_1)}{\|\hat{\varphi}_{10}^{20}\|_\nu}$. Following the ordering, we run over all functions of the form φ_{10}^{p0}, repeating the Gram-Schmidt process (2.29) to show that they satisfy (2.33) and (2.34).

The next functions involve coordinates that are different from $i = 1$. To prove that these functions satisfy (2.33) and (2.34), we run a recursive argument. Fix $i = 1$ and $j = 2$, and let us consider the orthogonal function for $p \in [r]$ using Gram-Schmidt process:

$$\hat{\varphi}_{20}^{p0}(x_2) = x_2^p - \mathbb{E}(x_2^p) - \sum_{l=1}^{r} \langle x_2^p, \varphi_{10}^{l0}\rangle \varphi_{10}^{l0}(x_1) - \sum_{l=1}^{p-1} \langle x_2^p, \varphi_{20}^{l0}\rangle \varphi_{20}^{l0}(x_2).$$

Note that if all inner products of the form $\langle x_2^p, \varphi_{10}^{l0}\rangle$ are zero, above equation satisfies (2.33). We state the following lemma:

Lemma 2.6.2. Let $i, j \in [N]$, $p, q \in [r]$. Suppose that φ_{i0}^{p0} is an orthonormal polynomial with respect to ν, i.e., it satisfies (2.33) and $\varphi_{i0}^{p0} \in \mathscr{L}_\nu$. Then,

$$\langle \phi_{j0}^{q0}, \varphi_{i0}^{p0}\rangle = 0.$$

whenever $i \neq j$.

Proof. By Fubini's theorem, we have that

$$\begin{aligned}\langle \phi_{j0}^{q0}, \varphi_{i0}^{p0}\rangle &= \int_{M^N} \phi_{j0}^{q0}(x_j)\varphi_{i0}^{p0}(x_i) d\nu(x_1, \ldots, x_N) \\ &= \langle \phi_{j0}^{q0}, 1\rangle \langle \varphi_{i0}^{p0}, 1\rangle.\end{aligned}$$

Since φ_{i0}^{p0} satisfies (2.33), it is orthonormal to 1, and the claim holds. □

We use above Lemma 2.6.2 to the inner product $\langle x_2^p, \varphi_{10}^{l0}\rangle$, where φ_{10}^{l0} satisfies (2.33). We conclude that for any $p \in [r]$: φ_{20}^{p0} also satisfies (2.33) and (2.34). We run iteratively, choosing $i \geq 2$ and $j = i + 1$, and repeating the argument to conclude the proof of Proposition 2.6.1.1.

For polynomials involving two variables, it is enough to construct them from the orthonormal polynomials in one variable as follows:

Proposition 2.6.2.1 (Formula of orthonormal functions in two variables)**.** Let $r \geq 2$, $i, j \in [N]$ with $i \neq j$ and $(p,q) \in \mathscr{V}_r$. Then

$$\varphi_{ij}^{pq}(x_i, x_j) = \varphi_{i0}^{p0}(x_i)\varphi_{j0}^{q0}(x_j). \tag{2.35}$$

Proof. The measure $\nu = \prod_{i=1}^{N} \nu_i$. For each marginal ν_i, let ρ_i be the density function. Then, we apply Proposition 2.5.10.1 for every distinct pair of nodes $i, j \in [N]$. □

To construct $\mathscr{L}_\nu$ we combine Proposition 2.6.1.1 and Proposition 2.6.2.1. The Gram-Schmidt process induces a set of orthonormal polynomials in one variable that satisfies the ordering of $\mathscr{L}$. The ordering of polynomials in two variables in $\mathscr{L}_\nu$ also satisfies, by construction, the ordering in $\mathscr{L}$. This proves the first part of Theorem 2.6.1.

To prove the second part of the theorem, we also use that $\mathscr{L}_\nu$ is constructed via the Gram-Schmidt process. Let $u, u_\nu \in \mathbb{R}^m$ be vectors with $m = \binom{N}{2}\binom{r}{2} + Nr + 1$ given by

$$u = (1, x_1, \ldots, x_1^r, x_2, \ldots, x_2^r, \ldots, x_1x_2, \ldots, x_{N-1}x_N^r)$$

and

$$u_\nu = (1, \varphi_{10}^{10}(x_1), \ldots, \varphi_{10}^{r0}(x_1), \varphi_{20}^{10}(x_2), \ldots, \varphi_{20}^{r0}(x_2), \ldots, \varphi_{12}^{11}(x_1, x_2), \ldots, \varphi_{N-1,N}^{r-1,1}(x_{N-1}, x_N)).$$

Each coordinate of u is an element of $\mathscr{L}$ that can be written as a linear combination of elements in $\mathscr{L}_\nu$. In fact, we rewrite (2.33) as

$$x_i^p = \|\hat{\varphi}_{i0}^{p0}\|_\nu \varphi_{i0}^{p0}(x_i) + \mathbb{E}(x_i^p) + \sum_{l=1}^{p-1} \langle x_i^p, \varphi_{i0}^{l0}\rangle \varphi_{i0}^{l0}(x_i), \tag{2.36}$$

which expresses the polynomials in one variable as a linear combination of orthonormal polynomials in one variable. For the two variables polynomials of the form $x_i^p x_j^q$, we replace each term in the multiplication by (2.36) and

1. Replace any multiplication of orthonormal polynomial of the form $\varphi_{i0}^{p0}(x_i)\varphi_{j0}^{q0}(x_j)$ by the orthonormal polynomial in two variables Equation (2.35).

2. Use these identities that follow from Fubini's theorem:

$$\langle x_i^p, \varphi_{i0}^{l0}\rangle \langle x_j^q, \varphi_{j0}^{k0}\rangle = \langle x_i^p x_j^q, \varphi_{ij}^{lk}\rangle,$$

$$\mathbb{E}(x_i^p)\langle x_j^q, \varphi_{j0}^{k0}\rangle = \langle x_i^p x_j^q, \varphi_{j0}^{k0}\rangle$$

and

$$\|\hat{\varphi}_{i0}^{p0}\|_\nu \|\hat{\varphi}_{j0}^{q0}\|_\nu = \|\hat{\varphi}_{ij}^{pq}\|_\nu.$$

Then, we can recast the Gram-Schmidt process as the following linear equation

$$u^T = u_\nu^T \mathbf{R}_\nu, \tag{2.37}$$

where T denotes the transpose and $\mathbf{R}_\nu \in \mathbb{R}^{m\times m}$ is a triangular matrix given as

$$\mathbf{R}_\nu = \begin{pmatrix} 1 & \mathbf{V}_1 & \mathbf{V}_2 \\ 0 & \mathbf{U}_1 & \mathbf{U}_2^1 \\ 0 & 0 & \mathbf{U}_2 \end{pmatrix}. \tag{2.38}$$

Here $\mathbf{V}_1 \in \mathbb{R}^{rN}$ and $\mathbf{V}_2 \in \mathbb{R}^{\binom{N}{2}\binom{r}{2}}$ are given by

$$\mathbf{V}_1 = \begin{pmatrix} v_1 & v_2 & \dots & v_N \end{pmatrix} \qquad \mathbf{V}_2 = \begin{pmatrix} v_{12} & v_{13} & \dots & v_{N-1,N} \end{pmatrix},$$

where for each $i, j \in [N]$ with $i \neq j$, $v_i \in \mathbb{R}^r$ and $v_{ij} \in \mathbb{R}^{\binom{r}{2}}$:

$$v_i = \Big(\mathbb{E}(x_i), \dots, \mathbb{E}(x_i^r)\Big)$$

and

$$v_{ij} = \Big(\mathbb{E}(x_i)\mathbb{E}(x_j), \mathbb{E}(x_i)\mathbb{E}(x_j^2), \mathbb{E}(x_i^2)\mathbb{E}(x_j), \dots, \mathbb{E}(x_i^{r-1})\mathbb{E}(x_j)\Big).$$

Also, $\mathbf{U}_1 \in \mathbb{R}^{Nr\times Nr}$ and $\mathbf{U}_2 \in \mathbb{R}^{\binom{N}{2}\binom{r}{2}\times\binom{N}{2}\binom{r}{2}}$ are block diagonal matrices defined as follows:

$$\mathbf{U}_1 = \operatorname{diag}\Big(U_1, \dots, U_N\Big) \qquad \mathbf{U}_2 = \operatorname{diag}\Big(U_{12}, \dots, U_{N-1,N}\Big),$$

where for each $i, j \in [N]$, $U_i \in \mathbb{R}^{r\times r}$ and $U_{ij} \in \mathbb{R}^{\binom{r}{2}\times\binom{r}{2}}$ are given by

$$U_i = \begin{pmatrix} \|\hat{\varphi}_{i0}^{10}\|_\nu & \langle x_i^2, \varphi_{i0}^{10}\rangle & \dots & \langle x_i^r, \varphi_{i0}^{10}\rangle \\ 0 & \|\hat{\varphi}_{i0}^{20}\|_\nu & \dots & \langle x_i^r, \varphi_{i0}^{20}\rangle \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \|\hat{\varphi}_{i0}^{r0}\|_\nu \end{pmatrix}$$

$$U_{ij} = \begin{pmatrix} \|\hat{\varphi}_{ij}^{11}\|_\nu & \|\hat{\varphi}_{i0}^{10}\|_\nu\langle x_i, \varphi_{j0}^{10}\rangle & \|\hat{\varphi}_{j0}^{10}\|_\nu\langle x_i^2, \varphi_{i0}^{10}\rangle & \dots & \|\hat{\varphi}_{j0}^{10}\|_\nu\langle x_i^{r-1}, \varphi_{i0}^{10}\rangle \\ 0 & \|\hat{\varphi}_{ij}^{12}\|_\nu & 0 & \dots & 0 \\ 0 & 0 & \|\hat{\varphi}_{ij}^{21}\|_\nu & \dots & \|\hat{\varphi}_{j0}^{10}\|_\nu\langle x_i^{r-1}, \varphi_{i0}^{20}\rangle \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \|\hat{\varphi}_{ij}^{r-1,1}\|_\nu \end{pmatrix},$$

and $\mathbf{U}_2^1 \in \mathbb{R}^{Nr\times\binom{N}{2}\binom{r}{2}}$ is a block matrix

$$\mathbf{U}_2^1 = \begin{pmatrix} U_{12}^1 & U_{13}^1 & \dots & 0 \\ U_{12}^2 & 0 & \dots & 0 \\ 0 & U_{13}^3 & \dots & \vdots \\ \vdots & \ddots & \ddots & U_{N-1,N}^{N-1} \\ 0 & \dots & 0 & U_{N-1,N}^N \end{pmatrix},$$

where for each $i \in [N]$, $j = i, \dots, N$:

$$U_{ij}^{i} = \begin{pmatrix} \|\hat{\varphi}_{i0}^{10}\|_{\nu}\mathbb{E}(x_j) & \|\hat{\varphi}_{i0}^{10}\|_{\nu}\mathbb{E}(x_j^2) & \langle x_i^2 x_j, \varphi_{i0}^{10}\rangle & \dots & \langle x_i^{r-1}x_j, \varphi_{i0}^{10}\rangle \\ 0 & 0 & \|\hat{\varphi}_{i0}^{20}\|_{\nu}\mathbb{E}(x_j) & \dots & \langle x_i^{r-1}x_j, \varphi_{i0}^{20}\rangle \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & \dots & \|\hat{\varphi}_{i0}^{(r-1)0}\|_{\nu}\mathbb{E}(x_j) \end{pmatrix} \in \mathbb{R}^{r\times\binom{r}{2}}$$

$$U_{ij}^{j} = \begin{pmatrix} \mathbb{E}(x_i)\|\hat{\varphi}_{j0}^{10}\|_{\nu} & \langle x_i x_j^2, \varphi_{j0}^{10}\rangle & \mathbb{E}(x_i^2)\|\hat{\varphi}_{j0}^{10}\|_{\nu} & \dots & \mathbb{E}(x_i^{r-1})\|\hat{\varphi}_{j0}^{10}\|_{\nu} \\ 0 & \|\hat{\varphi}_{i0}^{10}\|_{\nu}\mathbb{E}(x_j^2) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{r\times\binom{r}{2}}.$$

Linear equation (2.37) is valid for every point $x \in M^N$. Hence, evaluating along the trajectory $\{x(t)\}_{t=0}^{n}$ we obtain:

$$\Phi(X) = \Phi_{\nu}(X)\mathbf{R}_{\nu}. \tag{2.39}$$

Consider an s-sparse representation in $\mathscr{L}$, then there is an s-sparse vector $c \in \mathbb{R}^m$ such that

$$\bar{x} = \Phi(X)c,$$

where we dropped the dependence on the node $i \in [N]$ for a moment. Note that $\mathscr{L}_{\nu}$ is also a set of basis functions, so there exists a c_{ν} such that

$$\Phi(X)c = \Phi_{\nu}(X)c_{\nu}.$$

(2.39) implies that $c_{\nu} = \mathbf{R}_{\nu}c$. Since c_{ν} is the linear combination of s columns of $\mathbf{R}_{\nu}$, the sparsity level of c_{ν} is given by the number of non-zero entries of $\mathbf{R}_{\nu}$ multiplied by the sparsity level s of c.

The sparsity of $\mathbf{R}_{\nu}$ columns can be upper bounded by counting the non-zero entries of columns in the block matrices involving the pairwise interaction. It is enough to calculate the maximum number of elements in the multiplication $x_i^p x_j^q$ using (2.36) for all combinations of $p, q \in [r-1]$ with $p+q \leq r$. More precisely, for a p in (2.36) there is a linear combination of $p+1$ elements of $\mathscr{L}_{\nu}$. Then, in the multiplication $x_i^p x_j^q$ there are at maximum

$$\omega_r = \max_{p,q\in[r-1], p+q\leq r} (p+1)(q+1),$$

which has the following expression

$$\omega_r = \left\lfloor \frac{r}{2} \right\rfloor \left(r - \left\lfloor \frac{r}{2} \right\rfloor\right) + r + 1.$$

So, c_{ν} is an $\omega_r(s)$-sparse vector with $\omega_r(s) = \left(\left\lfloor \frac{r}{2} \right\rfloor \left(r - \left\lfloor \frac{r}{2} \right\rfloor\right) + r + 1\right)s$.

We repeat the same argument for each $i \in [N]$ separately, concluding the proof of Theorem 2.6.1.

2.6.3 Bounds for orthonormal polynomials

In the next section, we will need bounds of orthonormal polynomials with respect to the product measure ν. We focus on the one variable case because, as we have seen in the previous section, it suffices to analyze this case.

First, note that: consider a system of orthonormal polynomials $\{\psi_p(z)\}_{p\geq 0}$ with weight (density) function $\lambda(z)$ defined on the interval $[a_2,b_2]\subset\mathbb{R}$. The linear transformation $T(x)=\alpha x+\beta$ with $\alpha\neq 0$ maps an interval $[a_1,b_1]\subset\mathbb{R}$ onto the interval $[a_2,b_2]$, and $\lambda\circ T(x)$ into λ, then the polynomials

$$\{\operatorname{sgn}(\alpha)^p|\alpha|^{\frac{1}{2}}\psi_p\circ T(x)\}_{p\geq 0}$$

are orthonormal on $[a_1,b_1]$ with the weight function $\lambda\circ T(x)$.

Consider the set of Legendre polynomials $\{L_p(z)\}_{p\geq 0}$ which is defined on $[-1,1]$ with $\lambda(z)=1$. From the above remark, any Legendre polynomial $\{\hat{L}_p(x)\}_{p\geq 0}$ defined in an arbitrary interval $[a,b]$ is given by

$$\left\{\hat{L}_p(x):=\operatorname{sgn}\left(\frac{2}{b-a}\right)^p\left|\frac{2}{b-a}\right|^{\frac{1}{2}}L_p\left(\frac{2}{b-a}(x-b)+1\right)\right\}_{p\geq 0} \tag{2.40}$$

with weight $\lambda\left(\frac{2}{b-a}(x-b)+1\right)=1$. Note that $\|L_p\|_\infty\leq 1$ (SZEGÔ, 1939), consequently,

$$\|\hat{L}_p\|_\infty\leq\left(\frac{2}{b-a}\right)^{\frac{1}{2}}\|L_p\|_\infty\leq\left(\frac{2}{b-a}\right)^{\frac{1}{2}}.$$

We apply the above observation to our case, using the Korous inequality for orthonormal polynomials 2.5.10. See the following:

Proposition 2.6.2.2 (Supremum norm of orthonormal polynomials in one variable)**.** For a given $i\in[N]$ let $M_i=[a,b]\subset\mathbb{R}$ with $b>a$. Consider the one variable orthonormal polynomials $\{\varphi_{i0}^{p0}(x_i)\}_{p\in[r]}$ with respect to ν_i, which is the one-dimensional marginal of the product measure ν. Suppose that ν_i is absolutely continuous with respect to Lebesgue and its density ρ_i is at least Lipschitz with constant $\operatorname{Lip}(\rho_i)$. Moreover, $\rho_i(x_i)>0$ for any $x_i\in M_i$. The following holds:

$$\|\varphi_{i0}^{p0}\|_\infty\leq\left(\frac{1}{\rho_0}+2\frac{a_1\operatorname{Lip}(\rho)}{\rho_0^{3/2}}\right)\left(\frac{2}{b-a}\right)^{\frac{1}{2}},\quad p\in[r],$$

where $\rho_0=\min_{i\in[N]}\{\min_{x\in[a,b]}\rho_i(x)\}$, $a_1=\max\{|a|,|b|\}$ and $\operatorname{Lip}(\rho)=\max_{i\in[N]}\operatorname{Lip}(\rho_i)$. Moreover,

$$\|D\varphi_{i0}^{p0}\|_\infty\leq\left(\frac{1}{\rho_0}+2\frac{a_1\operatorname{Lip}(\rho)}{\rho_0^{3/2}}\right)\left(\frac{2}{b-a}\right)r^2.$$

Proof. Consider the system $\{\hat{L}_p(x)\}_{p\in[r]}$ of Legendre polynomials as in (2.40) defined on M_i. Also, consider the orthonormal polynomials $\{\varphi_{i0}^{p0}(x_i)\}_{p\in[r]}$ with respect to ν_i, which is given by

$d\nu_i(x) = \rho_i(x_i)d\text{Leb}(x_i) = \rho_i(x_i)\lambda\left(\frac{2}{b-a}(x_i-b)+1\right)dx_i$. Then, we apply Korous inequality for orthonormal polynomials 2.5.10. Additionally, using Markov's inequality for polynomials

$$\|D\varphi_{i0}^{p0}\|_\infty \leq \left(\frac{2}{b-a}\right)r^2\|\varphi_{i0}^{p0}\|_\infty,$$

and the result holds. □

Corollary 2.6.3 (Supremum norm of orthonormal polynomials in two variables)**.** Let $r \geq 2$, $i, j \in [N]$ with $i \neq j$ and $(p,q) \in \mathscr{V}_r$. Then

$$\|\varphi_{ij}^{pq}\|_\infty \leq \left(\frac{1}{\rho_0}+2\frac{a_1\text{Lip}(\rho)}{\rho_0^{3/2}}\right)^2\left(\frac{2}{b-a}\right)$$

and

$$\|D\varphi_{ij}^{pq}\|_\infty \leq 2\left(\frac{1}{\rho_0}+2\frac{a_1\text{Lip}(\rho)}{\rho_0^{3/2}}\right)^2\left(\frac{2}{b-a}\right)^{\frac{3}{2}}r^2.$$

Proof.

$$\|\varphi_{ij}^{pq}\|_\infty = \sup_{x_i,x_j\in(a,b)}|\varphi_{ij}^{pq}(x_i,x_j)| \leq \|\varphi_{i0}^{p0}\|_\infty\|\varphi_{j0}^{q0}\|_\infty,$$

and for the derivative, we calculate

$$\|D\varphi_{ij}^{pq}\|_\infty = \sup_{x_i,x_j\in(a,b)}|D\varphi_{ij}^{pq}(x_i,x_j)| \leq \|D\varphi_{i0}^{p0}\|_\infty\|\varphi_{j0}^{q0}\|_\infty + \|\varphi_{i0}^{p0}\|_\infty\|D\varphi_{j0}^{q0}\|_\infty.$$

The result holds applying Proposition 2.6.2.2. □

From here on, for short notation, we denote

$$K = K(\text{Lip}(\rho),\rho_0) \equiv \left(\frac{1}{\rho_0}+2\frac{a_1\text{Lip}(\rho)}{\rho_0^{3/2}}\right)^2\left(\frac{2}{b-a}\right). \tag{2.41}$$

2.7 Ergodic Basis Pursuit has a unique solution

In this section, we present our main result of the paper. We use the exponential mixing conditions of the network dynamics to estimate the minimum length of time series such that the ergodic basis pursuit has a unique solution. Here we avoid the multi-index notation in $\mathscr{L}$ and $\mathscr{L}_\nu$ used in the previous section and instead employ the notation that makes explicit the ordering index as $\phi_l : M^N \to \mathbb{R}$ with $l \in [m]$. More precisely, in an explicit form, we say that each ϕ_l corresponds to a function of the form $\phi_{ij}^{pq} \circ \pi_{\mathscr{J}}$ for a particular $\mathscr{J} \subset [N]$ with $i, j \in \mathscr{J}$. Also, we use the distance between probability measures introduced in 2.5.4.

Theorem 2.7.1. Let (F_α, μ_α) be an exponential mixing network dynamical system on M^N with decay exponent $\gamma > 0$ uniform on N. Let $\nu = \prod_{i\in[N]}\nu_i \in \mathscr{M}(M^N)$ be a product probability

measure and absolutely continuous w.r.t. Lebesgue. Let $\mathscr{L}_\nu$ be the orthonormal network library with respect to ν and cardinality $m = \binom{N}{2}\binom{r}{2} + Nr + 1$. Let $\mathscr{K} = (\mathscr{L}_\nu \cdot \mathscr{L}_\nu)$ and $\omega_r(s)$ satisfies (2.32). Suppose that given $\alpha > 0$ there is $\zeta \in (0, \frac{\sqrt{2}-1}{2\omega_r(s)+\sqrt{2}-2})$ such that $d_{\mathscr{K}}(\nu, \mu_\alpha) < \zeta$ and each one-dimensional marginal ν_i has Lipschitz density ρ_i with constant $\mathrm{Lip}(\rho) = \max_{i\in[N]} \mathrm{Lip}(\rho_i)$ and $\rho_0 = \min_{i\in[N]}\{\min_{x\in M_i} \rho_i(x)\} > 0$. Then:

1. *[$\Phi_\nu(X)$ satisfies RIP]* Given $\lambda \in (0,1)$ there exists set of initial conditions $\mathscr{G} \subset M^N$ with probability $\mu_\alpha(\mathscr{G}) \geq 1 - \lambda$ if the length of time series n satisfies
$$n \geq K_1 \frac{(2\omega_r(s) + \sqrt{2} - 2)^2}{\left(\sqrt{2} - 1 - \zeta(2\omega_r(s) + \sqrt{2} - 2)\right)^2} \ln\left(\frac{4m(m-1)}{\lambda}\right), \tag{2.42}$$
for some positive constant $K_1 = K_1(\mathrm{Lip}(\rho), \rho_0)$, then $\Phi_\nu(X)$ satisfies the RIP with constant $\delta_{2\omega_r(s)} \leq \sqrt{2} - 1$.

2. *[EBP has unique solution]* Consider that the length of time series n satisfies (2.42). Let $\bar{x} = \Phi_\nu(X)c_\nu$ where $c_\nu \in \mathbb{R}^m$ is an $\omega_r(s)$-sparse vector and consider the set $\mathscr{F}_{\bar{x}} = \{w \in \mathbb{R}^m \ : \ \Phi_\nu(X)w = \bar{x}\}$. Then c_ν is the unique minimizer of the ergodic basis pursuit:
$$\text{(EBP)} \quad \min_{u\in\mathscr{F}_{\bar{x}}} \|u\|_1. \tag{2.43}$$

The above theorem has an asymptotic expression for sufficient large networks and small ζ to a simpler condition on the length of time series:

Corollary 2.7.2. For sufficiently large $N > 0$, if the length of time series n satisfies
$$n \geq n_0 = \frac{20K_1\omega_r^2(s)}{(\sqrt{2}-1)^2} \ln(Nr) + \mathscr{O}(\zeta) + \mathscr{O}(\frac{1}{Nr}). \tag{2.44}$$
then with probability at least $1 - \frac{4}{Nr}$ the restricted isometry constant $\delta_{2\omega_r(s)} \leq \sqrt{2} - 1$.

Proof. Assume that (2.42) holds. Recall that $m = \binom{N}{2}\binom{r}{2} + Nr + 1$, then $m < (Nr+1)^2$. Given $\lambda \in (0,1)$ there exists $N_0 > 0$ such that for any $N \geq N_0$: $\frac{4}{Nr} \leq \lambda$. Then, the following holds
$$\begin{aligned}
\ln\left(\frac{4m(m-1)}{\lambda}\right) &< \ln\left(\frac{4(Nr+1)^2}{\lambda}\right) \\
&= \ln\left(\frac{4(Nr)^4(1+\frac{1}{Nr})^4}{\lambda}\right) \\
&\leq \ln(Nr)^5(1+\frac{1}{Nr})^4 \\
&= \ln(Nr)^5 + \mathscr{O}(\frac{1}{Nr}).
\end{aligned}$$

Also, for $\zeta \in (0, \frac{\sqrt{2}-1}{2\omega_r(s)+\sqrt{2}-2})$, we can expand in geometric series:
$$\begin{aligned}
\frac{1}{(\sqrt{2}-1-\zeta(2\omega_r(s)+\sqrt{2}-2))^2} &= \frac{1}{(\sqrt{2}-1)^2(1-\frac{\zeta(2\omega_r(s)+\sqrt{2}-2)}{\sqrt{2}-1})^2} \\
&= \frac{1}{(\sqrt{2}-1)^2}(1+\mathscr{O}(\zeta)).
\end{aligned}$$

So, we obtain the claim. □

We split proof of Theorem 2.7.1 in steps detailed in the sections below. First, we show that the Bernstein-like inequality applied to (F_α, μ_α) implies that there exists n_0 such that the library matrix $\Phi_\nu(X)$ associated to ν has the desired restricted isometry constant. Then, we apply Theorem 2.5.5 to demonstrate that the ergodic basis pursuit in (2.43) has a unique solution.

2.7.1 Network library matrix satisfies RIP

We begin this section by proving an auxiliary lemma that will be used later.

Lemma 2.7.3. Let $\nu = \prod_{i\in[N]} \nu_i \in \mathscr{M}(M^N)$ be a product probability measure. Suppose that each one-dimensional marginal ν_i is absolutely continuous w.r.t. Lebesgue and its density is Lipschitz with constant $\mathrm{Lip}(\rho)$ and $\rho_0 = \min_{i\in[N]}\{\min_{x\in M_i} \rho_i(x)\} > 0$. Let $\mathscr{L}_\nu$ be the orthonormal network library and $\mathscr{K} = (\mathscr{L}_\nu \cdot \mathscr{L}_\nu)$. Given $\alpha > 0$ and $\zeta > 0$ sufficiently small, suppose that $d_{\mathscr{K}}(\nu, \mu_\alpha) < \zeta$. Denote $(\psi_i \cdot \psi_j) = (\varphi_i \cdot \varphi_j) - \int_{M^N} (\varphi_i \cdot \varphi_j) d\mu_\alpha$. Then, the following holds:

1. $\max_{i,j} \|(\psi_i \cdot \psi_j)\|_\infty \leq 2\max\{1, K^2\}$.
2. $\max_{i,j} \|(\psi_i \cdot \psi_j)\|^2_{\mu_\alpha} \leq \max\{1, K^4\} + (1+\zeta)^2$,

where $K > 0$ is the positive constant in (2.41).

Proof. To prove item 1, note that:

$$\begin{aligned}\|(\psi_i \cdot \psi_j)\|_\infty &\leq \|(\varphi_i \cdot \varphi_j)\|_\infty + \int |(\varphi_i \cdot \varphi_j)| d\mu_\alpha \\ &\leq 2\|(\varphi_i \cdot \varphi_j)\|_\infty.\end{aligned}$$

To calculate the sup norm of the product of two orthonormal polynomials in $\mathscr{L}_\nu$, we consider the notation of the previous section in the following cases:

$$(\varphi_i \cdot \varphi_j) = \begin{cases} (1 \cdot 1), \\ (1 \cdot \varphi_{i0}^{p0}), \\ (1 \cdot \varphi_{ij}^{pq}), \\ (\varphi_{i0}^{p0} \cdot \varphi_{jk}^{ql}), \\ (\varphi_{ij}^{pq} \cdot \varphi_{km}^{ln}). \end{cases}$$

By Proposition 2.6.2.2 and Corollary 2.6.3,

$$\|(\varphi_i \cdot \varphi_j)\|_\infty = \begin{cases} \max\{K^{\frac{1}{2}}, K, K^{\frac{3}{2}}, K^2\}, & i \neq j \\ \max\{1, K, K^2\}, & i = j. \end{cases}$$

A priori, the constant K is a given positive number, so it is enough to consider $\|(\psi_i \cdot \psi_j)\|_\infty \leq \max\{1, K^2\}$, proving item 1.

To prove item 2, note that given

$$\left| \int_{M^N} (\varphi_i \cdot \varphi_j) d\mu_\alpha - \int_{M^N} (\varphi_i \cdot \varphi_j) d\nu \right| \leq d_{\mathscr{K}}(\nu, \mu_\alpha) \leq \zeta.$$

Consequently, by the triangular inequality

$$\left| \int_{M^N} (\varphi_i \cdot \varphi_j) d\mu_\alpha \right| \leq \begin{cases} 1+\zeta, & i = j \\ \zeta, & \text{otherwise.} \end{cases}$$

Then to prove the statement suffices to use item 1 above:

$$\begin{aligned} \|(\varphi_i \cdot \varphi_j)\|_{\mu_\alpha}^2 &= \left| \int (\varphi_i \cdot \varphi_j)^2 d\mu_\alpha - \left(\int \varphi_i^2 d\mu_\alpha \right)^2 \right| \\ &\leq \|(\varphi_i \cdot \varphi_j)\|_\infty^2 + (1+\zeta)^2 \\ &\leq \max\{1, K^4\} + (1+\zeta)^2. \end{aligned}$$

□

2.7.1.1 Proof of Theorem 2.7.1.1

The following proposition proves that the matrix Φ_ν attains the desired RIP constant once the length of time series is given by (2.45).

Proposition 2.7.3.1. Consider the setting of Theorem 2.7.1. Given $\delta \in (0,1)$ and $\alpha > 0$, suppose that there is $\zeta \in (0, \frac{\delta}{\delta + \omega_r(s) - 1})$ such that $d_{\mathscr{K}}(\nu, \mu_\alpha) < \zeta$. Then, given $\lambda \in (0,1)$ there exists a set of initial conditions $\mathscr{G} \subset M^N$ with probability $\mu_\alpha(\mathscr{G}) \geq 1 - \lambda$ such that

$$n \geq K_1 \frac{(\delta + \omega_r(s) - 1)^2}{(\delta - \zeta(\delta + \omega_r(s) - 1))^2} \ln\left(\frac{4m(m-1)}{\lambda}\right) \tag{2.45}$$

for a positive constant K_1, then the restricted isometry constant $\delta_{\omega_r(s)}$ of $\Phi_\nu(X)$ satisfies $\delta_{\omega_r(s)} \leq \delta$.

Proof. We develop the argument for a coordinate of F_α. Let

$$u_i := \frac{1}{\sqrt{n}} \begin{pmatrix} \varphi_i(x_0) \\ \vdots \\ \varphi_i(F_\alpha^{n-1}(x_0)) \end{pmatrix} \qquad u_j := \frac{1}{\sqrt{n}} \begin{pmatrix} \varphi_j(x_0) \\ \vdots \\ \varphi_j(F_\alpha^{n-1}(x_0)) \end{pmatrix}$$

be the i-th and j-th columns of the matrix $\Phi_\nu(X) \in \mathbb{R}^{n\times m}$ for an arbitrary initial condition $x_0 \in M^N$, and their inner product

$$\begin{aligned}\langle u_i, u_j\rangle &= \frac{1}{n}\sum_{k=0}^{n-1} \varphi_i(F_\alpha^k(x_0))\varphi_j(F_\alpha^k(x_0)) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} (\varphi_i\cdot\varphi_j)\circ(F_\alpha^k(x_0)) \\ &=: \frac{1}{n} S_n(\varphi_i\cdot\varphi_j)(x_0).\end{aligned}$$

We aim to estimate this inner product using the inner product in $L^2(\nu)$. By triangular inequality, we know that:

$$\begin{aligned}\left|\frac{1}{n}S_n(\varphi_i\cdot\varphi_j)(x_0) - \int_{M^N}(\varphi_i\cdot\varphi_j)d\nu\right| \le \left|\frac{1}{n}S_n(\varphi_i\cdot\varphi_j)(x_0) - \int_{M^N}(\varphi_i\cdot\varphi_j)d\mu_\alpha\right| + \\ \underbrace{\left|\int_{M^N}(\varphi_i\cdot\varphi_j)d\mu_\alpha - \int_{M^N}(\varphi_i\cdot\varphi_j)d\nu\right|}_{|h_{ij}|}.\end{aligned} \tag{2.46}$$

We introduce a variant of $(\varphi_i\cdot\varphi_j)$ to have zero mean with respect to μ_α, i.e., let us denote $(\psi_i\cdot\psi_j) = (\varphi_i\cdot\varphi_j) - \int_{M^N}(\varphi_i\cdot\varphi_j)d\mu_\alpha$ and by hypothesis,

$$\begin{aligned}|h_{ij}| &= \left|\int_{M^N}(\varphi_i\cdot\varphi_j)d\mu_\alpha - \int_{M^N}(\varphi_i\cdot\varphi_j)d\nu\right| \\ &\le d_{\mathscr{K}}(\nu,\mu_\alpha) \\ &\le \zeta.\end{aligned} \tag{2.47}$$

Then, we split into two distinct cases that run in parallel:

1. $i \neq j$: $\int_{M^N}(\varphi_i\cdot\varphi_j)d\nu = 0$, consequently, using (2.47) we conclude that follows
$$\left|\frac{1}{n}S_n(\varphi_i\cdot\varphi_j)(x_0)\right| \le \left|\frac{1}{n}S_n(\psi_i\cdot\psi_j)(x_0)\right| + |h_{ij}| \le \left|\frac{1}{n}S_n(\psi_i\cdot\psi_j)(x_0)\right| + \zeta. \tag{2.48}$$
2. $i = j$: we have $\int_{M^N}\varphi_i^2 d\nu = 1$, and consequently, in (2.46), we obtain
$$\left|\frac{1}{n}S_n(\varphi_i^2)(x_0) - 1\right| \le \left|\frac{1}{n}S_n(\psi_i^2)(x_0)\right| + \zeta.$$
By the triangular inequality, we conclude that
$$\left|\frac{1}{n}S_n(\varphi_i^2)(x_0)\right| \ge 1 - \left|\frac{1}{n}S_n(\psi_i^2)(x_0)\right| - \zeta. \tag{2.49}$$

Note that $(\psi_i\cdot\psi_j)$ and ψ_i^2 are given by a finite linear combination of elements in $\mathscr{L}$, and consequently, the subsets with cardinality $\binom{m}{2}$ and m, respectively, satisfy

$$\mathscr{K}_1 = \{(\psi_i\cdot\psi_j) \,:\, i,j = 1,\ldots,m, i\neq j\} \subset \mathscr{C}^1(M^N;\mathbb{R}) \tag{2.50}$$

and

$$\mathscr{K}_2 = \{\psi_i^2 \ : \ i = 1, \dots, m\} \subset \mathscr{C}^1(M^N; \mathbb{R}). \tag{2.51}$$

Choose $\varkappa > 0$, $\varsigma > 0$ and $\sigma > 0$ such that

$$\begin{aligned} \varkappa &:= \max\{\max_{i \neq j} \|(\psi_i \cdot \psi_j)\|_\infty, \max_{i \in [m]} \|\psi_i^2\|_\infty\}, \\ \varsigma &:= \max\{\max_{i \neq j} \|D(\psi_i \cdot \psi_j)\|_\infty, \max_{i \in [m]} \|D\psi_i^2\|_\infty\}, \\ \sigma^2 &:= \max\{\max_{i \neq j} \|(\psi_i \cdot \psi_j)\|_{\mu_\alpha}^2, \max_{i \in [m]} \|\psi_i^2\|_{\mu_\alpha}^2\}. \end{aligned} \tag{2.52}$$

By the Bernstein inequality in Theorem 2.5.8, for $\eta_0 > 0$ and $n \geq n_0(\varkappa, \varsigma, \sigma, \gamma)$, which is defined in (2.24), if we define

$$\begin{aligned} \mathscr{O}_1 &= \bigcup_{i \neq j} \left\{ x_0 \in M^N \ : \ \left| \frac{1}{n} S_n(\psi_i \cdot \psi_j)(x_0) \right| \geq \eta_0 \right\} \\ \mathscr{O}_2 &= \bigcup_{i \in [m]} \left\{ x_0 \in M^N \ : \ \left| \frac{1}{n} S_n(\psi_i^2)(x_0) \right| \geq \eta_0 \right\}, \end{aligned}$$

then

$$\mu_\alpha(\mathscr{O}_1) \leq 4 \binom{m}{2} e^{-\theta(\eta_0, n, \sigma, \varkappa)} \quad \text{and} \quad \mu_\alpha(\mathscr{O}_2) \leq 4me^{-\theta(\eta_0, n, \sigma, \varkappa)}.$$

We are interested in the case $\mu_\alpha(\mathscr{O}) = \mu_\alpha(\mathscr{O}_1 \cup \mathscr{O}_2)$

$$\begin{aligned} \mu_\alpha(\mathscr{O}) &\leq 4 \binom{m}{2} e^{-\theta(\eta_0, n, \sigma, \varkappa)} + 4me^{-\theta(\eta_0, n, \sigma, \varkappa)} \\ &\leq 8 \binom{m}{2} e^{-\theta(\eta_0, n, \sigma, \varkappa)}. \end{aligned}$$

For the given $\lambda \in (0,1)$ the set $\mathscr{O}^c \subset M^N$ of initial conditions, whose Birkhoff sum satisfies the desired precision η_1, has measure $\mu_\alpha(\mathscr{O}^c) \geq 1 - \lambda$ whenever

$$\frac{n}{(\ln n)^2} \geq \frac{8}{\eta_0^2} (\sigma^2 + \varkappa \frac{\eta_0}{3}) \ln \left(\frac{8}{\lambda} \binom{m}{2} \right). \tag{2.53}$$

Instead of (2.53), one usually prefers a condition that features only n on the left-hand side. First, note that whenever $n \geq n_0$ implies that $n \in \mathscr{N}$ in (2.24), and consequently, the function $t \mapsto t/(\ln t)^2$ is monotonic for values in $\mathscr{N}$. So, the condition in (2.53) is in fact implied by

$$n \geq \frac{8}{\eta_0^2} (\sigma^2 + \varkappa \frac{\eta_0}{3}) \ln \left(\frac{4m(m-1)}{\lambda} \right). \tag{2.54}$$

For any n satisfying the bound in (2.54), we can normalize any two distinct columns vectors u_i and u_j, which we denote v_i and v_j, respectively. So, we can estimate the coherence of the matrix

$\Phi_\nu(X)$ for any $x_0 \in \mathscr{O}^c$ using (2.48) and (2.49)

$$\begin{aligned}\eta(\Phi_\nu) &:= \max_{i\neq j} |\langle v_i, v_j\rangle| = \max_{i\neq j} \frac{|\langle u_i, u_j\rangle|}{\|u_i\|_2\|u_j\|_2} \\ &= \max_{i\neq j} \frac{\left|\frac{1}{n}S_n(\varphi_i\cdot\varphi_j)(x_0)\right|}{\left|\frac{1}{n}S_n(\varphi_i^2)(x_0)\right|^{\frac{1}{2}}\left|\frac{1}{n}S_n(\varphi_j^2)(x_0)\right|^{\frac{1}{2}}} \\ &\leq \frac{\eta_0+\zeta}{1-(\eta_0+\zeta)},\end{aligned}$$

which is valid for a η_0 such that $\eta_0+\zeta<1$. Finally, the desired restricted isometry constant is attained because the coherence upper bounds the restricted isometry constant of the matrix $\Phi_\nu(X)$ by Proposition 2.5.4.2. So, for the given $\delta\in(0,1)$, we choose

$$\eta_0 = \frac{\delta-\zeta(\delta+\omega_r(s)-1)}{\delta+\omega_r(s)-1} \tag{2.55}$$

that is positive as long as $\zeta\in(0,\frac{\delta}{\delta+\omega_r(s)-1})$. To obtain the desired bounds on the length of the time series, note that $\eta_0\in(0,1)$. Also, we use Lemma 2.7.3 in order to bound σ^2 and $\varkappa$ in (2.52). Consequently, the condition in (2.53) is also implied by

$$n \geq \frac{8}{\eta_0^2}(\max\{1,K^4\}+(1+\zeta)^2+\frac{2}{3}\max\{1,K^2\})\ln\left(\frac{4m(m-1)}{\lambda}\right),$$

which can also be implied by

$$n \geq \frac{K_1}{\eta_0^2}\ln\left(\frac{4m(m-1)}{\lambda}\right), \tag{2.56}$$

with $K_1 := 8(\max\{1,K^4\}+4+\frac{2}{3}\max\{1,K^2\})$. Replacing (2.55) in (2.56), we obtain the result. □

Lemma 2.7.4. Let $c_\nu\in\mathbb{R}^m$ be a $\omega_r(s)-$sparse vector. If the length of time series n satisfies (2.42), then EBP in (2.43) has c_ν as its unique solution.

Proof. Combining Proposition 2.7.3.1 with Theorem 2.5.5 suffices. □

2.7.2 Ergodic basis pursuit has a sufficient infeasibility condition

Since Theorem 2.7.1 ensures that EBP has a unique solution, we can also prove an additional result.

Proposition 2.7.4.1 (Sufficient infeasibility condition)**.** Consider that the length of time series n satisfies (2.42). Let $\bar{x}=\Phi_\nu(X)c_\nu$ where $c_\nu\in\mathbb{R}^m$ is an $\omega_r(s)$-sparse vector and consider the set $\mathscr{F}_{\bar{x}}=\{w\in\mathbb{R}^m:\Phi_\nu(X)w=\bar{x}\}$. Given a set $\mathscr{U}\subseteq[m]$ where $\mathscr{U}\cap\operatorname{supp}(c_\nu)\neq\emptyset$. Then $\mathscr{U}\subsetneq\operatorname{supp}(c_\nu)$ if and only if

$$\mathscr{F}_{\bar{x}}\bigcap\{w\in\mathbb{R}^m\ :\ \operatorname{supp}(w)=\mathscr{U}\}=\emptyset. \tag{2.57}$$

Proof. Let us assume that $\mathscr{I} \subsetneqq \text{supp}(c_\nu)$. We will prove this by contradiction. Suppose there is a vector $w \neq 0$ in the intersection (2.57) and is given by

$$w = (w_1, \ldots, w_{\omega_r(s)-1}, 0, \ldots, 0)$$

as opposed to the $\omega_r(s)$-sparse vector c,

$$c_\nu = (c_1, \ldots, c_{\omega_r(s)}, 0, \ldots, 0),$$

so, the vector w has $\omega_r(s) - 1$ nonzero entries. Since $w \in \mathscr{F}_{\bar{x}}$, we have

$$\Phi_\nu(X)w = \Phi_\nu(X)c_\nu.$$

Consequently,

$$\Phi_\nu(X)(w - c_\nu) = 0.$$

But the $\Phi_\nu(X)$ satisfies RIP with constant $\delta_{2\omega_r(s)} < \sqrt{2} - 1$. Since $w - c_\nu$ is an $\omega_r(s)$-sparse vector, we calculate

$$\|\Phi_\nu(X)(w - c_\nu)\|_2^2 \geq (1 - \delta_{2\omega_r(s)})\|w - c_\nu\|_2^2 > 0.$$

So, we conclude that this is only possible when $w = c$, which is a contradiction since c is not in the intersection, and the claim follows.

The other direction we prove by contrapositive. We contradict $\mathscr{I} \subsetneqq \text{supp}(c_\nu)$. Since $\mathscr{I}$ must have an intersection with $\text{supp}(c_\nu)$, then suppose that $\text{supp}(c_\nu) \subseteq \mathscr{I}$. The intersection (2.57) is non-empty, since the sparse vector c_ν is an element of the set. This proves the claim, and the statement follows. □

2.8 Noise measurement case

Here, we extend the Ergodic Basis Pursuit to reconstruct the network from corrupted measurements

$$y(t) = x(t) + z(t), \tag{2.58}$$

such that $(z_n)_{n\geq 0}$ corresponds to independent and identically distributed $[-\xi, \xi]^N$-valued noise process for $\xi \in (0, 1)$ with probability measure ρ_ξ. Let the convolution $\mu_{\alpha,\xi} = \mu_\alpha * \rho_\xi$ be the probability measure of the process $(y_n)_{n\geq 0}$ (FOLLAND, 2013) and the matrix $\bar{Y}$ be the noisy data

$$\bar{Y} = \begin{pmatrix} y_1(1) & \cdots & y_N(1) \\ \vdots & \ddots & \vdots \\ y_1(n) & \cdots & y_N(n) \end{pmatrix}. \tag{2.59}$$

The next theorem assumes that there exists a product measure ν_ξ sufficiently close to the measure $\mu_{\alpha,\xi}$. Thus, the s-sparse vector $c \in \mathbb{R}^m$ corresponding to the representation in $\mathscr{L}$ is mapped to c_{ν_ξ} which represents the network dynamics in $\mathscr{L}_{\nu_\xi}$, i.e., it satisfies $\bar{x} = \Phi_{\nu_\xi}(X)c_{\nu_\xi}$. Here, we also introduce another convex minimization problem in terms of $\Phi_{\nu_\xi}(Y)$ evaluated along the process $(y_n)_{n\geq 0}$. We show that the family of solutions of this minimization problem is parametrized by the noise level in such way that approximates the sparse vector c_{ν_ξ}.

Here, we rewrite the interval bounds: for a given $i \in [N]$ let $M_{i,\xi} = [a-\xi, b+\xi] \subset \mathbb{R}$ with $b > a$ and $\xi \in (0,1)$. Then, consider the following:

Hypothesis 2.8.1. Let $\nu_\xi = \prod_{i\in[N]} \nu_{i,\xi} \in \mathscr{M}(M^N + [-\xi,\xi]^N)$ be a product probability measure and absolutely continuous w.r.t. Lebesgue. Suppose that given a sufficiently small $\xi > 0$, each one-dimensional marginal $\nu_{i,\xi}$ has Lipschitz density $\rho_{i,\xi}$ with constant $\mathrm{Lip}(\rho_\xi) = \max_{i\in[N]} \mathrm{Lip}(\rho_{i,\xi})$ and $\rho_{0,\xi} = \min_{i\in[N]}\{\min_{x\in M_{i,\xi}} \rho_{i,\xi}(x)\} > 0$.

Theorem 2.8.2 (Noise reconstruction case)**.** Consider the setting of Theorem 2.7.1 and Hypothesis 2.8.1. Let $\mathscr{L}_{\nu_\xi} = \{\varphi_l\}_{l=1}^m$ be the orthonormal ordered network library with respect to ν_ξ and $\mathscr{K} = (\mathscr{L}_{\nu_\xi} \cdot \mathscr{L}_{\nu_\xi})$. Suppose that given $\alpha > 0$ there is $\zeta \in (0, \frac{\sqrt{2}-1}{2\omega_r(s)+\sqrt{2}-2} - K_1 r^2 \xi)$ such that $d_{\mathscr{K}}(\nu_\xi, \mu_{\alpha,\xi}) < \zeta$ for a positive constant $K_1 = K_1(\mathrm{Lip}(\rho_\xi), \rho_{0,\xi}, \xi)$. Then

1. *[$\Phi_\nu(Y)$ satisfies RIP]* Given $\lambda \in (0,1)$ there exists a set of initial conditions $\mathscr{G} \subset M^N$ with probability $\mu_\alpha(\mathscr{G}) \geq 1-\lambda$ such that if the length of time series n satisfies

$$n \geq K_2 \frac{(2\omega_r(s)+\sqrt{2}-2)^2}{(\sqrt{2}-1-(\zeta+K_1 r^2\xi)(2\omega_r(s)+\sqrt{2}-2))^2} \ln\Big(\frac{4m(m-1)}{\lambda}\Big), \tag{2.60}$$

   for a positive constant $K_2 = K_2(\mathrm{Lip}(\rho_\xi), \rho_{0,\xi}, \xi)$, then the restricted isometry constant $\delta_{2\omega_r(s)}$ of $\Phi_{\nu_\xi}(X)$ satisfies $\delta_{2\omega_r(s)} \leq \sqrt{2}-1$.

2. *[EBP is robust]* Let $\bar{y} \in M^n + [-\xi,\xi]^n$ be a column of $\bar{Y}$, $c_{\nu_\xi} \in \mathbb{R}^m$ be an $\omega_r(s)$-sparse vector with $\|c_{\nu_\xi}\|_\infty < \infty$ such that $\bar{x} = \Phi_{\nu_\xi}(X)c_{\nu_\xi}$. Consider that the length of time series n satisfies (2.60). Then the family of solutions $\{c^\star(\varepsilon)\}_{\varepsilon>0}$ to the convex problem

$$\min_{\tilde{u}\in\mathbb{R}^m} \|\tilde{u}\|_1 \text{ subject to } \|\Phi_{\nu_\xi}(Y)\tilde{u} - \bar{y}\|_2 \leq \varepsilon \tag{2.61}$$

   satisfies

$$\|c^\star(\varepsilon) - c_{\nu_\xi}\|_2 \leq K_3\varepsilon \tag{2.62}$$

   as long as

$$\varepsilon \geq \sqrt{n}\xi\Big(1 + mNr^2K_4\|c_{\nu_\xi}\|_\infty\Big), \tag{2.63}$$

   for positive constants $K_3 = K_3(\delta_{2\omega_s(r)})$ and $K_4 = K_4(\mathrm{Lip}(\rho_\xi), \rho_{0,\xi})$.

We prove the above theorem in steps detailed in the sections below. First, we adapt the estimate of the minimum length of time series such that the library matrix has the desired restricted isometry constant. Subsequently, we show that the unique solution of the ergodic basis pursuit in (2.43) is approximated in ℓ_2 by a family of solutions $\{c^\star(\varepsilon)\}_{\varepsilon\geq 0}$.

2.8.1 Perturbed network library matrix satisfies RIP

We begin estimating the distance between the product measure ν_ξ and the physical measure μ_α of the deterministic network dynamics. To this end, we use the auxiliary lemmas below. First, we show that

Lemma 2.8.3. $\mu_{\alpha,\xi} \to \mu_\alpha$ converges weakly as $\xi \to 0$.

Proof. Fix a continuous function $\varphi : M^N \to \mathbb{R}$ and $\xi > 0$. Using the definition $\int \varphi d\mu_{\alpha,\xi} = \int \varphi d\mu_\alpha * \rho_\xi = \int\int \varphi(x+z)d\mu_\alpha(x)d\rho_\xi(z)$ and $\int \varphi d\mu_\alpha = \int\int \varphi(x)d\mu_\alpha(x)d\rho_\xi(z)$, we obtain

$$\begin{aligned}\Big|\int \varphi d\mu_{\alpha,\xi} - \int \varphi d\mu_\alpha\Big| &= \Big|\int\int \varphi(x+z)d\mu_\alpha(x)d\rho_\xi(z) - \int\int \varphi(x)d\mu_\alpha(x)d\rho_\xi(z)\Big| \\ &\leq \int \Big(\int |\varphi(x+z)-\varphi(x)|d\rho_\xi(z)\Big)d\mu_\alpha(x) \\ &\leq \int \sup_{|z|\leq\xi} |\varphi(x+z)-\varphi(x)|d\mu_\alpha(x).\end{aligned}$$

Since M^N is a compact set, φ is uniformly continuous. Then, letting $\xi \to 0$ implies that the right-hand side converges to zero, and consequently, the integrals in the left-hand side converge. This is valid for any continuous function φ, concluding the statement. □

We address to estimate the distance $d_{\mathscr{K}}(\nu,\mu_\alpha)$. Since the product measure ν_ξ is defined on $(M^N+[-\xi,\xi]^N)$. Also, we define a variant of the constant (2.41) given by

$$K_\xi = K(\mathrm{Lip}(\rho_\xi),\rho_{0,\xi},\xi) \equiv \Big(\frac{1}{\rho_{0,\xi}} + 2\frac{a_1\mathrm{Lip}(\rho)}{\rho_{0,\xi}^{3/2}}\Big)^2\Big(\frac{2}{b-a+2\xi}\Big) \tag{2.64}$$

that satisfies $K_\xi \to K$ when $\xi \to 0$. Then, the following holds

Lemma 2.8.4. Let $r \geq 2$. Given $\alpha,\zeta,\xi \in (0,1)$ suppose that $d_{\mathscr{K}}(\nu,\mu_{\alpha,\xi}) < \zeta$ for the product measure $\nu_\xi \in \mathscr{M}(M^N+[-\xi,\xi]^N)$. Then, there exists $K_1 = K_1(\mathrm{Lip}(\rho_\xi),\rho_{0,\xi},\xi)$ such that

$$d_{\mathscr{K}}(\nu,\mu_\alpha) \leq \zeta + K_1 r^2 \xi.$$

Proof. First we calculate $d_{\mathscr{K}}(\mu_{\alpha,\xi},\mu_\alpha)$. Fix $\mathscr{J} \subset [N]$ and $\psi \in \mathscr{K}$. Since the projection $\pi_{\mathscr{J}} : M^N \to \prod_{i\in\mathscr{J}} M_i$ is Lipschitz with constant 1 and $\mathscr{K}$ is a set of product of polynomials, the composition $\psi\circ\pi_{\mathscr{J}}$ is also Lipschitz with constant $\mathrm{Lip}(\psi\circ\pi_{\mathscr{J}}) = \|D\psi\|_\infty$. Then, we obtain

$$\begin{aligned}\Big|\int_{M^N} \psi\circ\pi_{\mathscr{J}} d\mu_{\alpha,\xi} - \int_{M^N} \psi\circ\pi_{\mathscr{J}} d\mu_\alpha\Big| &\leq \int \sup_{|z|\leq\xi} \big|\psi\circ\pi_{\mathscr{J}}(x+z) - \psi\circ\pi_{\mathscr{J}}(x)\big| d\mu_\alpha(x) \\ &\leq \mathrm{Lip}(\psi\circ\pi_{\mathscr{J}})\xi.\end{aligned}$$

For each $\psi \in \mathscr{K}$ it corresponds to a pair $(\varphi_i \cdot \varphi_j)$, so we use Proposition 2.6.2.2 and Corollary 2.6.3 to calculate

$$\max_{\substack{\mathscr{J}\subset[N]\\ 1\leq|\mathscr{J}|\leq 4}} \max_{(\varphi_i\cdot\varphi_j)\in\mathscr{K}} \mathrm{Lip}((\varphi_i\cdot\varphi_j)\circ\pi_{\mathscr{J}}) \leq \max_{\substack{\mathscr{J}\subset[N]\\ 1\leq|\mathscr{J}|\leq 4}} \max_{(\varphi_i\cdot\varphi_j)\in\mathscr{K}} \|D\varphi_i\|_\infty\|\varphi_j\|_\infty + \|\varphi_i\|_\infty\|D\varphi_j\|_\infty$$

that is upper bounded by

$$\underbrace{K_\xi \max\left\{\frac{4}{b-a+2\xi}, 4K_\xi\left(\frac{2}{b-a+2\xi}\right)^{\frac{1}{2}}, 2K_\xi^{\frac{1}{2}}\left(\frac{2}{b-a+2\xi}\right)^{\frac{1}{2}} + \left(\frac{2}{b-a+2\xi}\right)K_\xi^{\frac{1}{2}}\right\}}_{K_1(\mathrm{Lip}(\rho),\rho_{0,\xi},\xi)} r^2.$$

This yields $d_{\mathscr{K}}(\mu_{\alpha,\xi},\mu_\alpha) \leq K_1 r^2 \xi$. Using the triangular inequality

$$d_{\mathscr{K}}(\nu,\mu_\alpha) \leq d_{\mathscr{K}}(\nu,\mu_{\alpha,\xi}) + d_{\mathscr{K}}(\mu_{\alpha,\xi},\mu_\alpha),$$

we conclude the lemma. □

Before we proceed, we extend $\mu_\alpha \in \mathscr{M}(M^N)$ to $\mathscr{M}(M^N + [-\xi,\xi]^N)$, defining the measure of a set $E \subseteq M^N + [-\xi,\xi]^N$ as $\mu_\alpha(E \cap M^N)$. We abuse notation and denote the measure as μ_α.

We can state a similar version of Proposition 2.7.3.1, making the appropriate changes. See below:

Proposition 2.8.4.1. Consider the setting of Theorem 2.8.2. Given $\delta,\lambda \in (0,1)$ there exists a set of initial conditions $\mathscr{G} \subset M^N$ with probability $\mu_\alpha(\mathscr{G}) \geq 1-\lambda$ such that

$$n \geq K_2 \frac{(\omega_r(s)+\delta-1)^2}{(\delta-(\zeta+K_1 r^2\xi)(\omega_r(s)+\delta-1))^2} \ln\left(\frac{4m(m-1)}{\lambda}\right) \tag{2.65}$$

for positive constants K_1 and K_2, then the restricted isometry constant $\delta_{\omega_r(s)}$ of $\Phi_\nu(X)$ satisfies $\delta_{\omega_r(s)} \leq \delta$.

Proof. The proof is similar to the proof of Proposition 2.7.3.1. Using Lemma 2.8.4 for the measures ν_ξ,μ_α, there is a constant $K_1(\mathrm{Lip}(\rho_\xi),\rho_{0,\xi},\xi)$ such that $d_{\mathscr{K}}(\nu_\xi,\mu_\alpha) \leq \zeta + K_1 r^2\xi =: \zeta'$, which we define so we can repeat the proof of Proposition 2.7.3.1 replacing ζ by ζ'. The new bounds of n_0 can be deduced as follows: we estimate a new condition that is implied by

$$n \geq \frac{8}{\eta_0^2}(\max\{1,K_\xi^4\} + (1+\zeta+K_1 r^2\xi)^2 + \frac{2}{3}\max\{1,K_\xi^2\}) \ln\left(\frac{4m(m-1)}{\lambda}\right).$$

This expression can also be implied by

$$n \geq \frac{K_2}{\eta_0^2} \ln\left(\frac{4m(m-1)}{\lambda}\right), \tag{2.66}$$

with $K_2 := 8(\max\{1,K_\xi^4\} + (2+K_1 r^2) + \frac{2}{3}\max\{1,K_\xi^2\})$. Using (2.55) replacing ζ by $\zeta + K_1 r^2\xi$ in the above expression, we obtain the result. □

Proof of Theorem 2.8.2.1. It suffices to use Proposition 2.7.3.1 for $\delta = \sqrt{2}-1$ and sparsity level $2\omega_r(s)$ in the expression of the length of time series in (2.60). □

2.8.2 Ergodic basis pursuit is robust against noise

We can write that $\bar{X} = \Phi_{\mathcal{V}_\xi}(X)C_{\mathcal{V}_\xi}$, where $C_{\mathcal{V}_\xi} \in \mathbb{R}^{m\times N}$ is the coefficient matrix associated to the network library $\mathscr{L}_{\mathcal{V}_\xi}$. We deduce that the noisy data in (2.59) satisfies

$$\bar{Y} = \Phi_{\mathcal{V}_\xi}(X)C_{\mathcal{V}_\xi} + \bar{Z}, \tag{2.67}$$

where

$$\bar{Z} = \begin{pmatrix} z_1(1) & \cdots & z_N(1) \\ \vdots & \ddots & \vdots \\ z_1(n) & \cdots & z_N(n) \end{pmatrix} \in [-\xi,\xi]^{n\times N}, \tag{2.68}$$

such that each column $\bar{z}$ of $\bar{Z}$ is bounded as $\|\bar{z}\|_2 \leq \sqrt{n}\xi$. The following lemma states that the library matrix can be evaluated at the noisy data:

Lemma 2.8.5.

$$\Phi_{\mathcal{V}_\xi}(Y) = \Phi_{\mathcal{V}_\xi}(X) + \Lambda(X,\bar{Z}), \tag{2.69}$$

where $\|\Lambda(X,\bar{Z})\|_\infty \leq mNr^2K_4\xi$ with $K_4 := \max\{K_\xi^{\frac{1}{2}}, 2K_\xi\left(\frac{2}{b-a+2\xi}\right)\}$.

Proof. For $l \in [m]$ let $\varphi_l \in \mathscr{L}_{\mathcal{V}_\xi}$. The Mean Value Theorem states that for each $t = 0,\ldots,n-1$:

$$\varphi_l(x(t)+z(t)) = \varphi_l(x(t)) + \left(\int_0^1 D\varphi_l(x(t)+sz(t))ds\right)\cdot z(t),$$

where the integral is understood component-wise. Repeating the calculation for each entry of $\Phi_{\mathcal{V}_\xi}(Y)$, by linearity we obtain $\Phi_{\mathcal{V}_\xi}(Y) = \Phi_{\mathcal{V}_\xi}(X) + \Lambda(X,\bar{Z})$, where $\Lambda(X,\bar{Z})$ is the matrix with entries

$$\Lambda_{j,k}(X,Z) = \left(\int_0^1 D\varphi_k(x(j)+sz(j))ds\right)\cdot z(j).$$

We use Proposition 2.6.2.2 and Corollary 2.6.3 for $M^N + [-\xi,\xi]^N$. Let us denote

$$\max_{l\in[m]}\|D\varphi_l\|_\infty \leq r^2\max\{K_\xi^{\frac{1}{2}}, 2K_\xi\left(\frac{2}{b-a+2\xi}\right)\} \equiv r^2K_4.$$

Using Cauchy-Schwarz inequality, note that each entry $\Lambda_{j,k}$ satisfies

$$\begin{aligned} |\Lambda_{j,k}| &= |\left(\int_0^1 D\varphi_k(x(j)+sz(j))ds\right)\cdot z(j)| \\ &\leq \|\int_0^1 D\varphi_k(x(j)+sz(j))ds\|_2\|z(j)\|_2 \\ &\leq Nr^2K_4\xi. \end{aligned}$$

So, this implies that $\|\Lambda(X,\bar{Z})\|_\infty \leq mNr^2K_4\xi$ and proves the lemma. □

Proof of Theorem 2.8.2.2. Using (2.67) and (2.69) we have

$$\begin{aligned}\bar{Y} &= \big(\Phi_{\nu_\xi}(Y) - \Lambda(X,\bar{Z})\big)C_{\nu_\xi} + \bar{Z} \\ &= \Phi_{\nu_\xi}(Y)C_{\nu_\xi} + \bar{Z} - \Lambda(X,\bar{Z})C_{\nu_\xi}.\end{aligned}$$

The above equation for each column $i \in [N]$ is given by an equation of the form $\bar{y} = \Phi_{\nu_\xi}(Y)c_{\nu_\xi} + \bar{u}_i$, where the perturbation is

$$\bar{u} = \bar{z} - \Lambda(X,\bar{Z})c_{\nu_\xi}.$$

Using Lemma 2.8.5 the perturbation vector $\bar{u}$ is bounded as

$$\begin{aligned}\|\bar{u}\|_2 &\leq \sqrt{n}\xi + \sqrt{n}mNr^2K_4\xi\|c_{\nu_\xi}\|_\infty \\ &= \sqrt{n}\xi\left(1 + mNr^2K_4\|c_{\nu_\xi}\|_\infty\right).\end{aligned}$$

We apply Theorem 2.5.6, and this concludes the proof. □

2.9 Closing remarks

In summary, we proposed a method to reconstruct sparse networks from noisy and limited data. Our approach blends ergodic theory of dynamical systems and compressive sensing to demonstrate that once a minimum length of time series is achieved, the EBP, particularly its extension QEBP, is a robust method to identify network structures from noisy data. The main advantage of this method is that it enables to use of a smaller amount of time series (quadratically in the degree and log of the system size) as opposed to a linear dependence on the system size of the classical Basis Pursuit method.

We introduced the relaxing path algorithm that reconstructs the network as a weighted graph parametrized by the bound of the noise. Without prior knowledge of the statistical properties of the noise corrupting the data, this algorithm can reveal the network structure in an optimal interval of the tuned parameter. Because a noisy and limited amount of length of time series arises typically in experimental settings, our findings apply to a wide range of chaotic systems.

Data and code availability. All data necessary for the reproduction of the results, all simulations and analysis scripts are available in the Ergodic Basis Pursuit repository (SANTOS, b).

CHAPTER

3

STABILITY UNDER BASIS EXTENSION

This chapter is devoted to presenting our work entitled *Recovering sparse networks: Basis adaptation and stability under extensions* (NOVAES; Roque dos Santos; PEREIRA, 2021).

Usually, the reconstruction problem may use prior expert knowledge of possible network structures. From these guesses, one may extend the reconstruction by identifying further interactions. In this chapter, we study the reconstruction of sparse networks. We start from a network seed that gives an approximation of the network to be recovered and extend the search for further connections.

We show that by adapting the reconstruction to the dynamics, the basis extension does not lead to prediction instability. We discuss the least square techniques are unstable under basis extension. A heuristic upshot of our study is that if the network is sparse and has $k \ll N$ links, where N is the number of nodes in the network, then using sparse recovery we need only $O(Nk)$ data points as opposed to least-square where we need $O(N^2)$.

We will focus on the case when isolated dynamics of the nodes have a stable periodic motion and the interaction is weak. This is an interesting case, as the phase itself is not observed and thus we need to preprocess the data.

3.1 Dynamics near a Hopf Bifurcation

We consider the isolated dynamics of each node in the network to be near a Hopf-Andronov bifurcation, modelled by the Stuart-Landau equation

$$\dot{z}_j = F(z_j) = (1 + i\omega_j)z_j - |z_j|^2 z_j, \tag{3.1}$$

where $i^2 = -1$ and z_j is a complex number. Each isolated oscillator has an exponentially attractive periodic orbit with amplitude 1 and frequency ω_j for $j = 1, \ldots, N$. The effect of a linear pairwise

interactions is modelled as

$$\dot{z}_j = F(z_j) + \alpha \sum_{k=1}^{N} A_{jk}(z_k - z_j) \tag{3.2}$$

for $j = 1,\ldots,N$. Here, α denotes the coupling strength, assumed small. The connectivity matrix A describes the interaction structure: A_{jk} is 1 if node j is influenced by node k and is 0 otherwise. Notice that in the absence of linear terms, if nonlinear terms are included in the coupling this could lead to higher order resonances (NIJHOLT *et al.*, 2022). However, we will consider only linear coupling which is enough to show how the recovery method works.

3.1.1 Phase Dynamics

By introducing polar coordinates $z_j = r_j e^{i\theta_j}$ we can obtain the dynamics of amplitudes r_j and phases θ_j. As α is small, the network effect on the amplitudes is small, in fact, $r_j(t) = 1 + O(\alpha)$. The relevant dynamics generated by the network is encoded in the phases. The coupled phase equations to leading order in α read as

$$\dot{\theta}_j = \omega_j + \alpha \sum_{k=1}^{n} A_{jk} \sin(\theta_k - \theta_j). \tag{3.3}$$

Extracting phase from data. In applications, we do not have direct access to $\theta_j(t)$ and may need to infer another phase variable from a time series. Let x_j and y_j denote, respectively, the real and imaginary parts of z_j and we assume that we only measure $x_j(t)$ for each oscillator. Thus, we have a multivariate time series for the network. To extract the phase from each time series we use the standard Hilbert transform

$$H(x_j(t)) = \frac{1}{\pi}\text{p.v.} \int_{-\infty}^{+\infty} \frac{x_j(\tau)}{t-\tau} d\tau. \tag{3.4}$$

Thus using the analytic signal

$$s_j(t) = x_j(t) + H(x_j(t)) = R_j(t) e^{i\vartheta_j(t)} \tag{3.5}$$

we can extract a phase $\vartheta_j(t)$ corresponding to the signal $x_j(t)$. Although this phase is a surrogate and not necessarily equal to $\theta_j(t)$, meaningful dynamical information can be obtained from it. Once we have the phases ϑ_j, their time derivatives are obtained numerically and a smoothing filter is applied to remove noise introduced in this process.

3.2 The recovery method

3.2.1 The basis functions

The idea is to express the time derivatives of the phases, obtained from data, as linear combinations of certain functions. Here as we deal with phases we use Fourier modes depending

on all variables and on the differences of all variables,

$$\dot{\vartheta}_j = \omega_j + \sum_\ell g^{(j)}(\vartheta_\ell) + \sum_{k,m} h^{(j)}(\vartheta_k, \vartheta_m), \tag{3.6}$$

where

$$g^{(j)}(\vartheta_\ell) = a_\ell^{(j)} \cos(\vartheta_\ell) + b_\ell^{(j)} \sin(\vartheta_\ell) \tag{3.7}$$

is the isolated component and the coupling function is

$$h^{(j)}(\vartheta_k, \vartheta_m) = c_{k,m}^{(j)} \cos(\vartheta_k - \vartheta_m) + d_{k,m}^{(j)} \sin(\vartheta_k - \vartheta_m), \quad k < m. \tag{3.8}$$

The choice of coupling function h, depending only on phase differences, is motivated by the theory of phase reduction.

The aim is to find the coefficients $\{a, b, c, d\}$ that provide a good approximation to the data $\dot{\vartheta}_j$. We have N time series for our ϑ_j variables, with n points each, obtained with a fixed known sampling rate. With this data we form time-series for the $m = 1 + 2N + N(N-1)/2$ Fourier modes and arrange them as columns of a $n \times m$ matrix, we denote it as Φ, so

$$\Phi = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & \sin(\vartheta_1(t_1)) & \cdots & \sin(\vartheta_N(t_1)) & \cos(\vartheta_1(t_1)) & \cdots & \cos(\vartheta_{N-1}(t_1) - \vartheta_N(t_1)) \\ 1 & \sin(\vartheta_1(t_2)) & \cdots & \sin(\vartheta_N(t_2)) & \cos(\vartheta_1(t_2)) & \cdots & \cos(\vartheta_{N-1}(t_2) - \vartheta_N(t_2)) \\ 1 & \sin(\vartheta_1(t_3)) & \cdots & \sin(\vartheta_N(t_3)) & \cos(\vartheta_1(t_3)) & \cdots & \cos(\vartheta_{N-1}(t_3) - \vartheta_N(t_3)) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \sin(\vartheta_1(t_n))) & \cdots & \sin(\vartheta_N(t_n)) & \cos(\vartheta_1(t_n)) & \cdots & \cos(\vartheta_{N-1}(t_n) - \vartheta_N(t_n)) \end{pmatrix}.$$

The problem of recovering the equations of motion can be formulated as the search for a $m \times n$ matrix of coefficients C such that the equation

$$\Phi C = V \tag{3.9}$$

is satisfied, where

$$V = \begin{pmatrix} \dot{\vartheta}_1(t_1) & \cdots & \dot{\vartheta}_N(t_1) \\ \vdots & \ddots & \vdots \\ \dot{\vartheta}_1(t_n) & \cdots & \dot{\vartheta}_N(t_n) \end{pmatrix} \tag{3.10}$$

is a $n \times N$ matrix of time series of derivatives, we apply smoothening to the derivatives.

The matrix Φ has all possible connections and because the network is sparse only a subset will contribute. We will denote by

– A a subset of columns of Φ that contain the expert guess.

– B further columns we wish to probe.

Without loss of generality (up to relabelling nodes) we assume that A correspond to the first p columns of Φ. Next we consider the concatenation of $[A, B]$ of the matrices A and B and consider the problem

$$[A, B]c = v$$

where v is one of the columns of the matrix V. The vector of coefficients c can be decomposed in terms of the action of A and B

$$c = \begin{pmatrix} x \\ y \end{pmatrix}.$$

The remaining exposition will address two problems: How to find the vector coefficients x, and the effect of the basis extension B on the solution x.

3.2.2 The minimization

Consider the problem of finding the vector of coefficients starting from the expert guess

$$Ax = v$$

The least squares approximation provides the vector x that minimizes the L_2 error

$$\min_{x \in \mathbb{R}^p} \|Ax - v\|_2.$$

A major advantage of this L_2 minimization is that the unique solution has a closed form,

$$x_0 = A^+ v \tag{3.11}$$

where $A^+ = (A^\dagger A)^{-1} A^\dagger$ is the pseudoinverse of A and $\dagger$ denotes the transpose.

Kraleman et al. (KRALEMANN *et al.*, 2008; KRALEMANN; PIKOVSKY; ROSENBLUM, 2011a) have used L_2 minimization to recover the topology of networks with up to nine oscillators. For a brief review, see (PIKOVSKY, 2018). Notice that, although this approach minimizes the Euclidean error, it may not be an optimal solution with respect to other criteria, specially when the smallest singular value of A becomes small. To obtain a well conditioned matrix the size of the time series needs to be significantly large.

Denote Im A the image of the matrix A. Let us consider the case $n > p$, if $v \in \text{Im } A$ the system of equations has a unique solution and it is independent of the minimization. As the data is subjected to fluctuations, in general

$$v = b + z$$

where $b \in \text{Im } A$ and $z \in (\text{Im } A)^\perp$, the orthogonal complement, with $\|z\|_2 \leq \varepsilon$, for some small $\varepsilon > 0$ capturing the fact that fluctuations are small.

3.2.3 Finding Sparse Solutions

For example, we may want a sparse solution, i.e. a vector x with a few non-zero elements. This will indeed be the case when the network has sparse connectivity (such as the star network we shall consider, which has only N connections out of a total of $N(N-1)$ possibilities).

Sparsity can be measured in terms of the condition where

$$\|x\|_0 = \text{number of nonzero elements of } x \tag{3.12}$$

should be as small as possible. Finding a sparse solution is a combinatorial NP-hard problem and not tractable. When the matrix Φ has some additional structure, namely it satisfies the restricted isometry property (RIP) (CANDES; ROMBERG; TAO, 2006), it is well known that a valid heuristics to obtain sparse solutions is to include in the minimization process a penalization on the L_1 norm,

$$\|x\|_1 = \sum_{i=1}^{m} |x_i| \tag{3.13}$$

and consider

$$\min_{\tilde{w}\in\mathbb{R}^m} \|\tilde{w}\|_1 \text{ subject to } \|\Phi\tilde{w} - v\|_2 < \varepsilon, \tag{3.14}$$

for some small ε, where we are still considering $\Phi = [A,B]$. This is known as basis pursuit denoising (DONOHO *et al.*, 2006). The solution to this problem can be obtained by quadratic programming. This is the idea behind the Matlab package "l1magic"[1]. However, there is a small technical drawback here, which is that to start the search for a minimal solution one needs a seed, and this is usually the L_2 solution similar to Equation (3.11). In situations when this L_2 solution is a poor choice (see at Section 3.4.1), the algorithm may not be successful (and finding other clever seeds is a challenging problem).

Another approach is the LASSO algorithm (least absolute shrinkage and selection operator), which we shall adopt. It works by computing solutions to

$$\min_{w\in\mathbb{R}^m} \|\Phi w - v\|_2^2 + \lambda\|w\|_1 \tag{3.15}$$

for a series of values of λ. When λ is large, the solution approaches the null vector. When λ is gradually decreased, each previous solution is a good seed for a new minimization process that finds sparse solutions. If λ becomes too small, sparsity is no longer promoted.

Intermediate values of λ therefore lead to solutions that come close to minimizing $\|\Phi x - v\|_2$, while at the same time being significantly sparse. The actual value of λ is selected by a process of k-fold cross validation, in which: the data is split into k equal-sized parts; a solution is found using all but the lth part; a prediction error is computed when predicting the behavior on the lth part; the errors are added for $1 \leq l \leq k$ to form the total prediction error; the value of λ is chosen to minimize the total prediction error. Later we will see that by our Theorem 3.4.3 once we establish an adapted basis, LASSO is not affected by the poor conditioning of Φ and performs significantly better when data acquisition time is short.

3.3 Numerical experiments

3.3.1 Results for a directed star

We consider a directed star motif for a paradigm. It consists of a central node driving to $N-1$ peripheral nodes, as shown in Figure 11. Since every node's dynamics is only influenced

[1] https://statweb.stanford.edu/ candes/l1magic/

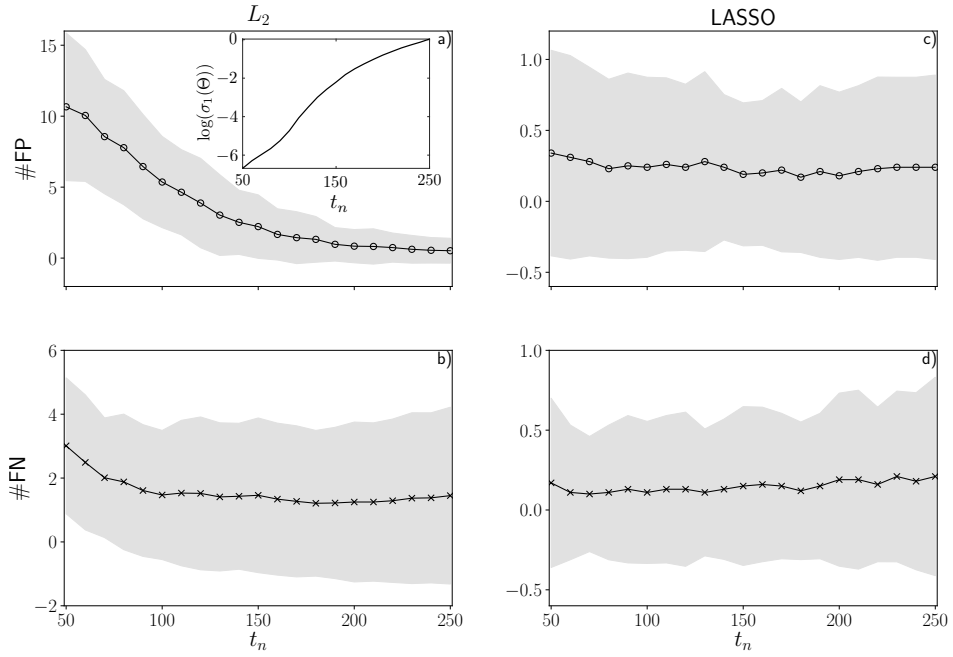

Figure 9 – **Influence of acquisition time t_n on the network recovery**. We consider a directed star graph with $N = 10$ and connections diverging from the hub. Panels a) and c) show the false positives #FP (circles) predicted by L_2 and LASSO, respectively, as number the acquisition time increases. Panels b) and d) show the false negatives #FN (crosses). Each point is an average over 100 random initial conditions and the shaded region is the standard deviation. The inset of panel a) shows the logarithm of the minimum singular value of Φ, averaged over the 100 random initial conditions.

by node 1, the center, we have that $c_{km}^{(j)}$ and $d_{km}^{(j)}$ vanish unless $k = 1$. In our simulations we choose a coupling strength $\alpha = 0.1$, and take the natural frequencies ω_j to be random with uniform distribution in the interval $[0, 2\pi]$ radians per second. Initial conditions are evolved with a fourth order Runge-Kutta integrator with variable step and time series of the phases ϕ_j are then collected with a rate of 10 points per second.

To measure the success of the recovery of methods L_2 and LASSO, we use the measures

#FP (false positives) consisting of connections that are not present in the true network;

#FN (false negatives) the connections that were missed by the recovery.

We do not take into account the strength of the recovered connection; instead we simply check whether a certain connection is present or not. We discard connections that are too weak, less than 10% of the largest entry of the coefficient vector.

3.3.1.1 Effects of the length of the time series

In Figure 9, we show $\#FP$ (circles) and $\#FN$ (crosses), for the L_2 minimization (left column) and for the solution obtained using LASSO (right column), as the acquisition time t_n is

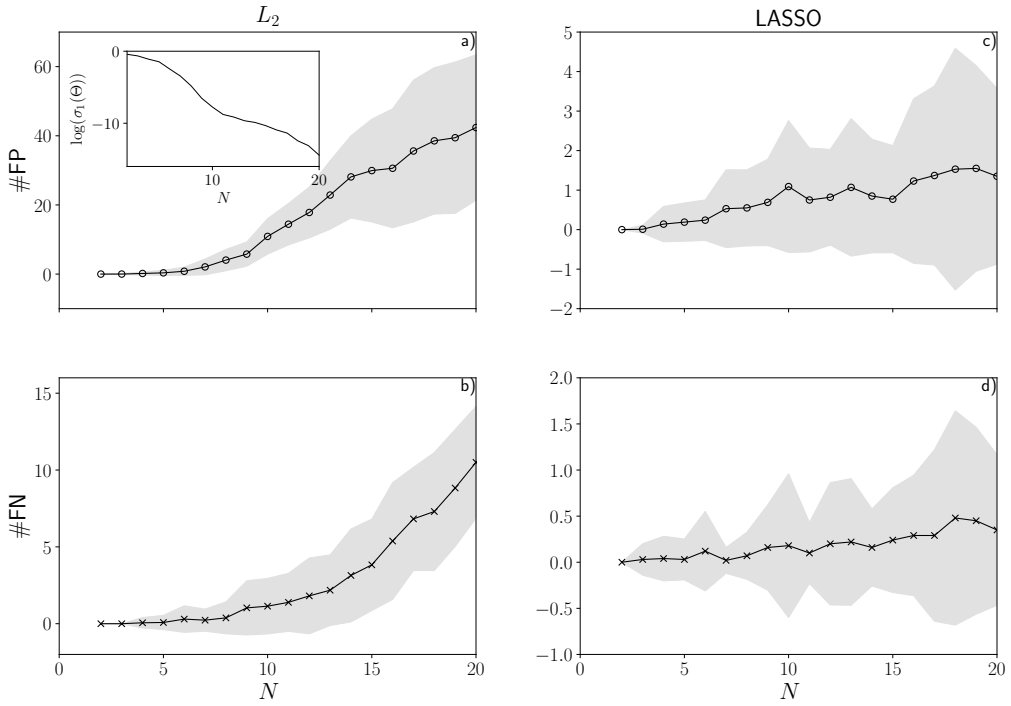

Figure 10 – **Influence network size on the recovery.** We consider a directed star graph of N nodes with connections diverging from. We fix acquisition time $t_n = 100$. Panels a) and c) show the false positives #FP (circles) recovered by L_2 and LASSO, respectively, as number of nodes N increases. Panels b) and d) show the false negatives #FN (crosses). Each point results from average over 100 random initial conditions and the shaded region is the standard deviation. In the inset of panel a) we show logarithm of the minimum singular value of Φ, averaged over the 100 random initial conditions.

varied. These values were averaged over 100 random initial conditions of our network system with $N = 10$ nodes. The LASSO solution is excellent for all values of t_n. The L_2 minimization performs relatively well if t_n is large, but for small values of t_n it predicts many wrong connections. Similar results were obtained by Napoletani and Sauer (NAPOLETANI; SAUER, 2008).

As discussed in the Section 3.4.1, the performance of L_2 minimization as a function of t_n seems to be related to $\sigma_1(\Phi)$, the smallest singular value of the matrix Φ, which can be small for small t_n, as shown in the inset. Subsequently, in Section 3.4.3, we show the reason the LASSO approximation is not affected as much by the poor conditioning of Φ.

3.3.1.2 *Effects of the size of the network with fixed length of time series*

For the directed star graph, in Figure 10 we show $\#FP$ (circles) and $\#FN$ (crosses) as functions of the total number of nodes, for both solutions of L_2 minimization and LASSO. The LASSO solution is stable while L_2 minimization is accurate for small networks, with $N \leq 7$, and is not able to handle the large-but-sparse configuration. In the inset, we show the corresponding average value of $\log(\sigma_1(\Phi))$. It suggest a correlation between between the poor performance of the L_2 minimization and ill condition of Φ captured by a small singular value.

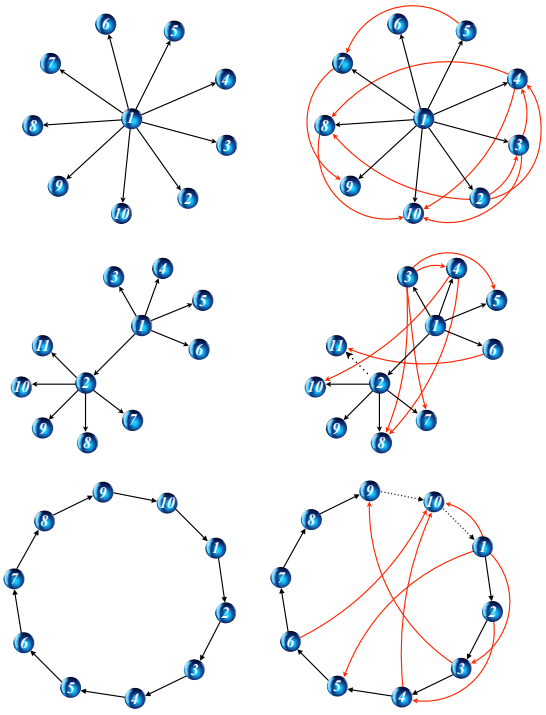

Figure 11 – **Comparison between LASSO and L_2 minimization for three paradigmatic networks**. In the upper panel the directed star, in the mid panel one connected directed star forcing another directed star, in the bottom panel a directed ring. The network recovered by the LASSO is presented in the left and shows perfect recovery and the L_2 minimization recover is presented in the right. Spurious connections are shown as thin red lines, missing connections are shown as dotted lines. We used a single random initial condition in each case and $t_n = 100$.

3.3.2 Results for other networks

In this section we briefly consider some other sparse networks: the twin stars, which consists of two stars joined by a single link, and a ring, both are illustrated in Figure 11. In the twin stars configuration node 1 drives nodes 2 to 6, while node 2 drives nodes 7 to 11. In the ring configuration node j drives its following neighbour $j+1$, and node N drives node 1.

We again use $\alpha = 0.1$ and take the natural frequencies ω_j to be random with uniform distribution in the interval $[0, 2\pi]$. Initial conditions are evolved with a Runge-Kutta integrator and time series of the phases ϕ_j are then collected with time steps of 0.1.

In Figure 4 we show the connections that were recovered by the two methods, L_2 minimization and LASSO, from a single random initial condition propagated for $t_n = 100$. We performed a kind of hard thresholding, by discarding connections that were too weak (we considered coupling strengths smaller that 10% of the largest one to be weak).

The results from LASSO are excellent in all cases, but the L_2 minimization does not perform so well: it fails both to recover existing connections (false negatives depicted with dotted lines in Figure 3) and recovers false positive (thin grey lines).

3.3.3 Effects of basis extension

We discuss how the inclusion of new functions in the basis can affect the recovery. Our first example is shown in Figure 10. Since the network is a directed star with connections diverging from the hub, the recovery of each node is independent as the hub acts as a master to the leaves. Thus, increasing the network size and recovering the connections of a given node has the same effect as including new (*a posteriori*) unnecessary functions in the basis. We could wrongly expect this basis extension would not influence the recovery method. Figure 10 shows that the L_2 recovery is strongly affected by such extensions as the inclusions of new functions, while keeping the length of the time-series fixed, makes the operator Φ ill-conditioned.

Next, we notice that the function g in Eq.(3.6) plays no role in the dynamics when phases are slow variables. We study the effect of the inclusion of such functions in the recovery process. Figure 12 shows the results of such basis extension for a directed ring. The basis extension is made using higher harmonics $\mathscr{E}_k = \{\sin(m\vartheta_k), \cos(m\vartheta_k)\}_{m=2}^{10}$ for each node k. Starting from $k = 0$ we include the new functions of a node k while keeping all previously included functions. Thus, in the first iterator $k = 1$ we include 16 new functions and at the end of the process $k = 10$ we include 160 functions. We observe that the number of false positives and negatives remains unaltered as the basis is increased either for L_2 and LASSO. We notice that LASSO remains stable under basis extension.

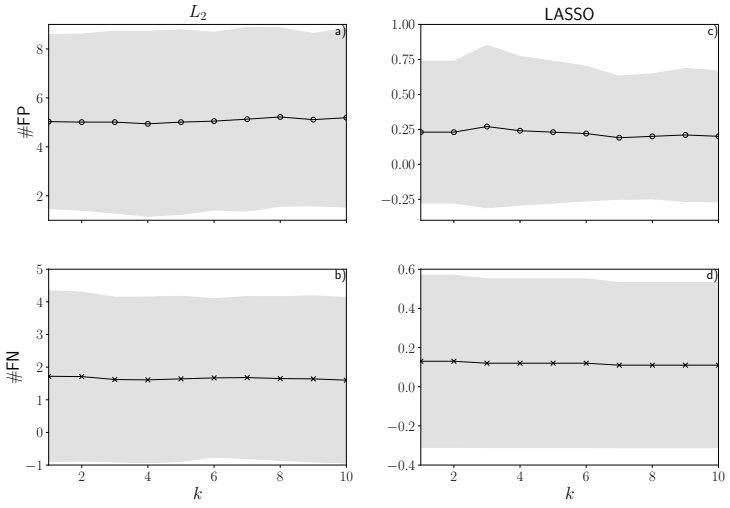

Figure 12 – **Effects of basis extension in the recovery**. We consider a fixed acquisition time $t_n = 100$ and a directed ring with $N = 10$. First, we recover the network without higher harmonics in the phases corresponding to k. Then, we extend the basis to include higher harmonics $\mathscr{E}_k = \{\sin(m\vartheta_k), \cos(m\vartheta_k)\}_{m=2}^{10}$ of a node k iteratively. Thus, for each k we include 16 new functions in the basis and apply the recovery methods while keeping the previously added functions. Panels a) and c) show the false positives #FP (circles) recovered by L_2 and LASSO, respectively, as a function of k. Panels b) and d) show the false negatives #FN (crosses). Each point is an average over 100 initial conditions and the shaded region is the standard deviation.

3.4 Stability of sparse networks under basis extension

3.4.1 L_2 is unstable under basis extension

When we extend the basis, probing new possible connections, we face a problem as $[A,B]$ may have small singular values, leading to instabilities. This means that even if

$$Ax = b$$

has a sparse solution, it may happen that

$$[A,B]w = b + z$$

has a solution w that is far from being sparse in its restriction to the components corresponding to x, here z captures small measurement errors. This would mean that the basis extension is unstable.

Our next proposition characterizes this situation. We prove it using the concept of principal angle between subspaces, in particular the largest principle angle between the orthogonal complement of the image of matrix A, $(\mathrm{Im}\, A)^{\perp}$, and the image of the matrix B, $\mathrm{Im}\, B$.

Proposition 3.4.0.1. Let $A \in \mathbb{R}^{n\times p}$ be a column full rank matrix, $b \in \mathrm{Im}\, A$ and $z \in (\mathrm{Im}\, A)^{\perp}\backslash\{0\}$. Let x^* be the unique solution of the problem

$$\min_{x\in\mathbb{R}^p} \|Ax - b - z\|_2.$$

Let $B \in \mathbb{R}^{n\times q}$ be such that the matrix concatenation $\Phi = [A,B]$ is also column full rank with $n > p+q$. Let $r = \min\{p,q\}$ and the principal angles between the subspaces $(\mathrm{Im}\, A)^{\perp}$ and $\mathrm{Im}\, B$ satisfy: $0 < \beta_1 < \cdots < \beta_r < \frac{\pi}{2}$. Let $\hat{w} = (\hat{w}_1, \hat{w}_2)$ be the unique solution of the problem

$$\min_{w\in\mathbb{R}^{p+q}} \|\Phi w - b - z\|_2.$$

Then for a generic $z > 0$ given a natural number $N_0 > 0$ there is a $\varepsilon > 0$ such that if $|\beta_r - \pi/2| < \varepsilon$ we obtain $\|x^* - \hat{w}_1\|_2 > N_0$.

We prove this proposition in Appendix B.2. The above proposition explains the instability we observed in the numerical results, which are also in agreement with the observations made by Napoletani and Sauer (NAPOLETANI; SAUER, 2008).

As a remark, when the dynamics is chaotic the columns of the matrix Φ behave as pseudorandom vectors. Let us assume that $p = q$ for the matrices A and B. Thus we can think of the column spaces of A and B as two p-dimensional vector spaces taken at random from a larger n-dimensional space, $n > 2p$. The principal angles between them have a joint multivariate beta distribution (ABSIL; EDELMAN; KOEV, 2006); from well known random matrix theory results, it then follows that, as $n \to \infty$ with $p = \xi n$, the average value of the smallest principal angle satisfies $\cos(\beta_1) = 4\xi(1-\xi)$. The value $\xi \to 1/2$ corresponds to the case $\Phi = [A,B]$, when the principal angles tends to 0. This indicates that, in the large basis limit, instability is generic.

3.4.2 Basis Adaptation guarantees coherence

Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ be the torus. From here on our theoretical formulation and analysis is described in terms of a map denoting the dynamics. This assumption is not harmful since the phase dynamics recovery on $\mathbb{T}^N$ is given by the time-one map f of the flow. This map is induced by the Euler approximation of the differential equations and the sampling procedure of the trajectories.

In the following exposition we will denote by X the metric space being either a compact subset of $\mathbb{R}^d$ or a parallelizable manifold such as the torus $\mathbb{T}^d$. We assume the map denoting the dynamical system is $C^r(X)$ with $r \geq 1$. This will contain all examples in the paper and avoid a technical detour. We denote ψ as basis functions representing the map f and the functions φ and ϕ are observables. We understand sparse representation of the map as

Definition 3.4.1 (Sparse Representation)**.** Let $f : X \to X$ and $\mathscr{L} = \{\psi_i\}_{i=1}^m$ be a set of basis functions with $\psi_i : X \to X$ for $i = \{1, \ldots, m\}$ such that

$$f = \sum_{i=1}^{m} c_i \psi_i.$$

We say that f has an s-sparse representation in $\mathscr{L}$ if the vector $x = (c_1, \ldots, c_m)^\dagger$ is s-sparse.

Dynamical information: Ergodicity. We focus our analysis on ergodic dynamical systems. A well-known property is that the time average of an observable evaluated at a typical orbit converges to the space average. This is more generally stated in the following Theorem 3.4.2

Theorem 3.4.2 (Birkhoff Ergodic theorem)**.** Let (f, μ) be a discrete ergodic dynamical system on the compact metric space X. Given any $\varphi \in L_1(\mu)$, there exists a set of initial conditions $E \subset X$ with $\mu(E) = 1$ such that for any $\varepsilon > 0$ and $x_0 \in E$ there exists $n_0(\varepsilon, x_0) > 0$ where the following holds

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k(x_0) - \int \varphi d\mu \right| < \varepsilon \qquad \forall\, n > n_0. \tag{3.16}$$

Birkhoff Ergodic theorem is the main ingredient to calculate the coherence for ergodic dynamical systems in Theorem 3.4.3. It introduces a change of inner product: instead of looking at the Euclidean inner product among vectors on $\mathbb{R}^n$, we approximate it by the inner product on the space of integrable functions with respect to the ergodic measure.

Theorem 3.4.3 goes beyond. It states that for any discrete ergodic dynamical system whose measure has density, we can construct a set of basis functions adapted to the ergodic measure via Gram-Schmidt procedure[2]. These adapted basis functions do not harm the sparsity representation of the map and has control over the coherence of the matrix for large enough data.

[2] These adapted basis functions are related to the Bounded Orthonormal System (BOS) in Foucart and Rauhut (FOUCART; RAUHUT, 2013). They differ in respect to the choice of the reference measure. BOS carries the measure given by the uniform sampling procedure whereas here it comes along the observed trajectory.

Our result is related to what was obtained by Tran and Ward (TRAN; WARD, 2017b). The authors use Central Limit Theorem applied for Lorenz systems perturbed over time to obtain the null-space property (which is a weaker property than RIP) for a similar version of the matrix Φ.

To our best knowledge, our results are one of the few examples to advance the search for basis functions adapted to the dynamical system generating the time series. Recently, Hamzi and Owhadi (HAMZI; OWHADI, 2021) proposed a kernal-based method in a similar direction.

Drawback. It is worth noting that Theorem 3.4.3 is an existence statement since it requires that the sparse representation of the dynamical system is known a priori. Besides it is valid for large enough data. To determine the minimum amount of data for controlling the coherence of Φ, it would be necessary to know the speed of convergence of the Birkhoff sums for the basis functions. This will be done in the near future.

Theorem 3.4.3 (Ergodic Coherence)**.** Let (f,μ) be an ergodic dynamical system with μ absolutely continuous with respect to Lebesgue (Leb(X)). Let $\mathscr{L}_0$ be a set of basis functions such that f has an s-sparse representation in $\mathscr{L}_0$. Given $\eta_0 > 0$ and $\varepsilon \in (0,1)$ there is a set of basis functions $\mathscr{L}$, $n_0 > 0$ and a good set of initial conditions $G \subset X$ such that

(i) $\mu(G) > 1-\varepsilon$, and for any $x_0 \in G$ and $n > n_0$ we have $\eta(\Phi(\mathscr{L})) < \eta_0$.

(ii) the representation of f in $\mathscr{L}$ is also s-sparse.

Proof. We develop the argument assuming that $X \subset \mathbb{R}$. To generalize for large dimensions or for $\mathbb{T}^d$ it is enough to break down the problem in terms of coordinates. The main ingredient in the proof is the Birkhoff's Ergodic Theorem 3.4.2. Having the ergodic theorem we split the proof in three steps.

Step 1: Ergodicity and basis adaptation. Let $\mathscr{L}_0 = \{\psi_1,\dots,\psi_m\}$ be a set of basis functions, where each $\psi_i : X \to X$. We perform a Gram-Schmidt process in $L_2(\mu)$ and obtain an orthogonal basis

$$\hat{\mathscr{L}} = \{\varphi_1,\dots,\varphi_m\}.$$

Notice that since $\mu = \nu$Leb we define

$$\phi_i = a_i\varphi_i$$

where $a_i^2 = 1/\int \varphi_i^2 \; \nu \; d\text{Leb}$ such that $\mathscr{L} = \{\phi_i\}_{i=1}^m$ is an orthonormal system with respect to $L_2(\mu)$ in the span of $\mathscr{L}_0$. For an arbitrary initial condition x_0, let

$$u_i := \frac{1}{\sqrt{n}}\begin{pmatrix}\phi_i(x_0)\\ \vdots \\ \phi_i(f^{n-1}(x_0))\end{pmatrix} \qquad u_j := \frac{1}{\sqrt{n}}\begin{pmatrix}\phi_j(x_0)\\ \vdots \\ \phi_j(f^{n-1}(x_0))\end{pmatrix} \tag{3.17}$$

be the ith and jth columns of the matrix $\Phi(\mathscr{L}) \in \mathbb{R}^{n\times m}$. Then notice that the inner product between columns i and j is

$$\begin{aligned}\langle u_i, u_j\rangle &= \frac{1}{n}\sum_{k=1}^{n}\phi_i(f^k(x_0))\phi_j(f^k(x_0)) \\ &= \frac{1}{n}\sum_{k=1}^{n}(\phi_i\cdot\phi_j)\circ(f^k(x_0)) \\ &=: \frac{1}{n}S_n(\phi_i\cdot\phi_j)(x_0).\end{aligned}$$

From the smoothness of the map f, $(\phi_i\cdot\phi_j)$ is integrable $L_1(\mu)$ so by Birkhoff Ergodic theorem there is a set G_{ij} such that $\mu(G_{ij})$ has full measure and for each $x_0 \in G_{ij}$ and $\varepsilon_1 > 0$ there is $n_0 > 0$ such that for any $n > n_0$ we have

$$\begin{aligned}\left|\langle u_i, u_j\rangle - \int \phi_i\cdot\phi_j d\mu\right| &\leq \varepsilon_1 \\ \left|\langle u_i, u_j\rangle - \delta_{ij}\right| &\leq \varepsilon_1\end{aligned}$$

where δ_{ij} is the Kronecker delta.

Step 2: Large measure of initial conditions for the basis. Hence, we are interested in the subset with cardinality $K = \frac{m(m-1)}{2}$

$$\mathscr{G} = \{(\phi_i\cdot\phi_j) \mid i,j = 1,\ldots,m\} \subset L_1(\mu) \tag{3.18}$$

where each element corresponds to pairwise multiplication of basis functions in $\mathscr{L}$. We aim at finding a good set G of initial conditions where the control of n_0 is uniform.

Using Egoroff's theorem (FOLLAND, 2013) we can make the Birkhoff sum $\frac{1}{n}S_n\phi$ converge uniformly on a large measure set G_ϕ of X instead of the "almost every point" convergence. Fix $\eta_0 > 0$ and take $\varepsilon/(2K)$. For each observable ϕ in the set of Equation (3.18), the precision $\varepsilon/(2K)$ determines a subset G_ϕ of X which by Egoroff's theorem has measure $\mu(G_\phi) > 1 - \frac{\varepsilon}{2K}$ where the convergence of $\frac{1}{n}S_n\phi$ is uniform. So, we take the set of initial conditions as

$$G = \bigcap_{\phi\in\mathscr{G}} G_\phi. \tag{3.19}$$

Using the complement of G we can calculate that

$$\mu(G) > 1 - \varepsilon.$$

This determines the set of initial conditions for which we can calculate the coherence of the matrix $\Phi(\mathscr{L})$. Due to uniformity of initial conditions in G, for each observable $(\phi_i\cdot\phi_j)$ in the set of Equation 3.18 there exists $n_{i,j} > 0$ such that the inner product of any two distinct normalized column vectors $|\langle v_i, v_j\rangle|$ has the following form for any $n > n_{i,j}$

$$|\langle v_i, v_j\rangle| = \left|\frac{1}{n}S_n(\phi_i\cdot\phi_j)(x_0)\right| \leq \eta_0.$$

Take $n_0 := \max_{i \neq j} n_{i,j}$ and this proves the statement.

Step 3: Sparsity. Thus we are only left to prove sparsity. We know by assumption that there is a sparse solution to

$$\Phi(\mathscr{L}_0)x_s = v.$$

Let us rearrange $\mathscr{L}_0$ such that x_s has only its first s entries nonzero. Next, the Gram-Schmit process reduces to a QR decomposition that is

$$\Phi(\mathscr{L}_0) = \Phi(\mathscr{L})R$$

thus,

$$\Phi(\mathscr{L}_0)x_s = \Phi(\mathscr{L})\hat{x}_s$$

where

$$\hat{x}_s = Rx_s$$

but R is upper triangular and thus by construction only the first s entries of $\hat{x}_s$ will be nonzero. □

3.4.3 Sparse Solutions are stable under basis extension

Next, we wish to prove that once the basis is adapted and the initial expert guess is meaningful, extending the basis is not harmful for the solution. Next proposition proves that, given a set of basis functions which represents f sparsely, the minimization problem from Candès Theorem 2.5.6 has a solution that approximates the true sparse solution. Moreover, using Theorem 3.4.3, which introduces a orthonormal set of basis functions $\mathscr{L}$ and a matrix $\Phi(\mathscr{L})$, we can find a sub-matrix of $\Phi(\mathscr{L})$, $A(\mathscr{L})$, which approximates the same solution in a smaller space.

It is worth noting that both LASSO and quadratically constrained basis pursuit are L_1 minimization problems related to each other. More precisely, for each solution $x^\star$ of LASSO there exists a $\varepsilon := \varepsilon_{x^\star} > 0$ such that $x^\star$ is solution of Equation (2.22), see Fourcart and Rauhut (FOUCART; RAUHUT, 2013, Proposition 3.2). So, our results using the quadratically constrained basis pursuit are extended to LASSO solutions as well.

Proposition 3.4.3.1 (Sparsity level is attained)**.** Let $\mathscr{L}_0$ be a set of basis functions with cardinality m such that f has a s-sparse representation in $\mathscr{L}_0$. Then, there is $n_0 > 0$, a large set of initial conditions and a basis $\mathscr{L}$ such that we find a matrix $A(\mathscr{L}) \in \mathbb{R}^{n \times p}$ where $s < p < m$ and the solution x^* of the reconstruction problem

$$\min_{\tilde{x} \in \mathbb{R}^p} \|\tilde{x}\|_1 \text{ subject to } \|A(\mathscr{L})\tilde{x} - v\|_2 < \varepsilon \tag{3.20}$$

attains the sparse representation of f.

Proof. We use Proposition 2.5.4.2 together with Theorem 3.4.3. Let $1 < s < m$ be the sparsity level of the representation of the map f with respect to the proposed set $\mathscr{L}_0$. By assumption we know there exists a sparse solution $x_s \in \mathbb{R}^m$ such that $\Phi(\mathscr{L}_0)x_s = v$. We rearrange $\mathscr{L}_0$ such that x_s has only its first s entries nonzero. Fix $0 < \eta_0 < (\sqrt{2}-1)/(2s-1)$. By Theorem 3.4.3 there exists an orthogonal basis $\mathscr{L}$, $n_0 > 0$ and a large set of initial conditions that $\eta(\Phi(\mathscr{L})) \leq \eta_0$ and $\hat{x}_s \in \mathbb{R}^m$. Thus from Proposition 2.5.4.2

$$\delta_{2s}(\Phi(\mathscr{L})) < \sqrt{2}-1. \tag{3.21}$$

By Theorem 2.5.6, the sparse solution $\hat{x}_s$ is approximated by the solution of the quadractically constrained basis pursuit problem.

Let $p,q \in \mathbb{N}$ be chosen such that $s < p,q < m$ and $m = p+q$. Without loss of generality, we can rearrange the basis elements in such way $\Phi(\mathscr{L}) = [A(\mathscr{L}), B(\mathscr{L})]$ where $A(\mathscr{L}) \in \mathbb{R}^{n\times p}$ and $B(\mathscr{L}) \in \mathbb{R}^{n\times q}$. Moreover, using $A(\mathscr{L})$ in the quadratically constrained basis pursuit problem the solution approximates the sparse solution $\hat{x}_s$ through a vector lying in $\mathbb{R}^p$. This is true because $\delta_{2s}(\Phi(\mathscr{L}))$ is an upper bound for $\delta_{2s}(A(\mathscr{L}))$ and $\delta_{2s}(B(\mathscr{L}))$.

For the noiseless case we could say that $A(\mathscr{L})$ is the minimum matrix such that the minimization problem attains the sparse solution.

□

The existence of a sub-matrix of $\Phi(\mathscr{L})$ in the above proposition indicates that we can use Theorem 3.4.3 in a different way to guarantee that sparse solutions are stable under basis extension. The following corollary states this stability more precisely.

Corollary 3.4.4 (Stability under basis extension)**.** Suppose $\mathscr{L}_0$ is a subset of basis functions with cardinality $p < m$ such that f has a s-sparse representation in $\mathscr{L}_0$. Denote $x_s \in \mathbb{R}^p$ the unique sparse solution of Equation (3.20). Then there is $n_0 > 0$, a large set of initial conditions and a basis $\mathscr{L}$ such that $w = (x_s, 0) \in \mathbb{R}^m$ is solution of

$$\Phi(\mathscr{L})w = [A(\mathscr{L}), B(\mathscr{L})]w = v$$

and the solution $w^* = (w_1^*, w_2^*)$ of

$$\min_{\tilde{w}\in\mathbb{R}^m} \|\tilde{w}\|_1 \text{ subject to } \|[A(\mathscr{L}), B(\mathscr{L})]\tilde{w} - v\|_2 < \varepsilon$$

satisfies

$$\|w_1^* - x_s\|_2 \leq C\varepsilon \quad \text{and} \quad \|w_2^*\|_2 \leq \hat{C}\varepsilon$$

for constants C and $\hat{C}$.

Proof. Thinking in the reverse direction as in the previous proposition we could assume $\mathscr{L}_0$ is a subset of basis functions with cardinality $p < m$ such that f has a s-sparse representation in $\mathscr{L}_0$.

Then by Theorem 3.4.3, Proposition 2.5.4.2 and Candès theorem 2.5.6 there is $n_1 > 0$, a large set of initial conditions and a basis $\mathscr{L}$ such that $A(\mathscr{L}) \in \mathbb{R}^{n\times p}$ satisfies Equation (3.20).

The key fact is the finiteness of the set of basis functions. Let us denote by $\mathscr{L}_0^c$ the complement of $\mathscr{L}_0$. If we take the union $\mathscr{L} \cup \mathscr{L}_0^c$ we can apply Theorem 3.4.3 for this set. Since $\mathscr{L}$ is already orthonormal, the Gram-Schmidt procedure is necessary only for the functions of $\mathscr{L}_0^c$. Adjusting $n_0 > n_1 > 0$ and the initial conditions, and using orthonormality we can guarantee continuity of the unique sparse solution of Equation (3.20) in the larger space. The estimate in Equation (3.4.4) is given by applying Theorem 2.5.6. □

3.5 Closing remarks

We considered the problem of recovering, from phase dynamics, the interaction structure of a sparse network of oscillators. We compared two different recovery methods, both based on a Fourier expansion of the interaction functions. One of them is the traditional least squares approximation, which finds the vector of coefficients that minimize the L_2 error of the approximation and has been successful in previous approaches. The other is LASSO. For small networks and when long data sets are available, both approaches are equivalent. But we have found LASSO to be much more apt to sparse network configurations and short times than the L_2 minimization. We showed that LASSO can perform remarkably well when dynamical information is taken into account and the basis functions are adapted. This adaptation leads to unique solutions to the minimization problem that are also stable under basis extension. Once the basis is adapted to the dynamics LASSO recovers sparse networks with excellent precision even when only relatively little data is available.

CHAPTER

4

DIVIDE-AND-CONQUER NETWORK RECONSTRUCTION

This chapter is devoted to presenting our work entitled *Ergodic Basis Pursuit induces Divide-and Conquer Network Reconstruction*, which is in preparation to be submitted.

Can we design an algorithm that reconstructs the network using only part of the input data? To answer this question we tackle the following perspective: Although the network structure is unknown, there is still a priori knowledge. For instance, neuronal networks contain groups of nodes that are expected to interact such as interneurons (TOWLSON *et al.*, 2013; SPORNS; BETZEL, 2016). The motif structure might be known as rich club motifs, where a group of nodes form a highly connected cluster mediating information to the rest of the network. Hence, our approach is to incorporate expert knowledge of the structure into the reconstruction method. Using expert knowledge we can break the reconstruction into subproblems and solve each subproblem locally and in parallel. Then, we combine information gathered for each local solution to form the global solution. That is, if we know which neighbors two distinct nodes have, we can combine this information and obtain a larger uncovered structure. The caveat of this approach is to guarantee the exact reconstruction of local network structure and how to combine them to obtain a global solution. Any instabilities could be obstructions to the search for local network solutions in parallel.

In this paper, we use a recent reconstruction method introduced by the authors so-called Ergodic Basis Pursuit (EBP) to solve the inverse problem (PEREIRA; SANTOS; STRIEN, 2023). EBP is devoted to ergodic network dynamics that exhibit the decay of correlations. Recently, the authors showed that once a minimum length of time series is attained, EBP guarantees successful network reconstruction. So, incorporating expert knowledge of the network structure into EBP, we prove the existence of a Divide-and-Conquer Network Reconstruction (DCNR) method by solving the local subproblem and combining the found local solution into a large solution. Local subproblems do not require more data than the estimated minimum, and as we probe new nodes

or connections the solution remains unchanged. In other words, the reconstruction is stable. The major advantage is that the search for different motifs can be performed in parallel speeding up the reconstruction tremendously.

4.1 Network dynamics

We will focus on a class of networks common in applications such as coupled lattice maps (CHAZOTTES; FERNANDEZ, 2005) and laser dynamics (HART *et al.*, 2019). We consider networks with N nodes given by

$$\bar{x}_i = f_i(x_i) + \alpha \sum_{j=1}^{N} A_{ij} h_{ij}(x_j, x_i), \quad i \in [N] := \{1, \dots, N\}, \tag{4.1}$$

where x_i represents the state of node i and $\bar{x}_i$ the evolved state, the adjacency matrix A satisfies A_{ij} equals 1 if node i receives a connection from j and 0 otherwise. $f_i \colon M_i \to M_i$ corresponds to the isolated map, $h_{ij} : M_i \times M_i \to M_i$ is the pairwise coupling function and α is the coupling strength. Here, M_i is compact subset of $\mathbb{R}$. We let $x = (x_1, \dots, x_N)$ and write the full network dynamics as $\bar{x} = F(x)$.

We assume the availability of a time series of each node over time, forming a multivariate time series $\{x(t)\}_{t=0}^{n}$. We aim to obtain the network structure from $\{x(t)\}_{t=0}^{n}$.

4.2 Reconstruction problem as a linear equation

The reconstruction problem can be recast as a linear equation. First, consider the assumption

(*a*) Denote a finite (ordered) set of basis functions $\mathscr{L} = \{\phi_1, \phi_2, \dots, \phi_m\}$, where $\phi_i : M^N \to \mathbb{R}$, with $M^N := \prod_{i=1}^{N} M_i$. We restrict ourselves to the case that the collection $\{f_i\}_{i \in [N]}$ and $\{h_{ij}\}_{i,j \in [N]}$ are in the span of $\mathscr{L}$. In other words, $\mathscr{L}$ contains functions that represent up to pairwise interactions. We refer to $\mathscr{L}$ as an ordered network library.

Then, we denote the library matrix as

$$\Phi(X) = \frac{1}{\sqrt{n}} \begin{pmatrix} \phi_1(x(0)) & \phi_2(x(0)) & \cdots & \phi_m(x(0)) \\ \phi_1(x(1)) & \phi_2(x(1)) & \cdots & \phi_m(x(1)) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(x(n-1)) & \phi_2(x(n-1)) & \cdots & \phi_m(x(n-1)) \end{pmatrix}.$$

Also, define the $n \times N$ matrix of multivariate data

$$\bar{X} = \begin{pmatrix} x_1(1) & \cdots & x_N(1) \\ \vdots & \ddots & \vdots \\ x_1(n) & \cdots & x_N(n) \end{pmatrix}.$$

Then the reconstruction problem becomes to solve the following linear equation

$$\bar{X} = \Phi(X)C \tag{4.2}$$

whose solution is the $m \times N$ matrix of coefficients C which encodes the graph structure.

When the length of time series n is sufficiently large ($n \gg m$), then the least square approximation $\arg\min \|\bar{X} - \Phi(X)C\|_2$ provides a good approximation of the solution C (KRALEMANN; PIKOVSKY; ROSENBLUM, 2011a; KRALEMANN; PIKOVSKY; ROSENBLUM, 2011b). In fact, $\Phi(X)$ is full column rank (LUENBERGER, 1997) and the least square minimization finds a unique solution, allowing to recover the network structure. However, the least square approximation applied to large networks requires a large amount of data. For example consider coupled logistic maps: let $M = [0,1]$, the logistic map $f(x) = ax(1-x)$ as isolated map and the pairwise coupling function given by

$$h(x_j, x_i) = x_j^3 x_i, \quad i, j = 1, \ldots, N. \tag{4.3}$$

Due to cross terms $x_i^p x_j^q$ (which are elements of the space of the pairwise polynomials with the degree at most $r = p+q$) the number of elements m in the basis will be large. Roughly speaking the number of functions grows as $m = \mathcal{O}(N^2 r^2)$. So, even in small networks of $N = 10$ nodes and $r = 4$, we have $m \approx 1600$.

4.3 Uniqueness of solutions

Large networks make $\Phi(X)$ ill-conditioned, and Equation (4.2) has infinitely many solutions. So, recovering C is challenging. The strategy is to use the sparsity of the network and search sparse matrices C that satisfy the linear equation. More precisely, we assume

(b) The network structure is undirected and sparse, that is only a few connections are realized out of all possible. So the network sparsity induces a sparse representation of the network dynamics F. More precisely, for each node i: $F_i = \sum_l c_i^l \phi_l$ where $c_i = (c_i^1, \ldots, c_i^m) \in \mathbb{R}^m$ is an s-sparse vector, that is, at most s of its entries are nonzero.

Network sparsity alone is not guaranteed to obtain the uniqueness of sparse solutions that satisfy the linear equation. The library matrix must satisfy extra conditions. Based on compressive sensing theory, when any $2s$ set of columns of $\Phi(X)$ forms a set of a near orthonormal system

of vectors, or more precisely, $\Phi(X)$ satisfies restricted isometry property (RIP) or has restricted isometry constant given by

$$\delta_s(\Phi) := \max_{\mathscr{S}\subset[m],\text{card}(\mathscr{S})\leq s} \|\Phi^*_{\mathscr{S}}\Phi_{\mathscr{S}} - \mathbf{1}_s\|_2 \in (0,1), \tag{4.4}$$

where $\Phi_{\mathscr{S}}$ is the submatrix of Φ composed by columns supported in $\mathscr{S}$, then each column of Equation (4.2) has a unique s-sparse solution (CANDES; TAO, 2005). Assumption (b) states that each node dynamics is s-sparsely represented in the library $\mathscr{L}$. In other words, C is the unique solution of Equation (4.2) whenever $\Phi(X)$ satisfies RIP. The caveat is that a priori $\Phi(X)$ does not satisfy this RIP property for the given library $\mathscr{L}$. Then, the strategy is instead of using $\mathscr{L}$, whose matrix fails in satisfying RIP, we introduce a new library whose associated matrix has the desired property.

First, the new library must preserve the sparse representation of F. To this end, we assume

(c) Also, we assume that the network physical measure μ is close to a product measure ν, i.e., statistically the network dynamics evolve almost such as there were N independent nodes. We can introduce a notion of distance d that calculates the maximum difference between integrals with respect to μ and ν over a pair of functions in a set of functions. This assumption is summarized as $d(\mu,\nu) < \zeta$ for small $\zeta > 0$ (PEREIRA; SANTOS; STRIEN, 2023). For instance, in the weak coupling regime, this assumption is fulfilled (EROGLU *et al.*, 2020).

This assumption (c) plays a fundamental role. We show that the Gram-Schmidt process in $L_2(\nu)$ in the span of $\mathscr{L}$ yields $\mathscr{L}_\nu = \{\varphi_i\}_{i=1}^m$ which contains orthonormal functions with respect to $L_2(\nu)$, i.e., $\int \varphi_i\varphi_j d\nu = \delta_{ij}$ with δ_{ij} is the Kronecker delta (PEREIRA; SANTOS; STRIEN, 2023). Moreover, F has a $\omega_r(s)$-sparse representation in $\mathscr{L}_\nu$ with $\omega_r(s) = \Big(\lfloor\frac{r}{2}\rfloor\big(r-\lfloor\frac{r}{2}\rfloor\big)+r+1\Big)s$. From here on, we call $\mathscr{L}_\nu$ the orthonormal network library, its associated library matrix is denoted $\Phi_\nu(X)$ and the associated coefficient matrix C_ν.

As a second step, dynamical information is relevant to make $\Phi(X)$ satisfy RIP. In fact, when the network dynamics is evolving in synchronous motion, $\Phi(X)$ trivially does not fulfill condition (4.4). Then, the final assumption is a dynamical regime that the network dynamics must satisfy. We assume that

(d) The data comes from an exponential mixing network dynamics (HANG; STEINWART, 2017), that is, it is chaotic and exhibits exponential decay of correlations. In particular, (F,μ) corresponds to an ergodic dynamical system with μ being the physical measure.

The main consequence of assumptions $(a)-(d)$ is that for large networks given a desired

$\delta \in (0,1)$ then we can determine a minimum length of time series that satisfies

$$n_0 \approx \frac{20K_1\omega_r(s)^2}{\delta^2}\ln(Nr), \tag{4.5}$$

such that $\delta(\Phi_v) \leq \delta$ with high probability (PEREIRA; SANTOS; STRIEN, 2023).

4.4 Divide-and-Conquer: all or nothing

Equation (4.5) guarantees that $\Phi_v(X)$ satisfies RIP with a desired RIP constant. Hence, we can apply a known result in compressive sensing theory (CANDES; TAO, 2005) in terms of network reconstruction. For any $\delta < 1$, once n_0 is attained, c_v is the unique solution of the equation

$$\bar{x} = \Phi_v(X)c_v. \tag{4.6}$$

Hence, we can describe in detail a divide-and-conquer network reconstruction: (i) Suppose based on expert knowledge, we know the group of nodes belonging to a motif. Then, when trying to reconstruct this motif, if we miss a link in the motif, the linear equation Equation (4.6) is inconsistent, i.e., there exists no solution. (ii) Once all neighbors of a given node are found, probing extra links does not spoil the connections already found. Looking at this problem as a basis extension, local solutions are stable under basis extension (NOVAES; Roque dos Santos; PEREIRA, 2021). A priori, this observation seems obvious, but this is only possible because the solution is unique. These observations can be summarized in the following result:

Proposition 4.4.0.1 (Sufficient infeasibility condition)**.** (PEREIRA; SANTOS; STRIEN, 2023) Consider that the length of the time series satisfies Equation (4.5). Let $\bar{x} = \Phi_v(X)c_v$ where $c_v \in \mathbb{R}^m$ is an $\omega_r(s)$-sparse vector and consider the set

$$\mathscr{F}_{\bar{x}} = \{w \in \mathbb{R}^m : \Phi_v(X)w = \bar{x}\}.$$

Given a set $\mathscr{U} \subseteq [m]$ where $\mathscr{U} \cap \text{supp}(c_v) \neq \emptyset$. Then $\mathscr{U} \subsetneqq \text{supp}(c_v)$ if and only if

$$\mathscr{F}_{\bar{x}} \bigcap \{w \in \mathbb{R}^m \ : \ \text{supp}(w) = \mathscr{U}\} = \emptyset. \tag{4.7}$$

See proof in (PEREIRA; SANTOS; STRIEN, 2023).

4.5 Divide-and-Conquer Network Reconstruction (DCNR)

Previous results can be used to construct an algorithm to reconstruct the connections inside a group of nodes such as motifs or clusters, instead of using all nodes in the library matrix. There are techniques to identify clusters in complex networks from data, including correlation

analysis (REVERTER; CHAN, 2008), Granger causality (LADROUE *et al.*, 2009), mutual information (BUTTE; KOHANE, 1999), causal entropy (SUN; BOLLT, 2014) or via dynamical systems approach (EROGLU *et al.*, 2020). However, identifying the connections within the clusters is a hard problem since one faces computational limitations and sensitive dependence on the network size.

Consider a clustered undirected network and assume that the nodes belonging to each cluster are known, but their connectivity structure is unknown. The algorithm attempts to find the intra-cluster connections among those nodes and uncover the *bridge nodes*, those nodes sharing connections outside their cluster.

Since we aim to reconstruct the neighbors of a node, note that the assumption (a) implies that there exists a set of indices $\mathscr{U}_f \subset [m]$ in which the isolated map $f \in \operatorname{span}\{\varphi_l \in \mathscr{L}_v : l \in \mathscr{U}_f\}$. So, by construction, the $\omega_r(s)$-sparse vector c_v of the probed node satisfies $\operatorname{supp}(c_v) \cap \mathscr{U}_f \neq \emptyset$. Besides, $\mathscr{U}_f \subset \mathscr{U}$ where $\mathscr{U}$ is the set of indices that identifies the subset of basis functions in $\mathscr{L}_v$ representing the dynamics of the probed node. Consequently, the hypothesis of Proposition 4.4.0.1 holds, $\mathscr{U} \cap \operatorname{supp}(c_v) \neq \emptyset$.

The Divide-and-Conquer Network Reconstruction (DCNR) algorithm initializes from an initial graph partition $\mathscr{P}_{[N]}$, whose each element P identifies a cluster of the graph and the zero coefficient matrix $C = 0 \in \mathbb{R}^{m \times N}$. Then, DCNR seeks whenever Equation (4.7) is violated:

I) for $P \in \mathscr{P}_{[N]}$, construct the subset of basis functions involving the coordinates of nodes in P. This induces a subset of indices $\mathscr{U} \subseteq [m]$, and consequently, a submatrix of $\Phi_v(X)$. Check:

   • if Equation (4.7) holds, classify the probed node as a bridge node of the cluster P.

   • Otherwise, recast the Ergodic Basis Pursuit (EBP) for node $i \in P$ and find the minimizer $u_i^\star$

$$\text{(EBP)} \quad u_i^\star = \underset{u \in \mathbb{R}^m}{\arg\min} \Big\{ \|u\|_1 \quad \text{subject to } \Phi_v(X)u = \bar{x}, \quad \operatorname{supp}(u) \subset \mathscr{U} \Big\}. \tag{4.8}$$

   Recast as the i-th column vector of the matrix C and construct the subgraph of node i inside the cluster P.

II) Use the resulting coefficient matrix C to factor out what is already known about the network structure, i.e., update the data matrix by $\bar{X} - \Phi_v(X)C$.

III) repeat I) for the set of bridge nodes and the updated data matrix $\bar{X} - \Phi_v(X)C$, see left panel of Figure 13 for an illustration.

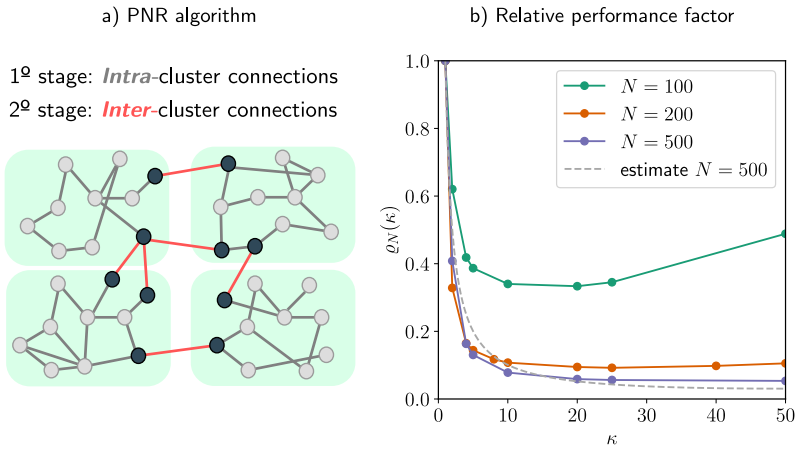

Figure 13 – **PNR algorithm speeds up the network reconstruction.** a) The PNR algorithm scheme contains two stages to find the global network structure. At first, the algorithm searches the intra-cluster connections, and then the remaining connections among the bridge nodes (in dark color). b) The relative performance factor $\rho_N(\kappa)$ for different clustered network of size κ in coupled optoelectronic units of Equation (4.10) with $\beta = 4.5, \alpha = 0.05, \theta = \frac{\pi}{4}$. The dashed line represents the solution in Equation (4.9) for $N = 500$ and $m(N) = 4N + 1$.

DCNR works because EBP uses $\mathscr{L}_V$. Once n_0 in Equation (4.5) is attained with $\delta(\Phi_V) < \sqrt{2} - 1$, EBP is guaranteed to have a unique $\omega_r(s)$-sparse solution for the noiseless case, and known estimates bounds for the noisy case (CANDÈS, 2008).

4.6 DCNR decreases the reconstruction computational time

The DCNR algorithm splits the reconstruction problem into smaller subproblems since fewer columns of $\Phi_V(X)$ are involved in the reconstruction, reducing the computational time and total length of time series to reconstruct the network. Let $m(N)$ be the number of columns of $\Phi_\mu(X)$ written as a function of N. Moreover, we assume a linear polynomial time complexity to solve the convex optimization problem for $m(N)$, see Appendix C.0.1. Suppose we aim to reconstruct an undirected network with κ clusters, where each cluster has equal size N/κ and one bridge node. The computational time of the DCNR algorithm τ_{DCNR} is:

$$\tau_{DCNR} = \kappa m\left(\frac{N}{\kappa}\right)\frac{N}{\kappa} + m(\kappa)\kappa,$$

where the first term is the computational time to check the intra-cluster connections: $m\left(\frac{N}{\kappa}\right)$ is the number of columns involved in the cluster reconstruction and N/κ is the number of nodes in a given cluster. The second term is the time to check the inter-cluster connections: since the network is undirected by assumption (b), step III) implies the time is reduced in searching connections exclusively among the κ bridge nodes. We calculate the relative performance factor

in terms of the number of clusters κ and network size N

$$\rho_N(\kappa) = \frac{\kappa m\left(\frac{N}{\kappa}\right)\frac{N}{\kappa} + \kappa m(\kappa)}{Nm(N)}, \tag{4.9}$$

where the denominator corresponds to the computation time when we include all nodes in the reconstruction.

To illustrate the performance of the PNR algorithm, we compute the relative performance factor numerically in different clustered networks. For each fixed network size N, we generate random networks with κ clusters. Each cluster corresponds to an ER network with N/κ nodes and mean degree 4 and has a unique bridge node. The set of bridge nodes forms an ER network with a mean degree 4.

We consider the coupled optoelectronic network modeled as

$$F_i(x(t)) = \beta I_\theta(x_i(t)) + \alpha \sum_{j=1}^{17} A_{ij}[I_\theta(x_j(t)) - I_\theta(x_i(t))] \bmod 2\pi, \quad i = 1,\ldots,N, \tag{4.10}$$

where the normalized intensity output of the Mach-Zehnder modulator is given by $I_\theta(x) = \sin^2(x+\theta)$, x represents the normalized voltage applied to the modulator, β is the feedback strength, and θ is the operating point set to $\frac{\pi}{4}$. We consider $\{x(t)\}_{t=0}^{n}$, for $N = 100$ using $n = 300$ while for $N = 200$ and $N = 500$ we use $n = 600$; and parameter values $\beta = 4.5$, $\alpha = 0.05$ and $\theta = \frac{\pi}{4}$. We represent it in terms of trigonometric functions up to second harmonics, $\mathscr{L} = \{1\} \cup \{\sin(kx_i), \cos(kx_i)\}_{k=1,i=1}^{2,N}$, so $m(N) = 4N+1$. We construct the product measure ν estimating from the data using Gaussian kernel density function. Orthonormalize $\mathscr{L}$ to obtain $\mathscr{L}_\nu$. The right panel of Figure 13 shows that the numerical relative performance factor and the expression in Equation (4.9) agree for $N = 500$.

Discussion and conclusions. We proposed the DCNR method for chaotic and sparse network dynamics. Our approach merges the ergodic theory of dynamical systems with sparse recovery to guarantee the network reconstruction once a minimum length of time series is attained. In principle, designed for clustered networks, DCNR checks for the connections inside the clusters and combines them into a global network structure.

Two relevant factors heavily influence DCNR: the initial graph partition $\mathscr{P}_{[N]}$, and the dependence of the numerical method to solve the EBP on quantities such as length of time series n and cardinality of the library m. For instance, assume the underlying graph is connected. So, if $\mathscr{P}_{[N]} = \{\{i\} : i \in [N]\}$ corresponds to the partition of singletons, the performance of solving DCNR is definitely worse than taking the entire graph at once since one performs N additional operations. We could use information theory to interpret how good is the selected partition, in particular the entropy of the partition. So, the problem is to maximize information for the reconstruction, which is measured by the entropy of the partition. Establishing a complete description of solving this variational problem of choosing the initial partition such that DCNR performs better is beyond the scope of this study and remains a promising subject for future research.

Here we discussed a simpler setting which gave us insights into the problem. We restrict our attention to the case regarding prior knowledge (or an expert) which describes that the network has community (clusters) structures. Regardless of the way to pre-process the data for identifying an initial partition, we obtain a proxy of the network structure. The DCNR algorithm improves this network estimate, allowing us to reconstruct the network from large experimental data sets. Although the global network structure can be obtained via the EBP as a single minimization problem, see Appendix C.0.2, for large networks this problem may become infeasible computationally due to the memory allocation limitations. Thus, our scheme makes the network reconstruction feasible in this scenario, saving processing demand and the search can be made in parallel.

CHAPTER

5

RECONSTRUCTION OF BURSTING NETWORKS

This chapter is devoted to presenting our work entitled *Reconstruction of bursting networks from data*, which is in preparation to be submitted.

Bursting dynamics is the alternation between trains of rapid spike-like oscillations followed by a moment of silence (RINZEL, 1987). While successful single neuron models capture bursting dynamics (IZHIKEVICH, 2000), bursting networks evolve in a high-dimensional and nonlinear system on top of an intricate connectivity structure (BORISYUK; COOKE, 2006; IZHIKEVICH, 2007). Bursting networks are found in different neuronal systems from cat primary visual cortex (NOWAK *et al.*, 2003), hippocampal regions (SANABRIA; SU; YAARI, 2001; BUZSáKI, 2015) to rodent trigeminal neurons (NEGRO *et al.*, 1998). There has been intense research to establish the influence of bursting on the onset of epilepsy seizures either in animals or humans (SANABRIA; SU; YAARI, 2001; HOFER *et al.*, 2022) (and references therein). Hence, understanding such bursting network dynamics opens the possibility to gain insights into the prediction of neurological disorders. Yet current technology only allows us to observe multivariate time series instead of the actual information about how neurons (nodes) evolve in time and interact among themselves (PARK; FRISTON, 2013). The challenge is to reverse engineer and extract the governing equations from data.

A great deal of effort has pushed forward sparse recovery methods to extract governing equations from data (BONGARD; LIPSON, 2007; SCHMIDT; LIPSON, 2009). Sparse recovery methods search for sparse representations of the input data and have been used to reconstruct different systems (NAPOLETANI; SAUER, 2008; WANG; LAI; GREBOGI, 2016; BRUNTON; PROCTOR; KUTZ, 2016). In neuronal networks, based on empirical observation, sparsity is a feature that permeates different neuronal systems (MASON; NICOLL; STRATFORD, 1991; WATERS; HELMCHEN, 2006; HE; CHEN; EVANS, 2007; GUZMAN *et al.*, 2016). For instance, the nematode *C.elegans*, which is one of the few nervous systems completely mapped

at the cellular scale, has a small connectivity density (about 10%) (SCHRÖTER; PAULSEN; BULLMORE, 2017). So, although the whole network is large (order of hundred to thousands of neurons), each node dynamics can be described by a sparse combination of basis functions. The connectivity structure has been successfully reconstructed using the basis pursuit method (BARRANCA; ZHOU, 2019; BARRANCA, 2023), but extracting the governing equations remains elusive.

In this paper, we utilize implicit-SINDy (MANGAN *et al.*, 2016b) to reconstruct bursting network dynamics modeled by coupled Rulkov maps in small motifs. We characterize the performance of implicit-SINDy in terms of length of time series and network size. We show that the minimum length of time series for a successful reconstruction scales exponentially with the network size. For sufficient enough data implicit-SINDy can be compared against the least square approximation. Although both methods have similar performance, the least square approximation does scales better to tackle large networks. In particular, we estimate the same amount of work to compute the reconstruction for both methods, $\mathcal{O}(N^6)$. While implicit-SINDy requires model selection, the least square approximation can be solved only once.

This paper is organized as follows. First, we introduce the dynamics of coupled Rulkov maps. Then, implicit-SINDy is reviewed in Section 5.2. Since implicit-SINDy searches sparse vectors in the kernel of a matrix, we study this kernel to deduce the minimum length of time series in Section 5.3.1. Section 5.4 compares the performance of implicit-SINDy against a formulation in terms of a least square approximation, which only assumes the definiteness of the denominator in the rational representation. In Section 5.5 using theoretical bounds developed in (Qu; Sun; Wright, 2016), we argue that although implicit-SINDy could be a viable option to tackle short time series regime, the computational cost is a bottleneck. Finally, in Section 5.6 we discuss potential directions to tackle large network structures.

5.1 Bursting dynamics

With the interest in investigating the effects of mutual synchronization and chaos regularization of bursting cells, Rulkov (RULKOV, 2001) considered the following two-dimensional map

$$\begin{aligned} f(u,v) &= \frac{\alpha}{1+u^2} + v \\ g(u,v) &= v - \beta(u+1), \end{aligned} \tag{5.1}$$

where u is the fast variable, and v is the slow variable due to the small magnitude of $\beta = 0.001$. The value of α determines the dynamical regime, either chaotic oscillations or chaotic spiking bursts. From here on, without loss of generality, $\alpha = 4.4$ is fixed and corresponds to a chaotic bursting regime. The electrical coupling function among cells is modeled as:

$$h_{ij}(x_i, x_j) = x_j - x_i, \tag{5.2}$$

which only occurs through the fast variables.

We consider the dynamics of N nodes given by

$$\begin{aligned} x_i(t+1) &= f_i(x_i(t), y_i(t)) + \frac{\lambda}{\Delta} \sum_{j=1}^{N} A_{ij}\, h_{ij}(x_i(t), x_j(t)), \\ y_i(t+1) &= g_i(x_i(t), y_i(t)), \end{aligned} \tag{5.3}$$

where the state of node $i \in [N] := 1, \ldots, N$ is $(x_i, y_i) \in M_i \subset \mathbb{R}^2$. The isolated map is the Rulkov map with components f_i and g_i in Equation (5.1) evaluated at (x_i, y_i). The coupling strength λ is divided by the maximum degree Δ to normalize the interaction for high-connected nodes. The adjacency matrix A defines who is connected to whom: $A_{ij} = 1$ if nodes i and j are connected and $A_{ij} = 0$ otherwise. Figure 14 b) displays the time series in the chaotic bursting regime of two nodes depicted in Figure 14 a), where we observe the slow-fast structure by $|y_i(t+1) - y_i(t)|$ being much less than one.

For the network structure, we consider motifs of neurologically relevant network structures. In particular, our main interest is a network featuring highly connected neurons, so-called *hub neurons*, that interconnect themselves in an integrating cluster. The phenomenon is named *rich-club* effect, where the hubs are densely connected to each other and the level of connectivity exceeds what would be expected by chance alone. Hence, we consider motifs that are building blocks of such rich-club networks. For instance, a reference network is the neuronal connectome of the *C.elegans* (WHITE *et al.*, 1986) that displays rich-club phenomenon (TOWLSON *et al.*, 2013).

From here on, the state of the network is denoted as

$$\mathbf{x} = \mathbf{vec}(x_1, y_1, \ldots, x_N, y_N) \in M^N = \prod_{i \in [N]} M_i \subset \mathbb{R}^{2N},$$

where **vec** corresponds to the vectorization by stacking the vectors (x_i, y_i) into a single column vector and write the network dynamics as

$$\begin{aligned} \mathbf{x}(t+1) &= F(\mathbf{x}(t)) \\ &:= (F_1(\mathbf{x}(t)), F_2(\mathbf{x}(t)), \ldots, F_N(\mathbf{x}(t)), F_{2N}(\mathbf{x}(t))) \end{aligned}$$

to make explicit the temporal dependence.

5.2 Reconstruction problem

Consider the multivariate time series $\{\mathbf{x}(t)\}_{t \geq 0}^{n}$ of the network dynamics introduced in Equation (5.3). We are interested to extract the governing equations of the network dynamics from the multivariate time series. Since the fast variable of the isolated map contains a rational function in Equation (5.1), it makes the reconstruction a non-trivial task. In fact, if one attempts

to reconstruct the dynamics using only polynomials (viewing as a Taylor expansion around zero), the resulting model only predicts the data inside certain vicinity of zero (EROGLU *et al.*, 2020; TOPAL; EROGLU, 2023), limiting its usage. So, we consider a class of network dynamics that can be represented by rational functions.

Consider the following assumptions:

(a). The time series of all nodes in the network are observed.

(b). Rational representation. The map F is given by

$$F_j(\mathbf{x}) = \frac{P_j(\mathbf{x})}{Q_j(\mathbf{x})}, \quad j \in [2N], \tag{5.4}$$

where the collection of numerators $\{P_j(\mathbf{x})\}_{j\in[2N]}$ and denominators $\{Q_j(\mathbf{x})\}_{j\in[N]}$ lie in the span of a library $\mathscr{L} = \{\phi_1, \phi_2, \ldots, \phi_m\}$ where $\phi_l \colon M^N \to \mathbb{R}$. We consider the polynomials of two variables with degree at most r

$$\begin{aligned}\mathscr{L} = \{1\} &\cup \{x_i^p\}_{i,p} \cup \{y_i^p\}_{i,p} \\ &\cup \{x_i^p x_j^q\}_{i,j,p,q} \cup \{x_i^p y_j^q\}_{i,j,p,q} \cup \{y_i^p y_j^q\}_{i,j,p,q},\end{aligned} \tag{5.5}$$

where $i, j \in [N]$ with $i \neq j$ and we remove any redundancy, $p \in [r], q \in [r-1]$, and $p+q \leq r$. The cardinality of $\mathscr{L}$ is given by $m = \binom{2N}{2}\binom{r}{2} + 2Nr + 1$.

The representation of F as in Equation (5.4) may allow for rational functions. In particular, we are interested in the class of rational functions that contains the x-coordinate of the Rulkov map in Equation (5.1). We fix maximum degree $r = 3$ because that is the least degree to represent coupled Rulkov maps in the rational representation in Equation (5.4). In fact, if we need to reconstruct the fast coordinate of the isolated map in Equation (5.1), we recast as

$$f(u,v) = \frac{\alpha + v(1+u^2)}{1+u^2}.$$

Note that automatically there is the term vu^2. The same observation can be done to network dynamics in Equation (5.3). Although we fix throughout all simulations $r = 3$, if the coupling function h had polynomials of the form $x_i^p x_j^q$, this would require higher degree polynomials.

(c). Sparse representation. Moreover, real-world neuronal networks are sparse (MASON; NICOLL; STRATFORD, 1991; WATERS; HELMCHEN, 2006; HE; CHEN; EVANS, 2007; GUZMAN *et al.*, 2016), then we assume that the network structure, which is encoded by the adjacency matrix A in (5.3), is sparse, i.e., out of all possible connections only a few are realized. So, the network dynamics F has a sparse representation in $\mathscr{L}$. Specifically, for each component $j \in [2N]$, $P_j = \sum_l a_j^l \phi_l$ and $Q_j = \sum_l b_j^l \phi_l$, where $a_j = (a_j^1, \ldots, a_j^m), b_j = (b_j^1, \ldots, b_j^m) \in \mathbb{R}^m$ are sparse vectors, that is, most of their entries are zero. From here on, we denote $c_j = \mathbf{vec}(a_j, b_j) \in \mathbb{R}^{2m}$.

We formulate the reconstruction problem in terms of solving a linear equation. Consider the library matrix

$$\Phi(X) = \begin{pmatrix} \phi_1(\mathbf{x}(0)) & \phi_2(\mathbf{x}(0)) & \cdots & \phi_m(\mathbf{x}(0)) \\ \phi_1(\mathbf{x}(1)) & \phi_2(\mathbf{x}(1)) & \cdots & \phi_m(\mathbf{x}(1)) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}(n-1)) & \phi_2(\mathbf{x}(n-1)) & \cdots & \phi_m(\mathbf{x}(n-1)) \end{pmatrix}, \tag{5.6}$$

where each column is an element of $\mathscr{L}$ evaluated along the network trajectory. Let the network trajectory be recast as

$$\bar{X} = \begin{pmatrix} x_1(1) & y_1(1) & \cdots & x_N(1) & y_N(1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1(n) & y_1(n) & \cdots & x_N(n) & y_N(n) \end{pmatrix} \tag{5.7}$$

where each column is denoted $\bar{\mathbf{x}}_j$ with $j \in [2N]$. Also, let $D_j = \text{diag}(\bar{\mathbf{x}}_j)$ be the diagonal matrix constructed from the vector $\bar{\mathbf{x}}_j$. Then, define the augmented library matrix

$$\Psi(\bar{\mathbf{x}}_j) := \begin{bmatrix} \Phi(X) & \Phi(X)D_j \end{bmatrix} \in \mathbb{R}^{n \times 2m}, \quad j = 1, \ldots, 2N. \tag{5.8}$$

Rewriting Equation (5.4) we obtain

$$P_j(\mathbf{x}) - F_j(\mathbf{x})Q_j(\mathbf{x}) = 0 \tag{5.9}$$

for each $j \in [2N]$. By assumption (c), P_j and Q_j can be written as a linear combination of polynomials in $\mathscr{L}$, then Equation (5.9) can be rewritten in terms of the augmented library matrix as

$$\Psi(\bar{\mathbf{x}}_j)c_j = 0, \quad j = 1, \ldots, 2N. \tag{5.10}$$

Hence, to reconstruct the network dynamics the sparse vectors $c_1, \ldots, c_{2N}$ that satisfy Equation (5.10) must be found. Since each c_j lies on the kernel of the augmented library matrix

$$\ker(\Psi(\bar{\mathbf{x}}_j)) = \{u \in \mathbb{R}^{2m} : \Psi(\bar{\mathbf{x}}_j)u = 0\},$$

implicit-SINDy is devoted to the following approach: to search the non-zero sparsest vector in the null space of $\Psi(\bar{\mathbf{x}}_j)$ (MANGAN *et al.*, 2016b). A priori, this problem could be formulated in terms of a ℓ_0 minimization problem, but it is intractable computationally (Qu; Sun; Wright, 2016). So the alternative formulation is a ℓ_1 minimization problem that is constrained to search solutions in the unit sphere, due to invariance by scaling of Equation (5.10).

Implicit-SINDy implements the so-called alternating direction method (ADM) algorithm proposed by Qu and co-authors (Qu; Sun; Wright, 2016). The main idea is to find a linear combination of the kernel's orthonormal basis such that the coefficients lie on the unit sphere and the spanned vector corresponds to the sparsest vector in the kernel, see Appendix D.1 for

details. More precisely, let $\Theta_j \in \mathbb{R}^{2m\times d}$ be the orthonormal basis of the kernel of $\Psi(\bar{\mathbf{x}}_j)$, where d is the dimension of $\ker(\Psi(\bar{\mathbf{x}}_j))$. The algorithm iterates the following:

$$\begin{aligned} u_j^{k+1} &= \text{soft}(\Theta_j w_j^k, \gamma), \\ w_j^{k+1} &= \frac{\Theta_j^T w_j^{k+1}}{\|\Theta_j^T w_j^{k+1}\|_2}, \end{aligned} \tag{5.11}$$

where $\text{soft}(u,\gamma) = \text{sign}(u)\max\{|u|-\gamma, 0\}$ is the soft-thresholding operator applied component-wise, T corresponds to the transpose and w_j^0 is initialized using normalized rows of Θ_j. Here we adopt a maximum number of iterations as 10000, see algorithm in 4.

The parameter γ penalizes the large entries of the vector on the null space. Equation (5.11) depends on γ, and consequently, the resulting sparse solution. Implicit-SINDy solves Equation (5.11) for different values of γ, and then performs a Pareto front to choose the γ that corresponds to the most parsimonious model. In particular, Pareto front selects the solution $u^\star$ that minimizes the error in satisfying Equation (5.10) and also has the least number of non-zero entries, see Appendix D.2 for details. Overall, implicit-SINDy formulates the network dynamics reconstruction for each node in the network: iterating (5.11) for each γ, and then finding the best γ parameters via a Pareto front.

5.2.1 Example: two-nodes motif

To illustrate the reconstruction of bursting network dynamics, let us consider two coupled nodes as depicted in Figure 14 a). We consider 5000 time steps as transients, so the network can attain the attractor, and iterate the system during n time steps. We repeat the same scheme for all motifs tested in this paper.

For $n = 200$, we run implicit-SINDy for each coordinate $j \in [2N]$ and obtain the resulting coefficient vector, which is plugged into Equation (5.10) to compare against the true model. Using symbolic language programming (MEURER *et al.*, 2017), we build the reconstructed model $\hat{F}$. Figure 14 c) displays the comparison between the isolated map in Equation (5.1) and the reconstructed model of the node map projected to its coordinate, i.e., $\hat{F}(0,0,\ldots,x_i,y_i,\ldots,0,0)$. We notice that both isolated maps are in agreement, showing that the reconstruction is successful. In Equation (5.12), we show the reconstructed model expression, where the isolated map, coupling function, and parameters such as α, β, and coupling strength δ/Δ are successfully recovered. The reconstructed model for the two nodes motif in Figure 14 c) is given below

$$\begin{aligned} \bar{x}_1 &= \frac{-0.010000184x_1^3 + 0.010000184x_1^2x_2 + 1.0x_1^2y_1 - 0.010000184x_1 + 0.010000184x_2 + 1.0y_1 + 4.4000005}{x_1^2 + 1.0} \\ \bar{y}_1 &= -0.0010000908x_1 + 1.0y_1 - 0.0010000908 \\ \bar{x}_2 &= \frac{0.010000184x_1x_2^2 + 0.010000184x_1 - 0.010000184x_2^3 + 1.0x_2^2y_2 - 0.010000184x_2 + 1.0y_2 + 4.4000005}{x_2^2 + 1.0} \\ \bar{y}_2 &= -0.0010000908x_2 + 1.0y_2 - 0.0010000908 \end{aligned} \tag{5.12}$$

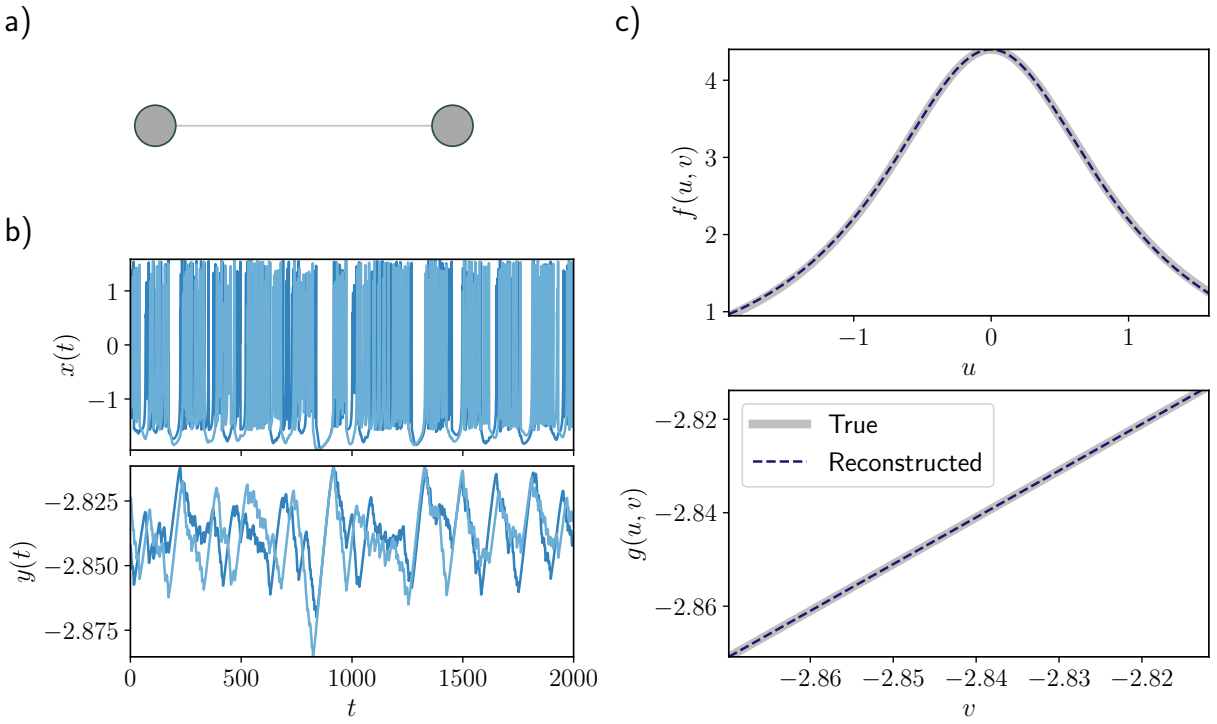

Figure 14 – **Reconstruction of a two-nodes motif.** a) Two nodes motif. b) Network time series where each color represents a node. The top panel displays the bursting dynamics of the fast variable while the bottom panel shows the slow dynamics. c) Reconstruction of the isolated map (5.1). The top panel is the isolated map projected in the u axis and the bottom panel is in the v axis. The solid line (in gray) represents the true expression and the dashed line (in blue) represents the reconstructed model. The network dynamics parameters are $\lambda = 0.01$ and $\Delta = 1$. The length of the time series is $n = 200$ and transient time is 5000 time steps.

To recover successfully the original network dynamics, we must perform a factorization in the numerator and denominator. The factorization can be done even by inspection for small networks, but it becomes intractable for large networks.

5.3 Trade-off: uniqueness versus performance

The above example demonstrates that the implicit-SINDy successfully reconstructs the governing equations for both fast and slow variables. Note that when the governing equations contain rational terms, by construction, if the time series is sufficiently long, then the true coefficient vector satisfying Equation (5.9) is the unique non-zero vector lying on the null space of $\Psi(\bar{\mathbf{x}}_j)$. Implicit-SINDy finds this vector, which is sparse by **Assumption (c)**, and exactly reconstructs the governing equations. Although for two nodes this is easily satisfied, we demonstrate that for larger motifs there is a trade-off: either implicit-SINDy requires a long time series for exact reconstruction or the reconstruction can be computationally expensive.

5.3.1 Mininum length of time series

The reconstruction heavily depends on the length of time series n. The natural question is: how long should be the time series to obtain a unique vector in the null space? We address to determine the minimum length of time series n_0 such that we have a successful reconstruction. This problem can be investigated in terms of the dimension of the kernel of $\Psi(\bar{\mathbf{x}})$, establishing the number of sparse vectors that populate the kernel of $\Psi(\bar{\mathbf{x}})$ as the length of time series n and network size N are varied.

Let us consider the augmented matrix $\Psi(\bar{\mathbf{x}}) \in \mathbb{R}^{n\times 2m}$. Let $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ be the dimension of $\ker(\Psi(\bar{\mathbf{x}}))$, also called defect or nullity (BERNSTEIN, 2009). To obtain the dependence of $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ with respect to the length of time series n, we consider two cases separately.

Fat matrix ($n < 2m$). $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ has a linear decay envelope as we increase n. By the dimension theorem (rank-nullity theorem) and the inequality $\mathrm{rank}(\Psi(\bar{\mathbf{x}})) \leq \min\{n, 2m\}$ (BERNSTEIN, 2009) the following holds:

$$2m \geq \mathrm{def}(\Psi(\bar{\mathbf{x}})) \geq 2m - n. \tag{5.13}$$

Tall matrix ($n \geq 2m$). $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ decays nonlinearly as n is increased. The defect is estimated using Singular Value Decomposition (SVD). SVD calculates the dimension of the orthonormal basis for the null space in terms of the spectra $\sigma(\Psi(\bar{\mathbf{x}}))$, selecting those orthonormal vectors associated with eigenvalues that are smaller than a given threshold. Since each $\Psi(\bar{\mathbf{x}})$ column is a vector whose entry corresponds to a ϕ_l (observable) evaluated at the network trajectory, spurious linear dependences may occur between columns of $\Psi(\bar{\mathbf{x}})$, and consequently, a nonlinear dependence on n. After transient, by the construction of $\Psi(\bar{\mathbf{x}})$ in Equation (5.8), we can deduce that $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ approaches 1 for the fast variables, which contain the rational term. This implies that as we increase n, eventually $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ converges to the constant 1, so from here on, we define

Definition 5.3.1. The minimum length of time series n_0 is the length of time series when $\mathrm{def}(\Psi(\bar{\mathbf{x}})) = 1$.

Figure 15 shows how $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ depends on the length of time series n for three distinct network structures with 10 nodes. At first, $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ decays linearly for both fast and slow variables, as shown by the agreement with the linear equation $n \mapsto 2m - n$ corresponding to the dashed orange curve. Then, once n is close to $2m$, which in this case is $2m = 1261$, the curves split into two. After the splitting point, $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ stabilizes in a plateau for the slow variables. The isolated map for the slow variable is not a rational function, so $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ is larger than one. On the other hand, for the fast variables, $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ decays nonlinearly, until it converges to 1 which is represented by the black dashed line. Although the network structures are disparate to each other, the curves approach 1 similarly for all cases over different initial conditions, showing that the curve behavior is not strongly related to the network structure itself.

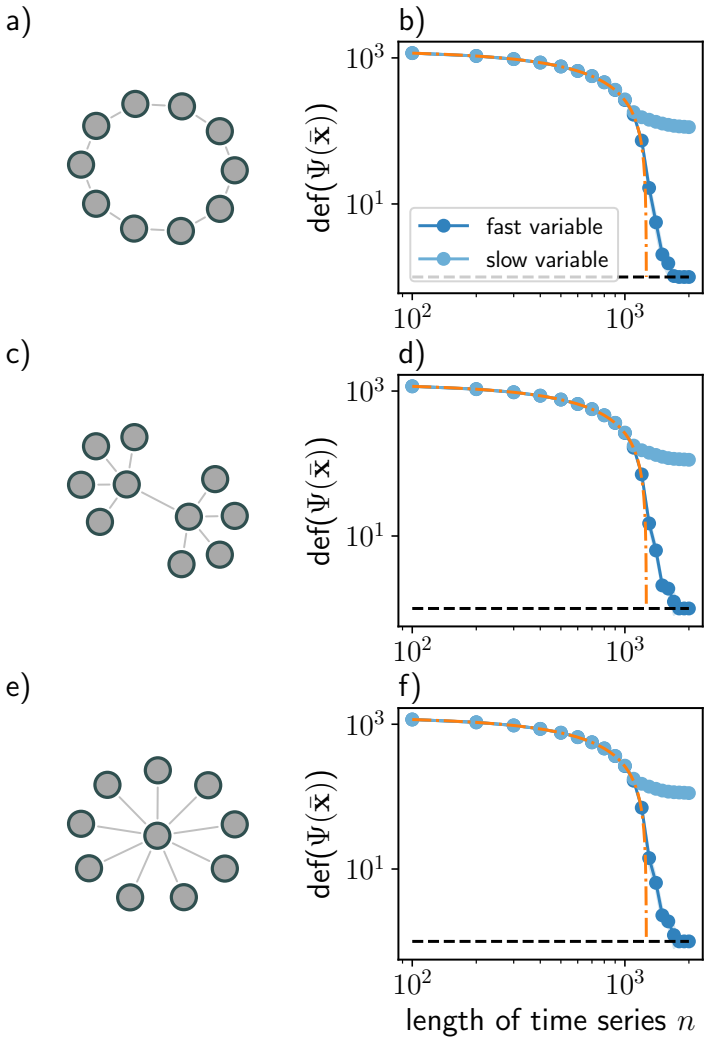

Figure 15 – **Dependence of defect of the augmented matrix for different network structures.** a) Ring graph with 10 nodes. b) Defect of $\Psi(\bar{\mathbf{x}})$ as we increase the length of time series for fast (dark blue) and slow (light blue) variables. Orange and black dashed lines correspond to the linear equation $n \mapsto 2m - n$ and $\text{def}(\Psi(\bar{\mathbf{x}})) = 1$, respectively. c) Motif with 10 nodes with two coupled hubs and e) Star motif with 10 nodes. d) and f) are similar to b). For all networks the curve behavior is similar. When the length of time series is smaller than the number of basis functions $2m$, the defect of $\Psi(\bar{\mathbf{x}})$ decays linearly. Once the length of the time series equals $2m$, there is a splitting. For the fast variable, the curve converges to 1 as opposed to the slow variable, which decays much slower than the fast variable. The dots correspond to the average over 10 initial conditions after we group all nodes' variables of the same type (fast or slow), while the shaded area is the standard deviation.

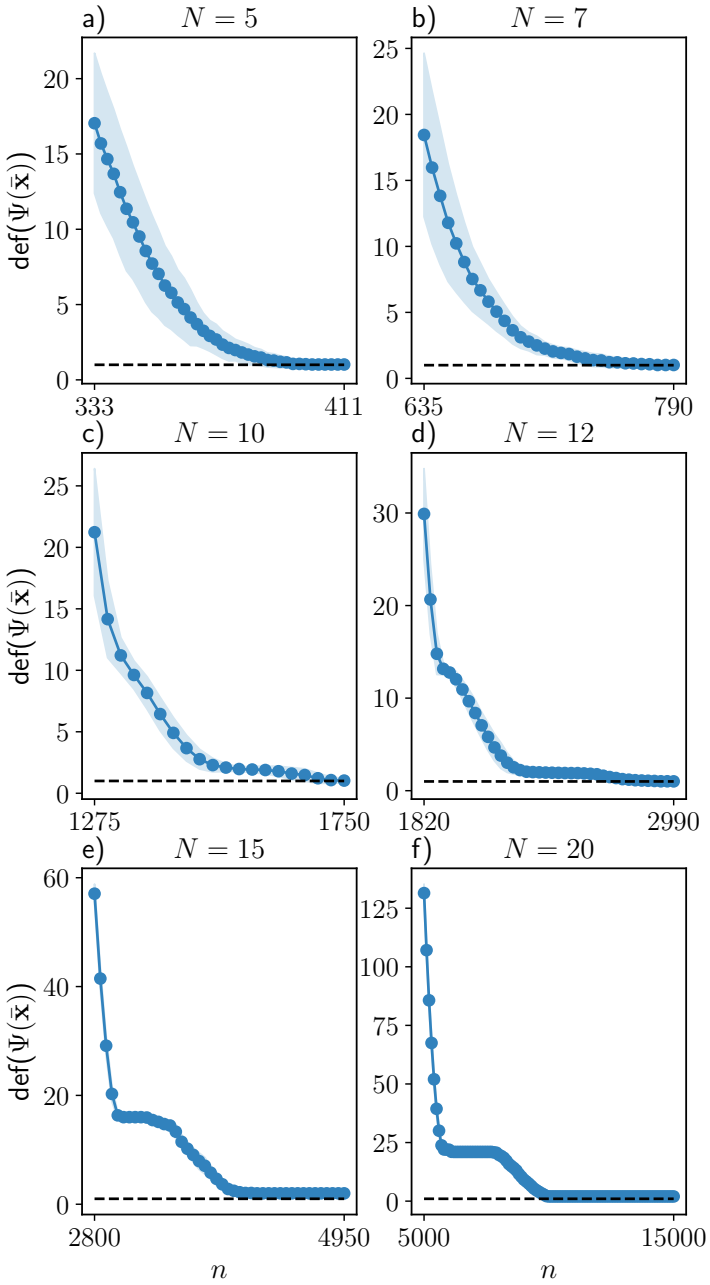

Figure 16 – **Dependence of defect of the augmented matrix for star graph with different network sizes.** a) - d) The dots correspond to the average over 10 initial conditions after we group all nodes' variables of the same type (fast or slow), while the shaded area is the standard deviation. For $N = 15$ and $N = 20$ the dots correspond to one realization.

Nevertheless, the nonlinear dependence depends significantly on the network size. For fast variables, fixing the star graph as the reference network, Figure 16 displays the nonlinear dependence of $\mathrm{def}(\Psi(\bar{\mathbf{x}}))$ with respect to the length of the time series under different network sizes. Note that the variation with respect to initial conditions decreases for larger networks, as shown in Figure 16 a) - d). All curves are shown in the interval when the length of the time series is larger than the number of $\Psi(\bar{\mathbf{x}})$ columns, i.e., the interval after the orange dashed curve in Figure 15. Although the curves have different profiles, the defect slowly decays with network size, consequently, the minimum length of time series n_0 increases with network size. In particular, for $N = 15$ and $N = 20$, the defect reaches 2 instead of 1 in the chosen interval. This shows that the spurious numerical linear dependencies among columns of $\Psi(\bar{\mathbf{x}})$ can be long-lasting as larger is the network, requiring large n_0.

The defect of $\Psi(\bar{\mathbf{x}})$ can also be used to estimate the computational cost of computing the solutions via the ADM algorithm as part of the implicit-SINDy method. Qu and co-authors (Qu; Sun; Wright, 2016) were able to establish conditions for successful reconstruction of the sparsest solution on the $\ker(\Psi(\bar{\mathbf{x}}))$ via ADM algorithm in terms of the defect of $\Psi(\bar{\mathbf{x}})$. In particular, under a specialized setting called the planted sparse model, which is defined as when $\ker(\Psi(\bar{\mathbf{x}}))$ contains the sparsest solution and a sequence of linearly independent random vectors filling the subspace, the algorithm requires a lower bound condition on the number of columns of $\Psi(\bar{\mathbf{x}})$: $2m \geq d^4 \log d$, where $d = \mathrm{def}(\Psi(\bar{\mathbf{x}}))$ is the short notation for defect of $\Psi(\bar{\mathbf{x}})$. Moreover, the formulation of the ADM algorithm (Qu; Sun; Wright, 2016) requires that Equation (5.11) should be iterated by $d^4 \log d$ times. Consequently, the ADM algorithm spends as many iterations to converge to a solution as larger the network size. By Equation (5.5) the number of basis functions $m(N,r)$ in the library $\mathscr{L}$ scales as $m = \mathscr{O}(N^2 r^2)$ where N is the network size and r is the maximum degree. The ADM algorithm may take a number of iterations that scales as $(2N^2r^2 - n)^4 \log(2N^2r^2 - n)$, i.e., it has leading order growing as $\mathscr{O}(N^8 r^8)$. For this reason, implicit-SINDy consumes a long computational time and may fail to reconstruct the slow variable dynamics, which does not have any rational term.

Figure 17 displays the dependence of the pair length of time series n and network size such that $\mathrm{def}(\Psi(\bar{\mathbf{x}})) = 15$. For smaller network sizes, $N < 10$, the length of the time series grows quadratically, following the number of columns of $\Psi(\bar{\mathbf{x}})$ with N, represented by the red solid line, $n = 2m$. Then, the curve changes growth behavior, scaling exponentially with network size, as confirmed by the fitting (blue dashed line). The inset in Figure 17 shows the dependence (n,N) for different defect d values, where as expected, smaller defect values require a large length of the time series. In particular, the smallest defect represents the behavior of the minimum length of time series n_0. The regime of length of time series $[2m, n_0)$, where the number of rows of $\Psi(\bar{\mathbf{x}})$ is larger than the number of columns, Equation (5.10) becomes overdetermined, but the defect is still larger than one for the fast variables. Finding the solution may be a computational issue, and other methods are available to compete with the ADM method to reconstruct the network dynamics such as least square approximation, as we will show in Section D.2.1.

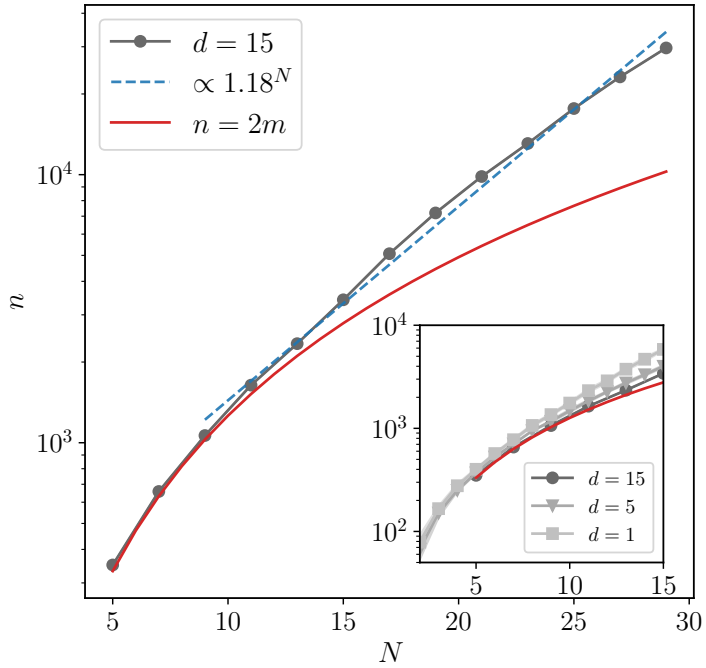

Figure 17 – **The length of time series grows exponentially with network size.** The length of the time series versus the network size for $d = 15$, where dots correspond to one initial condition. The red solid line corresponds to the case $n = 2m$. The blue (dashed) line corresponds to the regression of a linear fitting. The inset shows the same curve for different values of d. The dots correspond to an average over 10 initial conditions while the shaded area is the standard deviation. The maximum degree is $r = 3$. The coupling strength is $\lambda = 0.01$ and maximum degree $\Delta = N$.

Figure 17 illustrates the trade-off for reconstructing large networks: either the length of the time series should be sufficiently large, or the ADM method requires a sufficiently long computational time to compute solutions. Note that this performance estimate captures the reconstruction of a sparse solution for a given γ parameter, but it does not take into account probing different γ parameters, which also requires further computational time.

5.4 Reconstruction performance

For motifs that have more than 2 nodes, the interpretability of the reconstructed model becomes challenging. The expressions of the reconstructed model grow prohibitively large, making it difficult to interpret them by inspection. To quantify the reconstruction performance we introduce an error function for each node that evaluates one step prediction in fitting the trajectory of that node, and then average over evenly spaced locations within a certain period τ. In this way, this error function quantifies successfully whether the selected model captures the node dynamics in the network.

More precisely, let $\{\mathbf{x}(t)\}_{t\geq 0}^{n}$ be the trajectory of the network dynamics. And let $\hat{F}$ be the

$\bar{E}_1$	3.42×10^{-7}
$\bar{E}_2$	6.72×10^{-8}
$\bar{E}_3$	2.78×10^{-7}
$\bar{E}_4$	9.15×10^{-8}

Table 1 – Reconstruction performance of two nodes network for $n = 200$, $\tau = 2000$.

reconstructed model. Consider at time t

$$E_i(t) := |\mathbf{x}_i(t+1) - \hat{F}_i(\mathbf{x}_i(t))|, \quad i = 1, \ldots, 2N,$$

that is one step evaluation of the reconstructed map at the trajectory. Then, we evaluate the performance of the reconstructed model in 50 evenly spaced locations within a period τ, i.e., we consider the following error function

$$\bar{E}_i = \Big(\frac{1}{50}\sum_{k=1}^{50} E_i^2(t_k)\Big)^{\frac{1}{2}}, \quad t_k \in [0, \tau].$$

To illustrate that $\bar{E}_i$ captures a successful reconstruction, we calculate it for the coupled two-node network in Figure 14, see Table 1. We can observe that the order of magnitude of $\bar{E}_i$ for each node matches the precision on the coefficients obtained in the reconstructed model in Equation (5.12) in Section 5.2.1.

Because of Figure 17 regarding the reconstruction of large networks, we make two choices: we consider sufficient long time series, otherwise, the ADM algorithm takes a long time to converge to a solution. Moreover, since the loss function depends on the reconstruction of all nodes and the dynamics of the slow variables ($y-$coordinates) are costly to compute, to reconstruct the slow variables we use:

$$\min_{u \in \mathbb{R}^m} \|\Phi(X)u - \bar{x}\|_2,$$

whose solution is given by $u^* = \Phi(X)^+ \bar{x}$, where $\Phi(X)^+ := (\Phi(X)^T\Phi(X))^{-1}\Phi(X)^T$ is the Moore-Penrose pseudo-inverse. Note that using only the library matrix $\Phi(X)$ is enough because the slow variable does not contain the rational term.

To check implicit-SINDy performance, we compare it with another method. We introduce a least square formulation of the problem, see D.2.1 for details. We evaluate the performance of the reconstruction method for small motifs as we vary the length of the time series. Figure 18 b) shows the comparison between the implicit-SINDy and least square approximation to reconstruct the five node motif in Figure 18 a). The curve decays two orders of magnitude before reaching $n = 500$, then it attains a plateau where both reconstruction methods correctly identify the underlying network model. The ℓ_2 method outperforms implicit-SINDy using a single formula and does not require model selection.

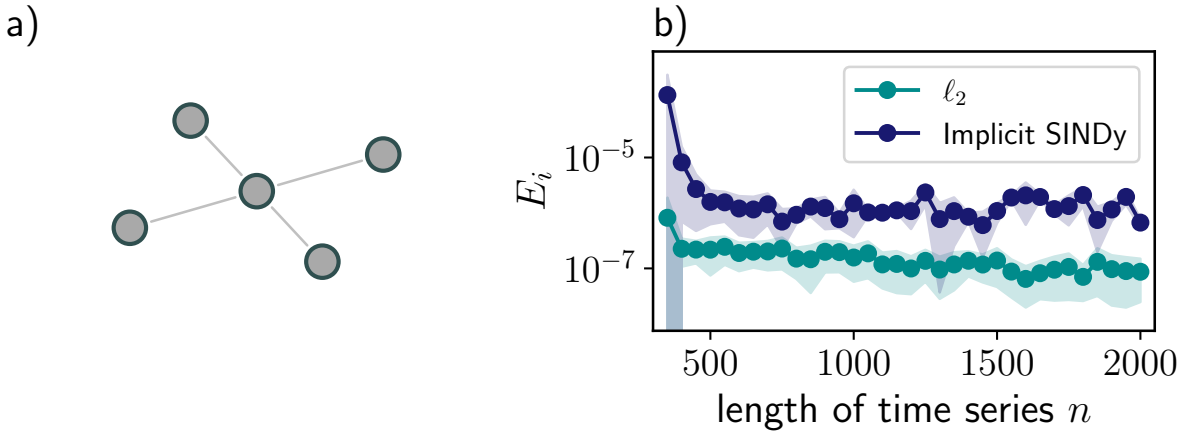

Figure 18 – **Comparison of reconstruction methods for varying length of time series.** a) Star motif with five nodes. b) Error function for least square approximation and implicit-SINDy. Dots corresponds to the average over 10 initial conditions and grouping all nodes' variables of the same type (fast or slow), while shaded area is the standard deviation. The coupling strength $\lambda = 0.01$ and $\Delta = 5$.

5.5 Which method can scale better to large networks?

As observed in Section 5.3, the performance of implicit-SINDy depends on the length of the time series and network size. But we also can deduce rough estimates of the total amount of work that implicit-SINDy and ℓ_2 methods require to reconstruct the network dynamics with respect to network size. The main interesting question is: which method does scale better to large networks?

ADM algorithm uses SVD to determine the defect of $\Psi(\bar{\mathbf{x}})$. To compute SVD the total amount of work (measured in flop counts) (GOLUB; LOAN, 1996) scales as $4nm^2 + 8m^3$. Since $m = \mathcal{O}(N^2 r^2)$, the amount of work scales as $\mathcal{O}(N^6 r^6)$. Using the formula ℓ_2 method also uses SVD, having the same performance. We could also estimate in terms of solving normal equations (GOLUB; LOAN, 1996) that requires $nm^2 + m^3/3$ to find a solution, consequently, also scales as $\mathcal{O}(N^6 r^6)$.

In the end, assuming that the length of the time series n is sufficiently large, at least $d = 15$ in Figure 17 we can roughly estimate how much work the two reconstruction methods, implicit-SINDy and ℓ_2, take to reconstruct a node dynamics. Implicit-SINDy requires $\mathcal{O}(N^6 r^6)$ times the number of γ parameters tested during the Pareto front. As opposed to the least square approximation that does not need the Pareto front. When comparing both of them ℓ_2 stands as a better choice for large networks.

5.6 Discussion and conclusions

Overall we numerically investigated the performance of implicit-SINDy for extracting governing equations of network dynamics. In particular, we evaluated the performance in terms

of the length of time series and network size. Implicit-SINDy requires at least a minimum amount of time series to reconstruct that scales exponentially with network size. When compared to least square approximation, least square approximation has better performance in all tests, being a better choice to tackle large networks. In particular, it can be solved using a single formula and does not require model selection.

There has been progress in developing methods to reconstruct rational functions from data using a mean-field reduction approach for chaotic dynamics (TOPAL; EROGLU, 2023), two-phase algorithm (GAO; YAN, 2022), or SINDy variants (ZHANG; LIN, 2018; KAHEMAN; KUTZ; BRUNTON, 2020), but none of these methods use advantageously the slow-fast structure of the dynamics. For instance, we showed that implicit-SINDy is not well adapted to the slow variable, due to the absence of rational terms. So, this poses an interesting research direction to be tackled in future research. We also point out that large networks require smarter reconstruction strategies. There are occasions when although the network structure is unknown, it is not completely unknown. For instance, in neuronal systems, a motif structure might be known such as in rich club motifs (REUS; HEUVEL, 2013; TOWLSON *et al.*, 2013), where a group of nodes form a highly connected cluster mediating information to the rest of the network. Then to reconstruct a large network one could incorporate such expert knowledge on the structure into the reconstruction method. Using the expert knowledge we could break the reconstruction into subproblems and solve reconstruction subproblems locally. Subsequently, combining information gathered for each local solution, the global solution is built. Another potential research direction is multidimensional regression using tensor representation (GELß *et al.*, 2019).

Data Availability Statement. The code to produce the data that supports the findings of this study is available in the rulkov-repository (SANTOS, a).

CHAPTER

6

CONCLUSIONS

Over recent years, significant progress has been made in the extraction of governing equations for complex network dynamics from data. This advancement opens up possibilities for predicting critical transitions, especially in large-dimensional systems. Leveraging the sparsity feature commonly found in real networks, sparse recovery methods emerge as potential options to recover these governing equations. Sparse recovery methods search for a sparse representation of the data, making them well-suited for reconstructing the network dynamics when the length of available time series data is much shorter than the network size. In this thesis, we explore the synergy of dynamical systems theory and ergodic theory with sparse recovery methods to obtain reconstruction guarantees.

Reconstruction of sparse networks requires a minimum length of time series. In situations where time series data is limited, the reconstruction problem can formulated as an underdetermined linear system, and sparse solutions become unique under certain conditions, such as when the matrix satisfies RIP (Restricted Isometry Property). Using the decay of correlations of the network dynamics, pairs of columns of the associated matrix become linear independent of each other once a minimum length of time series is attained, see Theorem 1. Our study evaluates the performance of EBP concerning time series length and network size, highlighting its robustness against noise and its applicability to experimental data.

Divide-and-conquer approach enabled by uniqueness. Exploiting the unique reconstruction of the EBP, we explore the reconstruction of clustered networks. We address the question of the best strategy for reconstructing network structures: dividing them into clusters or solving the entire network at once. The choice depends on the number of clusters and number of nodes within each cluster. We delved into the specifics of this dependency, demonstrating the reconstruction can be sped up tremendously, see 4.6. Moreover, least square approximation fails in implementing this divide-and-conquer approach within the regime of the short length of time series, see Section 3.4.1.

Sparse recovery methods are suitable for bursting network reconstruction. We also investigate the suitability of sparse recovery methods for reconstructing bursting network dynamics, particularly employing implicit-SINDy. Implicit-SINDy requires a minimum length of time series that scales quadratically with the network size to reconstruct the network dynamics. When compared to the least square approximation, both had similar performance in the reconstruction, although the latter offers computational efficiency.

6.1 Open problems and potential research directions

1. **Partial Network Observation.** Throughout this thesis the main assumption was that the multivariate time series of all nodes' states in the network were available. However, typically this is unrealistic.

   Problem. Assume that there are observations given by an observable $\psi : M \to \mathbb{R}$

   $$y_i(t) = \psi(x_i(t)),$$

   where ψ is a projection to a variable on which nodes interact to each other.

   Problem. More generally, assume that there are observations given by $\psi_k : M^N \to \mathbb{R}$

   $$y_k(t) = \psi_k(x(t)).$$

   Consider the case that $k < N$. In this scenario, instead of observing all nodes, the network is partially observed.

2. **Extension to maps on the Torus.** Although we introduced and demonstrated that EBP holds for exponential mixing network dynamics, we have not proved that is valid for an example. Moreover, all examples we have studied consist of isolated maps on a subset of the real line. In view of standard and recent results of statistical properties of hyperbolic systems on the torus, we consider

   Problem. Formulate a reconstruction problem for maps on $\mathbb{T}^N$. A priori, this can be challenging. In fact, consider the expanding map on the circle $x \mapsto 3x \mod 1$. The return map trajectory of this map is discontinuous in the real line, and any polynomial basis is not capable of representing the dynamics.

   Once the reconstruction problem can be formulated, we envision that skew-product systems are potential examples such as torus extensions over expanding maps (CHEN; HU, 2018). In terms of network terminology, torus extensions are directed star graphs, where the hub receives connections from the leaves. Moreover, uniformly expanding maps are potential examples to use a key step in our proof (TANZI, 2022): approximate the physical measure using a product measure.

3. **Robust Noise Magnitude Identification.** The Relaxing path algorithm A.1 searches the noise magnitude tuning the ε parameter to identify the correct structure. However, the current strategy is fragile. It heavily depends on the interval $\mathscr{E}$ and how each probed parameter is spaced to each other, i.e., which grid we use to probe the parameters.

   Problem. Formulate a more robust way to identify the noise magnitude in the reconstruction of noisy measurements.

4. **Cluster Detection.** DCNR algorithm depends on the initial partition. We focused on the reconstruction of clustered networks. We assumed the clusters are given a priori, but that is not general.

   Problem. Pre-process the data to find the cluster information. Consider a correlation analysis of the time series to find a starting initial partition for DCNR. In particular, Pearson correlation of the multivariate time series of all nodes in the network. DCNR will be responsible for identifying the connections inside the clusters and among clusters.

5. **Incorporating Slow-Fast Structures.** We have used the rational information of the coupled Rulkov maps to reconstruct the bursting network dynamics. However, the slow variable is challenging to be recovered using this method, because it does not contain the rational term. The idea is to exploit the slow-fast structure advantageously. Inspired by the idea of the critical manifold from Geometric Singular Perturbation Theory (FENICHEL, 1979), consider:

   Problem. Assume that the slow variable is a parameter for chunks of data within short time intervals and solve the reconstruction for the fast variable. Here, we reconstruct the fast dynamics on the critical manifold. Then, assuming that the fast dynamics is known, incorporate it in the reconstruction of the slow dynamics akin to the approach described in (LUCHINSKY *et al.*, 2008).

BIBLIOGRAPHY

ABSIL, P.-A.; EDELMAN, A.; KOEV, P. On the largest principal angle between random subspaces. **Linear Algebra and its applications**, Elsevier, v. 414, n. 1, p. 288–294, 2006. Citation on page 84.

AGRAWAL, A.; VERSCHUEREN, R.; DIAMOND, S.; BOYD, S. A rewriting system for convex optimization problems. **Journal of Control and Decision**, v. 5, n. 1, p. 42–60, 2018. Citation on page 34.

BAKER, G. A.; BAKER, G. A. G. A. **Padé Approximants Second Edition**. Cambridge: Cambridge University Press, 1996. (Encyclopedia of mathematics and its applications v. 59). ISBN 0-521-45007-1. Citation on page 152.

BARRANCA, V. J. Reconstruction of sparse recurrent connectivity and inputs from the nonlinear dynamics of neuronal networks. **Journal of Computational Neuroscience**, v. 51, n. 1, p. 43–58, Feb 2023. ISSN 1573-6873. Available: <https://doi.org/10.1007/s10827-022-00831-x>. Citation on page 102.

BARRANCA, V. J.; ZHOU, D. Compressive sensing inference of neuronal network connectivity in balanced neuronal dynamics. **Frontiers in Neuroscience**, v. 13, 2019. ISSN 1662-453X. Available: <https://www.frontiersin.org/articles/10.3389/fnins.2019.01101>. Citation on page 102.

BERNSTEIN, D. S. **Matrix Mathematics: Theory, Facts, and Formulas (Second Edition)**. Princeton: Princeton University Press, 2009. ISBN 0-691-14039-1. Citations on pages 108, 143, and 145.

BOHLAND, J. W.; WU, C.; BARBAS, H.; BOKIL, H.; BOTA, M. *et al.* A Proposal for a Coordinated Effort for the Determination of Brainwide Neuroanatomical Connectivity in Model Organisms at a Mesoscopic Scale. **PLOS Computational Biology**, Public Library of Science, v. 5, n. 3, p. 1–9, 03 2009. Available: <https://doi.org/10.1371/journal.pcbi.1000334>. Citation on page 27.

BONGARD, J.; LIPSON, H. Automated reverse engineering of nonlinear dynamical systems. **Proceedings of the National Academy of Sciences**, v. 104, n. 24, p. 9943–9948, 2007. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.0609476104>. Citation on page 101.

BORISYUK, R.; COOKE, T. Book review: "bursting: The genesis of rhythm in the nervous system", s. coombes and p. bressloff, eds., (2005). **Journal of Integrative Neuroscience**, Imperial College Press, v. 5, n. 3, p. 483–488, 2006. ISSN 0219-6352. Citation on page 101.

BRESSLER, S. L.; SETH, A. K. Wiener–granger causality: A well established methodology. **NeuroImage**, v. 58, n. 2, p. 323–329, 2011. ISSN 1053-8119. Available: <https://www.sciencedirect.com/science/article/pii/S1053811910002272>. Citation on page 28.

BROIDO, A. D.; CLAUSET, A. Scale-free networks are rare. **Nature Communications**, v. 10, n. 1, p. 1017, Mar 2019. ISSN 2041-1723. Available: <https://doi.org/10.1038/s41467-019-08746-5>. Citation on page 27.

BRUNTON, S. L.; PROCTOR, J. L.; KUTZ, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 113, n. 15, p. 3932–3937, 2016. ISSN 0027-8424. Available: <https://www.pnas.org/content/113/15/3932>. Citations on pages 29, 31, 41, 101, and 152.

BUDIŠIĆ, M.; MOHR, R.; MEZIĆ, I. Applied koopmanism. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, v. 22, n. 4, p. 047510, 2012. Available: <https://doi.org/10.1063/1.4772195>. Citation on page 29.

BUTTE, A. J.; KOHANE, I. S. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. In: ____. **Biocomputing 2000**. [s.n.]. p. 418–429. Available: <https://www.worldscientific.com/doi/abs/10.1142/9789814447331_0040>. Citation on page 28.

____. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. In: **Biocomputing 2000**. [S.l.]: World Scientific, 1999. p. 418–429. Citation on page 96.

BUZSáKI, G. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. **Hippocampus**, v. 25, n. 10, p. 1073–1188, 2015. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hipo.22488>. Citation on page 101.

CANDES, E.; TAO, T. Decoding by linear programming. **IEEE Transactions on Information Theory**, v. 51, n. 12, p. 4203–4215, 2005. Citations on pages 33, 41, 51, 94, and 95.

CANDÈS, E. J. The restricted isometry property and its implications for compressed sensing. **Comptes Rendus Mathematique**, v. 346, n. 9, p. 589 – 592, 2008. ISSN 1631-073X. Citations on pages 51 and 97.

CANDES, E. J.; ROMBERG, J. K.; TAO, T. Stable signal recovery from incomplete and inaccurate measurements. **Communications on Pure and Applied Mathematics**, Wiley Online Library, v. 59, n. 8, p. 1207–1223, 2006. Citation on page 79.

CHAZOTTES, J.-R.; FERNANDEZ, B. **Dynamics of coupled map lattices and of related spatially extended systems**. [S.l.]: Springer Science & Business Media, 2005. Citation on page 92.

CHEN, J.; HU, H. The spectral gap for transfer operators of torus extensions over expanding maps. **Nonlinearity**, IOP Publishing, v. 32, n. 1, p. 356, dec 2018. Available: <https://dx.doi.org/10.1088/1361-6544/aaea87>. Citation on page 118.

CHUNG, F.; LU, L. **Complex graphs and networks**. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006. viii+264 p. (CBMS Regional Conference Series in Mathematics, v. 107). ISBN 978-0-8218-3657-6; 0-8218-3657-9. Available: <http://dx.doi.org/10.1090/cbms/107>. Citation on page 44.

CLEVELAND, R. B.; CLEVELAND, W. S.; MCRAE, J. E.; TERPENNING, I. Stl: A seasonal-trend decomposition procedure based on loess (with discussion). **Journal of Official Statistics**, v. 6, p. 3–73, 1990. Citation on page 135.

DIAMOND, S.; BOYD, S. CVXPY: A Python-embedded modeling language for convex optimization. **Journal of Machine Learning Research**, v. 17, n. 83, p. 1–5, 2016. Citation on page 34.

DOMAHIDI, A.; CHU, E.; BOYD, S. ECOS: An SOCP solver for embedded systems. In: **European Control Conference (ECC)**. [S.l.: s.n.], 2013. p. 3071–3076. Citations on pages 13, 34, 49, and 147.

DONOHO, D.; ELAD, M.; TEMLYAKOV, V. Stable recovery of sparse overcomplete representations in the presence of noise. **IEEE Transactions on Information Theory**, v. 52, n. 1, p. 6–18, 2006. Citations on pages 33 and 43.

DONOHO, D.; HUO, X. Uncertainty principles and ideal atomic decomposition. **IEEE Transactions on Information Theory**, v. 47, n. 7, p. 2845–2862, 2001. Citation on page 43.

DONOHO, D. L. *et al.* Compressed sensing. **IEEE Transactions on information theory**, Citeseer, v. 52, n. 4, p. 1289–1306, 2006. Citation on page 79.

DUNKL, C. F.; XU, Y. **Orthogonal Polynomials of Several Variables**. 2. ed. [S.l.]: Cambridge University Press, 2014. (Encyclopedia of Mathematics and its Applications). Citations on pages 53 and 54.

ERNST, J.; BAR-JOSEPH, Z. Stem: a tool for the analysis of short time series gene expression data. **BMC Bioinformatics**, v. 7, n. 1, p. 191, Apr 2006. ISSN 1471-2105. Available: <https://doi.org/10.1186/1471-2105-7-191>. Citation on page 28.

EROGLU, D.; TANZI, M.; STRIEN, S. van; PEREIRA, T. Revealing dynamics, communities, and criticality from data. **Physical Review X**, American Physical Society, v. 10, p. 021047, 2020. Available: <https://link.aps.org/doi/10.1103/PhysRevX.10.021047>. Citations on pages 40, 94, 96, and 104.

FENICHEL, N. Geometric singular perturbation theory for ordinary differential equations. **Journal of Differential Equations**, v. 31, n. 1, p. 53–98, 1979. ISSN 0022-0396. Available: <https://www.sciencedirect.com/science/article/pii/0022039679901529>. Citation on page 119.

FIGUEIREDO, M. A. T.; NOWAK, R. D.; WRIGHT, S. J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. **IEEE Journal of Selected Topics in Signal Processing**, v. 1, n. 4, p. 586–597, 2007. Citation on page 135.

FOLLAND, G. **Real Analysis: Modern Techniques and Their Applications**. [S.l.]: Wiley, 2013. (Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts). ISBN 9781118626399. Citations on pages 44, 69, and 87.

FOUCART, S.; RAUHUT, H. **A Mathematical Introduction to Compressive Sensing**. [S.l.]: Birkhäuser Basel, 2013. Citations on pages 43, 48, 51, 85, 88, and 143.

FTOREK, B.; ORŠANSKY, P. Korous type inequalities for orthogonal polynomials in two variables. **Tatra Mountains Mathematical Publications**, v. 58, n. 1, p. 1–12, 2014. Available: <https://doi.org/10.2478/tmmp-2014-0001>. Citation on page 53.

FUENTE, A. de la; BING, N.; HOESCHELE, I.; MENDES, P. Discovery of meaningful associations in genomic data using partial correlation coefficients. **Bioinformatics**, v. 20, n. 18, p. 3565–3574, 07 2004. ISSN 1367-4803. Available: <https://doi.org/10.1093/bioinformatics/bth445>. Citation on page 28.

GAO, T.-T.; YAN, G. Autonomous inference of complex network dynamics from incomplete and noisy data. **Nature Computational Science**, v. 2, n. 3, p. 160–168, Mar 2022. ISSN 2662-8457. Available: <https://doi.org/10.1038/s43588-022-00217-0>. Citation on page 115.

GELß, P.; KLUS, S.; EISERT, J.; SCHüTTE, C. Multidimensional Approximation of Nonlinear Dynamical Systems. **Journal of Computational and Nonlinear Dynamics**, v. 14, n. 6, p. 061006, 04 2019. ISSN 1555-1415. Available: <https://doi.org/10.1115/1.4043148>. Citation on page 115.

GENIO, C. I. D.; GROSS, T.; BASSLER, K. E. All scale-free networks are sparse. **Physical Review Letters**, APS, v. 107, p. 178701, 2011. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.107.178701>. Citation on page 27.

GERVEN, M. van; FARQUHAR, J.; SCHAEFER, R.; VLEK, R.; GEUZE, J.; NIJHOLT, A.; RAMSEY, N.; HASELAGER, P.; VUURPIJL, L.; GIELEN, S.; DESAIN, P. The brain–computer interface cycle. **Journal of Neural Engineering**, v. 6, n. 4, p. 041001, jul 2009. Available: <https://dx.doi.org/10.1088/1741-2560/6/4/041001>. Citation on page 28.

GOLUB, G. H.; LOAN, C. F. V. **Matrix Computations**. Third. [S.l.]: The Johns Hopkins University Press, 1996. Citation on page 114.

GUZMAN, S. J.; SCHLöGL, A.; FROTSCHER, M.; JONAS, P. Synaptic mechanisms of pattern completion in the hippocampal ca3 network. **Science**, v. 353, n. 6304, p. 1117–1123, 2016. Available: <https://www.science.org/doi/abs/10.1126/science.aaf1836>. Citations on pages 27, 101, and 104.

HAMZI, B.; OWHADI, H. Learning dynamical systems from data: A simple cross-validation perspective, part i: Parametric kernel flows. **Physica D: Nonlinear Phenomena**, v. 421, p. 132817, 2021. ISSN 0167-2789. Available: <https://www.sciencedirect.com/science/article/pii/S0167278920308186>. Citation on page 86.

HAN, X.; SHEN, Z.; WANG, W.-X.; DI, Z. Robust reconstruction of complex networks from sparse data. **Physical Review Letters**, APS, v. 114, p. 028701, 2015. Citation on page 29.

HANG, H.; STEINWART, I. A Bernstein-type inequality for some mixing processes and dynamical systems with an application to learning. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 45, n. 2, p. 708 – 743, 2017. Available: <https://doi.org/10.1214/16-AOS1465>. Citations on pages 40, 43, 48, 52, and 94.

HANG, H.; STEINWART, I.; FENG, Y.; SUYKENS, J. A. Kernel density estimation for dynamical systems. **Journal of Machine Learning Research**, Microtome Publishing, v. 19, 2018. ISSN 1532-4435. Citations on pages 138 and 139.

HART, J. D.; ZHANG, Y.; ROY, R.; MOTTER, A. E. Topological control of synchronization patterns: Trading symmetry for stability. **Physical Review Letters**, APS, v. 122, n. 5, p. 058301, 2019. Citations on pages 30, 39, 47, 92, and 136.

HE, Y.; CHEN, Z. J.; EVANS, A. C. Small-World Anatomical Networks in the Human Brain Revealed by Cortical Thickness from MRI. **Cerebral Cortex**, v. 17, n. 10, p. 2407–2419, 01 2007. ISSN 1047-3211. Available: <https://doi.org/10.1093/cercor/bhl149>. Citations on pages 27, 101, and 104.

HEMPEL, S.; KOSESKA, A.; KURTHS, J.; NIKOLOSKI, Z. Inner composition alignment for inferring directed networks from short time series. **Phys. Rev. Lett.**, American Physical Society, v. 107, p. 054101, Jul 2011. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.107.054101>. Citation on page 28.

HOFER, K. T.; KANDRÁCS, Á.; TÓTH, K.; HAJNAL, B.; BOKODI, V.; TÓTH, E. Z.; ERŐSS, L.; ENTZ, L.; BAGÓ, A. G.; FABÓ, D.; ULBERT, I.; WITTNER, L. Bursting of excitatory cells is linked to interictal epileptic discharge generation in humans. **Scientific Reports**, v. 12, n. 1, p. 6280, Apr 2022. ISSN 2045-2322. Available: <https://doi.org/10.1038/s41598-022-10319-4>. Citation on page 101.

HORN, R. A.; HORN, R. A. **Topics in matrix analysis**. Cambridge: Cambridge University Press, 1991. ISBN 0-521-30587-X. Citation on page 144.

IZHIKEVICH, E. M. Neural excitability, spiking and bursting. **International Journal of Bifurcation and Chaos**, v. 10, n. 06, p. 1171–1266, 2000. Available: <https://doi.org/10.1142/S0218127400000840>. Citation on page 101.

____. **Dynamical systems in neuroscience the geometry of excitability and bursting**. Cambridge, Mass: MIT Press, 2007. (Computational neuroscience). ISBN 0262090430. Citations on pages 27 and 101.

KAHEMAN, K.; KUTZ, J. N.; BRUNTON, S. L. Sindy-pi: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 476, n. 2242, p. 20200279, 2020. Available: <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2020.0279>. Citation on page 115.

KRALEMANN, B.; CIMPONERIU, L.; ROSENBLUM, M.; PIKOVSKY, A.; MROWKA, R. Phase dynamics of coupled oscillators reconstructed from data. **Physical Review E**, APS, v. 77, n. 6, p. 066205, 2008. Citation on page 78.

KRALEMANN, B.; PIKOVSKY, A.; ROSENBLUM, M. Reconstructing phase dynamics of oscillator networks. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, AIP, v. 21, n. 2, p. 025104, 2011. Citations on pages 78 and 93.

____. Reconstructing phase dynamics of oscillator networks. **Chaos**, AIP, v. 21, p. 025104, 2011. Citation on page 93.

KURAMOTO, Y. (Ed.). **Chemical Oscillations, Waves and Turbulence**. Berlin: Dover, 1984. Citation on page 27.

LADROUE, C.; GUO, S.; KENDRICK, K.; FENG, J. Beyond element-wise interactions: Identifying complex interactions in biological processes. **PLOS ONE**, Public Library of Science, v. 4, n. 9, p. 1–14, 09 2009. Available: <https://doi.org/10.1371/journal.pone.0006899>. Citation on page 96.

LEVITIN, A. **Introduction to the design & analysis of algorithms**. 2nd ed.. ed. Boston: Pearson Addison-Wesley, 2007. ISBN 0321358287. Citation on page 30.

LU, Z.; HUNT, B. R.; OTT, E. Attractor reconstruction by machine learning. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, v. 28, n. 6, p. 061104, 06 2018. ISSN 1054-1500. Available: <https://doi.org/10.1063/1.5039508>. Citation on page 29.

LUCHINSKY, D. G.; SMELYANSKIY, V. N.; DUGGENTO, A.; MCCLINTOCK, P. V. E. Inferential framework for nonstationary dynamics. i. theory. **Phys. Rev. E**, American Physical Society, v. 77, p. 061105, Jun 2008. Available: <https://link.aps.org/doi/10.1103/PhysRevE.77.061105>. Citations on pages 29 and 119.

LUENBERGER, D. G. **Optimization by Vector Space Methods**. New York, NY, USA: John Wiley & Sons, Inc., 1997. Citations on pages 31 and 93.

MANGAN, N. M.; BRUNTON, S. L.; PROCTOR, J. L.; KUTZ, J. N. Inferring biological networks by sparse identification of nonlinear dynamics. **IEEE Transactions on Molecular, Biological and Multi-Scale Communications**, IEEE, v. 2, n. 1, p. 52–63, 2016. Citation on page 41.

____. Inferring biological networks by sparse identification of nonlinear dynamics. **IEEE Transactions on Molecular, Biological and Multi-Scale Communications**, v. 2, n. 1, p. 52–63, 2016. Citations on pages 102 and 105.

MASON, A.; NICOLL, A.; STRATFORD, K. Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. **Journal of Neuroscience**, Society for Neuroscience, v. 11, n. 1, p. 72–84, 1991. ISSN 0270-6474. Available: <https://www.jneurosci.org/content/11/1/72>. Citations on pages 27, 101, and 104.

MEURER, A.; SMITH, C. P.; PAPROCKI, M.; ČERTÍK, O.; KIRPICHEV, S. B.; ROCKLIN, M.; KUMAR, A.; IVANOV, S.; MOORE, J. K.; SINGH, S.; RATHNAYAKE, T.; VIG, S.; GRANGER, B. E.; MULLER, R. P.; BONAZZI, F.; GUPTA, H.; VATS, S.; JOHANSSON, F.; PEDREGOSA, F.; CURRY, M. J.; TERREL, A. R.; ROUČKA, Š.; SABOO, A.; FERNANDO, I.; KULAL, S.; CIMRMAN, R.; SCOPATZ, A. Sympy: symbolic computing in python. **PeerJ Computer Science**, v. 3, p. e103, Jan. 2017. ISSN 2376-5992. Available: <https://doi.org/10.7717/peerj-cs.103>. Citation on page 106.

MOLNAR, F.; NISHIKAWA, T.; MOTTER, A. E. Asymmetry underlies stability in power grids. **Nature Communications**, v. 12, n. 1, p. 1457, Mar 2021. ISSN 2041-1723. Available: <https://doi.org/10.1038/s41467-021-21290-5>. Citation on page 27.

NAPOLETANI, D.; SAUER, T. D. Reconstructing the topology of sparsely connected dynamical networks. **Phys. Rev. E**, American Physical Society, v. 77, p. 026103, 2008. Citations on pages 29, 33, 41, 81, 84, and 101.

NAWRATH, J.; ROMANO, M. C.; THIEL, M.; KISS, Z.; WICKRAMASINGHE, M.; TIMMER, J.; KURTHS, J.; SCHELTER, B. Distinguishing Direct from Indirect Interactions in Oscillatory Networks with Multiple Time Scales. **Physical Review Letters**, v. 104, p. 038701, 2010. Citation on page 28.

NEGRO, C. A. D.; HSIAO, C.-F.; CHANDLER, S. H.; GARFINKEL, A. Evidence for a novel bursting mechanism in rodent trigeminal neurons. **Biophysical Journal**, Elsevier, v. 75, n. 1,

p. 174–182, Jul 1998. ISSN 0006-3495. Available: <https://doi.org/10.1016/S0006-3495(98)77504-6>. Citation on page 101.

NIJHOLT, E.; OCAMPO-ESPINDOLA, J. L.; EROGLU, D.; KISS, I. Z.; PEREIRA, T. Emergent hypernetworks in weakly coupled oscillators. **Nature Communications**, v. 13, n. 1, p. 4849, Aug 2022. ISSN 2041-1723. Available: <https://doi.org/10.1038/s41467-022-32282-4>. Citation on page 76.

NOVAES, M.; Roque dos Santos, E.; PEREIRA, T. Recovering sparse networks: Basis adaptation and stability under extensions. **Physica D: Nonlinear Phenomena**, v. 424, p. 132895, 2021. ISSN 0167-2789. Available: <https://www.sciencedirect.com/science/article/pii/S0167278921000531>. Citations on pages 29, 33, 36, 41, 75, and 95.

NOWAK, L. G.; AZOUZ, R.; SANCHEZ-VIVES, M. V.; GRAY, C. M.; MCCORMICK, D. A. Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. **Journal of Neurophysiology**, v. 89, n. 3, p. 1541–1566, 2003. PMID: 12626627. Available: <https://doi.org/10.1152/jn.00580.2002>. Citation on page 101.

O'DONOGHUE, B. Operator splitting for a homogeneous embedding of the linear complementarity problem. **SIAM Journal on Optimization**, v. 31, p. 1999–2023, August 2021. Citation on page 147.

O'DONOGHUE, B.; CHU, E.; PARIKH, N.; BOYD, S. Conic optimization via operator splitting and homogeneous self-dual embedding. **Journal of Optimization Theory and Applications**, v. 169, n. 3, p. 1042–1068, June 2016. Available: <http://stanford.edu/~boyd/papers/scs.html>. Citation on page 147.

____. **SCS: Splitting Conic Solver, version 3.0.0**. 2019. Https://github.com/cvxgrp/scs. Citation on page 147.

PARK, H.-J.; FRISTON, K. Structural and functional brain networks: From connections to cognition. **Science**, American Association for the Advancement of Science, v. 342, n. 6158, 2013. ISSN 0036-8075. Citations on pages 27 and 101.

PATHAK, J.; HUNT, B.; GIRVAN, M.; LU, Z.; OTT, E. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. **Phys. Rev. Lett.**, American Physical Society, v. 120, p. 024102, Jan 2018. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.120.024102>. Citation on page 29.

PEREIRA, T.; SANTOS, E. R. dos; STRIEN, S. van. **Robust reconstruction of sparse network dynamics**. 2023. Citations on pages 36, 39, 91, 94, and 95.

PEREIRA, T.; STRIEN, S. van; TANZI, M. Heterogeneously coupled maps: hub dynamics and emergence across connectivity layers. **J. Eur. Math. Soc.**, v. 22, p. 2183–2252, 2020. ISSN 7. Citation on page 27.

PIKOVSKY, A. Reconstruction of a random phase dynamics network from observations. **Physics Letters A**, v. 382, p. 147–152, 2018. Citation on page 78.

Qu, Q.; Sun, J.; Wright, J. Finding a sparse vector in a subspace: Linear sparsity using alternating directions. **IEEE Transactions on Information Theory**, v. 62, n. 10, p. 5855–5880, Oct 2016. Citations on pages 102, 105, 111, and 149.

RACHEV, S. **Probability Metrics and the Stability of Stochastic Models**. [S.l.]: Wiley, 1991. (Wiley Series in Probability and Statistics - Applied Probability and Statistics Section). ISBN 9780471928775. Citation on page 52.

REUS, M. A. de; HEUVEL, M. P. van den. Rich club organization and intermodule communication in the cat connectome. **Journal of Neuroscience**, Society for Neuroscience, v. 33, n. 32, p. 12929–12939, 2013. ISSN 0270-6474. Available: <https://www.jneurosci.org/content/33/32/12929>. Citation on page 115.

REVERTER, A.; CHAN, E. K. F. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. **Bioinformatics**, v. 24, n. 21, p. 2491–2497, 09 2008. ISSN 1367-4803. Available: <https://doi.org/10.1093/bioinformatics/btn482>. Citation on page 96.

RINZEL, J. A formal classification of bursting mechanisms in excitable systems. In: ____. **Mathematical Topics in Population Biology, Morphogenesis and Neurosciences: Proceedings of an International Symposium held in Kyoto, November 10–15, 1985**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1987. p. 267–281. ISBN 978-3-642-93360-8. Available: <https://doi.org/10.1007/978-3-642-93360-8_26>. Citation on page 101.

RULKOV, N. F. Regularization of synchronized chaotic bursts. **Physical Review Letters**, American Physical Society, v. 86, p. 183–186, Jan 2001. Citation on page 102.

SANABRIA, E. R. G.; SU, H.; YAARI, Y. Initiation of network bursts by ca2+-dependent intrinsic bursting in the rat pilocarpine model of temporal lobe epilepsy. **The Journal of Physiology**, v. 532, n. 1, p. 205–216, 2001. Available: <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-7793.2001.0205g.x>. Citation on page 101.

SANTOS, E. R. dos. Code available at <https://github.com/edmilson-roque-santos/rulkov_reconstruction>. Citation on page 115.

SANTOS, E. Roque dos. **Ergodic Basis Pursuit**. Code available at <https://github.com/edmilson-roque-santos/Ergodic-basis-pursuit>. Citation on page 74.

SCANNELL, J.; BLAKEMORE, C.; YOUNG, M. Analysis of connectivity in the cat cerebral cortex. **Journal of Neuroscience**, Society for Neuroscience, v. 15, n. 2, p. 1463–1483, 1995. ISSN 0270-6474. Available: <https://www.jneurosci.org/content/15/2/1463>. Citation on page 35.

SCANNELL, J. W.; YOUNG, M. P. The connectional organization of neural systems in the cat cerebral cortex. **Current Biology**, v. 3, n. 4, p. 191 – 200, 1993. ISSN 0960-9822. Available: <http://www.sciencedirect.com/science/article/pii/096098229390331H>. Citation on page 35.

SCHAEFFER, H.; TRAN, G.; WARD, R. Extracting sparse high-dimensional dynamics from limited data. **SIAM Journal on Applied Mathematics**, v. 78, n. 6, p. 3279–3295, 2018. Available: <https://doi.org/10.1137/18M116798X>. Citations on pages 29, 33, and 41.

SCHAEFFER, H.; TRAN, G.; WARD, R.; ZHANG, L. Extracting structured dynamical systems using sparse optimization with very few samples. **Multiscale Modeling & Simulation**, v. 18, n. 4, p. 1435–1461, 2020. Available: <https://doi.org/10.1137/18M1194730>. Citations on pages 33 and 41.

SCHÄFER, B.; WITTHAUT, D.; TIMME, M.; LATORA, V. Dynamically induced cascading failures in power grids. **Nature Communications**, v. 9, n. 1, p. 1975, May 2018. ISSN 2041-1723. Available: <https://doi.org/10.1038/s41467-018-04287-5>. Citation on page 27.

SCHMIDT, M.; LIPSON, H. Distilling free-form natural laws from experimental data. **Science**, v. 324, n. 5923, p. 81–85, 2009. Available: <https://www.science.org/doi/abs/10.1126/science.1165893>. Citation on page 101.

SCHRÖTER, M.; PAULSEN, O.; BULLMORE, E. T. Micro-connectomics: probing the organization of neuronal networks at the cellular scale. **Nature Reviews Neuroscience**, v. 18, n. 3, p. 131–146, Mar 2017. ISSN 1471-0048. Available: <https://doi.org/10.1038/nrn.2016.182>. Citations on pages 27 and 102.

SLIPANTSCHUK, J.; BANDTLOW, O. F.; JUST, W. Dynamic mode decomposition for analytic maps. **Communications in Nonlinear Science and Numerical Simulation**, v. 84, p. 105179, 2020. ISSN 1007-5704. Available: <https://www.sciencedirect.com/science/article/pii/S1007570420300149>. Citation on page 29.

SPORNS, O.; BETZEL, R. F. Modular brain networks. **Annual Review of Psychology**, v. 67, n. 1, p. 613–640, 2016. PMID: 26393868. Available: <https://doi.org/10.1146/annurev-psych-122414-033634>. Citation on page 91.

SPORNS, O.; TONONI, G.; KÖTTER, R. The human connectome: a structural description of the human brain. **PLoS Comput Biol**, Public Library of Science, v. 1, n. 4, p. e42, 2005. Citation on page 27.

STANKOVSKI, T.; DUGGENTO, A.; MCCLINTOCK, P. V. E.; STEFANOVSKA, A. Inference of time-evolving coupled dynamical systems in the presence of noise. **Phys. Rev. Lett.**, American Physical Society, v. 109, p. 024101, Jul 2012. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.109.024101>. Citation on page 29.

STANKOVSKI, T.; PEREIRA, T.; MCCLINTOCK, P. V.; STEFANOVSKA, A. Coupling functions: universal insights into dynamical interaction mechanisms. **Reviews of Modern Physics**, APS, v. 89, n. 4, p. 045001, 2017. Citation on page 27.

SUN, J.; BOLLT, E. M. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. **Physica D: Nonlinear Phenomena**, v. 267, p. 49–57, 2014. ISSN 0167-2789. Evolving Dynamical Networks. Available: <https://www.sciencedirect.com/science/article/pii/S0167278913001991>. Citation on page 96.

SUN, J.; TAYLOR, D.; BOLLT, E. M. Causal network inference by optimal causation entropy. **SIAM Journal on Applied Dynamical Systems**, v. 14, n. 1, p. 73–106, 2015. Available: <https://doi.org/10.1137/140956166>. Citation on page 28.

SZEGÔ, G. **Orthogonal polynomials**. New York city: American mathematical society, 1939. (Colloquium publications (American Mathematical Society) v. 23). Citations on pages 53 and 61.

TANZI, M. **Uniformly Expanding Coupled Maps: Self-Consistent Transfer Operators and Propagation of Chaos**. 2022. Citation on page 118.

THIBEAULT, V.; ST-ONGE, G.; DUBÉ, L. J.; DESROSIERS, P. Threefold way to the dimension reduction of dynamics on networks: An application to synchronization. **Phys. Rev. Research**, American Physical Society, v. 2, p. 043215, Nov 2020. Available: <https://link.aps.org/doi/10.1103/PhysRevResearch.2.043215>. Citation on page 29.

TOPAL, I.; EROGLU, D. Reconstructing network dynamics of coupled discrete chaotic units from data. **Phys. Rev. Lett.**, American Physical Society, v. 130, p. 117401, Mar 2023. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.130.117401>. Citations on pages 104 and 115.

TOWLSON, E. K.; VÉRTES, P. E.; AHNERT, S. E.; SCHAFER, W. R.; BULLMORE, E. T. The rich club of the c. elegans neuronal connectome. **Journal of Neuroscience**, Society for Neuroscience, v. 33, n. 15, p. 6380–6387, 2013. ISSN 0270-6474. Available: <https://www.jneurosci.org/content/33/15/6380>. Citations on pages 91, 103, and 115.

TRAN, G.; WARD, R. Exact recovery of chaotic systems from highly corrupted data. **Multiscale Modeling & Simulation**, v. 15, n. 3, p. 1108–1129, 2017. Available: <https://doi.org/10.1137/16M1086637>. Citation on page 29.

____. Exact recovery of chaotic systems from highly corrupted data. **Multiscale Modeling & Simulation**, v. 15, n. 3, p. 1108–1129, 2017. Available: <https://doi.org/10.1137/16M1086637>. Citation on page 86.

VIANA, M. **Stochastic dynamics of deterministic systems**. [S.l.]: IMPA Rio de Janeiro, 1997. Citation on page 29.

WANG, S.; HERZOG, E. D.; KISS, I. Z.; SCHWARTZ, W. J.; BLOCH, G.; SEBEK, M.; GRANADOS-FUENTES, D.; WANG, L.; LI, J.-S. Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 115, n. 37, p. 9300–9305, 2018. Citation on page 41.

WANG, W.-X.; LAI, Y.-C.; GREBOGI, C. Data based identification and prediction of nonlinear and complex dynamical systems. **Physics Reports**, Elsevier, v. 644, p. 1–76, 2016. Citations on pages 29, 33, 41, and 101.

WANG, W.-X.; YANG, R.; LAI, Y.-C.; KOVANIS, V.; GREBOGI, C. Predicting catastrophes in nonlinear dynamical systems by compressive sensing. **Phys. Rev. Lett.**, American Physical Society, v. 106, p. 154101, Apr 2011. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.106.154101>. Citation on page 29.

WANG, X.; WU, M.; LI, Z.; CHAN, C. Short time-series microarray analysis: Methods and challenges. **BMC Systems Biology**, v. 2, n. 1, p. 58, Jul 2008. ISSN 1752-0509. Available: <https://doi.org/10.1186/1752-0509-2-58>. Citation on page 28.

WATERS, J.; HELMCHEN, F. Background synaptic activity is sparse in neocortex. **Journal of Neuroscience**, Society for Neuroscience, v. 26, n. 32, p. 8267–8277, 2006. ISSN 0270-6474. Available: <https://www.jneurosci.org/content/26/32/8267>. Citations on pages 27, 101, and 104.

WHITE, J.; SOUTHGATE, E.; THOMSON, J.; BRENNER, S. The structure of the nervous system of the nematode caenorhabditis elegans (the mind of a worm). **Phil. Trans. R. Soc. Lond. B**, v. 314, p. 1–340, 1986. Citation on page 103.

WINFREE, A. T. **The geometry of biological time**. [S.l.]: Springer Science & Business Media, 2001. Citation on page 27.

WITTHAUT, D.; HELLMANN, F.; KURTHS, J.; KETTEMANN, S.; MEYER-ORTMANNS, H.; TIMME, M. Collective nonlinear dynamics and self-organization in decentralized power grids. **Rev. Mod. Phys.**, American Physical Society, v. 94, p. 015005, Feb 2022. Available: <https://link.aps.org/doi/10.1103/RevModPhys.94.015005>. Citation on page 27.

ZAMORA-LÓPEZ, G.; ZHOU, C.; KURTHS, J. Exploring brain function from anatomical connectivity. **Frontiers in Neuroscience**, v. 5, n. JUN, p. 1–11, 2011. ISSN 16624548. Citation on page 27.

ZHANG, S.; LIN, G. Robust data-driven discovery of governing physical laws with error bars. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 474, n. 2217, p. 20180305, 2018. Available: <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2018.0305>. Citation on page 115.

ZHU, P.; KNYAZEV, A. Angles between subspaces and their tangents. **Journal of Numerical Mathematics**, v. 21, n. 4, p. 325–340, 2013. Available: <https://doi.org/10.1515/jnum-2013-0013>. Citations on pages 143, 144, and 145.

APPENDIX

A

ERGODIC BASIS PURSUIT: NOISY DATA AND ESTIMATING MEASURE

A.1 Relaxing path algorithm

Here we detail the relaxing path algorithm used to reconstruct the coupled optoelectronic network in Section 2.4. This algorithm tunes the parameter ε to identify correct connections in the underlying network, see Figure 19 for an illustration of the network reconstruction scheme using relaxing path algorithm.

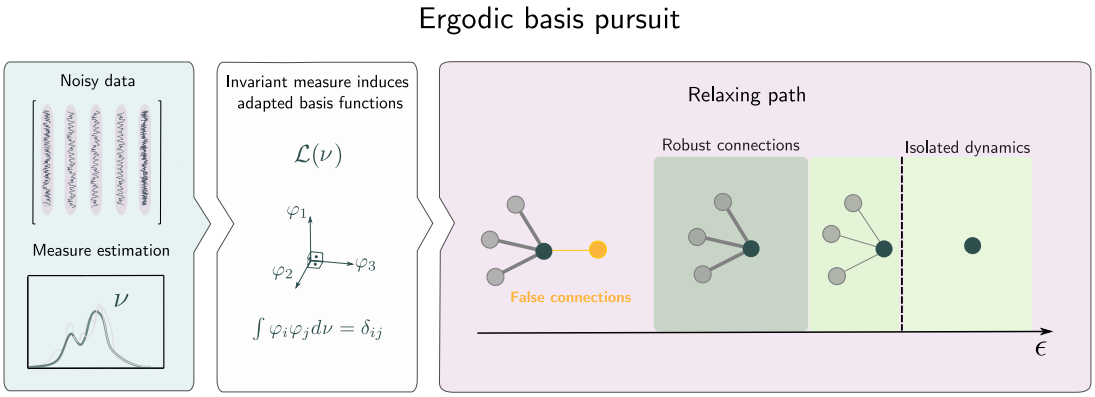

Figure 19 – **Robust network reconstruction scheme using ergodic basis pursuit.** The noisy data is generated from a network dynamics whose underlying measure is μ_ξ. Using its estimated measure ν, we induce an orthonormal set of basis functions $\mathscr{L}(\nu)$ representing the dynamics. Under the assumption that the network dynamics is sparse, the noisy data and $\mathscr{L}(\nu)$ are recast as a minimization problem, whose solution encodes a proxy of the network. Although the noise level may be unknown, the relaxing path algorithm searches the connections of each node varying the noise level ε as a parameter. The true connections remain robust over an interval of ε.

Relaxing path algorithm uses the Theorem 7. To gain insight into what is happening, it is

worth to consider a special case. Consider the case we probe a node in a weakly coupled network via the relaxing path algorithm. In this case, the true coefficient vector has the entries' magnitude relative to the isolated dynamics f larger than those relative to the coupling function h. This difference also appears in the family of solutions $\{c_v^\star(\varepsilon)\}_\varepsilon$ as we increase ε. For small ε, we may observe a few false connections; see Figure 19. These false connections disappear for larger ε, and only the robust connections remain. If we keep increasing ε, the robust connections have their magnitude reduced until we only observe those entries relative to the isolated dynamics, i.e., there are no incoming connections. Thus, the parameter ε identifies the magnitude difference among the entries of the true solution, capturing the correct incoming connections to the node inside an appropriate interval of ε.

Here, we detail the three main stages of the algorithm.

A.1.1 Model selection (MS)

Model selection selects the coefficient vector of a particular node for a given parameter value ε. The current version of our algorithm encompasses the criterion used in the main text, where we search the robust connections in consecutive parameter values. Alternative approaches are valid and can be incorporated. For short we denote the model selection step as $MS(c^\star(\varepsilon))$ for a given coefficient vector $c^\star(\varepsilon)$.

A.1.2 Network selection

Network selection is a map that obtains the network structure from the coefficient matrix. Although each coefficient matrix C with column vectors $\{c_1,\dots,c_N\}$ is mapped to a directed multigraph (a graph that permits multiple edges among the nodes), the Network Selection introduces a map NS that yields a graph structure instead.

Let $\mathscr{S}_i \subset [m]$ be the set of indices corresponding to basis functions in $\mathscr{L}$ that depend on node i. Thus, the directed multigraph $G_m = ([N], E_m)$ from the representation of the network dynamics in $\mathscr{L}$ is defined by

$$E_m = \{(i,j) \ : \ w_{ij}^k = c_i^k \neq 0, k \in \mathscr{S}_j\}, \tag{A.1}$$

where (i,j) corresponds to an edge from node j to i and w_{ij}^k is the edge weight. Self-loops are excluded since the graph only carries information about the coupling structure. The Network Selection step constructs a weighted graph from this multigraph, defined as

$$W_{ij} = \max_{k \in \mathscr{S}_j} \{w_{ij}^k\}, \tag{A.2}$$

such that W corresponds to an adjacency matrix whose entries are the edge weights in (A.2). We construct a directed subgraph of node i picking the i-th row of W and denote it by subgraph(c_i).

The same construction can be replicated for all nodes, so the graph structure from the coefficient matrix C is

$$G = \bigcup_{i=1}^{N} \text{subgraph}(c_i),$$

characterizing the network selection map $NS : \mathbb{R}^{m\times N} \to \mathbb{R}^{N\times N}$ defined as $C \mapsto W$. When the edge weights are unnecessary, all edge weights are set to 1.

A.1.3 Algorithm

Equation (2.17) quantifies the approximation accuracy w.r.t. to the sparse vector c_ν. We can use it to estimate the entries' magnitude lying outside the support set of this sparse vector, $\mathscr{S} = \text{supp}(c_\nu)$. Let us denote $u_\mathscr{S}$ as the vector equal to u on the index set $\mathscr{S}$ and zero on its complement $\mathscr{S}^c$. We can decompose $c^\star_\nu(\varepsilon)$ into the sum of $c^\star_{\nu,\mathscr{S}}(\varepsilon)$ and $c^\star_{\nu,\mathscr{S}^c}(\varepsilon)$. Note that $\|c^\star_{\nu,\mathscr{S}}(\varepsilon) - c_\nu\|_2^2 + \|c^\star_{\nu,\mathscr{S}^c}(\varepsilon)\|_2^2 = \|c^\star_\nu(\varepsilon) - c_\nu\|_2^2$ since $\mathscr{S}$ and $\mathscr{S}^c$ are disjoint, and it implies that $\|c^\star_{\nu,\mathscr{S}^c}(\varepsilon)\|_2 \leq K_4\varepsilon$. Hence, assuming the wrong entries are assigned at random, we consider that any entry of $c^\star_\nu(\varepsilon)$ with a magnitude less than $\mathscr{O}(\varepsilon/\sqrt{m})$ is zero.

Since the entries' magnitude supported in $\mathscr{S}^c$ are bounded by $K_2\varepsilon$, we discard the irrelevant connections (to represent the node dynamics) encoded in $c^\star_\nu(\varepsilon)$ as we tune ε. The idea is to tune ε and find the connections that are robust over different parameter values, the relevant connections. The challenge is that ξ is unknown, as well as the other quantities that bound the error in (2.15). We look at this problem as a one-parameter family, searching the support set that persists over different ε and reconstructing the sparse network. We propose the relaxing path algorithm:

1. Select a set of equally spaced values ε_k within the interval $\mathscr{E} = [\varepsilon_{\min}, \varepsilon_{\max}]$. A pre-processing analysis can estimate the interval bounds (CLEVELAND *et al.*, 1990).
2. For each $\varepsilon_k \in \mathscr{E}$ find the optimal solution to the (2.16), the support $\mathscr{S}_k = \text{supp}(c^\star_\nu(\varepsilon_k))$ and $\mathscr{T}_k = \mathscr{S}_k \Delta \mathscr{S}_{k-1}$, where Δ corresponds to the symmetric difference of the two sets and checks the change in their cardinality (FIGUEIREDO; NOWAK; WRIGHT, 2007).
3. If $|\mathscr{T}_k| = 0$, the support has not changed, then stop, and the corresponding solution $c^\star_\nu(\varepsilon_k)$ is returned. Otherwise, iterate $k \mapsto k+1$ and repeat Step 2.

The quadratically constrained Ergodic Basis Pursuit method is given in Algorithm 1. Let the hard-threshold function be given as

$$\text{hard}(u, \lambda) = u\chi_{|u|>\lambda},$$

where χ_A is the characteristic function on the set A. For a vector $u \in \mathbb{R}^m$, we consider that the hard-threshold function is evaluated coordinate-wise. The relaxing path algorithm reconstructs the networks as described in Algorithm 2.

Algorithm 1 – Quadractically constrained Ergodic Basis Pursuit (QEBP)

Input: $\bar{y} \in \mathbb{R}^n$, $\Phi_\nu(Y) \in \mathbb{R}^{n\times m}$, ε

/* $\bar{y} \in \mathbb{R}^n$: time series (measurement) vector */

/* $\Phi_\nu(Y) \in \mathbb{R}^{n\times m}$: adapted library matrix */

/* ε: relaxing parameter */

Output: $c^\star(\varepsilon)$ coefficient vector

Output.

$$c_\nu^\star(\varepsilon) = \arg\min_{\tilde{u}\in\mathbb{R}^m}\{\|\tilde{u}\|_1 \quad \text{subject to } \|\Phi_\nu(Y)\tilde{u} - \bar{y}\|_2 \leq \varepsilon\}. \tag{A.3}$$

Algorithm 2 – Relaxing path algorithm

Input: $\bar{Y} = (\bar{y}_1, \ldots, \bar{y}_N) \in \mathbb{R}^{n\times N}$, $\Phi_\nu(Y) \in \mathbb{R}^{n\times m}$, $X \in \mathbb{R}^{m\times N}$

/* $\bar{Y}$: matrix of the noisy multivariate time series */

/* criterion: select criterion for Model Selection */

/* $\Phi_\nu(Y)$: adapted library matrix */

Output: Reconstructed graph G

Initialization: $G^0 = ([N], \emptyset)$, $\mathscr{F}^0 = \emptyset$, $\varepsilon_{\min}, \varepsilon_{\max} \in \mathbb{R}$ **for** $j \in [N]$ **do**

  for $\varepsilon \in [\varepsilon_{\min}, \varepsilon_{\max}]$ **do**

    $c_\nu^\star(\varepsilon) = \text{QEBP}(\bar{y}_j, \varepsilon)$ $c_\nu^\star(\varepsilon) = \text{hard}(c_\nu^\star(\varepsilon), \varepsilon/\sqrt{m})$ $c_j^\star(\varepsilon) = R_\nu^{-1} c_\nu^\star(\varepsilon)$

    if $MS(c_j^\star(\varepsilon), \texttt{criterion})$ *is satisfied* **then**

      $H_j = \text{subgraph}(c_j^\star(\varepsilon))$ $G^{k+1} = G^k \cup H_j$ $\mathscr{F}^{k+1} = \mathscr{F}^k \cup \{j\}$

      Stop and go to next node

    end

  end

end

Output. Coefficient matrix $C^\star$.

A.2 Optoelectronic experimental data

The experimental data corresponds to a network of optoelectronic oscillators whose nonlinear component is a Mach-Zehnder intensity modulator. Each node can be modeled as

$$x_i(t+1) = \beta I_\theta(x_i(t)) + \alpha \sum_{j=1}^{17} A_{ij}[I_\theta(x_j(t)) - I_\theta(x_i(t))] \mod 2\pi, \quad i = 1, \ldots, N, \tag{A.4}$$

where the normalized intensity output of the Mach-Zehnder modulator is given by

$$I_\theta(x) = \sin^2(x + \theta),$$

x represents the normalized voltage applied to the modulator, β is the feedback strength, and δ is the operating point set to $\frac{\pi}{4}$. We obtained the experimental multivariate time series $\{y_1(t), \cdots, y_{17}(t)\}_{t=1}^{16385}$ corresponding to $N = 17$ units coupled through a network, where the following parameters were fixed and known experimentally: $\beta = 4.5$ and coupling strength $\alpha = 0.171875$ (HART *et al.*, 2019). Let us denote the optoelectronic network dynamics as F.

A.2.1 Localization on a subset of the phase space

Figure 20 illustrates the extraction of experimental data used in the main text.

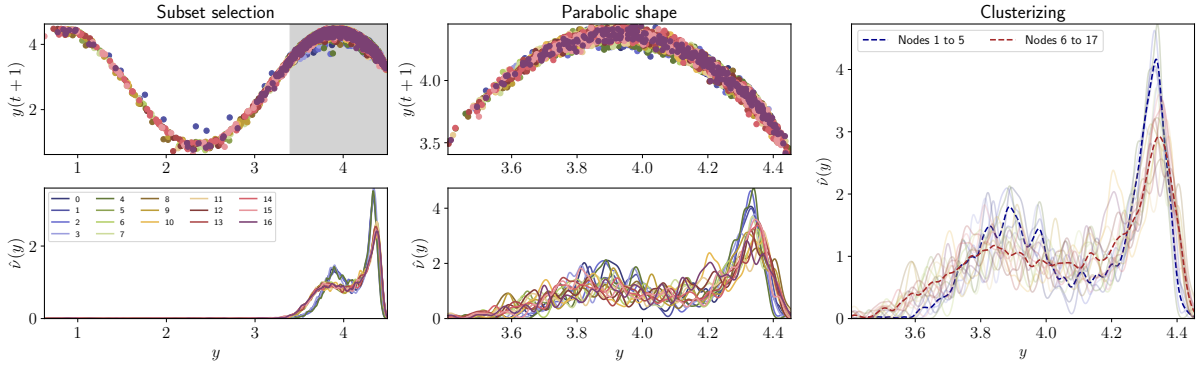

Figure 20 – **Step-by-step description of the extraction of the experimental data**. The top middle panel displays the return map of the resulting network dynamics' trajectory with 264-time steps for coupling $\alpha = 0.171875$.

Subset selection. In Figure 20, the upper left panel shows the return map for all nodes in the network measured experimentally (according to private communication, the initial $\sim$1000 points can be regarded as transient. Hence we removed it). The shaded gray area represents the phase space subset where nodes spend more time compared to other regions; in this case, it corresponds to the interval $\mathscr{I} = [3.4, 4.5]$. This is confirmed by the bottom left panel, which depicts the estimated density function for each node. All densities are concentrated inside the same gray area but decay rapidly to zero over the complement of the interval.

Parabolic shape of the return map. We extracted the most significant sample of the network dynamics' trajectory, satisfying that all nodes should remain inside the gray region. This subset selection yields a sample with a length of 224 time steps and is shown in the top middle panel. The shape of the return map changes from a sinusoidal to a parabolic shape, suggesting using polynomial basis functions up to degree two in the library for the reconstruction procedure. The bottom middle panel shows the estimated density for the trajectories inside the subset. Observe that in view of the densities, either the bottom left and bottom middle panels, there are roughly two groups of nodes, represented by nodes 1 to 5 and 6 to 17 (near $y = 4.4$ this observation is clear).

Clustering. For each group of nodes, we consider all trajectories as different initial conditions of the same map and estimate the density function, as described in (A.7). In Figure 20, right panel depicts the density function $\rho_1(y_1, \ldots, y_5)$ for group 1 to 5 in blue, and $\rho_2(y_6, \ldots, y_{17})$ for group 6 to 17 in red. Note both density functions are smoother than the individual densities since we used more data to estimate them. The final estimated density we use to orthonormalize the polynomial basis functions is written as the product density function of both group density functions, i.e.,

$$\rho(y_1, \ldots, y_{17}) \approx \rho_1(y_1, \ldots, y_5) \times \rho_2(y_6, \ldots, y_{17}). \tag{A.5}$$

A.2.2 Basis functions selection preserves the network structure

The parabolic shape of the return map in the top middle panel in Figure 20 corresponds to the restriction of the optoelectronic network dynamics F onto $\mathscr{A}$, which we denote $\tilde{F} = F|_{\mathscr{A}}$. Hence, $\tilde{F}$ lies in the span of the quadratic polynomials. Taylor expanding (A.4) at the maximum point inside the interval $\mathscr{A} = [3.4, 4.5]$, there are no crossed terms in the expansion. Hence, we only included monomials up to degree 2, so for each $j = 1, \ldots, N$, the basis functions are $\phi_k(y_j) = y_j^k$ where $k = 1, 2$. The independent term $\phi_0(y) = 1$ is added with a single column in the library matrix.

A.2.3 Reconstruction of the network structure

To quantify the overall reconstruction performance, we introduce a weighted false link proportion for each node. Let $\mathscr{M}_i$ and $\hat{\mathscr{M}}_i$ be the subset of edges node i shares with its neighbors of the original and estimated graph, respectively. We assume that any edge has a weight equal to 1, but those estimated using the relaxing algorithm and denoted by $\{w_{ij}\}_{i,j}$. So, we calculate the proportion of false positive (FP) and false negative (FN) at node i as:

$$FP_i = \frac{\sum_{j=1}^{N} w_{ij} \chi_{\hat{\mathscr{M}}_i \cap \mathscr{M}_i^c}\big((i,j)\big)}{\sum_{j=1}^{N} \Big(w_{ij} \chi_{\hat{\mathscr{M}}_i \cap \mathscr{M}_i^c}\big((i,j)\big) + \chi_{\hat{\mathscr{M}}_i^c \cap \mathscr{M}_i^c}\big((i,j)\big) \Big)},$$

$$FN_i = \frac{\sum_{j=1}^{N} \chi_{\hat{\mathscr{M}}_i^c \cap \mathscr{M}_i}\big((i,j)\big)}{\sum_{j=1}^{N} \chi_{\mathscr{M}_i}\big((i,j)\big)},$$

where $\chi_{\mathscr{U}}$ is the indicator function of the subset $\mathscr{U}$.

A.3 Approximating invariant measures from multivariate time series

In this section, we discuss how we estimate a proxy for the physical measure μ_α from the multivariate time series $\{x(t)\}_{t=0}^{n}$ (we write it for the noiseless case, but the same approach is used for the noisy measurements $\{y(t)\}_{t=0}^{n}$).

Let $\{x(t)\}_{t=0}^{n}$ be the multivariate time series on M^N and consider the empirical measure $P_n = \frac{1}{n}\sum_{t=0}^{n-1} \delta_{x(t)}$ on M^N. To estimate the physical measure, we employ kernel density estimators. The kernel density estimator consists of taking the convolution (which we denote by $*$) of the empirical measure P_n and the kernel $G_\chi : [0,\infty) \to [0,\infty)$ (HANG *et al.*, 2018), where χ is the bandwidth parameter. We consider the Gaussian kernel throughout our results, i.e., $G_\chi(x) = e^{-x^2/\chi^2}$.

Hang and co-authors (HANG *et al.*, 2018) evaluated the convergence rate of kernel estimators for dynamical systems satisfying certain mixing conditions. Following there results, assume that μ_α is absolutely continuous with respect to Lebesgue with density ρ_α. The rate of

convergence of the estimated density with respect to ρ_α in L_1-norm depends on the regularity of ρ_α. If ρ_α is $\beta-$Hölder continuous then, it behaves asymptotically for the length of time series as

$$\|\rho_\alpha - G_\chi * P_n\|_1 \leq \mathscr{O}\Big(\big(\frac{\log n^3}{n}\big)^{\frac{\beta}{2\beta+N}}\Big). \tag{A.6}$$

Hence, in the case of large networks, the convergence is slow, which requires a large amount of time series. In (HANG *et al.*, 2018) the authors also consider more general regularity conditions, for instance pointwise β-Hölder controllable condition. However, there are no explicit expressions of the convergence rates. So, for sake of exposition we consider the case in (A.6).

A.3.1 Product measure

Following our assumption in the paper, we assume that the physical measure μ_α is close to a product measure ν. So, instead of estimating μ_α, we estimate ν from the data. We consider the multivariate time series $\{x_i(t)\}_{t=0,i=1}^{n,N}$ as the observation from the same system, i.e., $\{x(t)\}_{t=0}^{nN}$. More precisely, the estimation from data is formulated as follows: we consider the empirical measure $\frac{1}{nN}\sum_{t=0}^{nN-1}\delta_{x(t)}$. Hence, for a fixed $\chi > 0$, the proxy of ρ_α is given by the convolution of the empirical measure and the kernel G_χ

$$\rho_{n,N,\chi}(x) = \frac{1}{nN\chi}\sum_{t=0}^{nN-1} G_\chi(x - x(t)). \tag{A.7}$$

In other words, the density of the product measure ν is given by $\prod_{i\in[N]}\rho_{n,N,\chi}$. Since the expression of $\rho_{n,N,\chi}$ corresponds to a sum of Gaussian kernels, it fulfils restrictions on the shape given in (HANG *et al.*, 2018). In particular, we can deduce an upper bound of the Lipschitz constant of $\rho_{n,N,\chi}$ as follows:

Lemma A.3.1. For a given $i \in [N]$ let $M_i = [a,b] \subset \mathbb{R}$ with $b > a$, and let $a_1 = \max\{|a|,|b|\}$. Then

$$\mathrm{Lip}(\prod_{i\in[N]}\rho_{n,N,\chi}) = \mathrm{Lip}(\rho_{n,N,\chi}) \leq \frac{4}{\chi^2}a_1.$$

Proof. As before, here we consider $\mathrm{Lip}(\prod_{i\in[N]}\rho_{n,N,\chi}) = \max_{i\in[N]}\mathrm{Lip}(\rho_{n,N,\chi})$ then the first equality holds. By Mean-value theorem, the Lipschitz constant $\mathrm{Lip}(\rho_{n,N,\chi})$ is given by

$$\begin{aligned}
\mathrm{Lip}(\rho_{n,N,\chi}) &= \|D\rho_{n,N,\chi}\|_\infty \\
&= \max_{x\in[a,b]}|D\rho_{n,N,\chi}(x)| \\
&= \max_{x\in[a,b]}\frac{2}{nN\chi^3}\left|\sum_{t=0}^{nN-1}(x-x(t))e^{-\frac{(x-x(t))^2}{\chi^2}}\right| \\
&\leq \frac{2}{\chi^2}(\max\{|a|,|b|\} + \frac{1}{Nn}\sum_{t=0}^{nN-1}|x(t)|) \\
&\leq \frac{4}{\chi^2}a_1.
\end{aligned}$$

where we used that $\frac{1}{\chi}e^{-\frac{(x-x(t))^2}{\chi^2}} \leq 1$ for any $x \in [a,b]$, and the claim follows. □

We also guarantee that $\rho_{n,N,\chi}(x) > 0$ for any $x \in [a,b]$ then $\rho_0 = \min_{i\in[N]}\{\min_{x\in[a,b]} \rho_{n,N,\chi}(x)\} > 0$. More importantly, the $\|\prod_{i\in[N]} \rho_{n,N,\chi} - \rho_\alpha\|_1$ has similar rate of convergence as (A.6), but for the $d = 1$. So, the speed of convergence does not depend on the dimension of the phase space.

APPENDIX

B

STABILITY UNDER BASIS EXTENSION: NOISE AND PROOF

B.1 Stability of Lasso under noise

We analyse the effect of noise by adding, to the original equations of motion, Equation (3.1), a term $\sqrt{2D}\dot{B}_j(t)$, with a homogeneous complex Wiener process $B_j(t) = \xi_j(t) + i\zeta_j(t)$ with $\langle \xi_k(t)\xi_j(s)\rangle = \langle \zeta_k(t)\zeta_j(s)\rangle = \delta_{ik}\delta(t-s)$, and D =diag(η,η) and we obtain

$$\dot{\phi}_j = \omega_j + \alpha \sum_{k=1}^{N} A_{jk} \sin(\phi_k - \phi_j) + \mathcal{N}_j(\phi_j, t) \tag{B.1}$$

where the noise term is given by

$$\mathcal{N}_j(\phi,t) = \sqrt{2\eta}(\cos(\phi)\xi_j(t) + \sin(\phi)\zeta_j(t)). \tag{B.2}$$

The noisy equations of motion are integrated by Euler's method with a time step of 0.1, the time series for the phases are obtained by Hilbert transform and a Savitzky-Golay filter is applied to them, before the time derivative is calculated. The filtered phases are then used in the matrix Φ.

We use as a measure of performance the number of recovered connections. Suppose we have found equations of motion for the variables $\phi_j(t)$ in the form of a vector of coefficients $c^{(j)}$, $1 \le j \le N$, where $c^{(j)}$ denotes the jth column of C. The norm of the function $h^{(j)}$, we can recover the strength of the coupling between node j and the central node 1 as

$$\kappa_j = \|h^{(j)}\|_2 = \sqrt{(c_{1j}^{(j)})^2 + (d_{1j}^{(j)})^2}. \tag{B.3}$$

We define a quantity playing the role of effective total number of connections as

$$\kappa = \frac{1}{\alpha} \sum_{j=2}^{N} \kappa_j, \tag{B.4}$$

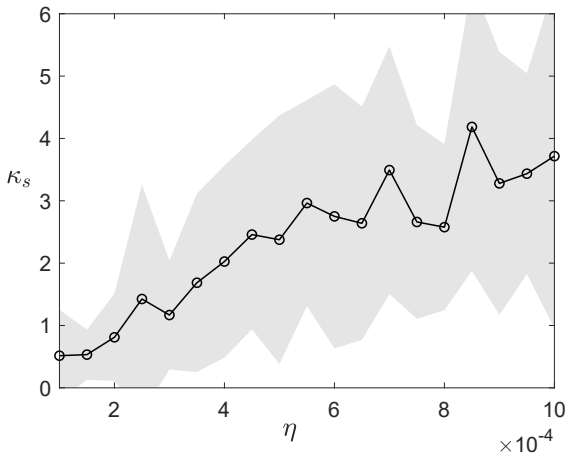

Figure 21 – **Reconstruction against noise.** Number of spurious connections κ_s predicted by the LASSO in the presence of noise of intensity η, for a star network with $N = 10$ and averaged over 20 random initial conditions (shaded region represents the corresponding variance).

and the effective number of spurious connections (in general not an integer number),

$$\kappa_s = \kappa - N. \tag{B.5}$$

In Figure 21 we show how the performance of the method deteriorates as the amplitude of the noise increases, by plotting the effective number of spurious connections as a function of the noise intensity η, averaged over 50 random initial conditions (shaded region corresponds to standard deviation).

Equation (B.1) can be recast in the linear form

$$v^{(j)} + \sqrt{2\eta}\, z^{(j)} = \Phi w^{(j)}, \quad j = 1, \ldots, N, \tag{B.6}$$

where $v^{(j)} \in \mathbb{R}^n$ corresponds to the Euler approximation of the time-derivative. Besides $z^{(j)} \in \mathbb{R}^n$ is a random variable whose each entry has the form of Equation (B.2). Equation (B.6) written in this form is similar to the noisy recovery case estimated by Candès. Again, we take advantage of the relation between LASSO and the quadractically constrained basis pursuit problem.

Proposition B.1.0.1. Assume that $\delta_{2s} < \sqrt{2} - 1$. Given $\eta > 0$, $n > 0$ and $i = 1, \ldots, N$, for each $\varepsilon > 0$ the following holds

$$\mathbb{P}\left(\|z^{(j)}\|_2 \leq \frac{\varepsilon}{\sqrt{2\eta}} \right) \geq 1 - e^{-\frac{1}{4}\left(\frac{\varepsilon}{\sqrt{2\eta}} - \sqrt{\frac{2}{\pi}}\sqrt{n} \right)^2} \tag{B.7}$$

So with high probability there exists the solution $w^\star$ of Equation (2.22) satisfies

$$\|w^* - w\|_2 \leq K_0 s^{-1/2} \|w - w_s\|_1 + K_1 \varepsilon$$

for constants K_0 and K_1.

Proof. Fix a $\eta > 0$. We need to estimate how probable $\sqrt{2\eta}z^{(j)}$ is L_2 bounded by a constant ε. We drop the dependence of (j). Note that

$$\sqrt{2\eta}\|z\|_2 \leq \sqrt{2\eta}(\|\xi\|_2 + \|\zeta\|_2) \tag{B.8}$$

where each ξ and ζ are Gaussian random vectors. So, we can estimate the expected value of this L_2 norm: $\mathbb{E}(\|\xi\|_2) \geq \sqrt{\frac{2}{\pi}n}$ (FOUCART; RAUHUT, 2013, Proposition 8.1) and the same for ζ. We use the concentration of measure for Gaussian random vector (FOUCART; RAUHUT, 2013, Theorem 8.34). Since the norm L_2 is a Lipschitz function with constant 1, the estimate follows. Then $\|\xi\|_2 \leq \frac{\varepsilon}{2\sqrt{2\eta}}$ holds with probability

$$1 - \exp\left(-\frac{1}{4}\left(\frac{\varepsilon}{\sqrt{2\eta}} - \sqrt{\frac{2}{\pi}}\sqrt{n}\right)^2\right). \tag{B.9}$$

Consequently, we can apply Candès' estimate and the statement is proved. □

B.2 Proof of Proposition 3.4.0.1

We first need two preliminary results

Remark B.2.1. Given $S \in \mathbb{R}^{s_1 \times s_2}$ with $1 \leq rank(S) \leq \min\{s_1, s_2\}$, then

$$\|Sx\|_2 \geq \sigma_{min}(S)\|x\|_2, \forall\, x \in \mathbb{R}^{s_2} \backslash \{\ker S\},$$

where $\sigma_{\min}(S)$ is the minimum singular value of S (BERNSTEIN, 2009, Fact 9.13.1).

Remark B.2.2. Let $A \in \mathbb{R}^{n\times p}$ and $B \in \mathbb{R}^{n\times q}$ be column full rank matrices with $n > p+q$, and $\varepsilon \in (0,1)$. Let $r = \min\{p,q\}$. We have the following:

i) For generic $z \in \mathbb{R}^n$: $z \notin (\operatorname{Im} B)^{\perp}$. The map $H(z) = B^{\dagger}z$ is not constant thus $\operatorname{Leb}(H^{-1}(0)) = 0$ and $\mathbb{R}^n \backslash H^{-1}(0)$ is a generic set.

ii) For generic $z \in (\operatorname{Im} A)^{\perp}$ with $\|z\|_2 = \varepsilon$, there exists $K(B,\varepsilon) > 0$ such that

$$\|B^{\dagger}z\|_2 \geq K\cos(\beta_r),$$

where β_r is the largest principle angle between $(\operatorname{Im} A)^{\perp}$ and $\operatorname{Im} B$. Indeed, let $Q_B R_B$ be the QR decomposition of the matrix B and $Q_{A^{\perp}}$ an orthonormal matrix whose columns form an orthonormal basis to $(\operatorname{Im} A)^{\perp}$. Hence, there exists a unique $v \neq 0$ such that $z = Q_{A^{\perp}}v$ and $\|v\|_2 = \varepsilon$. Applying Remark B.2.1 and previous item (i), generically, we have

$$\|B^{\dagger}z\|_2 \geq \sigma_{\min}(R_B)\sigma_{\min}(Q_B^{\dagger}Q_{A^{\perp}})\|v\|_2. \tag{B.10}$$

By (ZHU; KNYAZEV, 2013, Theorem 2.1) the principal angles β's between subspaces $(\operatorname{Im} A)^{\perp}$ and $\operatorname{Im} B$ are

$$(\cos(\beta_1), \ldots, \cos(\beta_r)) = (\sigma_{\max}(Q_B^{\dagger}Q_{A^{\perp}}), \ldots, \sigma_{\min}(Q_B^{\dagger}Q_{A^{\perp}})), \tag{B.11}$$

where $\beta_k \in [0, \frac{\pi}{2}]$, $k = 1, \ldots, r$ and $\beta_k < \beta_{k+1}$, $k = 1, \ldots, r-1$. In particular, the cosine of the largest angle between $(\operatorname{Im} A)^{\perp}$ and $\operatorname{Im} B$ is given as follows

$$\cos \beta_r = \sigma_{\min}(Q_B^{\dagger} Q_{A^{\perp}}). \tag{B.12}$$

So, there exists $K(B, \varepsilon) > 0$ such that

$$\|B^{\dagger} z\|_2 \geq K \cos(\beta_r). \tag{B.13}$$

Lemma B.2.3. Let $A \in \mathbb{R}^{n \times p}$ and $B \in \mathbb{R}^{n \times q}$ be column full rank matrices with $n > p + q$. Let $r = \min\{p, q\}$ then $\beta_r \neq \frac{\pi}{2}$ be the largest principle angle between the subspaces $(\operatorname{Im} A)^{\perp}$ and $\operatorname{Im} B$. Consider

$$M = B^{\dagger} B - B^{\dagger} A (A^{\dagger} A)^{-1} A^{\dagger} B \tag{B.14}$$

then there exists a constant $K > 0$ such that

$$\sigma_{\min}\left(M^{-1}\right) \geq \frac{1}{K \cos^2(\beta_r)}. \tag{B.15}$$

Proof. Let E and F be the orthogonal projection onto the $\operatorname{Im} A$ and $(\operatorname{Im} A)^{\perp}$, respectively. So, using that A is column full rank we can write M as follows:

$$\begin{aligned} M &= B^{\dagger} B - B^{\dagger} A (A^{\dagger} A)^{-1} A^{\dagger} B \\ &= B^{\dagger} (\mathbf{I} - A A^{+}) B \\ &= B^{\dagger} F B \in \mathbb{R}^{q \times q}. \end{aligned}$$

The orthogonal projections have the following formulas: $E = Q_A Q_A^{\dagger}$ and $F = Q_{A^{\perp}} Q_{A^{\perp}}^{\dagger}$, where Q_A and $Q_{A^{\perp}}$ are orthonormal matrices whose columns form an orthonormal basis to $\operatorname{Im} A$ and $(\operatorname{Im} A)^{\perp}$, respectively. Besides, let us denote $B = Q_B R_B$ the QR decomposition of the matrix B. Using this notation and inequality of the singular value (HORN; HORN, 1991, Theorem 3.3.14) we can split up the maximum singular value as follows:

$$\sigma_{\max}(B^{\dagger} F B) \leq \sigma_{\max}(R_B^{\dagger}) \sigma_{\max}(Q_B^{\dagger} Q_{A^{\perp}}) \sigma_{\max}(Q_{A^{\perp}}^{\dagger} Q_B) \sigma_{\max}(R_B). \tag{B.16}$$

Again by (ZHU; KNYAZEV, 2013, Theorem 2.1) the least angle between $(\operatorname{Im} A)^{\perp}$ and $\operatorname{Im} B$ is given as follows

$$\cos \beta_1 = \sigma_{\max}(Q_B^{\dagger} Q_{A^{\perp}}) = \sigma_{\max}(Q_{A^{\perp}}^{\dagger} Q_B), \tag{B.17}$$

Also $\beta_1 < \beta_r$ so $\beta_1 \neq \frac{\pi}{2}$ and Equation (B.17) is not zero. If we use that $\sigma_{\max}(B^{\dagger} F B) = \frac{1}{\sigma_{\min}((B^{\dagger} F B)^{-1})}$ and Equation (B.17) into Equation (B.16) we obtain

$$\sigma_{\min}((B^{\dagger} F B)^{-1}) \geq \frac{1}{K \cos^2(\beta_1)}, \tag{B.18}$$

where we use that $\sigma_{\max}(R_B)$ can be bounded by a constant $K > 0$. Since cos is decreasing in the interval $[0, \frac{\pi}{2}]$, we can replace β_1 by the largest principle angle β_r, and the claim follows. □

Proof of Proposition 3.4.0.1. Note that since Φ is column full rank, we can write the solution of Equation (3.16) as

$$\begin{pmatrix} \hat{w}_1 \\ \hat{w}_2 \end{pmatrix} = \Phi^+(b+z) = \begin{pmatrix} A^\dagger A & A^\dagger B \\ B^\dagger A & B^\dagger B \end{pmatrix}^{-1} \begin{pmatrix} A^\dagger b \\ B^\dagger b + B^\dagger z \end{pmatrix}, \tag{B.19}$$

where we use that $z \in (\operatorname{Im} A)^\perp$ implies that $A^\dagger z = 0$. Using the analytic inversion formula (BERNSTEIN, 2009), we obtain

$$(\Phi^\dagger \Phi)^{-1} = \begin{pmatrix} (A^\dagger A)^{-1} + (A^\dagger A)^{-1}(A^\dagger B)M^{-1}(B^\dagger A)(A^\dagger A)^{-1} & -(A^\dagger A)^{-1}A^\dagger B M^{-1} \\ -M^{-1}(B^\dagger A)(A^\dagger A)^{-1} & M^{-1} \end{pmatrix} \tag{B.20}$$

where $M^{-1} = (B^\dagger B - B^\dagger A(A^\dagger A)^{-1}A^\dagger B)^{-1}$. Since A and B are column full rank, we can use the formula of A^+ and AA^+ is a projector onto the $\operatorname{Im} A$. So, we obtain

$$\begin{pmatrix} \hat{w}_1 \\ \hat{w}_2 \end{pmatrix} = \Phi^+(b+z) = \begin{pmatrix} A^+ b - A^+ B M^{-1} B^\dagger z \\ M^{-1} B^\dagger z \end{pmatrix} \tag{B.21}$$

where we used that $b \in \operatorname{Im} A$.

We aim at calculating how much the solution x^* is perturbed, so

$$\|x^* - \hat{w}_1\|_2 = \|A^+ B M^{-1} B^\dagger z\|_2. \tag{B.22}$$

Since $\beta_1 > 0$ we have $BM^{-1}B^\dagger z \notin (\operatorname{Im} A)^\perp$, so using Remark B.2.1 for A^+ we obtain

$$\|x^* - \hat{w}_1\|_2 \geq \sigma_{\min}(A^+)\|BM^{-1}B^\dagger z\|_2. \tag{B.23}$$

By item (i) of Remark B.2.2 for generic $z \neq 0$ we have $B^\dagger z \neq 0$. Recall that $M^{-1}B^\dagger z \in (\ker B)^\perp$. Thus, Remark B.2.1 is valid for B and M^{-1} and we obtain

$$\|x^* - \hat{w}_1\|_2 \geq \sigma_{\min}(A^+)\sigma_{\min}(B)\sigma_{\min}(M^{-1})\|B^\dagger z\|_2.$$

remarkerve that $\sigma_{\min}(A^+), \sigma_{\min}(B) > 0$ from the full rank condition on the matrices A and B, so there exists $K_1 > 0$ given by

$$K_1 = \min\{\sigma_{\min}(A^+), \sigma_{\min}(B)\}. \tag{B.24}$$

Moreover, by item (ii) of Lemma B.2.2 we have: $\|B^\dagger z\|_2 \geq K_2 \cos(\beta_r)$. Hence, by (ZHU; KNYAZEV, 2013, Property 2.1) if $|\beta_r - \pi/2| < \varepsilon$ for sufficiently small ε, using Lemma B.2.3 there exists $K > 0$ we obtain

$$\|x^* - \hat{w}_1\|_2 \geq \frac{K_1 K_2}{K \cos \beta_r} > 0 \tag{B.25}$$

and the statement holds. □

APPENDIX

C

GREEDY NETWORK RECONSTRUCTION: FURTHER DETAILS

C.0.1 Convex optimization solvers: polynomial time complexity

To illustrate that polynomial time complexity is valid for the solvers embedded in the `CVXPY` package, we compare the performance of ECOS (DOMAHIDI; CHU; BOYD, 2013) and SCS (O'DONOGHUE *et al.*, 2019; O'DONOGHUE *et al.*, 2016; O'DONOGHUE, 2021) for an example. We selected to solve the quadratically constrained basis pursuit

$$\arg\min_{\tilde{u}\in\mathbb{R}^m} \left\{ \|\tilde{u}\|_1 \text{ subject to } \|\Phi\tilde{u} - b\|_2 \leq \varepsilon \right\} \tag{C.1}$$

for a dense library matrix Φ. More precisely, let $\Phi \in \mathbb{R}^{n\times m}$ be a random matrix whose entries $\phi_{i,j}$ are independent realizations of standard Gaussian random variables

$$\phi_{i,j} \sim \mathcal{N}(0,1).$$

The measurement vector b follows the same distribution, whose entries $b_i \sim \mathcal{N}(0,1)$ are independent realizations of a standard Gaussian random variable. In all our tests we fixed the relaxing parameter $\varepsilon = 0.01$. Figure 22 displays the running times for different n values. Note that the scaling in terms of the number of columns m is $\mathcal{O}(m^q)$ with $q \approx 1$ for all cases.

C.0.2 Global network reconstruction

Note that the global network structure can be obtained via the EBP: let $\mathbf{1}_N$ be the identity matrix and $\mathbf{c} = \mathbf{vec}(c_1, \ldots, c_N)$, where **vec** denotes the vectorization formed by stacking the columns vectors c_i into a single column vector. Define the problem

$$\min_{\mathbf{u}\in\mathbb{R}^{mN}} \|\mathbf{u}\|_1 \text{ subject to } (\mathbf{1}_N \otimes \Phi_\nu(X))\mathbf{u} = (\mathbf{1}_N \otimes \Phi_\nu(X))\mathbf{c}, \tag{C.2}$$

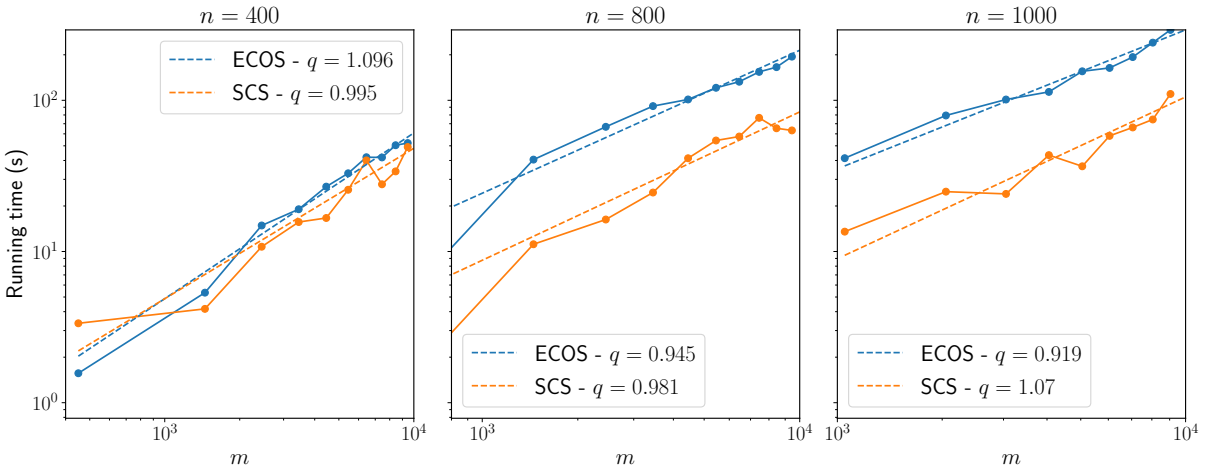

Figure 22 – **Convex minimization solvers performance.** Running times with respect to the number of columns m are displayed for different values of rows. The scope of the fitting is shown for each solver. All scope values are around 1, showing the performance is nearly linear.

where $\otimes$ is the Kronecker product. Since the library matrix $\Phi_\nu(X)$ satisfies the RIP property, then $(\mathbf{1}_N \otimes \Phi_\nu(X))$ satisfies as well, ensuring the uniqueness of solutions for problem Equation (C.2).

C.0.3 DCNR algorithm

Algorithm 3 – Divide-and-conquer network reconstruction algorithm

Input: $\bar{X} = (\bar{x}_1, \ldots, \bar{x}_N) \in \mathbb{R}^{n \times N}$, $\Phi \in \mathbb{R}^{n \times m}$, $\mathscr{P}_{[N]}$

/* $\bar{X}$: matrix of multivariate time series */

/* criterion: select criterion for Model Selection */

/* Φ_ν: adapted library matrix evaluated to the training time series */

/* $\mathscr{P}^{(0)}_{[N]}$: initial graph partition. */

Output: G reconstructed graph

Initialization: $G^0 = ([N], \emptyset)$, $\mathscr{F}^0 = \emptyset$, $\mathscr{T}^0 = P_1 \in \mathscr{P}^{(0)}_{[N]}$, $C^0 = 0 \in \mathbb{R}^{m \times N}$

for $P \in \mathscr{P}^{(0)}_{[N]}$ **do**

  $H^s, \mathscr{U}, C^{s+1} = \mathscr{B}_\varepsilon(P, C^s)$

  $G^{s+1} = G^s \cup H^s$

  $\mathscr{F}^{s+1} = \mathscr{F}^s \cup \left(P \cap \mathscr{U}^c\right)$

end

$B = B - \Phi C$

$H^s, \mathscr{U}, C^{s+1} = \mathscr{B}_\varepsilon(\mathscr{F}^c, X^s)$

$G^{s+1} = G^s \cup H^s$

$\mathscr{F}^{s+1} = \mathscr{F}^s \cup \left(P \cap \mathscr{U}^c\right)$

Output. Reconstructed graph $G^\star$.

APPENDIX

D

BURSTING NETWORK DYNAMICS: ADM METHOD AND PARETO FRONT

D.1 Alternating direction method (ADM)

Suppose a linear space $\mathscr{S} \subset \mathbb{R}^m$ of dimension s contains a sparse vector $u^\star \neq 0$. Given an arbitrary $\Psi \in \mathbb{R}^{s\times m}$ with $\ker(\Psi) = \mathscr{S}$, we aim to find (efficiently) a nonzero sparse vector u such that $\Psi u = 0$. A minimization formulation is searching a sparse solution of the problem

$$\min_{w\in\mathbb{R}^m\setminus 0} \|w\|_0 \qquad \text{subject to } \Psi w = 0,$$

but the problem is NP-hard to solve. Then, Qu and co-authors (Qu; Sun; Wright, 2016) introduced an alternative minimization problem which consists of relaxed version in terms of ℓ_1 norm, $\|u\|_1 = \sum_{i=1}^m |u_i|$. Since the desired solution is invariant by scaling, the constraint is replaced by searching vectors that lie on the unit embedded sphere $\mathbb{S}^{s-1} := \{u \in \mathbb{R}^s \mid \|u\|_2 = 1\}$, and the alternative problem is given by:

$$\min_{w\in\mathbb{R}^s} \|\Theta w\|_1 \qquad \text{subject to } \|w\|_2 = 1,$$

where $\Theta \in \mathbb{R}^{m\times p}$ is any orthonormal basis for $\mathscr{S}$. Although the problem is still nonconvex due the unit sphere is a nonconvex subset of $\mathbb{R}^s$, the authors built an algorithm for finding the sparse solution on the subspace $\mathscr{S}$, and also, obtain exact reconstruction guarantees (Qu; Sun; Wright, 2016).

We introduce an auxiliary variable $u \approx \Theta w$:

$$\min_{w\in\mathbb{R}^s, u\in\mathbb{R}^m} \frac{1}{2}\|\Theta w - u\|_2^2 + \gamma\|u\|_1 \qquad \text{subject to } \|w\|_2 = 1, \tag{D.1}$$

where γ is a penalty parameter. The algorithm consists of minimizing the problem (D.1) alternating over different directions until reach the sparse solution. In particular, It starts with a initial

guess $w^{(0)}$ and it alternates in minimizing with respect to (w.r.t.) x and minimizing w.r.t. to w:

$$u^{k+1} = \arg\min_{u\in\mathbb{R}^s} \frac{1}{2}\|\Theta w^k - u\|_2^2 + \gamma\|u\|_1,$$
$$w^{k+1} = \arg\min_{w\in\mathbb{S}^{s-1}} \frac{1}{2}\|\Theta w^k - x^{k+1}\|_2^2,$$

where $u^{(k)}$ and $w^{(k)}$ denote the values of u and w in the $k-$th iteration. The closed form of the above minimization problem is given in Equation (5.11).

Algorithm 4 – Modified Alternating Direction Method (ADM) algorithm

Input: Θ, γ, w^0 `maxiter`
`/* Θ: orthonormal basis for the null space. We use singular value decomposition. */`
`/* γ: regularizer parameter. */`
`/* w^0: normalized row of Θ. */`
`/* maxiter: maximum number of iterations. We fix 10000. */`
Output: The sparsest vector $u^\star(\gamma) = \Theta w^\star$
for $k = 0,\ldots,$ *maxiter* **do**
  $u_j^{k+1} = \text{soft}(\Theta_j w_j^k, \gamma)$ $w_j^{k+1} = \frac{\Theta_j^T w_j^{k+1}}{\|\Theta_j^T w_j^{k+1}\|_2}$
  if $\|\Theta^T w^k\|_2 \leq 0$ **then**
    stop.
  end
  if $\|w^{k+1} - w^k\|_2 \leq$ *tol* **then**
    Output. $u^\star(\gamma) = \Theta w^\star$
    stop.
  end
end

D.2 Implicit-SINDy and Pareto front

Implicit-SINDy algorithm is described in Algorithm 5. The idea is to solve ADM method for different regularizer parameters γ, and select the most parsimonious via Pareto front. Indeed, as larger γ becomes sparser the solution $u^\star(\gamma)$ of the ADM method (it decreases the number of terms). Left panel of Figure 23 shows the sparsity of $u^\star(\gamma)$ as γ is increased. Pareto front selects the most parsimonious model, as shown right panel of Figure 23. The coefficient with minimum error $\|\Psi(\bar{\mathbf{x}})u^\star(\gamma)|$ is chosen as the most parsimonious model. Surprisingly, this method succeeds at identifying the presence of rational terms.

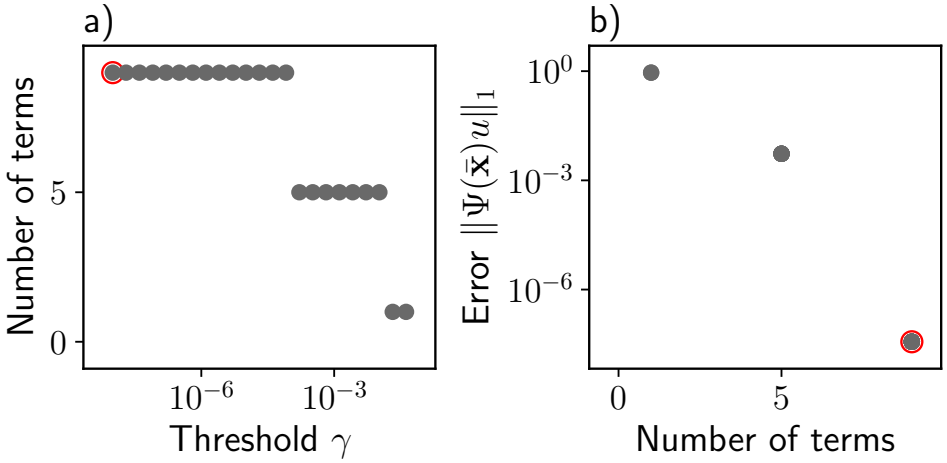

Figure 23 – **Pareto front for model selection.** a) Monotonic behaviour of the number of terms in the sparse solution as we increase the penalty parameter γ. b) The error $\|\Psi(\bar{\mathbf{x}})u^\star(\gamma)\|_1$ versus the number of non-zero entries of $u^\star(\gamma)$. The most parsimonious model attains the minimum error $\|\Psi(\bar{\mathbf{x}})u^\star(\gamma)\|_1$, because we aim at identifying the sparsest vector that lies on the kernel of the augmented matrix. Minimizing ℓ_1 norm is an option to have an approximation of such vector.

Algorithm 5 – Implicit-SINDy

Input: $\Psi(\bar{\mathbf{x}})$, γ^0, `maxiter`, `nrows`
`/*` $\Psi(\bar{\mathbf{x}})$`: augmented library matrix.` `*/`
`/*` γ_0`: Initial regularizer parameter for the Pareto front. We fix` $\gamma^0 = 10^{-8}$. `*/`
`/* maxiter: maximum number of regularizer points to be probed in the Pareto front. We fix maxiter` $= 30$. `*/`
`/* nrows: number of` Θ`'s rows to be tested. We select uniformly 10 percent of the` Θ`'s row.` `*/`
Output: The model selected $u^\star$
for $k = 0, \ldots,$ *maxiter* **do**
  $\Theta = \text{null space}(\Psi(\bar{\mathbf{x}}))$
  for $i = 0, \ldots,$ *nrows* **do**
    Pick a row of Θ, w^0 $u_i^\star(\gamma^k) = \text{ADM}(\Theta, \gamma^k, w^0)$
  end
  Select the $u_i^\star(\gamma^k)$ with most non-zero elements
  if $\|u^\star(\gamma^k)\|_0 \leq 0$ **then**
    stop.
  end
  $\gamma^{k+1} = 2\gamma^k$
end
$u^\star = \arg\min_k \|\Psi(\bar{\mathbf{x}})u^\star(\gamma^k)\|_1$ *Output.* $u^\star$

D.2.1 Alternative reconstruction method

Instead of the denominator polynomial Q_j in the rational representation in Equation (5.4), we consider

$$F_j(\mathbf{x}) = \frac{P_j(\mathbf{x})}{1+\tilde{Q}_j(\mathbf{x})}, \tag{D.2}$$

where the constant term is explicitly chosen 1 for definiteness akin to a Padé approximant (BAKER; BAKER, 1996), and $\tilde{Q}_j(\mathbf{x})$ does not contain the constant term in its polynomial expansion. We rearrange Equation (D.2) as

$$P_j(\mathbf{x}) - F_j(\mathbf{x})\tilde{Q}_j(\mathbf{x}) = F_j(\mathbf{x}).$$

Expanding $P_j(\mathbf{x})$ and $\tilde{Q}_j(\mathbf{x})$ in terms of polynomials in $\mathscr{L}$, denote the library matrix without the columns corresponding to the constant term as $\Phi(X|1)$. Also, denote the augmented library matrix as

$$\Omega(\bar{\mathbf{x}}_j) := \begin{bmatrix}\Phi(X) & \Phi(X|1)D_j\end{bmatrix} \in \mathbb{R}^{n\times(2m-1)}, \tag{D.3}$$

where $D_j = \mathrm{diag}(\bar{\mathbf{x}}_j)$ is the diagonal matrix constructed from the vector $\bar{\mathbf{x}}_j$. Then, we obtain the linear equation

$$\Omega(\bar{\mathbf{x}}_j)\tilde{c}_j = \bar{\mathbf{x}}_j, \tag{D.4}$$

which consists of a linear equation where the coefficients $\tilde{c}_j \in \mathbb{R}^{2m-1}$ are unknown . Here, the alternative reconstruction formulation is to find the coefficients that satisfy Equation (D.4). To solve the linear equation we consider the least-square approximation ℓ_2

$$\min_{\tilde{u}\in\mathbb{R}^m} \|\Omega(\bar{\mathbf{x}}_j)\tilde{u} - \bar{\mathbf{x}}_j\|_2, \quad j = 1,\ldots,2N,$$

whose solution is given by $u^* = \Omega(\bar{\mathbf{x}}_j)^+\bar{x}$. Other methods could also be employed such as SINDy (BRUNTON; PROCTOR; KUTZ, 2016) that sequentially thresholds a least-square approximation that fits the data.

—


ICMC
USP
SÃO CARLOS