Nonlinearity

PAPER • OPEN ACCESS

Robust reconstruction of sparse network dynamics

To cite this article: Tiago Pereira *et al* 2025 *Nonlinearity* **38** 055031

OPEN ACCESS
IOP Publishing | London Mathematical Society Nonlinearity
Nonlinearity **38** (2025) 055031 (41pp) https://doi.org/10.1088/1361-6544/add3b0

Robust reconstruction of sparse network dynamics

Tiago Pereira[1,2], **Edmilson Roque dos Santos**[1,2,*] and **Sebastian van Strien**[2]

[1] Instituto de Ciências Matemáticas e Computação, Universidade de São Paulo, São Carlos, Brazil
[2] Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom

E-mail: edmilson.roque.usp@gmail.com

Received 18 July 2024; revised 11 March 2025
Accepted for publication 2 May 2025
Published 15 May 2025

Recommended by Dr Lorenzo J Diaz

Abstract
Reconstructing the network interaction structure from multivariate time series is an important problem in multiple fields of science. When the network dynamics is represented as a linear combination of multivariate polynomials, the reconstruction can be formulated as an optimisation problem. For large networks, this optimisation problem does not always have a unique solution, leading to wrong reconstruction. We propose the Ergodic Basis Pursuit (EBP) method, which leverages the statistical properties of the network dynamics to accurately reconstruct sparse networks. The key idea is that the restricted isometry property of the associated library matrix—a crucial condition for ensuring unique reconstruction—can be derived from the ergodic properties of the network dynamics. We show that when the data length scales quadratically with node degree and logarithmically with network size the reconstruction is unique. Compared to traditional methods, the EBP reconstructs sparse networks

* Author to whom any correspondence should be addressed.

using significantly less data and is robust to noise. We validate its effectiveness using experimental time series from optoelectronic networks.

Supplementary material for this article is available online

Keywords: network dynamics, compressive sensing, basis pursuit, ergodic theory

Mathematics Subject Classification numbers: 37A25, 05C82, 37M10, 93B30

Contents

1. Introduction

Networks of coupled dynamical systems are successful models in diverse fields of science ranging from biology [Win01], chemistry [Kur84] to physics [SPMS17] and neuroscience [ZLZK11]. The network interaction structure impacts the dynamics [STK05, PvST20], in fact, many malfunctions are associated with disorders in the network structure [BWB+09]. While in general one cannot measure the network structure, one might have access to a multivariate time series of nodes' states. Therefore, reconstructing the network structure from multivariate data has attracted much attention by merging dynamical systems techniques [Tak06, Guc14] and optimisation [CNHT17, WHK+18, ETvSP20].

When the network is moderately large, the amount of data required for successful reconstruction is large, making the network reconstruction from data a non-trivial task. Indeed, in general, the network reconstruction becomes ill-posed and unstable [NS08, NRP21]. Recent strategies aim at incorporating the sparsity in the network interaction structure to formulate a minimisation problem that searches sparse representations of the input data [NS08, CGT14, HSWD15, MBPK16, WLG16, PYGS16]. The key idea is that sparsity may promote a decrease in the data length required for the reconstruction [TW17, STW18]. Nevertheless, ensuring exact network reconstruction, even in the presence of sparse interactions, remains an important open problem.

Here, we put forward a new method, which we call 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 BP problem adapted to the exponential mixing of the network dynamics.

1.1. Dynamics on complex networks and main assumptions

We consider network dynamics as

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

for each $i \in [N] := \{1, \ldots, N\}$, where x_i represents the state of node $i, f_i : M_i \to M_i$ 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 $h_{ij}: M_i \times M_j \to M_i$ is the pairwise coupling function. We denote the state of the full network as $x = (x_1, \ldots, x_N) \in M^N \equiv \prod_{i \in [N]} M_i$, and $x(t+1) = F(x(t))$. This class of networks can be generalised to higher dimensions and is common in applications such as laser dynamics [HZRM19]. We consider five assumptions on the network dynamics.

Assumption (o) The multivariate time series of all nodes $\{x(t)\}_{t \geqslant 0}^{n}$ is known. The triple (f, A, h) that defines the network dynamics in equation (1) is unknown.

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

$$\mathcal{L} = \{1\} \cup \{x_i^p\}_{i,p} \cup \left\{x_i^p x_j^q\right\}_{i,j,p,q}, \tag{2}$$

where $i, j \in [N]$ with $i \neq j$ and we remove any redundancy, $p \in [r], q \in [r-1]$, and $p + q \leqslant r$. The cardinality of $\mathcal{L}$ is given by $m = \binom{N}{2}\binom{r}{2} + Nr + 1$. We discuss the ordering of $\mathcal{L}$ in section 5.1. We say that an ordered library $\mathcal{L}$ is a network library when the functions depend at most on pairs of coordinates, see definition 1. Thus, nonzero coefficients in $\mathcal{L}$ can be identified with directed edges in the network structure.

Assumption (ii) Sparse network. We assume that the network structure represented by A is sparse. This implies a sparse representation of the network dynamics in equation (1) in terms of the network library defined in assumption (i): for each node i,

$$x_i(t+1) = \sum_{l=1}^{m} c_i^l \phi_l(x(t)) \quad \text{for all } t \geqslant 0, \tag{3}$$

where $c_i = (c_i^1, \dots, c_i^m) \in \mathbb{R}^m$ is an unknown s-sparse vector, that is, at most s of its entries are nonzero, see definition 3.

Assumption (iii) Exponential mixing. We assume (F, μ) satisfies exponential mixing conditions [HS17] for the physical measure μ: given a constant $\gamma > 0$ for all $\psi \in \mathcal{C}^1(M^N; M)$ and μ-integrable function φ, there exists $K(\psi, \varphi) > 0$ such that for any $t \geqslant 0$

$$\left| \int \psi \cdot (\varphi \circ F^t) \, \mathrm{d}\mu - \int \psi \, \mathrm{d}\mu \int \varphi \, \mathrm{d}\mu \right| \leqslant K(\psi, \varphi) \, \mathrm{e}^{-\gamma t}. \tag{4}$$

This assumption is typical for chaotic dynamical systems.

Assumption (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 5.4 for the formal definition. We assume that each marginal of ν is absolutely continuous with respect to Lebesgue, and the corresponding density is Lipschitz. In the weak coupling regime, this assumption is fulfilled [ETvSP20, Tan22]. However, this assumption also holds in other scenarios as in experimental data.

Notation. We introduce the notation $[m] := \{1, 2, \dots, m\}$. We denote $\lfloor \beta \rfloor$ as is the largest number $p \in \mathbb{N}$ satisfying $p \leqslant \beta$. We use Landau's notation $\mathcal{O}(\varepsilon)$ such that: there exists $\varepsilon_0 \geqslant 0$ and $K \geqslant 0$ such that $0 \leqslant |\mathcal{O}(\varepsilon)| \leqslant K\varepsilon$ for $0 \leqslant \varepsilon \leqslant \varepsilon_0$.

The ℓ_p-norm on $\mathbb{R}^m$ is defined for $1 \leqslant p < \infty$ as $\|u\|_p = (\sum_{l=1}^m |u_l|^p)^{1/p}$. The Euclidean inner product on $\mathbb{R}^m$ is defined by $\langle u, v \rangle = \sum_{l=1}^m u_l v_l$ for $u, v \in \mathbb{R}^m$. For a matrix $\Phi \in \mathbb{R}^{n \times m}$ and a subset $\mathcal{S} \subseteq [m]$, $\Phi_{\mathcal{S}}$ indicates the column submatrix of Φ consisting of the columns indexed by $\mathcal{S}$. We denote the transpose of Φ as Φ^{T}. We denote the inner product in $L^2(\mu)$ as $\langle \phi, \psi \rangle_\mu = \int \phi \psi \, \mathrm{d}\mu$, the induced norm as $\|\psi\|_\mu^2 = \langle \psi, \psi \rangle_\mu$, and $\|\psi\|_\infty := \sup_{x \in M^N} |\psi(x)|$. Let $\{M_i\}_{i \in [N]}$ be a collection of subsets of $\mathbb{R}$. For $\mathcal{J} \subset [N]$ denote the canonical projection by

$$\pi_{\mathcal{J}} : M^N \to \prod_{i \in \mathcal{J}} M_i. \tag{5}$$

1.2. Reconstruction problem

The network dynamics is encoded on the unknown sparse coefficient vectors $\{c_1, c_2, \dots, c_N\}$. To obtain these coefficients from the multivariate time series $\{x(t)\}_{t \geqslant 0}^n$. We recast the problem as a linear equation. Indeed, we consider the library matrix

$$\Phi(X) = \frac{1}{\sqrt{n}} \begin{pmatrix} \phi_1(x(0)) & \cdots & \phi_m(x(0)) \\ \phi_1(x(1)) & \cdots & \phi_m(x(1)) \\ \vdots & \ddots & \vdots \\ \phi_1(x(n-1)) & \cdots & \phi_m(x(n-1)) \end{pmatrix} \tag{6}$$

and arrange the trajectories into a matrix

$$\bar{X} = \begin{pmatrix} x_1(1) & \cdots & x_N(1) \\ \vdots & \ddots & \vdots \\ x_1(n) & \cdots & x_N(n) \end{pmatrix}. \tag{7}$$

We aim to find the $m \times N$ matrix of coefficients C, which has column vectors $\{c_1, c_2, \ldots, c_N\} \subset \mathbb{R}^m$ from equation (3), such that

$$\bar{X} = \Phi(X)\, C. \tag{8}$$

We identify edges from the reconstructed coefficient vectors. Here, we use the term reconstruction to refer as obtaining the network dynamics and the network structure.

When the amount of data is large in comparison to the network size, the library matrix $\Phi(X)$ might be full column rank (slim and tall) and equation (8) is overdetermined. So, unique least square solutions can be sparsely approximated [BPK16, MBPK16, WHK+18], see SI S-IV for details. For short time series, the matrix $\Phi(X)$ is at most full row rank (fat and short), then equation (8) is underdetermined, having infinite number of solutions if at least one exists. Sparse approximations can not be found by least-square, requiring another approach. One approach that exploits the sparsity information is solving for each node i the BP problem

$$(\mathrm{BP}) \quad \min_{u \in \mathbb{R}^m} \|u\|_1 \quad \text{subject to } \Phi(X)\, u = \bar{x}_i, \tag{9}$$

where $\bar{x}_i$ is the ith column $\bar{X}$. This implementation was used for networks of moderate size [WLG16, STW18, STWZ20]. For large networks, this may lead to spurious linear dependencies among the columns $\Phi(X)$ [NS08, NRP21], and (9) does not have a unique sparse solution. In figures 1(b) and (c) although the network is sparse, we show that the 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. Hence, the BP method is inappropriate for large-scale networks.

Consider the reconstruction problem for the noiseless case as in equation (8). To establish conditions for the uniqueness of the reconstruction of s-sparse solutions, successful approaches ensure that any set of $2s$ columns of $\Phi(X)$ is nearly orthonormal, what is known as the *restricted isometry property* (RIP) [CT05]. Our strategy is to introduce a new library matrix that satisfies this property for a large set of initial conditions. To do this, we first introduce a new library $\mathcal{L}_\nu$ by applying the Gram–Schmidt (GS) process in the span of $\mathcal{L}$, using an inner product $\langle \cdot, \cdot \rangle_\nu$ defined with respect to the product measure ν (from assumption iv). Then, we observe that the inner product between pairs of distinct columns of $\Phi_\nu(X)$ can be recast as Birkhoff sums between product of basis functions in $\mathcal{L}_\nu$. Since the network dynamics is ergodic and ν is close to the network physical measure μ, these inner products are small when the length of time series is sufficiently long. Finally, we use a concentration inequality to estimate the minimal length of the time series such that $\Phi_\nu(X)$ is RIP at the desired sparsity level.

This strategy is implemented with the following main results:

- *Constructing the adapted network library.* The new network library $\mathcal{L}_\nu$, which is orthonormal with respect to the product measure ν, preserves the appropriated sparsity of the original problem, see theorem 5 in section 6.2.
- *EBP.* Using this and exploring the decay of correlations of the dynamics, we establish the minimum length of time series n_0 such that $\Phi_\nu(X)$ has the desired RIP constant. Then, EBP can be formulated as a BP problem replacing the original library matrix $\Phi(X)$ by the new library matrix $\Phi_\nu(X)$. Theorem 6 shows that the reconstruction is unique, i.e. EBP exactly reconstructs the network structure, see section 7.
- *Robust reconstruction.* Finally, theorem 7 shows that the reconstruction via EBP is robust against additive measurement noise, see section 8. EBP enables us to treat the noise level as a tuning parameter to identify the network structure robustly.

Organisation of the paper. The paper is organised as follows. In section 2, we state the informal statements of the aforementioned results, describing the key steps of the proof. In section 3 we show numerical experiments in which the EBP outperforms the BP in coupled logistic maps. Section 4 shows one can apply the EBP also to data coming from experimental optoelectronic networks; this is possible because EBP has robustness against measurement noise. In section 5, we briefly recall some established results in compressive sensing [FR13], exponential mixing dynamics [HS17] and orthogonal polynomials [Sze39], providing all the necessary background in these three fields. Then, the remainder of the paper is devoted to proving the main results of the paper.

2. Main results: informal statements

2.1. Constructing the adapted network library

The first step is to introduce the new network library $\mathcal{L}_\nu$ via GS process with respect to the inner product $\langle \phi_k, \phi_l \rangle_\nu = \int_{M^N} \phi_k \phi_l \mathrm{d}\nu$. More precisely, we perform a GS process in the span of $\mathcal{L}$ and obtain a basis $\hat{\mathcal{L}} = \{\hat{\varphi}_1, \ldots, \hat{\varphi}_m\}$. We define $\varphi_i = a_i \hat{\varphi}_i$, where $a_i^2 = 1/\int \hat{\varphi}_i^2 \mathrm{d}\nu$, so the new basis $\mathcal{L}_\nu = \{\varphi_i\}_{i=1}^m$ is an orthonormal system with respect to product measure ν. We call $\mathcal{L}_\nu$ the adapted network library, which preserves the sparse representation of the network dynamics as follows:

Theorem 5 (Network library is preserved). *The GS process maps an s-sparse representation of F in $\mathcal{L}$ to an $\omega_r(s)$-sparse representation in the orthonormal network library $\mathcal{L}_\nu$, where $\omega_r(s) = \left(\lfloor \frac{r}{2} \rfloor \left(r - \lfloor \frac{r}{2} \rfloor\right) + r + 1\right) s$.*

The proof uses that the GS process is a recursive method involving projections onto preceding functions. Since ν is a product probability measure, the projections of the GS are split into products of integrals. Thus, $\mathcal{L}_\nu$ does not have functions that depend on more than two variables and characterise a network library.

Remark 1. The GS process using the measure μ to obtain a new basis leads to the loss of sparsity in the representation. Indeed, the new orthonormal basis would contain functions that depend on all coordinates because μ is not a product measure.

2.2. EBP

The library matrix associated with $\mathcal{L}_\nu$ is denoted as $\Phi_\nu(X) = \Phi(\mathcal{L}_\nu, X)$. The sth restricted isometry constant $\delta_s = \delta_s(\Phi_\nu(X))$ is defined as the smallest $\delta \geqslant 0$ such that

$$(1-\delta)\|u\|_2^2 \leqslant \|\Phi_\nu(X)u\|_2^2 \leqslant (1+\delta)\|u\|_2^2 \tag{10}$$

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 6.i ($\Phi_\nu(X)$ satisfies RIP). *Consider $d(\mu,\nu) < \zeta$ for sufficiently small ζ. For large network sizes and a large set of initial conditions if the length of time series n is at least*

$$n_0 = \mathcal{O}\left(\omega_r(s)^2 \ln(Nr)\right), \tag{11}$$

then $\Phi_\nu(X)$ satisfies (10) with $\delta_{2\omega_r(s)} \leqslant \sqrt{2}-1$.

The proof is presented in section 7 and the key steps are as follows. First, we note that the Euclidean inner product between pairs of distinct columns of $\Phi_\nu(X)$ is given by

$$\langle u_i, u_j\rangle = \frac{1}{n}\sum_{t=0}^{n-1}(\varphi_i\cdot\varphi_j)\circ(F^t(x(0))), \tag{12}$$

where u_i is the ith column of the matrix $\Phi_\nu(X)$, and the right-hand side is the Birkhoff sum of the observable $(\varphi_i\cdot\varphi_j)$. Then, we use the coherence [DH01, DET06] defined as

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

over distinct pairs of normalised $v_i = \frac{u_i}{\|u_i\|_2}$ (Euclidean norm) columns of $\Phi_\nu(X)$, where $\langle\cdot,\cdot\rangle$ is given in equation (12). Using that μ and ν are close, by triangular inequality and the Bernstein-type inequality, see [HS17], we control the coherence $\eta(\Phi_\nu)$ by approximating it by $\int\varphi_i\cdot\varphi_j\mathrm{d}\mu$. Since we know that $\delta_s \leqslant \eta(\Phi_\nu)(s-1)$ for any $s\geqslant 2$ [FR13], 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 7.1.1.

Remark 2 (The GS in $\mathcal{L}$ versus QR decomposition of $\Phi(X)$). The key idea is to control the Euclidean inner product $\langle\cdot,\cdot\rangle$ between distinct columns of the library matrix using the inner product $\langle\cdot,\cdot\rangle_\nu$ in $L^2(\nu)$. This allows us to orthogonalise functions, ensuring that $\Phi_\nu(X)$ has a nearly orthonormal set of columns under the Euclidean inner product. A QR decomposition directly on $\Phi(X)$ using the Euclidean inner product $\langle\cdot,\cdot\rangle$ is insufficient to guarantee the RIP condition or uniqueness of s-sparse solutions.

Since $\Phi_\nu(X)$ is RIP, we obtain

Theorem 6.ii (EBP has unique solution). *The convex problem that we call the EBP*

$$(\text{EBP})\quad \min_{u\in\mathbb{R}^m}\|u\|_1 \quad \text{subject to } \Phi_\nu(X)u = \bar{x}, \tag{13}$$

has a unique $\omega_r(s)$-sparse solution. That is, c_ν is the only solution of this minimisation problem when $\bar{x} = \Phi_\nu(X)c_\nu$.

The proof follows from theorem 6.i.

Remark 3 (Minimum length of time series for networks). Note that $\omega_r(s)$ is a linear function with the sparsity level s, and consequently, it is a linear function of the degree k_i of the node i. Also, $\omega_r(s) < (r+1)^2 s$. The degree distribution and the condition in (11) can be used to estimate the amount of data that ensures the network reconstruction using the EBP method. Applying equation (11) we obtain the following estimates:

- **Erdős–Rényi (ER) networks.** The degree distribution is given by a Poisson distribution, so by concentration inequality [CL06], most nodes have their degree close to the mean degree $\langle k\rangle$. Hence, to reconstruct a typical node in ER networks requires (in $\mathcal{O}(1)$) the minimum length of time series given by

$$n_0 = \mathcal{O}\left((r+1)^4 \langle k\rangle^2 \ln(Nr)\right). \tag{14}$$

  Note that $\langle k\rangle = pN$, where p is the probability of including an edge in the graph. In the phase where the ER network becomes almost sure connected, $p = K\ln N/N$ with $K \geqslant 1$ [CL06]. So,

$$n_0 = \mathcal{O}\left((r+1)^4 \ln(N)\ln(Nr)\right).$$

- **Scale-free networks.** In scale-free networks, the same growth scaling (14) is valid for low-degree nodes. However, hubs in Barabási–Albert networks have their degree proportional to $N^{\frac{1}{2}}$, so it requires

$$n_0 = \mathcal{O}\left((r+1)^4 N\ln(Nr)\right).$$

- **Regular networks.** All nodes have the same degree. So, the same growth scaling (14) is valid for any node in the network.

2.3. Robust reconstruction

We now extend the EBP to measurements corrupted by noise

$$y(t) = x(t) + z(t), \tag{15}$$

where $(z_n)_{n\geqslant 0}$ corresponds to independent and identically distributed $[-\xi,\xi]^N$-valued noise process, with probability measure ρ_ξ. The probability measure of the process $(y_n)_{n\geqslant 0}$ is the convolution $\mu_\xi := \mu * \rho_\xi$ [Fol13], 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 $\delta_{2\omega_r(s)} \leqslant \sqrt{2}-1$.

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

$$\Phi_\nu(Y) = \Phi_\nu(X) + \Lambda(X,\bar{Z}), \tag{16}$$

where $\|\Lambda(X,\bar{Z})\|_\infty \leqslant mNr^2K_1\xi$ and K_1 depends on the density of the marginals of ν. The noisy observation in (15) can be recast as a perturbed version of the orthonormal version of (8) column-wise

$$\bar{y} = \Phi_\nu(Y)c_\nu + \bar{u}, \tag{17}$$

where c_ν is the coefficient vector associated to the network library $\mathcal{L}_\nu$ and $\bar{u}$ is ℓ_2 bounded, see section 8. Thus, we can state our final result:

Theorem 7 (EBP is robust). *If the length of time series* $n \geqslant \tilde{n}_0$*, then the family of solutions* $\{c^\star_\nu(\epsilon)\}_{\epsilon>0}$ *to the convex problem (which we call the quadratically constrained EBP)*

$$\text{(QEBP)} \quad \min_{\tilde{u}\in\mathbb{R}^m} \|\tilde{u}\|_1 \text{ subject to } \|\Phi_\nu(Y)\tilde{u} - \bar{y}\|_2 \leqslant \epsilon \tag{18}$$

satisfies

$$\|c^\star_\nu(\epsilon) - c_\nu\|_2 \leqslant K_2\epsilon \tag{19}$$

for some $K_2 > 0$ as long as $\epsilon \geqslant \epsilon^\star(n, N, m, c_\nu)$, where

$$\epsilon^\star := \sqrt{n}\xi\left(1 + mNr^2K_1\|c_\nu\|_\infty\right).$$

We probe the performance of the EBP against different synthetic data of coupled logistic maps, and experimental data. In these experiments, ν is obtained using Gaussian kernel density estimation from $\{x(t)\}_{t\geqslant 0}^n$, see details in SI S-III. The GS process in the span of $\mathcal{L}$ is computed using the inner product $\langle\cdot,\cdot\rangle_\nu$ with respect to the estimated product measure ν. Then, we construct $\Phi_\nu(X)$ using the library $\mathcal{L}_\nu$.

To quantify the reconstruction performance, we use different metrics. First, the relative error $\|c^\star_\nu - c_\nu\|_2$ The second metric is defined in terms of the estimated graph encoded in $\{c_1,\ldots,c_N\}$. We create a weighted edge between node i and j using

$$W_{ij} = \max_{k\in\mathcal{S}_j} c_i^k. \tag{20}$$

We reconstruct a weighted subgraph using the node i, its neighbours, and the entry's magnitude of c_i as the edge weight; see details in the SI S-I.2. Then, we introduce a weighted false link proportion for each node. Let $\mathcal{M}_i$ and $\hat{\mathcal{M}}_i$ be the subset of edges node i shares with its neighbours of the original and estimated graph, respectively. The weights are denoted by $\{W_{ij}\}_{i,j}$. So, we calculate the proportion of false positive (FP) and false negative (FN) at node i as:

$$\begin{aligned} \mathrm{FP}_i &= \frac{\sum_{j=1}^N W_{ij}\chi_{\hat{\mathcal{M}}_i\cap\mathcal{M}_i^c}((i,j))}{\sum_{j=1}^N\left(W_{ij}\chi_{\hat{\mathcal{M}}_i\cap\mathcal{M}_i^c}((i,j)) + \chi_{\hat{\mathcal{M}}_i^c\cap\mathcal{M}_i^c}((i,j))\right)},\\ \mathrm{FN}_i &= \frac{\sum_{j=1}^N \chi_{\hat{\mathcal{M}}_i^c\cap\mathcal{M}_i}((i,j))}{\sum_{j=1}^N \chi_{\mathcal{M}_i}((i,j))}, \end{aligned} \tag{21}$$

where $\chi_\mathcal{U}$ is the indicator function of the subset $\mathcal{U}$. We denote the average over nodes of the FP and FN proportions as $\langle\mathrm{FP}\rangle$ and $\langle\mathrm{FN}\rangle$, respectively.

To quantify reconstruction performance of the network *discarding* the weights, we utilise false link proportions given by

$$\begin{aligned} \mathrm{FP} &= \frac{\text{number of edges } (i,j) \text{ with } \hat{A}_{ij} = 1 \text{ and } A_{ij} = 0}{\text{number of edges } (i,j) \text{ with } A_{ij} = 0},\\ \mathrm{FN} &= \frac{\text{number of edges } (i,j) \text{ with } \hat{A}_{ij} = 0 \text{ and } A_{ij} = 1}{\text{number of edges } (i,j) \text{ with } A_{ij} = 1}, \end{aligned} \tag{22}$$

where A and $\hat{A}$ correspond to the adjacency matrix of the original and estimate network structure, respectively. We say that the reconstruction was successful when $\text{FP}_i = \text{FN}_i = 0$ for all $i = 1, \dots, N$ or $\text{FP} = \text{FN} = 0$.

3. Numerical experiment: coupled logistic maps

To compare the reconstruction performance of the EBP against the classical BP, we consider coupled logistic maps, $f_i(x_i) = ax_i(1 - x_i)$ with $a = 3.990$, via the pairwise coupling function $h_{ij}(x_i, x_j) = x_i x_j$ with overall coupling strength $\alpha = 5 \times 10^{-4}$. Figure 1(a) illustrates a ring network with $N = 10$ nodes.

In figure 1(b), we evaluate the reconstruction performance employing the BP and the EBP as we increase the length of time series n. The convex minimisation problem is solved by employing the `CVXPY` package [DB16, AVDB18], in particular, `ECOS` solver [DCB13]. We consider the network library in (2) with the degree at most 3, so by construction, there exists a sparse representation of the network dynamics in this library. Let us denote n_0 as the length of time series such that we have a successful reconstruction, i.e. $\text{FP} = \text{FN} = 0$, see equation (22). We observe that the FP of the BP goes to zero when $n_0 \approx 400$, roughly tenfold the system size. On the other hand, EBP outperforms the BP 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 1(c), we confirm that n_0 scales with the system size for BP instead of $\ln N$ of the EBP method. In section 3.1, we demonstrate that our estimates of n_0 predict the numerical observation when we vary the maximum degree of different network structures.

3.1. Coupled logistic maps under different network structures

Here, we consider a different coupling function given by $h_{ij}(x_i, x_j) = x_j^2$ and analyse for distinct network structures, see figure 2. We observe that the EBP method outperforms the BP 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 less data to reconstruct the network structure. This phenomenon was predicted by our estimate in the expression (11). Since the maximum degree is larger, the sparsity level s of the target sparse vector is also larger, implying that n_0 grows.

3.2. Coupled logistic maps under noise

We evaluate the performance of both methods when the network trajectory is corrupted by noise, as described in equation (15). To quantify the performance, we define the relative error $E(\epsilon)$ defined to be $\|c^{\star}(\epsilon) - c\|_2$ for the quadratically BP and $\|c^{\star}_{\nu}(\epsilon) - c_{\nu}\|_2$ for the quadratically EBP (QEBP). The relative error is node-dependent, we report the average over nodes in the network, denoted as $\langle E(\epsilon) \rangle$.

The left panel in figure 3 shows the mean relative error with respect to ϵ. By theorem 7, the mean relative error should change dependence with respect to ϵ, satisfying equation (19) once ϵ is at least equal to the noise level, which is quantified by $\|\bar{u}\|_2$ in equation (17). We observe that QEBP correctly captures the change of behaviour as illustrated by the inset of the left panel. The vertical dashed line represents $\epsilon^{\star}$, the value of ϵ at which the mean relative error reaches its minimum over the interval $[10^{-4}, 10^{-2}]$. The right panel of figure 3 displays

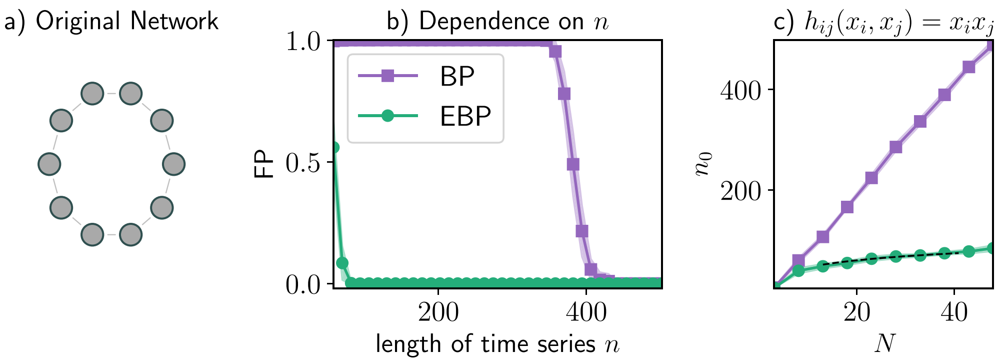

Figure 1. Ergodic Basis Pursuit performance requires only short time series. (a) Illustration of a ring graph with $N = 10$. (b) False positive (FP) of the reconstructed ring network with respect to the length of time series n for a network size $N = 40$. (c) The minimum length of time series n_0 for 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.

$\epsilon^\star$ as the length of time series n increases. The noise level $\|\bar{u}\|_2$ is computed plugging the true coefficient c_ν into equation (17). The behaviour of $\epsilon^\star$ confirms the bounds established in theorem 7.

4. Reconstruction of experimental optoelectronic networks

The data is generated from a network of optoelectronic units whose nonlinear component is a Mach–Zehnder modulator [HZRM19]. The network is modelled as

$$x_i(t+1) = \beta I_\theta(x_i(t)) - \alpha \sum_{j=1}^{17} L_{ij} I_\theta(x_j(t)) \mod 2\pi, \quad i = 1, \ldots, N, \tag{23}$$

where the normalised intensity output of the Mach–Zehnder modulator is given by $I_\theta(x) = \sin^2(x+\theta)$, x represents the normalised 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 ith node degree. The experiments were done by varying the coupling strength between the nonlinear elements in an undirected network, depicted in figure 4(a). We will show results for coupling $\alpha = 0.171875$.

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 4(b). Thus, we are naturally in the setting of (18) the randomly perturbed version of the EBP. For experimental data the noise level ξ is unknown. So, we use the constraint ϵ in (18) 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 $c_\nu^\star(\epsilon) = 0$ is a solution to (18). Next, for moderate values of ϵ, the coefficients corresponding to the isolated dynamics appear

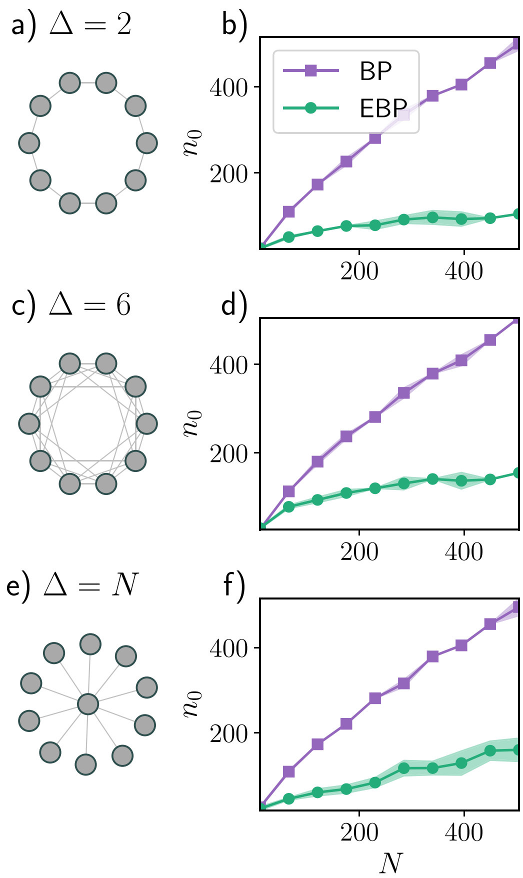

Figure 2. Comparison between BP and EBP under different network structures. (a) Ring graph with maximum degree $\Delta = 2$. (b) The minimum length of time series n_0 for a successful reconstruction versus system size N, and similarly in (d) and (f). (c) Lattice graph with maximum degree $\Delta = 6$. (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 normalised 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.

in $c^{\star}_{\nu}(\epsilon)$. 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 (18) for multiple values of ϵ while checking which entries of $c^{\star}_{\nu}(\epsilon)$ 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 SI S-II. After this procedure, we obtain a parabolic shape of the return map that corresponds to the

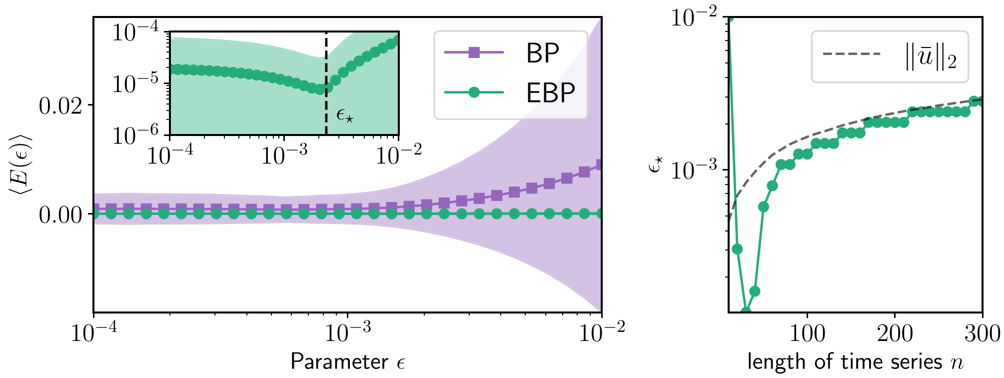

Figure 3. Ergodic Basis Pursuit is robust against noise. The left panel displays the mean relative error $E(\epsilon)$ over the set of nodes as the parameter ϵ is increased. The performance error of the quadratically Basis Pursuit (BP) grows more rapidly than the quadratically Ergodic Basis Pursuit (EBP), shown in (purple) squares and (green) circles, respectively. The inset panel shows that the $\langle E(\epsilon)\rangle$ for the EBP attains a minimum at $\epsilon^\star$. Each point corresponds to the average over 10 distinct initial conditions, and the shaded area is the standard deviation. The network is a ring with $N = 16$ nodes. The right panel shows $\epsilon^\star$ and the noise level $\|\bar{u}\|_2$ as the length of the time series is increased. The kernel density estimation of ν is used with bandwidth $\chi = 0.05$. The multivariate time series is generated without noise, and the noise is generated as i.i.d. withdrawn by a uniform distribution from the interval $[-\xi,\xi]^N$ with $\xi = 10^{-4}$.

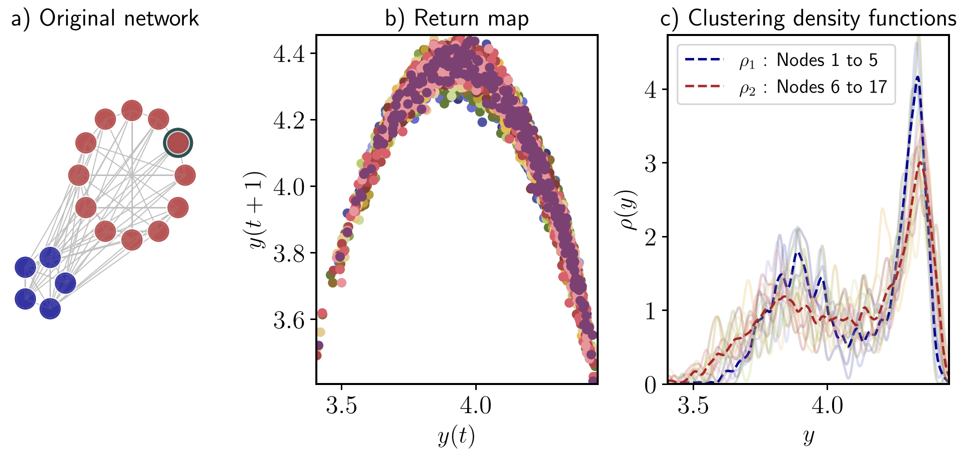

Figure 4. Network dynamics of experimental optoelectronic data. (a) Original optoelectronic network with two groups of nodes—dark grey node is marked for future reference. (b) Return map for all nodes in the network. (c) Densities function ρ_i for each node i (in light colour) estimated using each node's time series. Clustering density estimation displays two resulting densities corresponding to two groups of nodes, in blue and red. The density estimation utilises a Gaussian kernel with bandwidth $\chi = 0.05$.

restriction of the original optoelectronic network dynamics F onto the interval $\mathcal{A} = [3.4, 4.5]$ over 264-time steps, which we denote $\tilde{F} = F|_{\mathcal{A}}$. Hence, $\tilde{F}$ lies in the span of the quadratic polynomials, and we use $\mathcal{L} = \{\phi_i^p(x_i) = x_i^p : p = 0, 1, 2\}$. To perform a GS process, we estimate the ν using all trajectories of a group of nodes through kernel density estimator, improving

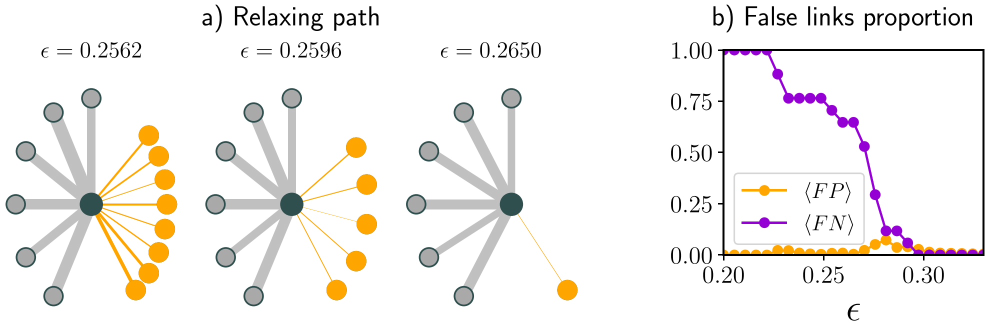

Figure 5. Reconstruction of the original network from experimental data. (a) Relaxing path algorithm is performed in the node (in dark grey) from the left panel. There are three different relaxing parameter values, where the edges are coloured accordingly: the true edges (in grey) and false positives (in orange) while the thickness is the edge weight, see equation (20). (b) The average over nodes of the false positive ⟨FP⟩ (in orange) and false negative ⟨FN⟩ (in purple) proportions of the reconstructed network versus the parameter ϵ. We varied the ϵ parameter through 25 values equally spaced in the interval $\mathcal{E} = [0.20, 0.33]$. We employ ECOS convex optimisation solver [DCB13] to solve (18).

the estimate accuracy. We assume $\mathrm{d}\nu = \prod_{i=1}^{5} \rho_1(x_i)\mathrm{d}x_i \times \prod_{i=6}^{17} \rho_2(x_i)\mathrm{d}x_i$ is a product of two densities, where ρ_1 is one-dimensional Gaussian kernel density using the trajectories of nodes 1–5, illustrated in blue, and ρ_2 is the corresponding density of the remaining nodes, illustrated in red in the right panel of figure 4.

The left panel of figure 5 displays the relaxing path algorithm probing a node (the marked dark grey node in figure 4) for three distinct ϵ values. For each ϵ, we use (20) to construct from $c_\nu^\star(\epsilon)$ the weighted subgraph corresponding to the probed node's neighbours. 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 grey). 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 equation (21), and then average over all 17 nodes. The right panel of figure 5 shows that the algorithm identifies the original network structure successfully within an interval of the parameter ϵ.

5. Mathematical analysis and preliminaries

In the remainder of this paper, we prove our main results theorems 5–7. To this end, we briefly recall some definitions and established results from compressive sensing [FR13] and exponentially mixing dynamical systems [HS17].

5.1. Network library

Consider a network dynamics in (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

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

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

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

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

Definition 1 (Network library). We call *network library* the collection of functions

$$\mathcal{L} = \{1\} \cup \mathcal{I} \cup \mathcal{P} \tag{24}$$

that represent the network dynamics map F_α in (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 $\phi_{ij}^{pq} \in \mathcal{L}$ 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 labelled coordinate on M_i. The following definition identifies the edge:

Definition 2 (Edge via network library). Let $i \in [N]$ and F_i has a representation in $\mathcal{L}$. Let $\mathcal{L}_i \subset \mathcal{L}$ be a subset that contains all necessary basis functions such that $F_i \in \operatorname{span} \mathcal{L}_i$. If $\phi_{ij}^{pq} \in \mathcal{L}_i$ 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 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 $\mathcal{I}$, which are ordered fixing the $i \in [N]$ and letting run the index $p \in [m_i]$. Finally, the set $\mathcal{P}$ is ordered, fixing an element of the index set $\{(i,j) \in [N] \times [N]\}$ (which is organised in lexicographic order) and running through the index set $\{(p,q) \in [m_i] \times [m_j]\}$ (also organised in lexicographic order), i.e.

$$\begin{aligned}\mathcal{L}^o = \{&1, \phi_1^1(x_1), \ldots, \phi_1^{m_1}(x_1), \phi_2^1(x_2), \ldots, \phi_2^{m_2}(x_2), \ldots, \phi_N^1(x_N), \ldots, \phi_N^{m_N}(x_N),\\ &\phi_{11}^{11}(x_1,x_1), \ldots, \phi_{NN}^{m_N m_N}(x_N,x_N)\}.\end{aligned} \tag{25}$$

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

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 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.

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

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 $\mathcal{L} = \{\phi_l : M^N \to \mathbb{R} : l \in [m]\}$ be the network library, where m is its cardinality.

Definition 4 (Sparse network dynamics representation). $F_\alpha : M^N \to M^N$ has an s-sparse representation in $\mathcal{L}$ 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 $c_i = (c_i^1, \ldots, c_i^m) \in \mathbb{R}^m$.

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.

Proposition 1. *The sth restricted isometry constant δ_s is given by*

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

$\Phi_{\mathcal{S}}$ is the submatrix of Φ composed by the columns supported in $\mathcal{S} \subset [m]$. Moreover, given $0 < \varepsilon < 1$ such that $\|E_{\mathcal{S}}\|_2 / \|\Phi_{\mathcal{S}}\|_2 \leqslant \varepsilon$ with $\mathcal{S} \subset [m], \mathrm{card}(\mathcal{S})$, then $\hat{\Phi} = \Phi + E$ has sth restricted isometry constant $\hat{\delta}_s = \hat{\delta}_s(\hat{\Phi})$ given by

$$\hat{\delta}_s \leqslant (1 + \delta_s)(1 + \varepsilon)^2 - 1.$$

Proof. See proof in [FR13, HS10]. □

Let the coherence of a matrix Φ be given by $\eta(\Phi) := \max_{i \neq j} |\langle v_i, v_j \rangle|$ defined over distinct pairs of normalised (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 (Coherence bounds restricted isometry constant). *If the matrix $\Phi \in M^{n \times m}$ has ℓ_2-normalised columns $\{v_1, \ldots, v_m\}$, then*

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

Proof. See proof in [FR13]. □

The uniqueness of solutions of the EBP is a consequence of the following results.

Theorem 1 (Uniqueness of noiseless recovery [Can08, FR13]). *Suppose $y = \Phi c$ where $c \in \mathbb{R}^m$ is an $s-$sparse vector. Also, suppose that the 2sth restricted isometry constant of the matrix $\Phi \in M^{n \times m}$ satisfies $\delta_{2s} < \sqrt{2} - 1$. Then c is the unique minimiser of*

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

Proof. See proof in [CT05, FR13]. □

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

Theorem 2 (Noisy recovery). *Suppose $y = \Phi c + z$ with $\|z\|_2 \leqslant \varepsilon$, and denote $c^\star$ the solution to the convex minimisation problem*

$$\min_{\tilde{u} \in \mathbb{R}^m} \|\tilde{u}\|_1 \text{ subject to } \|y - \Phi \tilde{u}\|_2 \leqslant \varepsilon. \tag{26}$$

Assume that $\delta_{2s} < \sqrt{2} - 1$. Then the solution to (26) obeys

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

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 [Can08, FR13]. □

5.3. Exponential mixing condition

We consider a class of chaotic dynamical systems—exponentially mixing systems—that satisfies a concentration inequality obtained in [HS17]. Here, we state this result applied to network dynamics.

Definition 5 (Exponential mixing condition). The network dynamics (F,μ) satisfies the exponential mixing condition for some constant $\gamma>0$ if for all $\psi\in\mathcal{C}^1(M^N;\mathbb{R})$ and $\varphi\in L^1(\mu)$ there exists a constant $K(\psi,\varphi)>0$ such that

$$\Big|\int_{M^N}\psi\cdot(\varphi\circ F^n)\,\mathrm{d}\mu-\int_{M^N}\psi\,\mathrm{d}\mu\int_{M^N}\varphi\,\mathrm{d}\mu\Big|\leqslant K(\psi,\varphi)\,\mathrm{e}^{-\gamma n},\quad n\geqslant 0. \tag{27}$$

We state an adapted version for network dynamics of the concentration inequality [HS17] for $\mathcal{C}^1(M^N;\mathbb{R})$ observables.

Theorem 3 (Bernstein inequality for exponential mixing network dynamics [HS17].). *Let (F,μ) be an exponential mixing network dynamical system on M^{N} for some constant $\gamma>0$. Moreover, let $\psi\in\mathcal{C}^1(M^N;\mathbb{R})$ be a function such that $\int_{M^N}\psi\,\mathrm{d}\mu=0$ and assume that there exist $\varsigma>0$, $\varkappa>0$ and $\sigma\geqslant 0$ such that $\|D\psi\|_\infty\leqslant\varsigma$, $\|\psi\|_\infty\leqslant\varkappa$, and $\|\psi^2\|_\mu^2\leqslant\sigma^2$. Let $\mathcal{N}\subset\mathbb{N}$ be defined as*

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

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

$$n\geqslant n_0:=\max\left\{\mathrm{e}^{\frac{3}{\gamma}},\min_{\mathcal{N}}p\right\}, \tag{28}$$

we have

$$\mu\left(x_0\in M^N\;:\;\left|\frac{1}{n}\sum_{k=0}^{n-1}\psi\circ F^k(x_0)\right|\geqslant\varepsilon\right)\leqslant 4\mathrm{e}^{-\theta(n,\varepsilon,\sigma,\varkappa)}, \tag{29}$$

where

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

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 $\mathcal{M}(M^N)$ be the set of probability measures on M^N. We introduce a probability semimetric [Rac91] between measures on $\mathcal{M}(M^N)$ over a reference finite set of functions $\mathcal{K}$ that is composed by functions on the given network library $\mathcal{L}$. In other words, elements of $\mathcal{K}$ are of the form $\phi_{ij}^{pq}\circ\pi_{\mathcal{J}}$ with $i,j\in\mathcal{J}\subset[N]$ and $\pi_{\mathcal{J}}$ is the canonical projection on the subset $\mathcal{J}$ and defined in equation (5). 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 $\mathcal{M}(M^N)$.

Definition 6. For any $\mu, \nu \in \mathcal{M}(M^N)$ we define the semimetric over a reference finite set of functions $\mathcal{K}$ as

$$d_{\mathcal{K}}(\mu,\nu) = \max_{\psi\in\mathcal{K}} \left| \int_{M^N} \psi \,\mathrm{d}\mu - \int_{M^N} \psi \,\mathrm{d}\nu \right|. \tag{30}$$

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

$$\mathcal{K} = \left\{ \psi_i : M_i \to \mathbb{R} : i \in [N], \int \psi_i \,\mathrm{d}x_i = 0, \psi_i(0) = 0 \right\},$$

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 $d_{\mathcal{K}}(\mu,\nu) = 0$ but $\mu \neq \nu$.

In what follows in section 7, it is useful to consider the following finite set $\mathcal{K} = (\mathcal{L}\cdot\mathcal{L})$, where $(\mathcal{L}\cdot\mathcal{L}) = \{(\psi_i \cdot \psi_j) : \psi_i, \psi_j \in \mathcal{L}\}$, removing any redundancy.

5.5. *Orthogonal polynomials*

We recall some results for orthonormal polynomials. First, let us state an inequality for orthonormal polynomials in one variable [Sze39, FO14]. Here we consider a system of orthonormal polynomials $\{\varphi_p(x)\}_{p\geqslant 0}$ 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 4 (One variable Korous inequality [Sze39, FO14]). *Let $\{\varphi_p(x)\}_{p\geqslant 0}$ be a generalised system of orthonormal polynomials w.r.t. the density $\lambda(x)$ and $\{\tilde{\varphi}_p(x)\}_{p\geqslant 0}$ be a system of orthonormal polynomials w.r.t. the density $\tilde{\lambda}(x)$ such that*

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

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

$$|\varphi_p(x)| \leqslant \frac{1}{\rho_0}|\tilde{\varphi}_p(x)| + \frac{K\mathrm{Lip}(\rho)}{\rho_0^{3/2}}\left(|\tilde{\varphi}_p(x)| + |\tilde{\varphi}_{p-1}(x)|\right), \tag{31}$$

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 [DX14].

Proposition 3 (Proposition 2.2.1 in [DX14]). *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 $\{\varphi_1^p(x_1)\}_{p\geqslant 0}^{\infty}$ and $\{\varphi_2^q(x_2)\}_{q\geqslant 0}^{\infty}$ 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:*

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

Furthermore, if $\{\varphi_1^p(x_1)\}_{p\geqslant 0}^{\infty}$ *and* $\{\varphi_2^q(x_2)\}_{q\geqslant 0}^{\infty}$ *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*.*

5.6. GS process

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

$$\langle \phi_k, \phi_l\rangle_\nu = \int_{M^N} \phi_k \phi_l \mathrm{d}\nu \qquad \|\phi_l\|_\nu^2 = \langle \phi_l, \phi_l\rangle_\nu. \tag{32}$$

We consider the GS process, which is a recursive method given as

$$\begin{aligned} \hat{\varphi}_1 &= \phi_1 \\ \hat{\varphi}_{k+1} &= \phi_{k+1} - \sum_{l=1}^{k} \langle \phi_{k+1}, \varphi_l\rangle_\nu \varphi_l, \\ \varphi_k &:= \frac{\hat{\varphi}_k}{\|\hat{\varphi}_k\|_\nu}, \quad k \geqslant 1. \end{aligned} \tag{33}$$

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

6. Network library is preserved under the GS process

To ensure that the EBP has a unique solution, the library matrix used in the reconstruction must satisfy the RIP, as defined in equation (10). However, *a priori*, the library matrix associated with the network library $\mathcal{L}$, in which F_α has a sparse representation, does not satisfy RIP. Our strategy is to introduce a new library $\mathcal{L}_\nu$ that is orthonormal with respect to a suitable measure ν in $L^2(\nu)$ such that the library $\Phi_\nu(X)$ satisfies RIP.

6.1. The set of pairwise polynomials of degree at most r

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

$$\mathcal{V}_r := \left\{(p,q) \in [r-1]^2 : p+q \leqslant r\right\}, \tag{34}$$

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

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

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

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

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 $\mathcal{I}_r$ is Nr because for each $i\in[N]$ there are r polynomials in the subset $\{\varphi^{p0}_{i0}\}_{p\in[r]}$. Finally, for $\mathcal{P}_r$ 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 $\mathcal{P}_r$ 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 $\{(i,j)\in[N]^2 : i\in[N], j = i+1,\ldots,N\}$ (organised in lexicographic order), we run through the exponent vector set $\mathcal{V}_r$.

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 orthonormalise with respect to μ_α itself. However, it does not necessarily preserve the sparsity of the representation of F_α in the network library $\mathcal{L}$. The next theorem states that the GS process over $\mathcal{L}$ with respect to a product measure ν introduces a new network library $\mathcal{L}_\nu$, and also, F_α is still sparsely represented in $\mathcal{L}_\nu$. In this new basis, the sparsity level depends on the maximum degree r and the sparsity level of the representation in $\mathcal{L}$.

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

$$\mathbb{E}\left(x_i^p\right) = \int_M x_i^p \,\mathrm{d}\nu_i(x_i), \quad i\in[N]. \tag{35}$$

Consider the following

Theorem 5 (Network library is preserved). *Let ν be a product measure on M^{N} that is absolutely continuous with respect to the Lebesgue measure. The Gram-Schmidt process maps an* *s-sparse representation of F_α in the network library $\mathcal{L}$ to an $\omega_r(s)$-sparse representation in the orthonormal network library $\mathcal{L}_\nu$ in $L^2(\nu)$, with*

$$\omega_r(s) = \left(\lfloor \frac{r}{2} \rfloor \left(r - \lfloor \frac{r}{2} \rfloor\right) + r + 1\right) s. \tag{36}$$

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

6.2.1. Proof of theorem 5. When we perform the GS process in $L^2(\nu)$ as in (33), to orthonormalise $\mathcal{L}$ with respect to the measure ν, the first element in $\mathcal{L}_\nu$ is evidently 1. Following the order in the network library $\mathcal{L}$ in (2), we can show that a general form of all polynomials that depend on only one variable is given by the proposition below.

Proposition 4 (Formula of orthonormal functions in one variable). *Let $i \in [N]$ and $p \in [r]$. Then any $\varphi^{p0}_{i0} \in \mathcal{L}_\nu$ is given by*

$$\begin{aligned} \hat{\varphi}^{p0}_{i0}(x_i) &= x_i^p - \mathbb{E}(x_i^p) - \sum_{l=1}^{p-1} \langle x_i^p, \varphi^{l0}_{i0} \rangle_\nu \varphi^{l0}_{i0}(x_i), \\ \varphi^{p0}_{i0}(x_i) &= \frac{\hat{\varphi}^{p0}_{i0}(x_i)}{\|\hat{\varphi}^{p0}_{i0}\|_\nu}, \end{aligned} \tag{37}$$

and

$$\mathbb{E}\left(\hat{\varphi}^{p0}_{i0}(x_i)\right) = \mathbb{E}\left(\varphi^{p0}_{i0}(x_i)\right) = 0. \tag{38}$$

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

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

and consequently, $\varphi^{10}_{10}(x_1) = \frac{\hat{\varphi}^{10}_{10}(x_1)}{\|\hat{\varphi}^{10}_{10}\|_\nu}$, which satisfies (37) and (38). To calculate the next element $\varphi^{20}_{10}(x_1)$, we follow (33):

$$\begin{aligned} \hat{\varphi}^{20}_{10}(x_1) &= \phi^{20}_{10}(x_1) - \langle \phi^{20}_{10}, 1 \rangle_\nu 1 - \langle \phi^{20}_{10}, \varphi^{10}_{10} \rangle_\nu \varphi^{10}_{10}(x_1) \\ &= x_1^2 - \mathbb{E}\left(x_1^2\right) - \langle x_1^2, \varphi^{10}_{10} \rangle_\nu \varphi^{10}_{10}(x_1), \end{aligned}$$

and consequently, $\varphi^{20}_{10}(x_1) = \frac{\hat{\varphi}^{20}_{10}(x_1)}{\|\hat{\varphi}^{20}_{10}\|_\nu}$. Following the ordering, we run over all functions of the form φ^{p0}_{10}, repeating the GS process (33) to show that they satisfy (37) and (38).

The next functions involve coordinates that are different from $i = 1$. To prove that these functions satisfy (37) and (38), we run a recursive argument. Fix $i = 1$ and $j = 2$, and let us consider the orthogonal function for $p \in [r]$ using GS process:

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

Note that if all inner products of the form $\langle x_2^p, \varphi^{l0}_{10} \rangle_\nu$ are zero, above equation satisfies (37). We state the following lemma:

Lemma 1. *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 (37) and $\varphi_{i0}^{p0} \in \mathcal{L}_\nu$. Then,*

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

whenever $i \neq j$.

Proof. By Fubini's theorem, we have that

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

Since φ_{i0}^{p0} satisfies (37), it is orthonormal to 1, and the claim holds. □

We use above lemma 1 to the inner product $\langle x_2^p, \varphi_{10}^{l0} \rangle_\nu$, where φ_{10}^{l0} satisfies (37). We conclude that for any $p \in [r]$: φ_{20}^{p0} also satisfies (37) and (38). We run iteratively, choosing $i \geqslant 2$ and $j = i+1$, and repeating the argument to conclude the proof of proposition 4.

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

Proposition 5 (Formula of orthonormal functions in two variables). *Let $r \geqslant 2$, $i,j \in [N]$ with $i \neq j$ and $(p,q) \in \mathcal{V}_r$. Then*

$$\varphi_{ij}^{pq}(x_i, x_j) = \varphi_{i0}^{p0}(x_i)\, \varphi_{j0}^{q0}(x_j). \tag{39}$$

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 3 for every distinct pair of nodes $i,j \in [N]$. □

To construct $\mathcal{L}_\nu$ we combine propositions 4 and 5. The GS process induces a set of orthonormal polynomials in one variable that satisfies the ordering of $\mathcal{L}$. The ordering of polynomials in two variables in $\mathcal{L}_\nu$ also satisfies, by construction, the ordering in $\mathcal{L}$. This proves the first part of theorem 5.

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

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

and

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

Each coordinate of u is an element of $\mathcal{L}$ that can be written as a linear combination of elements in $\mathcal{L}_\nu$. In fact, we rewrite (37) as

$$x_i^p = \|\hat{\varphi}_{i0}^{p0}\|_\nu \varphi_{i0}^{p0}(x_i) + \mathbb{E}(x_i^p) + \sum_{l=1}^{p-1} \langle x_i^p, \varphi_{i0}^{l0}\rangle_\nu \varphi_{i0}^{l0}(x_i), \tag{40}$$

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 (40) and

(i) 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 (39).
(ii) Use these identities that follow from Fubini's theorem:

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

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

and

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

Then, we can recast the GS process as the following linear equation

$$u^{\mathrm{T}} = u_\nu^{\mathrm{T}} \mathbf{R}_\nu, \tag{41}$$

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

$$\mathbf{R}_\nu = \begin{pmatrix} 1 & \mathbf{V}_1 & \mathbf{V}_2 \\ 0 & \mathbf{U}_1 & \mathbf{U}_2^1 \\ 0 & 0 & \mathbf{U}_2 \end{pmatrix}. \tag{42}$$

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 $v_{ij} \in \mathbb{R}^{\binom{r}{2}}$:

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

and

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

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:

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

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

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

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

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

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

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

Linear equation (41) is valid for every point $x \in M^N$. Hence, evaluating along the trajectory $\{x(t)\}_{t=0}^n$ we obtain:

$$\Phi(X) = \Phi_\nu(X)\mathbf{R}_\nu. \tag{43}$$

Consider an s-sparse representation in $\mathcal{L}$, 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 $\mathcal{L}_\nu$ is also a set of basis functions, so there exists a c_ν such that

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

([43](#)) 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 nonzero 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 nonzero 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 (40) for all combinations of $p, q \in [r-1]$ with $p+q \leqslant r$. More precisely, for a p in (40) there is a linear combination of $p+1$ elements of $\mathcal{L}_\nu$. Then, in the multiplication $x_i^p x_j^q$ there are at maximum

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

which has the following expression

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

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

We repeat the same argument for each $i \in [N]$ separately, concluding the proof of theorem 5.

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 $\{\psi_p(z)\}_{p\geqslant 0}$ 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

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

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

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

$$\left\{ \hat{L}_p(x) := \operatorname{sgn}\left(\frac{2}{b-a} \right)^p \left| \frac{2}{b-a} \right|^{\frac{1}{2}} L_p\left(\frac{2}{b-a}(x-b) + 1 \right) \right\}_{p\geqslant 0} \tag{44}$$

with weight $\lambda\left(\frac{2}{b-a}(x-b) + 1 \right) = 1$. Note that $\|L_p\|_\infty \leqslant 1$ [Sze39], consequently,

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

We apply the above observation to our case, using the Korous inequality for orthonormal polynomials theorem 4. See the following:

Proposition 6 (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 $\mathrm{Lip}(\rho_i)$. Moreover, $\rho_i(x_i) > 0$ for any $x_i \in M_i$. The following holds:*

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

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

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

Proof. Consider the system $\{\hat{L}_p(x)\}_{p\in[r]}$ of Legendre polynomials as in (44) 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 $\mathrm{d}\nu_i(x) = \rho_i(x_i) d\mathrm{Leb}(x_i) = \rho_i(x_i)\lambda\big(\frac{2}{b-a}(x_i - b) + 1\big)\mathrm{d}x_i$. Then, we apply Korous inequality for orthonormal polynomials theorem 4. Additionally, using Markov's inequality for polynomials

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

and the result holds. □

Corollary 1 (Supremum norm of orthonormal polynomials in two variables). *Let $r \geqslant 2$, $i,j \in [N]$ with $i \neq j$ and $(p,q) \in \mathcal{V}_r$. Then*

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

and

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

Proof.

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

and for the derivative, we calculate

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

The result holds applying proposition 6. □

From here on, for short notation, we denote

$$K = K(\mathrm{Lip}(\rho), \rho_0) \equiv \left(\frac{1}{\rho_0} + 2\frac{a_1 \mathrm{Lip}(\rho)}{\rho_0^{3/2}}\right)^2 \left(\frac{2}{b-a}\right). \tag{45}$$

7. EBP has a unique solution

In this section, we present the 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 EBP has a unique solution. Here we avoid the multi-index notation in $\mathcal{L}$ and $\mathcal{L}_\nu$ 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 $\phi_{ij}^{pq} \circ \pi_{\mathcal{J}}$ for a particular $\mathcal{J} \subset [N]$ with $i, j \in \mathcal{J}$. Also, we use the distance between probability measures introduced in section 5.4.

Theorem 6. *Let (F_α, μ_α) be an exponential mixing network dynamical system on M^{N} with decay exponent $\gamma > 0$ uniform on N. Let $\nu = \prod_{i \in [N]} \nu_i \in \mathcal{M}(M^N)$ be a product probability measure and absolutely continuous w.r.t. Lebesgue. Let $\mathcal{L}_\nu$ be the orthonormal network library with respect to ν and cardinality $m = \binom{N}{2}\binom{r}{2} + Nr + 1$. Let $\mathcal{K} = (\mathcal{L}_\nu \cdot \mathcal{L}_\nu)$ and $\omega_r(s)$ satisfies (36). Suppose that given $\alpha > 0$ there is $\zeta \in (0, \frac{\sqrt{2}-1}{4(2\omega_r(s)-1)})$ such that $d_{\mathcal{K}}(\nu, \mu_\alpha) < \zeta$ 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:*

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

$$n \geqslant K_1 \frac{(2\omega_r(s) - 1)^2}{\left(\sqrt{2} - 1 - 4\zeta(2\omega_r(s) - 1)\right)^2} \ln\left(\frac{4m(m-1)}{\lambda}\right), \tag{46}$$

for some positive constant $K_1 = K_1(\mathrm{Lip}(\rho), \rho_0)$, then $\Phi_\nu(X)$ satisfies the RIP with constant $\delta_{2\omega_r(s)} \leqslant \sqrt{2} - 1$.

(ii) *[EBP has unique solution] Consider that the length of time series n satisfies (46). 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 $\mathcal{F}_{\bar{x}} = \{w \in \mathbb{R}^m \; : \; \Phi_\nu(X) w = \bar{x}\}$. Then c_ν is the unique minimiser of the Ergodic Basis Pursuit:*

$$(\mathrm{EBP}) \quad \min_{u \in \mathcal{F}_{\bar{x}}} \|u\|_1. \tag{47}$$

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. *For sufficiently large $N > 0$, if the length of time series n satisfies*

$$n \geqslant n_0 = \frac{20 K_1 \omega_r^2(s)}{\left(\sqrt{2} - 1\right)^2} \ln(Nr) + \mathcal{O}(\zeta) + \mathcal{O}\left(\frac{1}{Nr}\right). \tag{48}$$

then with probability at least $1 - \frac{4}{Nr}$ the restricted isometry constant $\delta_{2\omega_r(s)} \leqslant \sqrt{2} - 1$.

Proof. Assume that (46) 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 $N \geqslant N_0$: $\frac{4}{Nr} \leqslant \lambda$. Then, the following holds

$$
\begin{aligned}
\ln\left(\frac{4m(m-1)}{\lambda}\right) &< \ln\left(\frac{4(Nr+1)^2}{\lambda}\right)\\
&= \ln\left(\frac{4(Nr)^4\left(1+\frac{1}{Nr}\right)^4}{\lambda}\right)\\
&\leqslant \ln(Nr)^5\left(1+\frac{1}{Nr}\right)^4\\
&= \ln(Nr)^5 + \mathcal{O}\left(\frac{1}{Nr}\right).
\end{aligned}
$$

Also, for $\zeta \in (0, \frac{\sqrt{2}-1}{4(2\omega_r(s)-1)})$, we can expand in geometric series:

$$
\begin{aligned}
\frac{1}{\left(\sqrt{2}-1-4\zeta(2\omega_r(s)-1)\right)^2} &= \frac{1}{\left(\sqrt{2}-1\right)^2\left(1-\frac{4\zeta(2\omega_r(s)-1)}{\sqrt{2}-1}\right)^2}\\
&= \frac{1}{\left(\sqrt{2}-1\right)^2}\left(1+\mathcal{O}(\zeta)\right).
\end{aligned}
$$

So, we obtain the claim. □

We split proof of theorem 6 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 1 to demonstrate that the EBP in (47) has a unique solution.

7.1. Network library matrix satisfies RIP

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

Lemma 2. *Let $\nu = \prod_{i\in[N]} \nu_i \in \mathcal{M}(M^N)$ 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 $\mathcal{L}_\nu$ be the orthonormal network library and $\mathcal{K} = (\mathcal{L}_\nu \cdot \mathcal{L}_\nu)$. Given $\alpha > 0$ and $\zeta > 0$ sufficiently small, suppose that $d_{\mathcal{K}}(\nu, \mu_\alpha) < \zeta$. Denote $(\psi_i \cdot \psi_j) = (\varphi_i \cdot \varphi_j) - \int_{M^N} (\varphi_i \cdot \varphi_j) \mathrm{d}\mu_\alpha$. Then, the following holds:*

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

where $K > 0$ is the positive constant in (45).

Proof. To prove item 1, note that:

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

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

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

By proposition 6 and corollary 1,

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

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

To prove item 2, note that given

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

Consequently, by the triangular inequality

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

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

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

□

7.1.1. Proof of theorem 6.i. The following proposition proves that the matrix $\Phi_\nu(X)$ attains the desired RIP constant once the length of time series is given by (49).

Proposition 7. *Consider the setting of theorem 6. Given $\delta \in (0, \frac{1}{2})$ and $\alpha > 0$, suppose that there is $\zeta \in (0, \frac{\delta}{4(\omega_r(s)-1)})$ such that $d_{\mathcal{K}}(\nu, \mu_\alpha) < \zeta$. Then, given $\lambda \in (0,1)$ there exists a set of initial conditions $\mathcal{G} \subset M^N$ with probability $\mu_\alpha(\mathcal{G}) \geqslant 1 - \lambda$ such that*

$$n \geqslant K_1 \frac{(\omega_r(s) - 1)^2}{(\delta - 4\zeta(\omega_r(s) - 1))^2} \ln\left(\frac{4m(m-1)}{\lambda}\right) \tag{49}$$

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

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

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

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

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

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

$$\begin{aligned}\left|\frac{1}{n}S_n(\varphi_i\cdot\varphi_j)(x_0) - \int_{M^N}(\varphi_i\cdot\varphi_j)\,\mathrm{d}\nu\right| \leqslant{}& \left|\frac{1}{n}S_n(\varphi_i\cdot\varphi_j)(x_0) - \int_{M^N}(\varphi_i\cdot\varphi_j)\,\mathrm{d}\mu_\alpha\right|\\ &+ \underbrace{\left|\int_{M^N}(\varphi_i\cdot\varphi_j)\,\mathrm{d}\mu_\alpha - \int_{M^N}(\varphi_i\cdot\varphi_j)\,\mathrm{d}\nu\right|}_{|h_{ij}|}.\end{aligned} \tag{50}$$

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

$$\begin{aligned}|h_{ij}| &= \left|\int_{M^N}(\varphi_i\cdot\varphi_j)\,\mathrm{d}\mu_\alpha - \int_{M^N}(\varphi_i\cdot\varphi_j)\,\mathrm{d}\nu\right|\\ &\leqslant d_{\mathcal{K}}(\nu,\mu_\alpha)\\ &\leqslant \zeta.\end{aligned} \tag{51}$$

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

(i) $i \neq j$: $\int_{M^N}(\varphi_i\cdot\varphi_j)\mathrm{d}\nu = 0$, consequently, using (51) we conclude that follows

$$\left|\frac{1}{n}S_n(\varphi_i\cdot\varphi_j)(x_0)\right| \leqslant \left|\frac{1}{n}S_n(\psi_i\cdot\psi_j)(x_0)\right| + |h_{ij}| \leqslant \left|\frac{1}{n}S_n(\psi_i\cdot\psi_j)(x_0)\right| + \zeta. \tag{52}$$

(ii) $i = j$: we have $\int_{M^N}\varphi_i^2\mathrm{d}\nu = 1$, and consequently, in (50), we obtain

$$\left|\frac{1}{n}S_n\left(\varphi_i^2\right)(x_0) - 1\right| \leqslant \left|\frac{1}{n}S_n\left(\psi_i^2\right)(x_0)\right| + \zeta. \tag{53}$$

By the triangular inequality, we conclude that

$$\left|\frac{1}{n}S_n\left(\varphi_i^2\right)(x_0)\right| \geqslant 1 - \left|\frac{1}{n}S_n\left(\psi_i^2\right)(x_0)\right| - \zeta. \tag{54}$$

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

$$\mathcal{K}_1 = \{(\psi_i \cdot \psi_j) \ : \ i,j = 1,\ldots,m, i \neq j\} \subset \mathcal{C}^1\left(M^N;\mathbb{R}\right) \tag{55}$$

and

$$\mathcal{K}_2 = \left\{\psi_i^2 \ : \ i = 1,\ldots,m\right\} \subset \mathcal{C}^1\left(M^N;\mathbb{R}\right). \tag{56}$$

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

$$\begin{aligned} \varkappa &:= \max\left\{\max_{i\neq j}\|(\psi_i\cdot\psi_j)\|_\infty, \max_{i\in[m]}\|\psi_i^2\|_\infty\right\}, \\ \varsigma &:= \max\left\{\max_{i\neq j}\|D(\psi_i\cdot\psi_j)\|_\infty, \max_{i\in[m]}\|D\psi_i^2\|_\infty\right\}, \\ \sigma^2 &:= \max\left\{\max_{i\neq j}\|(\psi_i\cdot\psi_j)\|_{\mu_\alpha}^2, \max_{i\in[m]}\|\psi_i^2\|_{\mu_\alpha}^2\right\}. \end{aligned} \tag{57}$$

By the Bernstein inequality in theorem 3, for $\eta > 0$ and $n \geqslant n_0(\varkappa,\varsigma,\sigma,\gamma)$, which is defined in (28), if we define

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

then

$$\mu_\alpha(\mathcal{O}_1) \leqslant 4\binom{m}{2}\mathrm{e}^{-\theta(\eta,n,\sigma,\varkappa)} \quad \text{and} \quad \mu_\alpha(\mathcal{O}_2) \leqslant 4m\mathrm{e}^{-\theta(\eta,n,\sigma,\varkappa)}.$$

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

$$\begin{aligned} \mu_\alpha(\mathcal{O}) &\leqslant 4\binom{m}{2}\mathrm{e}^{-\theta(\eta,n,\sigma,\varkappa)} + 4m\mathrm{e}^{-\theta(\eta,n,\sigma,\varkappa)} \\ &\leqslant 8\binom{m}{2}\mathrm{e}^{-\theta(\eta,n,\sigma,\varkappa)}. \end{aligned}$$

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

$$\frac{n}{(\ln n)^2} \geqslant \frac{8}{\eta^2}\left(\sigma^2 + \varkappa\frac{\eta}{3}\right)\ln\left(\frac{8}{\lambda}\binom{m}{2}\right). \tag{58}$$

Instead of (58), one usually prefers a condition that features only n on the left-hand side. First, note that whenever $n \geqslant n_0$ implies that $n \in \mathcal{N}$ in (28), and consequently, the function $t \mapsto t/(\ln t)^2$ is monotonic for values in $\mathcal{N}$. So, the condition in (58) is in fact implied by

$$n \geqslant \frac{8}{\eta^2}\left(\sigma^2 + \varkappa\frac{\eta}{3}\right)\ln\left(\frac{4m(m-1)}{\lambda}\right). \tag{59}$$

For any n satisfying the bound in (59), columns vectors u_i of $\Phi_\nu(X)$ can be normalised, $v_i = \frac{u_i}{\|u_i\|_2}$. This introduces a normalised $\hat{\Phi}_\nu(X) = \left(v_1, v_2, \ldots, v_m\right)$. So, we can estimate the coherence of the matrix $\Phi_\nu(X)$ for any $x_0 \in \mathcal{O}^c$ using (52) and (54)

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

which is valid such that $\eta+\zeta<1$.

By proposition 2 (relating coherence with RIP constant): $\hat{\delta}_{\omega_r(s)}(\hat{\Phi}_\nu) \leqslant \frac{\eta+\zeta}{1-(\eta+\zeta)}(\omega_r(s)-1)$. Moreover, note that $\Phi_\nu(X)$ and $\hat{\Phi}_\nu(X)$ are related by $\Phi_\nu(X) = \hat{\Phi}_\nu(X)\Lambda$, where $\Lambda = \mathrm{diag}(\|u_1\|_2, \|u_2\|_2, \ldots, \|u_m\|_2)$, and they satisfy $\Phi_\mathcal{S} = \hat{\Phi}_\mathcal{S}\Lambda_\mathcal{S}$. From equations (52) and (53), the following holds

$$\Lambda = \mathbf{1}_m + D, \quad D = \mathrm{diag}\left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m\right), \tag{60}$$

where $|\varepsilon_i| \leqslant \eta+\zeta$ for all $i \in [m]$. Also, $E = \Phi_\nu(X) - \hat{\Phi}_\nu(X)$. Then, by equation (60)

$$\|E_\mathcal{S}\|_2 \leqslant (\eta+\zeta)\,\|\hat{\Phi}_\mathcal{S}\|_2.$$

Using proposition 1

$$\delta_{\omega_r(s)} \leqslant \left(1+\frac{\eta+\zeta}{1-(\eta+\zeta)}\left(\omega_r(s)-1\right)\right)\left(1+(\eta+\zeta)\right)^2 - 1. \tag{61}$$

For the given $\delta \in (0, 1/2)$, we can choose any

$$\eta < \eta_0(\delta) = \frac{\delta}{4\left(\omega_r(s)-1\right)} - \zeta \tag{62}$$

as long as $\zeta \in (0, \frac{\delta}{4(\omega_r(s)-1)})$. Inequality (62) ensures that the right-hand side of the inequality (61) is upper bounded by δ:

$$\begin{aligned}&2(\eta+\zeta) + \frac{\eta+\zeta}{1-(\eta+\zeta)}\left(\omega_r(s)-1\right) + (\eta+\zeta)^2 + 2\frac{(\eta+\zeta)^2}{1-(\eta+\zeta)}\left(\omega_r(s)-1\right)\\ &\quad + \frac{(\eta+\zeta)^3}{1-(\eta+\zeta)}\left(\omega_r(s)-1\right) \leqslant \delta.\end{aligned}$$

We use lemma 2 in order to bound σ^2 and $\varkappa$ in (57). Consequently, if

$$n \geqslant \frac{8}{\eta^2}\left(\max\left\{1, K^4\right\} + (1+\zeta)^2 + \frac{2}{3}\max\left\{1, K^2\right\}\right)\ln\left(\frac{4m(m-1)}{\lambda}\right), \tag{63}$$

$\Phi_\nu(X)$ satisfies the RIP with constant $\delta_{\omega_r(s)} \leqslant \delta$. Replacing $\eta_0(\delta)$ of equation (62) in (63) and defining $K_1 := 128(\max\{1, K^4\} + 4 + \frac{2}{3}\max\{1, K^2\})$, we obtain the result. □

Lemma 3. *Let $c_\nu \in \mathbb{R}^m$ be a $\omega_r(s)-$sparse vector. If the length of time series n satisfies (46), then EBP in (47) has c_ν as its unique solution.*

Proof. Combining proposition 7 with theorem 1 suffices. □

7.2. EBP has a sufficient infeasibility condition

Since theorem 6 ensures that EBP has a unique solution, we can also prove an additional result.

Proposition 8 (Sufficient infeasibility condition). *Consider that the length of time series n satisfies (46). 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 $\mathcal{F}_{\bar{x}} = \{w \in \mathbb{R}^m : \Phi_\nu(X)w = \bar{x}\}$. Given a set $\mathcal{U} \subseteq [m]$ where $\mathcal{U} \cap \mathrm{supp}(c_\nu) \neq \emptyset$. Then $\mathcal{U} \subsetneq \mathrm{supp}(c_\nu)$ if and only if*

$$\mathcal{F}_{\bar{x}} \bigcap \{w \in \mathbb{R}^m \ : \ \mathrm{supp}\,(w) = \mathcal{U}\} = \emptyset. \tag{64}$$

Proof. Let us assume that $\mathcal{F}_{\bar{x}} \subsetneq \mathrm{supp}(c_\nu)$. We will prove this by contradiction. Suppose there is a vector $w \neq 0$ in the intersection (64) and is given by

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

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

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

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

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

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 $\mathcal{F}_{\bar{x}} \subsetneq \mathrm{supp}(c_\nu)$. Since $\mathcal{F}_{\bar{x}}$ must have an intersection with $\mathrm{supp}(c_\nu)$, then suppose that $\mathrm{supp}(c_\nu) \subseteq \mathcal{F}_{\bar{x}}$. The intersection (64) is non-empty because the sparse vector c_ν is an element of the set. This proves the claim, and the statement follows. □

8. Noise measurement case

Here, we extend the EBP to reconstruct the network from corrupted measurements

$$y(t) = x(t) + z(t), \tag{65}$$

such that $(z_n)_{n\geqslant 0}$ 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 $\mu_{\alpha,\xi} = \mu_\alpha * \varrho_\xi$ be the probability measure of the process $(y_n)_{n\geqslant 0}$ [Fol13] and the matrix $\bar{Y}$ be the noisy data

$$\bar{Y} = \begin{pmatrix} y_1(1) & \cdots & y_N(1) \\ \vdots & \ddots & \vdots \\ y_1(n) & \cdots & y_N(n) \end{pmatrix}. \tag{66}$$

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 $\mathcal{L}$ is mapped to c_{ν_ξ} which represents the network dynamics in $\mathcal{L}_{\nu_\xi}$, i.e. it satisfies $\bar{x} = \Phi_{\nu_\xi}(X)c_{\nu_\xi}$. Here, we also introduce another convex minimisation problem in terms of $\Phi_{\nu_\xi}(Y)$ evaluated along the process $(y_n)_{n\geqslant 0}$. We show that the family of solutions of this minimisation problem is parametrised 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 1. Let $\nu_\xi = \prod_{i\in[N]} \nu_{i,\xi} \in \mathcal{M}(M^N + [-\xi,\xi]^N)$ 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 7 (Noise reconstruction case). *Consider the setting of theorem 6 and hypothesis 1. Let $\mathcal{L}_{\nu_\xi} = \{\varphi_l\}_{l=1}^m$ be the orthonormal ordered network library with respect to ν_ξ and $\mathcal{K} = (\mathcal{L}_{\nu_\xi} \cdot \mathcal{L}_{\nu_\xi})$. Suppose that given $\alpha > 0$ there is $\zeta \in (0, \frac{\sqrt{2}-1}{4(2\omega_r(s)-1)} - K_1 r^2 \xi)$ such that $d_{\mathcal{K}}(\nu_\xi, \mu_{\alpha,\xi}) < \zeta$ for a positive constant $K_1 = K_1(\mathrm{Lip}(\rho_\xi), \rho_{0,\xi}, \xi)$. Then*

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

$$n \geqslant K_2 \frac{(2\omega_r(s)-1)^2}{\left(\sqrt{2}-1-4(\zeta+K_1 r^2\xi)(2\omega_r(s)-1)\right)^2} \ln\left(\frac{4m(m-1)}{\lambda}\right), \tag{67}$$

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 $\delta_{2\omega_r(s)} \leqslant \sqrt{2}-1$.

(ii) *[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 (67). Then the family of solutions $\{c^\star(\epsilon)\}_{\epsilon>0}$ to the convex problem*

$$\min_{\tilde{u}\in\mathbb{R}^m} \|\tilde{u}\|_1 \text{ subject to } \|\Phi_{\nu_\xi}(Y)\tilde{u} - \bar{y}\|_2 \leqslant \epsilon \tag{68}$$

satisfies

$$\|c^\star(\epsilon) - c_{\nu_\xi}\|_2 \leqslant K_3\epsilon \tag{69}$$

as long as

$$\epsilon \geqslant \sqrt{n}\xi\left(1 + mNr^2K_4\|c_{\nu_\xi}\|_\infty\right), \tag{70}$$

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 EBP in (47) is approximated in ℓ_2 by a family of solutions $\{c^\star(\epsilon)\}_{\epsilon\geqslant 0}$.

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 4. $\mu_{\alpha,\xi} \to \mu_\alpha$ *converges weakly as* $\xi \to 0$.

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

$$\begin{aligned}\left|\int \varphi \mathrm{d}\mu_{\alpha,\xi} - \int \varphi \mathrm{d}\mu_\alpha\right| &= \left|\int\int \varphi(x+z)\mathrm{d}\mu_\alpha(x)\mathrm{d}\varrho_\xi(z) - \int\int \varphi(x)\mathrm{d}\mu_\alpha(x)\mathrm{d}\rho_\xi(z)\right| \\ &\leqslant \int\left(\int |\varphi(x+z) - \varphi(x)|\mathrm{d}\varrho_\xi(z)\right)\mathrm{d}\mu_\alpha(x) \\ &\leqslant \int \sup_{|z|\leqslant\xi} |\varphi(x+z) - \varphi(x)|\mathrm{d}\mu_\alpha(x).\end{aligned}$$

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 $d_{\mathcal{K}}(\nu, \mu_\alpha)$. Since the product measure ν_ξ is defined on $(M^N + [-\xi,\xi]^N)$. Also, we define a variant of the constant (45) given by

$$K_\xi = K(\mathrm{Lip}(\rho_\xi), \rho_{0,\xi}, \xi) \equiv \left(\frac{1}{\rho_{0,\xi}} + 2\frac{a_1\mathrm{Lip}(\rho)}{\rho_{0,\xi}^{3/2}}\right)^2\left(\frac{2}{b-a+2\xi}\right) \tag{71}$$

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

Lemma 5. *Let* $r \geqslant 2$. *Given* $\alpha, \zeta, \xi \in (0,1)$ *suppose that* $d_{\mathcal{K}}(\nu, \mu_{\alpha,\xi}) < \zeta$ *for the product measure* $\nu_\xi \in \mathcal{M}(M^N + [-\xi,\xi]^N)$. *Then, there exists* $K_1 = K_1(\mathrm{Lip}(\rho_\xi), \rho_{0,\xi}, \xi)$ *such that*

$$d_{\mathcal{K}}(\nu, \mu_\alpha) \leqslant \zeta + K_1r^2\xi.$$

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

$$\left| \int_{M^N} \psi \circ \pi_{\mathcal{J}} \mathrm{d}\mu_{\alpha,\xi} - \int_{M^N} \psi \circ \pi_{\mathcal{J}} \mathrm{d}\mu_\alpha \right| \leqslant \int \sup_{|z| \leqslant \xi} |\psi \circ \pi_{\mathcal{J}}(x+z) - \psi \circ \pi_{\mathcal{J}}(x)| \,\mathrm{d}\mu_\alpha(x) \leqslant \mathrm{Lip}(\psi \circ \pi_{\mathcal{J}})\xi.$$

For each $\psi \in \mathcal{K}$ it corresponds to a pair $(\varphi_i \cdot \varphi_j)$, so we use proposition 6 and corollary 1 to calculate

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

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 $d_{\mathcal{K}}(\mu_{\alpha,\xi}, \mu_\alpha) \leqslant K_1 r^2 \xi$. Using the triangular inequality

$$d_{\mathcal{K}}(\nu, \mu_\alpha) \leqslant d_{\mathcal{K}}(\nu, \mu_{\alpha,\xi}) + d_{\mathcal{K}}(\mu_{\alpha,\xi}, \mu_\alpha),$$

we conclude the lemma. □

Before we proceed, we extend $\mu_\alpha \in \mathcal{M}(M^N)$ to $\mathcal{M}(M^N + [-\xi, \xi]^N)$, 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 7, making the appropriate changes. See below:

Proposition 9. *Consider the setting of theorem 7. Given $\delta \in (0, \frac{1}{2})$ and $\lambda \in (0,1)$ there exists a set of initial conditions $\mathcal{G} \subset M^N$ with probability $\mu_\alpha(\mathcal{G}) \geqslant 1 - \lambda$ such that*

$$n \geqslant K_2 \frac{(\omega_r(s) - 1)^2}{(\delta - 4(\zeta + K_1 r^2 \xi)(\omega_r(s) - 1))^2} \ln\left(\frac{4m(m-1)}{\lambda} \right) \tag{72}$$

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

Proof. The proof is similar to the proof of proposition 7. Using lemma 5 for the measures ν_ξ and μ_α, there is a constant $K_1(\mathrm{Lip}(\rho_\xi), \rho_{0,\xi}, \xi)$ such that $d_{\mathcal{K}}(\nu_\xi, \mu_\alpha) \leqslant \zeta + K_1 r^2 \xi =: \zeta'$, which we define so we can repeat the proof of proposition 7 replacing ζ by ζ'. The new bounds of n_0 can be deduced as follows: we estimate a new condition that is implied by

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

This expression can also be implied by

$$n \geqslant \frac{K_2}{\eta^2} \ln\left(\frac{4m(m-1)}{\lambda} \right), \tag{73}$$

with $K_2 := 128(\max\{1, K_\xi^4\} + (2 + K_1 r^2) + \frac{2}{3}\max\{1, K_\xi^2\})$. Using (62) replacing ζ by $\zeta + K_1 r^2 \xi$ in the above expression, we obtain the result. □

Proof of theorem 7.i. It suffices to use proposition 7 for $\delta = \sqrt{2} - 1$ and sparsity level $2\omega_r(s)$ in the expression of the length of time series in (67). □

8.2. EBP is robust against noise

We can write that $\bar{X} = \Phi_{\nu_\xi}(X)C_{\nu_\xi}$, where $C_{\nu_\xi} \in \mathbb{R}^{m\times N}$ is the coefficient matrix associated to the network library $\mathcal{L}_{\nu_\xi}$. We deduce that the noisy data in (66) satisfies

$$\bar{Y} = \Phi_{\nu_\xi}(X)\,C_{\nu_\xi} + \bar{Z}, \tag{74}$$

where

$$\bar{Z} = \begin{pmatrix} z_1(1) & \cdots & z_N(1) \\ \vdots & \ddots & \vdots \\ z_1(n) & \cdots & z_N(n) \end{pmatrix} \in [-\xi,\xi]^{n\times N}, \tag{75}$$

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

Lemma 6.

$$\Phi_{\nu_\xi}(Y) = \Phi_{\nu_\xi}(X) + \Lambda(X,\bar{Z}), \tag{76}$$

where $\|\Lambda(X,\bar{Z})\|_\infty \leqslant mNr^2K_4\xi$ *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 $\varphi_l \in \mathcal{L}_{\nu_\xi}$. The mean value theorem states that for each $t = 0,\ldots,n-1$:

$$\varphi_l(x(t)+z(t)) = \varphi_l(x(t)) + \left(\int_0^1 D\varphi_l(x(t)+sz(t))\,\mathrm{d}s\right)\cdot z(t),$$

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

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

We use proposition 6 and corollary 1 for $M^N + [-\xi,\xi]^N$. Let us denote

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

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

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

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

Proof of theorem 7.2. Using (74) and (76) we have

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

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 6 the perturbation vector $\bar{u}$ is bounded as

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

We apply theorem 2, and this concludes the proof. □

9. Conclusions

In summary, we proposed a method to reconstruct sparse networks from noisy and limited data. Our approach blends the 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 BP method.

We introduced the relaxing path algorithm that reconstructs the network as a weighted graph parametrised 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 availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.15007165 [Roq25].

Acknowledgment

We are indebted to Joseph Hart and Raj Roy for sharing the experimental data. We thank Arkady Pikovsky, Jeroen Lamb, Narcicegi Kiran, and Zheng Bian for enlightening discussions. E R S acknowledges support by FAPESP Grant 2018/10349-4 and Collaborative Research in Computational Neuroscience (CRCNS) through R01-AA029926. This work was supported by the FAPESP CEMEAI Grant No. 2013/07375-0, Serrapilheira Institute (Grant No.Serra-1709-16124), CNPq Grant 312287/2021-6, and EPSRC-FAPESP Grant No. 2023/13706.

Author contributions

All authors designed and performed research; E.R.S. wrote the code and made the figures. E R S and T P analysed data; all authors wrote the paper.

Conflict of interest

The authors declare no competing interest.

Appendix

A.1. Relaxing path algorithm for noisy data

Equation (19) 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, $\mathcal{S} = \text{supp}(c_\nu)$. Let us denote $u_\mathcal{S}$ as the vector equal to u on the index set $\mathcal{S}$ and zero on its complement $\mathcal{S}^c$. We can decompose $c^\star_\nu(\epsilon)$ into the sum of $c^\star_{\nu,\mathcal{S}}(\epsilon)$ and $c^\star_{\nu,\mathcal{S}^c}(\epsilon)$. Note that $\|c^\star_{\nu,\mathcal{S}}(\epsilon) - c_\nu\|_2^2 + \|c^\star_{\nu,\mathcal{S}^c}(\epsilon)\|_2^2 = \|c^\star_\nu(\epsilon) - c_\nu\|_2^2$ since $\mathcal{S}$ and $\mathcal{S}^c$ are disjoint, and it implies that $\|c^\star_{\nu,\mathcal{S}^c}(\epsilon)\|_2 \leqslant K_4\epsilon$. Hence, assuming the wrong entries are assigned at random, we consider that any entry of $c^\star_\nu(\epsilon)$ with a magnitude less than $\mathcal{O}(\epsilon/\sqrt{m})$ is zero.

Since the entries' magnitude supported in $\mathcal{S}^c$ are bounded by $K_2\epsilon$, we discard the irrelevant connections (to represent the node dynamics) encoded in $c^\star_\nu(\epsilon)$ 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 (17). 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:

(i) Select a set of equally spaced values ϵ_k within the interval $\mathcal{E} = [\varepsilon_{\min}, \varepsilon_{\max}]$. A pre-processing analysis can estimate the interval bounds [CCMT90].
(ii) For each $\epsilon_k \in \mathcal{E}$ find the optimal solution to the (18), the support $\mathcal{S}_k = \text{supp}(c^\star_\nu(\epsilon_k))$ and $\mathcal{T}_k = \mathcal{S}_k \Delta \mathcal{S}_{k-1}$, where Δ corresponds to the symmetric difference of the two sets and checks the change in their cardinality [FNW07].
(iii) If $|\mathcal{T}_k| = 0$, the support has not changed, then stop, and the corresponding solution $c^\star_\nu(\epsilon_k)$ is returned. Otherwise, iterate $k \mapsto k+1$ and repeat Step 2.

ORCID iDs

Edmilson Roque dos Santos https://orcid.org/0000-0002-1901-8273
Sebastian van Strien https://orcid.org/0000-0002-8079-3500

References

[AVDB18] Agrawal A, Verschueren R, Diamond S and Boyd S 2018 A rewriting system for convex optimization problems *J. Control Decis.* **5** 42–60
[BPK16] Brunton S L, Proctor J L and Nathan Kutz J 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems *Proc. Natl Acad. Sci.* **113** 3932–7

[BWB+09] Bohland J W *et al* 2009 A proposal for a coordinated effort for the determination of brain-wide neuroanatomical connectivity in model organisms at a mesoscopic scale *PLOS Comput. Biol.* **5** 1–9

[Can08] Candés E J 2008 The restricted isometry property and its implications for compressed sensing *C. R. Math.* **346** 589–92

[CCMT90] Cleveland R B, Cleveland W S, McRae J E and Terpenning I 1990 STL: a seasonal-trend decomposition procedure based on loess (with discussion) *J. Official Stat.* **6** 3–73

[CGT14] Chang Y H, Gray J W and Tomlin C J 2014 Exact reconstruction of gene regulatory networks using compressive sensing *BMC Bioinform.* **15** 400

[CL06] Chung F and Lu L 2006 *Complex Graphs and Networks* (*Cbms Regional Conference Series in Mathematics* vol 107) Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI

[CNHT17] Casadiego J, Nitzan M, Hallerberg S and Timme M 2017 Model-free inference of direct network interactions from nonlinear collective dynamics *Nat. Commun.* **8** 2192

[CT05] Candes E J and Tao T 2005 Decoding by linear programming *IEEE Trans. Inf. Theory* **51** 4203–15

[DB16] Diamond S and Boyd S 2016 CVXPY: a python-embedded modeling language for convex optimization *J. Mach. Learn. Res.* **17** 1–5

[DCB13] Domahidi A, Chu E and Boyd S 2013 ECOS: an SOCP solver for embedded systems *European Control Conf. (ECC)* pp 3071–6

[DET06] Donoho D L, Elad M and Temlyakov V N 2006 Stable recovery of sparse overcomplete representations in the presence of noise *IEEE Trans. Inf. Theory* **52** 6–18

[DH01] Donoho D L and Huo X 2001 Uncertainty principles and ideal atomic decomposition *IEEE Trans. Inf. Theory* **47** 2845–62

[DX14] Dunkl C F and Xu Y 2014 *Orthogonal Polynomials of Several Variables* (*Encyclopedia of Mathematics and its Applications*) 2nd edn (Cambridge University Press)

[ETvSP20] Eroglu D, Tanzi M, van Strien S and Pereira T 2020 Revealing dynamics, communities and criticality from data *Phys. Rev. X* **10** 021047

[FNW07] Figueiredo M A T, Nowak R D and Wright S J 2007 Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems *IEEE J. Sel. Top. Signal Process.* **1** 586–97

[FO14] Ftorek B and Orsansky P 2014 Korous type inequalities for orthogonal polynomials in two variables *Tatra Mt. Math. Publ.* **58** 1–12

[Fol13] Folland G B 2013 *Real Analysis: Modern Techniques and Their Applications* (*Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts*) (Wiley)

[FR13] Foucart S and Rauhut H 2013 *A Mathematical Introduction to Compressive Sensing* (Birkhäuser)

[Guc14] Guckenheimer J 2014 From data to dynamical systems *Nonlinearity* **27** R41

[HS10] Herman M A and Strohmer T 2010 General deviants: an analysis of perturbations in compressed sensing *IEEE J. Sel. Top. Signal Process.* **4** 342–9

[HS17] Hang H and Steinwart I 2017 A Bernstein-type inequality for some mixing processes and dynamical systems with an application to learning *Ann. Stat.* **45** 708–43

[HSWD15] Han X, Shen Z, Wang W-X and Di Z 2015 Robust reconstruction of complex networks from sparse data *Phys. Rev. Lett.* **114** 028701

[HZRM19] Hart J D, Zhang Y, Roy R and Motter A E 2019 Topological control of synchronization patterns: trading symmetry for stability *Phys. Rev. Lett.* **122** 058301

[Kur84] Kuramoto Y (ed) 1984 *Chemical Oscillations, Waves and Turbulence* (Dover)

[MBPK16] Mangan N M, Brunton S L, Proctor J L and Nathan Kutz J 2016 Inferring biological networks by sparse identification of nonlinear dynamics *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2** 52–63

[NRP21] Novaes M, dos Santos E R and Pereira T 2021 Recovering sparse networks: basis adaptation and stability under extensions *Physica D* **424** 132895

[NS08] Napoletani D and Sauer T D 2008 Reconstructing the topology of sparsely connected dynamical networks *Phys. Rev. E* **77** 026103

[PvST20] Pereira T, van Strien S and Tanzi M 2020 Heterogeneously coupled maps: hub dynamics and emergence across connectivity layers *J. Eur. Math. Soc.* **22** 2183–252

[PYGS16] Pan W, Yuan Y, Gonçalves J and Stan G-B 2016 A sparse bayesian approach to the identification of nonlinear state-space systems *IEEE Trans. Autom. Control* **61** 182–7

[Rac91] Rachev S T 1991 *Probability Metrics and the Stability of Stochastic Models* (*Wiley Series in Probability and Statistics - Applied Probability and Statistics section*) (Wiley)
[Roq25] Roque dos Santos E 2025 Ergodic Basis Pursuit *Zenodo* https://doi.org/10.5281/zenodo.15007165
[SPMS17] Stankovski T, Pereira T, McClintock P V E and Stefanovska A 2017 Coupling functions: universal insights into dynamical interaction mechanisms *Rev. Mod. Phys.* **89** 045001
[STK05] Sporns O, Tononi G and Kötter R 2005 The human connectome: a structural description of the human brain *PLoS Comput. Biol.* **1** e42
[STW18] Schaeffer H, Tran G and Ward R 2018 Extracting sparse high-dimensional dynamics from limited data *SIAM J. Appl. Math.* **78** 3279–95
[STWZ20] Schaeffer H, Tran G, Ward R and Zhang L 2020 Extracting structured dynamical systems using sparse optimization with very few samples *Multiscale Model. Simul.* **18** 1435–61
[Sze39] Szegô G 1939 *Orthogonal Polynomials* (*Colloquium Publications* vol 23) (American Mathematical Society)
[Tak06] Takens F 2006 Detecting strange attractors in turbulence *Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80* (Springer) pp 366–81
[Tan22] Tanzi M 2022 Uniformly expanding coupled maps: self-consistent transfer operators and propagation of chaos
[TW17] Tran G and Ward R 2017 Exact recovery of chaotic systems from highly corrupted data *Multiscale Model. Simul.* **15** 1108–29
[WHK+18] Wang S, Herzog E D, Kiss I Z, Schwartz W J, Bloch G, Sebek M, Granados-Fuentes D, Wang L and Li J-S 2018 Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks *Proc. Natl Acad. Sci.* **115** 9300–5
[Win01] Winfree A T 2001 *The Geometry of Biological Time* vol 12 (Springer Science & Business Media)
[WLG16] Wang W-X, Lai Y-C and Grebogi C 2016 Data based identification and prediction of nonlinear and complex dynamical systems *Phys. Rep.* **644** 1–76
[ZLZK11] Zamora-López G, Zhou C and Kurths J 2011 Exploring brain function from anatomical connectivity *Front. Neurosci.* **5** 1–11