Reconstruction of Physical Interactions in Stationary Stochastic Network Dynamics

Masterarbeit

zum Abschluss des Masterstudiengangs Physik am

Institut für Nicht-Lineare Dynamik
Fakultät für Physik
Georg-August-Universität Göttingen

geschrieben in der

Max-Planck-Forschungsgruppe
Netzwerk-Dynamik
am
Max-Planck-Institut
für Dynamik und Selbstorganisation

verfasst von

Benedict Johannes Lünsmann

Matrikelnummer:	20850142
Eingereicht am :	20. März 2015
Kontakt:	b.luensmann@stud.uni-goettingen.de
Betreuer:	Prof. Dr. Marc Timme
Erstprüfer:	Prof. Dr. Marc Timme
Zweitprüfer:	Prof. Dr. Ulrich Parlitz

Abstract

The reconstruction of interconnections in networks of interacting units is a transdisciplinary challenge that has not been met satisfactorily yet. The reconstruction of stochastic processes is of particular interest because many important and puzzling systems, like gene regulatory networks, neural circuits and the climate system, are exposed to or even driven by noise. Two approaches to reconstruct networks for these systems have been established over the last decade: The Bayesian network approach that focuses on the inference of conditional dependency networks and the thresholding approach that reconstructs effective networks of pairwise statistical dependence. Yet, it is still an open question if and how these networks reflect physical interactions of the underlying dynamical process. In this thesis, we address this fundamental question by studying the capability of both methods to infer physical interaction under idealized conditions. First, we address Bayesian inference methods and show that time-continuity leads to spurious correlations. Second, we investigate the performance of thresholding approaches and show which parameters of the system determine the reconstruction quality. And last, we introduce a new method to infer physical relationships in continuous-time random processes and test its performance on surrogate data. Our research reveals conceptual challenges of present reconstruction methods and aims on meeting them effectively by introducing a new perspective.

Contents

1. Introduction

From the gene regulatory networks in our cells to neural circuits in the brain, from power grids supplying electricity to social networks with our friends or colleagues. Networks - networks of interacting units pervade our lives. Most of them exhibit peculiar collective dynamical properties, yet we are just beginning to understand them properly.
In recent years, technological and scientific progress made it feasible to establish network science. A rapidly growing transdisciplinary branch in the science of complex systems that attempts to unravel the secretes of these irregular discrete systems. Systems that are successfully modeled by a network. A network is an abstract structure that is solely defined by entities equipped with dynamical properties and their interrelations.
The main question in this field is:

How does the structural connectivity of these networks relate to their dynamical properties?

The structural connectivity, or topology, of a network is often believed (or build) to be closely linked to its function. The most prominent example is the mammalian brain. Its astonishing properties like cognition, memory and learning are directly related to the dynamical network formed by its neurons and their physical interconnections [11]. Other examples are the gene regulatory network and the protein-protein interaction network. Both are chemical networks of inhibitory (repressive) and excitatory (inductive) influences formed by genes and proteins [10, 21]. These networks regulate large parts of the cellular metabolism and control the cell differentiation process [38]. Further examples exist in disciplines like climate science [12, 41, 42], engineering [45], epidemiology [25] and economics [19].

These examples illustrate that important systems in nature and society are strongly dependent on the topology of their underlying network structure. Yet, the structure of many networks is still unknown because (a) their connections are not directly accessible and (b) it is still an open question how to infer topologies only from dynamical features, e.g., time series. This is known as the inverse problem of network science [40]:

Does the dynamics of a system yield its topology?

For many systems even the interaction with the network is unpractical or impossible making the task of reverse engineering its topology even more challenging: Perturbing and measuring

single genes in the cellular gene regulatory network is time intensive and costly. Controlled perturbations of many neurons in a large populations is technically difficult. Significant perturbations of the earth's climate system or the stock market are close to impossible and ethically questionable. Still, knowing the structure of those systems would lead to tremendous breakthroughs like new insights into the inner workings of our brain, new medical treatments or better prediction of climate change.

Reconstruction methods often tackle two distinct types of systems: deterministic systems (e.g., [7]) and stationary stochastic processes. The reconstruction of stochastic processes is of particular interest because most systems are governed by fluctuations or noise. These fluctuations originate from hidden variables, chaos or intrinsic stochasticity and render the system stochastic in total. For some systems, like the brain or gen regulatory networks, noise is even an essential part of its function [27,39]. Present methods to inference structures in stationary random processes can be divided in two classes: Bayesian network approaches and thresholding approaches [43].

The Bayesian network approach focuses on the reconstruction of the network of conditional statistical dependencies. For this purpose, the space of possible network structures has to be searched for the best topology that explains the measured data. This can either be done by machine learning algorithms and scoring strategies [6, 14, 30] or by iterative logical reduction of the solution space [33, 35]. Since logical deduction involves estimation of high dimensional probability densities, which is problematic and computationally costly, machine learning algorithms are usually preferred in the literature. Nevertheless, both strategies face the challenge of searching a solution space that grows super-exponentially with the network size.
The Bayesian approach is quite successful for discrete-time models [46], in which the dynamical Bayesian network coincides with the structural network defined by the process. And even though it was already applied for continuous-time models in [35], it is still unclear if and how time-continuity effects Bayesian strategies.
Hence, the first question we address in this thesis is:

How does time-continuity interfere with Bayesian network reconstruction approaches?

Thresholding approaches include all forms of thresholding of measures of statistical dependence (e.g., Pearson correlation [32,41] or mutual information [5,12,29]). Among all measures, Pearson correlation is by far the most popular used in the literature [43, p. 2]. These methods generate undirected 'effective' networks. However, despite their popularity, one basic question regarding this approach has not been answered yet:

Which types of physical interactions networks are reconstructible by thresholding methods?

In this thesis, we concentrate on the reconstruction of physical interactions in stationary stochastic processes without interfering with the system dynamics.
We use the very general inductive causation algorithm, which has been proposed by Pearl [33, p. 52 ff.] and has been applied to dynamical systems by Runge et al. [35], and test its performance analytically using a continuous-time model (Sec. 3.1). It is revealed, that the reconstruction algorithm does not perform as predicted due to the time-continuity of the chosen process. But, even though the concept of a Bayesian network fails in theory to describe continuous-time dynamical systems, it might still be a good approximation for practical use.
We investigate the dynamical and topological dependencies of correlation thresholding in an idealized setting, both analytically and numerically (Sec. 3.2). Our results indicate that thresholding methods have two major downsides: First, even in the idealized case, its reconstruction performance strongly depends on the network topology. Secondly, we show that all topologies are theoretically reconstructible in the weak coupling limit for infinite time series. In practice however, statistical fluctuations caused by finite time series prevent the reconstruction precisely in this regime.
We develop a novel inference method for linear systems based on inversion of the covariance matrix (Sec. 3.3). Our approach exhibits some similarities with [8, 34] but constitutes an extension to directed networks. We show that our method greatly outperforms correlation thresholding. Moreover, unlike correlation thresholding covariance inversion is able to reconstruct directed networks and is applicable for systems with correlated noise. For infinite time series, even the noise strengths are reconstructible.

This thesis is divided in three parts: In the first part we provide the necessary theoretical foundation, introduce notations and discuss established reconstruction methods (Sec. 2). The

second part contains our findings and a detailed discussion of each result (Sec. 3). In the last part we summarize our results and list the consequences for future scientific research (Sec. 4). Moreover, since many questions had to be left unanswered, we provide ideas for further research in this field based on our deliberations.

2. Theoretical Background

In this section we present the necessary theoretic basis of graph theory, dynamical systems, stochastic processes and numerics. Furthermore, we present and discuss two state of the art reconstruction methods.

2.1. Graph Theory

It will become clear that networks with different topological features are differently difficult to reconstruct. The definitions and algorithms in this section help to discuss and to investigate this phenomenon.

Graph theory is a mathematical field that is concerned with the characteristics of graphs. Graphs are structures that laymen would call webs due to everyday life experience with arachnoid legacies. These structures consist of entities, called vertices, and their interconnections, called edges. Edges can be directed if the connection between vertices has only one direction, or undirected, if the connection is bidirectional. A graph with only directed edges is called directed and a graph with only undirected edges is called undirected.
Every undirected edge can be represented as a tuple of directed edges pointing in opposite directions. Also, every directed graph has an undirected representation which is generated by replacing all connections between two connected vertices with an undirected edge. Directed graphs are more general since the first transformation is injective and the latter transformation is surjective.
To summarize:

Definition 2.1.1 (directed graph)**.** A directed graph $G = (V, E)$ is a tuple of a set of vertices V and a set of edges E, so that every edge $e = (v, w) \in E$ is a tuple of two vertices $v, w \in V$. An edge $e = (v, w)$ denotes a directed connection from w to v.

2.1.1. Basic Definitions

Some rather basic definitions of graph theory are necessary to determine quantitative differences between the networks we are about to examine.

Adjacency Matrix

Definition 2.1.2 (adjacency matrix)**.** The adjacency matrix $A \in \{0,1\}^{N\times N}$ of a graph $G = (V, E)$ with
vertices $i \in V = \{1, \ldots, N\}$ is a $N \times N$ matrix in which every entry $A_{ij} = 1$ if and only if the edge $(i, j) \in E$, i.e. if there exists a connection from j to i. All other entries of A are zero.

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Degree and Degree Distribution

Definition 2.1.3 (degree)**.** Let $G = (V, E)$ be an undirected graph. Then the degree k_i of a vertex $i \in V$ is the number of edges attached to i.

Definition 2.1.4 (degree distribution)**.** The degree distribution p_k is the histogram of k_i over all $i \in V$.

Since the number of incoming and outgoing connections of a vertex is generally not identical in a directed graph , we define the indegree and outdegree of a vertex analogously. Both degrees have their own degree distribution.

Knowing the degree distributions of a graph is not sufficient to determine its topology. However, the distributions can serve as a valuable indicator what types of structures are present in the graph.
For instance, graphs with scale-free degree distribution (which follow a power-law for large k) possess many nodes of low degree, so called leaves, which are connected to nodes of high degree, so called hubs. These graph are likely to contain clusters or hierarchical structures.

Hence, the indegree and outdegree distribution serve as the first topological fingerprint of a graph.

Laplace Matrix

The Laplace matrix $L \in \mathbb{N}^N$ of a graph is a special representation of its adjacency matrix $A \in \{0,1\}^N$ with many useful properties.

Definition 2.1.5 (Laplace matrix)**.** The Laplace matrix L of a graph with adjacency matrix A is given by

$$L_{ij} = -A_{ij} + \delta_{ij} \sum_{l=0}^{N} A_{il} \ , \tag{2.2}$$

where δ_{ij} is the Kronecker delta.

Hence, the diagonal entries of L contain the indegree $k_i = L_{ii}$ of node i.
The Laplace matrix L is singular: $\det(L) = 0$.

Clustering Coefficient

Another useful indicator for topological features is the clustering coefficient c_i of a vertex i.

Definition 2.1.6 (clustering coefficient)**.** Let $G = (V, E)$ be an undirected graph with adjacency matrix A. Then the local clustering coefficient c_i of $i \in V$ is defined as

$$c_i = \frac{\text{numbers of connected neighbors of } i}{\text{maximal number of possible connections among neighbors of } i} \tag{2.3}$$

The clustering coefficient is a measure for the percentage of connected neighbors of i and is thus an indicator for the local connectedness of a graph. It is only defined for undirected graphs and can be computed using the adjacency matrix A via

$$\begin{aligned} c_i &= \frac{\sum_{j,k} A_{ij} A_{jk} A_{ki}}{\left(\sum_j A_{ij}\right)\left(\sum_j A_{ij} - 1\right)} \\ &= \frac{(A^3)_{ii}}{k_i^2 - k_i} \ . \end{aligned} \tag{2.4}$$

The mean clustering coefficient $c := \frac{\sum_i c_i}{|V|}$ of a graph corresponds to the probability that two neighbors of a randomly chosen vertex are connected.

Regular Topology: k-Rings

k-rings are one-dimensional periodic lattices. We introduce two different types of k-rings:

Definition 2.1.7 (undirected k-ring)**.** Let $k = 2n$ with $n \in \mathbb{N}$, then an undirected k-ring $R = (V, E)$ is a graph whose vertices can be ordered $(v_1, v_2, \ldots, v_N)$ such that every vertex is connected to the $k/2$ next and the $k/2$ preceding vertices in the ordering counted periodically.

The directed equivalent is only slightly modified.

Definition 2.1.8 (directed k-ring)**.** A directed k-ring $R = (V, E)$ is a graph whose vertices can be ordered $(v_1, v_2, \ldots, v_N)$ such that every vertex v_i has incoming connections from the next k vertices in the ordering counted periodically.

All vertices of an undirected k-ring have degree k and all vertices of a directed k-ring have in- and outdegree k.

2.1.2. Random Graphs

Random graphs are used to generate ensembles of networks with certain topological features. Here, we introduce the Erdős-Rényi model (ER), Watts-Strogatz model (WS) and Barabási-Albert model (BA).
These random graph algorithms only generate undirected graphs. Since we need to create directed graphs, we introduce appropriate modifications for each model.

Definition 2.1.9 (random graph)**.** A random graph is a probability distribution over graphs, or an algorithm how to construct a graph involving randomness.

In this sense, random graphs are implicit graphs without explicit topology.

Definition 2.1.10 (graph ensemble)**.** A set of graphs generated by the same random graph algorithm under the same conditions (e.g., using the same parameters) is called an ensemble. Equivalently, an ensemble is a set of graphs drawn from the same probability distribution over graphs.

Erdős-Rényi model

An Erdős-Rényi graph (ER graph) is the simplest version of a random graph: Every connection is present with a probability p, thus connections are statistically independent. In theory

every graph is possible but of course not equally likely.
The typical directed ER graph has binomial in- and outdegrees degrees distributions with mean $k = \frac{1}{N}\sum_i k_i = p(N-1)$ and variance $\sigma_k^2 = \frac{1}{N}\sum_i (k_i - k)^2 = p(1-p)(N-1)$.
ER graphs are well discussed in the literature (e.g., [18]).

> ER Algorithm: In this thesis, we construct directed ER graphs by distributing M connections randomly in an empty graph with N vertices.

The probability for two randomly chosen vertices i, j to be connected via a directed edge $i \leftarrow j$ is $p = \frac{M}{N(N-1)}$. This way we guarantee a predefined number of connections in the graph.
For our studies, it is irrelevant that edges in this model are not uncorrelated for finite M.

Watts-Strogatz model

Network scientists usually want to generate graphs that are very close to those observed in nature. Real networks are sometimes grid-like (e.g. regular atom structures in crystals) and few are ER-like, but most graphs are something in between. The regime between fully regular and fully random is well covered by the Watts-Strogatz model (WS model) [44].
The following random graph algorithm was introduced by Watts and Strogatz after interesting topological features of social networks had been found. Research suggest that networks of human relations posses a feature that is now well known as small-world property: a high clustering coefficient (friends of friends are often also friends) in combination with a small average shortest path length, the average distance between individuals (the famous "six degrees of separation").
Since rings show a high clustering coefficient but also large shortest path length and ER graphs display the opposite properties, the natural idea was to generate hybrids.

> WS algorithm: Start with an undirected k-ring with size N. Then, delete every edge with rewiring probability q. Afterwards, redistribute the deleted edges randomly in the graph.

For $q = 0$ the k-ring topology is not changed and for $q = 1$ an ER graph is generated. All graphs produced by intermediate q are neither fully regular nor fully random. For small

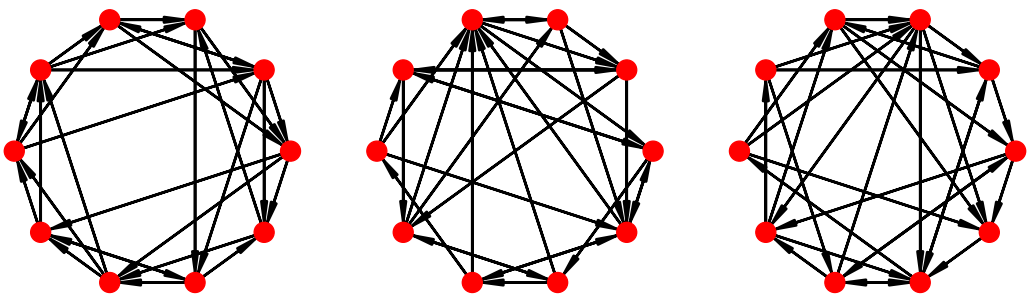
rewiring probability $0.003 < q < 0.01$ the graphs show small-world behavior. In this thesis the small-world regime is not of interest; we use the WS model to generate the full spectrum between rings and ER graphs.

The classical WS model has to be modified to generate directed graphs. Also, we like to exclude graphs whose k-ring topology was not changed, even though $q \neq 0$. We introduce the following changes:

> WS1 (WS2) algorithm: Start with a directed k-ring of size N and disconnect the source (target) of M edges. Afterwards, redistribute the loose ends randomly within the graph.
>
> WS3 algorithm: Start with a directed k-ring of size N and delete M edges. Afterwards, redistribute these edges randomly within the graph.

The WS1 model keeps the indegree constant while the WS2 model kepps the outdegree constant.

Ensembles of the WS1, WS2 and WS3 model are defined by three parameters: the number of nodes N, the connection density $\rho = \frac{k}{N-1} \in [0, 1]$ and the rewiring probability $q = \frac{M}{kN} \in [0, 1]$. Fig. 2.1 shows all three modified SW algorithms for $N = 10$ and different values of ρ and q.



(a) **(b)** **(c)**

Figure 2.1: Examples for the modified Watts-Strogatz models. graphs of size $N = 10$, density $\rho = 0.3$ and rewiring probability $q = 0.5$ generated by (a) the WS1 model , (b) the WS2 model and (c) the WS3 model.

Alternatively, it is also possible to modify the WS algorithm such that in- and outdegrees are randomized individually. By detaching targets of edges with probability q_{in} and sources with

probability q_{out} the in- and outdegree distribution are randomized differently. This defines the WS4 algorithm.

> WS4 algorithm: Start with a directed k-ring of size N. Then detach the target of M_{in} edges and the source of M_{out} edges. Then redistribute the loose ends randomly within the graph.

An ensemble of the WS4 algorithm is defined by four parameters: the number of nodes N, the connection density of the graph $\rho = \frac{k}{N-1}$ and the rewiring probabilities $q_{\text{in}} = \frac{M_{\text{in}}}{Nk}$ and $q_{\text{out}} = \frac{M_{\text{out}}}{Nk}$.

Barabási-Albert model

Another very important class of random graphs are those with power law degree distribution. These so called scale-free graph can be found in nature quite often (e.g., the Internet, social networks, protein-protein interaction networks [2]). The most popular model for generating graphs with this property is the Barabási-Albert model (BA model) [1, p. 71].
The desired shape of the distribution is achieved by introducing preferential attachment and grows: In each step an edge is added to the graph to which new edges are drawn randomly. The probability for a vertex to get a connection to the new vertex increases linearly with its degree.

> BA model: Start with a completely connected graph of N_0 nodes. Successively, add a new vertex tho the graph and connect it randomly to k vertices.
> The probability p_i to select vertex i is given by
>
> $$p_i = \frac{k_i}{\sum_j k_j} \tag{2.5}$$
>
> Stop when the graph has N vertices.

For our purposes it is necessary to extend the BA model such that directed instead of undirected graphs are created. In addition, we here introduce modifications to enable a continuous switching between ER graphs and scale-free graphs.

BA1 (BA2) model: Start with an empty graph of size N. Chose each vertex once as a the target (source) of k edges. The sources (targets) for these k edges are chosen randomly among the other $N-1$ vertices. The probability p_i to chose vertex i to be among the k sources (targets) is given by

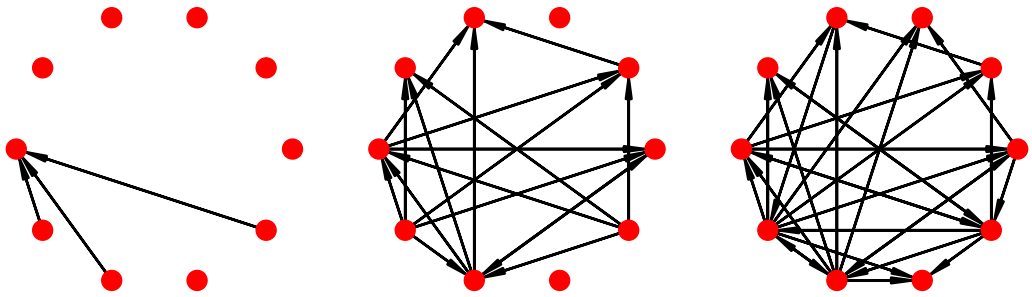
$$p_i(s) = \frac{(1-s) + s(k_{i,\text{in}} + k_{i,\text{out}})}{(1-s)(N-1) + s\sum_j (k_{j,\text{in}} + k_{j,\text{out}})} , \tag{2.6}$$

where $k_{i,\text{in}}$ is the indegree and $k_{i,\text{out}}$ is the outdegree of node i. $s \in [0,1]$ is the "scale-freeness" of the graph.
The first k connections are always chosen randomly.

The BA1 model keeps the indegree, the BA2 model the outdegree constant. For $s = 0$ an ER-like graph with constant indegree (outdegree) is generated. For $s = 1$ a scale-free graph is produced similar to the classic BA model but with a star-shaped graph in its core. Unlike the classic BA model, connectedness can not be guaranteed.
For graphs with randomized in- and outdegree, we introduce another algorithm.

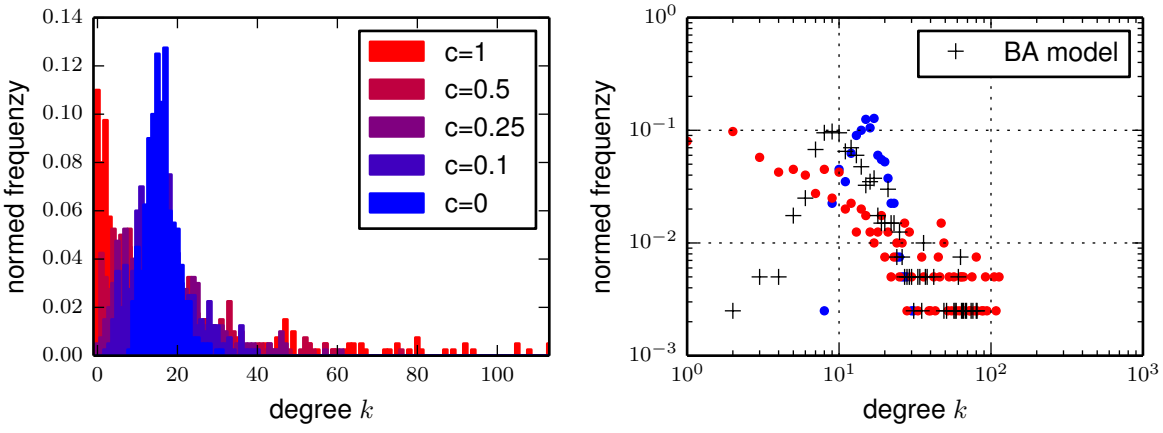
BA3 model: Generate a BA1 or BA2 graph. Flip each connection with probability $\frac{1}{2}$ if possible.

Ensembles generated by these three algorithms depend on the number of nodes N, the connection density $\rho = \frac{k}{(N-1)} \in [0,1]$ and the coefficient $s \in [0,1]$ which determines the "scale-freeness" of the graph.
Fig. 2.2 shows the *BA1* algorithm in different states. Fig 2.3 confirms that the modified BA model returns graphs with almost scale-free degree distribution.



(a) (b) (c)

Figure 2.2: Three stages of the BA1 model. Different states of the *BA1* algorithm for $(s, \rho) \approx (1, 0.3)$. (a) The first connection is chosen randomly, (b) after the fifth iteration, (b) end result after $\lfloor \rho N(N-1) \rfloor$ connections have been generated. The result displays strong clustering, a feature well known for scale-free graphs.



(a) (b)

Figure 2.3: The BA1 model returns scale-free degree distributions for $s = 1$. (a) For increasing s the outdegree distribution transforms smoothly from binomial to scale-free. (b) The log-log plot shows approximately a power law distribution for $s = 1$ (red). The BA1 model is more accurate for small degrees than the BA model (plus-markers). Shown are graphs of size $N = 400$ with a connection density of $\rho = 0.04$.

2.2. Dynamical Systems and Networks of Physical Interaction

We introduce dynamical systems and specify the concept of dynamical processes on graphs. In the literature, there is no agreement on the definition of the term "network", since its a very pictorial concept that serves multiple purposes. Hence, when speaking about network reconstruction it is necessary to clarify what type of network is meant. In this thesis, network reconstruction refers to the inference of the "network of physical interaction", a network which is defined by the graph of interactions between the state space variables of a dynamical system.

Dynamical System

A dynamical system is a tuple (T, M, Φ) of a set of times $T \subseteq \mathbb{R}$, a set M called state space or phase space, and an evolution function Φ that determines how a state $x \in M$ evolves in time

$$\Phi : \quad M \times T \to M \; . \tag{2.7}$$

This function has to fulfill the intuitive conditions

$$\Phi(x, 0) = x \quad \Phi(\Phi(x, t_1), t_2) = \Phi(x, t_1 + t_2) \tag{2.8}$$

associated with time evolution. Given an initial state $x_0 \in M$ and a time $t \in T$ the evolution function $\Phi(x, t) = x_t$ returns the state $x_t \in M$ that x_0 evolved into after the time t has passed. Classical examples of dynamical systems are iterated maps and systems of first order ordinary differential equations (ODEs).

A system of first order ODEs is a function $\varphi(x)$ that defines the derivative of $x \in M$ in respect to time $t \in \mathbb{R}$

$$\dot{x} := \frac{\mathrm{d}}{\mathrm{d}t} x = \varphi(x) \; . \tag{2.9}$$

Any system of ODEs can be reformulated as a system of first order ODEs.

The derivative φ defines an evolution function via

$$\Phi(x, t) = x + \int_0^t \varphi(x) \mathrm{d}t \; . \tag{2.10}$$

Networks of Physical Interactions

Consider a dynamical system $D = (\mathbb{R}, \mathbb{R}^N, \Phi)$ defined by a system of first order ODEs $\dot{\boldsymbol{x}} = \varphi(\boldsymbol{x})$ which can be expressed such that

$$\dot{x}_i = \varphi(\boldsymbol{x})_i = f_i(x_i) + \sum_{i \neq j} A_{ij} g_{ij}(x_i, x_j) \, . \tag{2.11}$$

f_i is called the auto-dependency function of x_i, $g_{ij} \neq 0$ is called the coupling function of x_i and x_j, and $A \in \{0,1\}^{N \times N}$ is called adjacency matrix.

Then, we define a graph $G = (V, E)$ with a vertex set $V = \{1, \ldots, N\}$ and an edge set E in which each vertex $i \in V$ is associated with exactly one element x_i of the state space $(x_1, \ldots, x_N) \in \mathbb{R}^N$. Two vertices in $i, j \in V$, $i \neq j$, are connected via a directed edge $i \leftarrow j$ if and only if the evolution of x_i depends explicitly on x_j, i.e. $A_{ij} \neq 0$. Hence, A is the adjacency matrix of the graph G.

In this ways, the dynamical system D can be understood as a dynamical process on the graph G. G is called the network of physical interactions or structural network of the dynamical system D.

The simplest class of networks are defined by systems of linear first order ODEs

$$\dot{\boldsymbol{x}} = J\boldsymbol{x} \, , \quad J \in \mathbb{R}^{N \times N} \, , \tag{2.12}$$

$$\dot{x}_i = \underbrace{(J_{ii} + \sum_{j \neq i} J_{ij}) x_i}_{f_i(x_i)} + \sum_{j \neq i} \underbrace{J_{ij}(x_j - x_i)}_{A_{ij} g_{ij}(x_i, x_j)} \, , \tag{2.13}$$

where J is the Jacobian matrix

$$J_{ij} = \frac{\partial(\dot{x}_i)}{\partial x_j} \, . \tag{2.14}$$

Despite their simplicity, these systems are of general importance, because according to the Hartman-Grobman theorem any dynamical system can be approximated by a linear dynamical system in the vicinity of a fixed point, a point x_0 in state space with $\varphi(x_0) = 0$ [37].

2.3. Probability and Stochasticity

The reader will find definitions and examples of the most basic concepts of probability and stochastic theory in this section. This summary includes the basic knowledge for the understanding and the analysis of random processes.

2.3.1. Random Variable

A random variable is a well defined mathematical concept. However, like many mathematical concept, its exact definition involves an extensive explanation of the structures in which it is embedded (sample space, σ-algebra, etc.). For this reason, instead of providing detailed explanations, we refer to a standard textbook (e.g., [24]) and introduce the term "random variable" in a physically more useful manner.

Let Ω be a set of atomic outcomes of a random experiment, an experiment involving randomness. Then any subset $\omega \subseteq \Omega$ is called an event, it is equipped with a probability $P(\omega) \in [0,1]$. A random variable X is a function mapping an event ω to a numerical representation $X(\omega) = x$, e.g. $x \in \mathbb{N}$, $x \in \mathbb{R}$ or $x \in \mathbb{R}^3$. This way we can assign probabilities for the numerical values x.

For example, imagine two dice. The atomic outcomes of throwing two dice are the combination of their pips, hence, $\Omega = \{\{\boxdot,\boxdot\},\{\boxdot,\boxdot\},\ldots,\{\boxplus,\boxplus\},\{\boxplus,\boxplus\}\}$. The probability of the outcome $\omega = \Omega/\{\boxplus,\boxplus\}$ is $P(\Omega/\{\boxplus,\boxplus\}) = \frac{35}{36}$. The sum of pips x shown after a throw is a random variable X. The probability for $x < 12$ is $P(x<12) = P(X(\Omega/\{\boxplus,\boxplus\})) = \frac{35}{36}$.

In this thesis the set of atomic outcomes Ω is usually uncountable and random variables X are usually continuous. Hence, probabilities for atomic outcomes $\omega \in \Omega$ are zero for any function P. In this case, one defines a probability density $p(X(\omega)=x)$ for an appropriate subset $A \subseteq \Omega$.

$$P(\omega \in A) = \int\limits_{\omega' \in A} p(\omega')\,\mathrm{d}\omega' \quad \Rightarrow \quad P(x \in X(A)) = \int\limits_{x' \in X(A)} p(x')\,\mathrm{d}x' \tag{2.15}$$

As from now on neither the meaning nor the structure of Ω is of any interest. X itself has become a random experiment defined by its probability density $p(X=x)$ or short $p(x)$.

2.3.2. Probabilities and Statistic Independence

Let X, Y be two random variables, then the probability for both $X = x$ *and* $Y = y$ is called joint probability $P(x, y)$ while the probabilities for each $X = x$ and $Y = y$ individually are called marginal probabilities $P(x)$ and $P(y)$.
The probability for $X = x$ given $Y = y$ is called conditional probability $P(x|y)$.
Joint, marginal and conditional probability are related by

$$P(x, y) = P(x|y) \cdot P(y) = P(y|x) \cdot P(x) \; . \tag{2.16}$$

This important relationship is often called the *chain rule* of probability calculus and can be generalized to

$$P(x_1, x_2, \ldots, x_n) = P(x_n) \cdot \prod_{i=1}^{n-1} P(x_i | x_{i+1}, \ldots, x_n) \; . \tag{2.17}$$

Definition 2.3.1 (statistic independence)**.** Two random variables X and Y are called statistically independent (short: $X \perp\!\!\!\perp Y$) if and only if $P(x|y) = P(x)$ and thus if their joint probability factorizes.

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad P(x, y) = P(x) \cdot P(y) \tag{2.18}$$

Definition 2.3.2 (conditional independence)**.** Two random variables X, Y are conditionally independent given a third random variable if their joint conditional probability factorizes.

$$(X \perp\!\!\!\perp Y) \mid Z \quad \Leftrightarrow \quad P(x, y|z) = P(x|z) \cdot P(y|z) \tag{2.19}$$

Only continuous random variables will be considered in this thesis. It is necessary to formulate the statements made above for probability densities instead of probabilities.
Fortunately, it is easy to show that two random variables X, Y are statistically independent

if and only if their joint probability distribution factorizes:

$$A, B \subseteq \Omega \tag{2.20}$$

$$\begin{aligned} P(x \in X(A), y \in Y(B)) &= \int_{X(A)} \int_{Y(B)} p(x,y)\, \mathrm{d}y\, \mathrm{d}x \\ &= \int_{X(A)} \int_{Y(B)} p(x)p(y)\, \mathrm{d}y\, \mathrm{d}x \\ &= \int_{X(A)} p(x)\, \mathrm{d}x \int_{Y(B)} p(y)\, \mathrm{d}y = P(x \in X(A)) \cdot P(y \in Y(B)) \end{aligned} \tag{2.21}$$

Hence, all statements made for the probability P, are also true for the probability density p.

2.3.3. Stochastic Processes and Markov Property

A random variable depending on time is called a stochastic process.
Let $X(t)$ be such a stochastic process. This means for times $(t_1, t_2, \ldots)$ with $t_1 > t_2 > \ldots$ $X(t)$ produces a sequence $(x_1, x_2, \ldots)$. The behavior of $X(t)$ is fully described by the joint probability density $p(x_1, t_1; x_2, t_2; ...)$. This also defines a conditional probability

$$p(x_1, t_1; x_2, t_2; \ldots; x_n, t_n | y_1, \tau_1; y_2, \tau_2 \ldots) = \frac{p(x_1, t_1; x_2, t_2; \ldots; x_n, t_n; y_1, \tau_1; y_2, \tau_2; \ldots)}{p(y_1, \tau_1; y_2, \tau_2; \ldots)} \tag{2.22}$$

where for all n, $t_n > \tau_1 > \tau_2 > \ldots$. We conceive the conditional probability as the probability of $X(t)$ to be x_1 at t_1, x_2 at t_2 and so on, on condition that $X(t)$ was y_1 at τ_1, y_2 at τ_2 et cetera.
In many cases it is sufficient to reduce the number of former states the conditional probability depends on. This means that the future states of $X(t)$ do not depend on the whole history of the random variable but only on a finite subset of k former states.

$$p(x_1, t_1; \ldots; x_n, t_n | y_1, \tau_1; y_2, \tau_2 \ldots) = p(x_1, t_1; \ldots; x_n, t_n | y_1, \tau_1; \ldots; y_k, \tau_k) \tag{2.23}$$

This feature is called kth order Markov property and the corresponding processes kth order Markov processes.
In physics dynamic systems are usually defined by differential equations. Since the deter-

ministic trajectories of these systems do not cross in state space, the future evolution is fully determined by one point in the past. Hence, these processes are first order Markov processes.

$$p(x_1, t_1; x_2, t_2; ...; x_n, t_n | y_1, \tau_1; y_2, \tau_2 \ldots) = p(x_1, t_1; x_2, t_2; ...; x_n, t_n | y_1, \tau_1) \tag{2.24}$$

2.3.4. Average

The average or mean of a function $f(X)$ of a random variable X is computed by integrating over all possible states weighted by their probability, i.e.

$$\langle f(X) \rangle_x := \int f(x) p(x) \, \mathrm{d}x \, . \tag{2.25}$$

$\langle \cdot \rangle_x$ is called average over state space. A stationary random process $X(t)$ is called ergodic, if the time average $\langle \cdot \rangle_t$ equals the average over state space $\langle \cdot \rangle_x$ for all functions $f(X(t))$. Hence,

$$\langle f(X(t)) \rangle_t = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(x(t)) \, \mathrm{d}t \tag{2.26}$$

$$= \int f(x(t)) p(x) \, \mathrm{d}t = \langle f(X(t)) \rangle_x \, . \tag{2.27}$$

2.3.5. Langevin equation, White Noise and Stochastic Differential Equation

Named after the French physicist Paul Langevin (*1872, †1946), the Langevin equation is a system of first order differential equations involving randomness. It is a so called stochastic differential equation (SDE).
For $\boldsymbol{x} \in \mathbb{R}^N$, the most general form can be written as

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, t) + g(\boldsymbol{x}, t) \boldsymbol{\eta}(t) \tag{2.28}$$

with $g(\boldsymbol{x}, t) \in \mathbb{R}^{N \times N}$. $\boldsymbol{\eta}(t) \in \mathbb{R}^N$ is a "white noise" vector, that is a strongly fluctuating multivariate random process with zero mean and an autocorrelation of

$$\langle \eta_i(t) \eta_j(t) \rangle_x = \delta_{ij} \delta(t - t') \tag{2.29}$$

where δ_{ij} is the Kronecker delta, $\delta(t-t')$ is the δ-distribution and $\langle\cdot\rangle_x$ is the average over state space. Since the covariance of this non-continuous process diverges, $\boldsymbol{\eta}(t)$ is ergodic.
The Langevin equation for x defines a random process X.

SDEs are often written in differential form. The differential form of the Langevin equation is

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x},t)\mathrm{d}t + g(\boldsymbol{x},t)\mathrm{d}\boldsymbol{W}(t)\ , \tag{2.30}$$

where $\mathrm{d}\boldsymbol{W}(t) := \boldsymbol{\eta}(t)\mathrm{d}t$. $\boldsymbol{W}(t)$ is the multivariate Wiener process (see below).

2.3.6. Wiener Process

The Wiener process is the basis of many stochastic differential equations. It is named after the American mathematician Norbert Wiener (*1894, †1964).
We need the properties of the Wiener process for numerical simulations.

The Wiener process is defined by the Langevin equation

$$\dot{W} = \eta(t) \quad \Rightarrow \quad W(t) = w_0 + \int_{t_0}^{t} \eta(t')\ \mathrm{d}t'\ . \tag{2.31}$$

in which $\eta(t)$ is white noise. Hence, the Wiener process is integrated white noise.
Since $|\eta(t)| < \infty$, $W(t)$ is a continuous random process. The evolution of its probability distribution is expressed by the partial differential equation (a Fokker-Planck equation) [15, Eq. 3.8.1]

$$p(w,t|w_0,t_0) = \frac{1}{2}\partial_w^2\ p(w,t|w_0,t_0) \tag{2.32}$$

also known as diffusion equation.
For precisely known initial conditions $p(w_0,t_0) = \delta(w-w_0)$ the solution of this partial differential equation is [15, Eq. 3.8.7]

$$p(w,t|w_0,t_0) = \frac{1}{\sqrt{2\pi(t-t_0)}} e^{-\frac{1}{2}\frac{(w-w_o)^2}{(t-t_0)}}\ . \tag{2.33}$$

Thus, the probability density function of the Wiener process is a Gaussian with variance $\sqrt{t - t_0}$. Since the variance of this process diverges as $t \to \infty$, the Wiener process is ergodic.

2.3.7. The Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck process (OU process) is a continuous random process named after the Dutch physicists Leonard Ornstein (*1880, †1941) and Georg Uhlenbeck (*1900, †1988). It describes a linear dynamical system driven by white noise. Hence, it is a first order approximation in the vicinity of a fixed point of any dynamical system that is exposed to rapid fluctuations. For this reason, the OU process is of general importance. It has been used to model stock market dynamics [31] and gene-regulatory networks [13].

The OU process for $\boldsymbol{x} \in \mathbb{R}^N$ is defined by the Langevin equation

$$\dot{\boldsymbol{x}} = J\boldsymbol{x} + B\boldsymbol{\eta}(t), \tag{2.34}$$

here $J, B \in \mathbb{R}^{N \times N}$ are constant matrices. Hence, the process is linear. J is the Jacobi matrix of the underlying dynamical system. B is called noise input matrix.
For known initial conditions $\boldsymbol{x}_0 := \boldsymbol{x}(t_0)$, the process has an analytic solution [15, Eq. 4.4.43]

$$\boldsymbol{x}(t) = e^{J(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} e^{J(t-t')} B \, \mathrm{d}\boldsymbol{W}(t') \,, \tag{2.35}$$

where $\boldsymbol{W}(t)$ is the Wiener process (see Sec. 2.3.6).
If the process initially has a Gaussian shaped probability density, all higher order momenta of the probability distribution vanish. Thus, it will stay Gaussian. The same holds for deterministic initial conditions. Also, the Ornstein-Uhlenbeck is ergodic, because all states of state space can be reached from all initial conditions.
Next, we derive joint and conditional probability densites of the OU process.

Mean

The mean of $\boldsymbol{X}(t)$ is computed using Eq. 2.35. If the mean of $\boldsymbol{X}_0 := \boldsymbol{X}(t_0)$ is $\boldsymbol{x}_0$, the mean of $\boldsymbol{X}(t)$ is given by

$$\langle \boldsymbol{X}(t) \rangle_{\boldsymbol{x}} = e^{J(t-t_0)} \underbrace{\langle \boldsymbol{X}_0 \rangle_{\boldsymbol{x}}}_{\boldsymbol{x_0}} + \int_{t_0}^{t} e^{J(t-t')} B \underbrace{\langle \boldsymbol{\eta}(t') \rangle_{\boldsymbol{x}}}_{=0} \, \mathrm{d}t' = U_t \boldsymbol{x}_0 \, . \tag{2.36}$$

Here, we defined $U_t := e^{Jt}$.

Covariance Matrix

The conditional covariance matrix $\Sigma(t, t \,|\, t_0)$ is the covariance matrix of $\boldsymbol{X}(t)$ on condition that the probability distribution of $\boldsymbol{X}_0 := \boldsymbol{X}(t_0)$ has a covariance matrix of $\Sigma(t_0)$. It has an analytic solution which is again calculated using Eq. 2.35:

$$\begin{aligned}
\Sigma(t, t \,|\, t_0) &= \langle \boldsymbol{X}(t) \boldsymbol{X}(t)^\intercal \rangle_{\boldsymbol{x}} - \langle \boldsymbol{X}(t) \rangle_{\boldsymbol{x}} \langle \boldsymbol{X}(t)^\intercal \rangle_{\boldsymbol{x}} \\
&= U_{t-t_0} \underbrace{\left(\langle \boldsymbol{X}_0 \boldsymbol{X}_0^\intercal \rangle_{\boldsymbol{x}} - \langle \boldsymbol{X}_0 \rangle_{\boldsymbol{x}} \langle \boldsymbol{X}_0^\intercal \rangle_{\boldsymbol{x}} \right)}_{\Sigma_0} U_{t-t_0}^\intercal + \int_{t_0}^{t} \underbrace{\langle \boldsymbol{X_0 \eta}^\intercal(t') \rangle_{\boldsymbol{x}}}_{=0} B^\intercal U_{t-t'}^\intercal \, \mathrm{d}t' \\
&\quad + \int_{t_0}^{t} U_{t-t'} B \underbrace{\langle \boldsymbol{\eta}(t') \boldsymbol{X_0}^\intercal \rangle_{\boldsymbol{x}}}_{=0} \, \mathrm{d}t' + \int_{t_0}^{t} \int_{t_0}^{t} U_{t-t'} B \underbrace{\langle \boldsymbol{\eta}(t') \boldsymbol{\eta}(t'')^\intercal \rangle_{\boldsymbol{x}}}_{=\delta(t'-t'')\mathbb{1}} B^\intercal U_{t-t''}^\intercal \, \mathrm{d}t' \mathrm{d}t'' \\
&= U_{t-t_0} \Sigma(t_0) U_{t-t_0}^\intercal + \int_{t_0}^{t} U_{t-t'} B B^\intercal U_{t-t'}^\intercal \, \mathrm{d}t' \, ,
\end{aligned} \tag{2.37}$$

where the matrix $U_t := e^{Jt}$ is defined as in the previous paragraph.
The delayed covariance matrix $\Sigma(t + \tau, t \mid t_0)$, or autocorrelation function, is the covariance matrix between $\boldsymbol{X}(t)$ and its future values $\boldsymbol{X}(t + \tau)$. If the probability distribution at $t = t_0$ has a covariance matrix of $\Sigma(t_0)$, it is given by

$$\begin{aligned}
\Sigma(t + \tau, t \mid t_0) &= \langle \boldsymbol{X}(t + \tau) \boldsymbol{X}(t)^\intercal \rangle_{\boldsymbol{x}} - \langle \boldsymbol{X}(t + \tau) \rangle_{\boldsymbol{x}} \langle \boldsymbol{X}(t)^\intercal \rangle_{\boldsymbol{x}} \\
&= U_{t+\tau-t_0} \Sigma(t_0) U_{t-t_0}^\intercal + \int_{t_0}^{t} U_{t+\tau-t'} B B^\intercal U_{t-t'}^\intercal \, \mathrm{d}t' \, .
\end{aligned} \tag{2.38}$$

with $\tau \geq 0$.

If all eigenvalues of J have negative real part, the dynamical system has a stable fixed point at $\boldsymbol{x} = 0$. In this case, the process becomes stationary after infinite time. To compute the covariance matrix of the stationary state Σ_0, we only have to take the limit $t_0 \to -\infty$ for $\Sigma(t, t \,|\, t_0)$:

$$\begin{aligned} \Sigma_0 &= \int_{-\infty}^{t} U_{t-t'} B B^\intercal U_{t-t'}^\intercal \, \mathrm{d}t' \\ &= \int_{0}^{\infty} e^{Jt'} B B^\intercal e^{J^\intercal t'} \, \mathrm{d}t' \end{aligned} \tag{2.39}$$

Partial integration leads to a fundamental algebraic relation between J, B and Σ_0 [15, Eq. 4.4.51], the Lyapunov equation

$$J\Sigma_0 + \Sigma_0 J^\intercal = -BB^\intercal \; . \tag{2.40}$$

Taking the same limit for $\Sigma(t + \tau, t \,|\, t_0)$ leads to the time shifted covariance matrix of the stationary state

$$\Sigma_\tau := \lim_{t_0 \to -\infty} \Sigma(t + \tau, t \mid t_0) = U_\tau \Sigma_0 \; . \tag{2.41}$$

The covariance matrix for negative time shifts is defined by $\Sigma_{-\tau} := \Sigma_\tau^\intercal$.

Probability Densities

Here, we compute important probability densities. The calculations can also be found in [26]. We derive the joint probability density $p(\boldsymbol{x}, \tau; \boldsymbol{x}_0, 0)$ for being in state $\boldsymbol{x}$ at time $t = \tau$ and at $\boldsymbol{x}_0$ at time $t = 0$ in the stationary state by applying the chain rule

$$p(\boldsymbol{x}, \tau; \boldsymbol{x}_0, 0) = p(\boldsymbol{x}, \tau \mid \boldsymbol{x}_0, 0) \cdot p(\boldsymbol{x_0}) \; . \tag{2.42}$$

Since

$$p(\boldsymbol{x}, \tau \mid \boldsymbol{x}_0, 0) \propto \exp\left\{ -\frac{1}{2} \left[\boldsymbol{x}^\intercal(\tau) - \boldsymbol{x_0}^\intercal U_\tau^\intercal \right] \Sigma(\tau, \tau \,|\, 0)^{-1} \left[\boldsymbol{x}(\tau) - U_\tau \boldsymbol{x_0} \right] \right\} \tag{2.43}$$

and

$$p(\boldsymbol{x_0}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}_0^\intercal \Sigma_0^{-1} \boldsymbol{x}_0^\intercal\right\} , \tag{2.44}$$

the joint probability distribution is given by

$$p(\boldsymbol{x}, \tau; \boldsymbol{x}_0, 0) \propto \exp\left\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}_0 \\ \boldsymbol{x}\end{pmatrix}^\intercal \begin{pmatrix}\Sigma_0^{-1} + U_\tau^\intercal \Sigma(\tau,\tau\,|\,0)^{-1} U_\tau & U_\tau^\intercal \Sigma(\tau,\tau\,|\,0) \\ \Sigma(\tau,\tau\,|\,0)^{-1} U_\tau & \Sigma(\tau,\tau\,|\,0)^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_0 \\ \boldsymbol{x}\end{pmatrix}\right\} . \tag{2.45}$$

Thus, the covariance matrix Σ of the joint distribution has to fulfill

$$\Sigma^{-1} = \begin{pmatrix}\Sigma_0^{-1} + U_\tau^\intercal \Sigma(\tau,\tau\,|\,0)^{-1} U_\tau & U_\tau^\intercal \Sigma(\tau,\tau\,|\,0) \\ \Sigma(\tau,\tau\,|\,0)^{-1} U_\tau & \Sigma(\tau,\tau\,|\,0)^{-1}\end{pmatrix} . \tag{2.46}$$

A matrix inversion yields

$$\Sigma = \begin{pmatrix}\Sigma_0 & \Sigma_\tau^\intercal \\ \Sigma_\tau & \Sigma_0\end{pmatrix} . \tag{2.47}$$

2.4. Global Thresholding of Pairwise Measures of Statistic Dependence

In this section, we present a class of reconstruction methods that has become a standard method in the statistical analysis of networks: Thresholding of pairwise measures of statistical dependence.

We present the basic idea of the method and introduce Pearson correlation and mutual information, popular measures of statistical dependence for linear and nonlinear systems. Then, we show that Pearson correlation sufficiently describes statistical dependencies between variables of the Ornstein-Uhlenbeck process.
In the last part, we explain instantaneous thresholding and maximal thresholding, two different thresholding methods.

2.4.1. The Method

Analyzing dynamical systems by global thresholding of pairwise measures of statistical dependence works exactly like indicated by the name.
First, the statistical dependence between pairs of state space variables is calculated using a suitable measure (e.g., Pearson correlation, mutual information or transfer entropy). Subsequently, an appropriate global threshold is chosen to classify pairs as connected or unconnected: Any pair of variables correlated less than the threshold is considered unconnected and any pair whose statistical dependence lies above the threshold is said to be connected. This defines a so called "effective network".
Slightly different networks are constructed depending on which measure is chosen as well as how and where the values are thresholded. Hence, thresholding statistical dependencies is rather a family of reconstruction methods than just a single method.
The effective network reflects the most correlated variable pairs of a system. Correlations help to estimate the reaction of a dynamical system to changes. Even though, the exact changes in the behavior can not be predicted explicitly, at least it is possible to predict which parts of the system are likely not to change.
In order to estimate the statistical dependence between variables their joint and marginal probability distributions have to be estimated. Often, the necessary statistic is taken from the time series of the process implying that the process is in a stationary state.

The question remains how the effective network (given by the most correlated state space variables) is related to the network of physical interactions (given by the dynamical system itself). Arguments exist that support that effective and physical network have to be connected: If two units are physically interacting, they are very unlikely to be statistically independent. Thus, assuming that real interaction lead to statistical dependence is a reliable approximation of reality (see: stability [33, p. 31]). Given this assumption, the logical opposite should also be true: If two variables are not statistically dependent they can not be physically connected in any way.
In addition, if the interaction channels are noisy, directly interacting units should be statistically more dependent than indirectly communicating units (see: data processing inequality [9, Sec. 2.8]). Hence, if the pairwise statistical dependencies are thresholded at a suitable

value, it could be possible to separate non-existing from existing connections.
At least, this is the chain of reasoning. If this was true, the problem of reconstructing the network of physical interactions would be relaxed to finding the right threshold of statistical dependence. Unfortunately, heterogeneous connection strengths and second order effects can render physically unconnected variables more dependent than physically connected variables. Thus, the effective network and the network of physical interactions do not coincide in general.

Nevertheless, this method has convincing advantages:
First, it is simple and intuitive. For many measures (e.g., Pearson correlation) it is even possible to estimate confidence intervals, so that only a few data points are needed to generate useful results.
Secondly, even though the resulting effective network may not be the actual physical network behind the process, the numbers themselves have an interpretable meaning. The edges reconstructed by thresholding are usually called functional links and the matrices of pairwise statistical dependence effective connectivity matrices.
And last, thresholding is model free. As long as the analyzed time series are stationary and a suitable measure of statistical dependence is used, the results reflect the effective interaction within the network. Consequently, this method is widely used in many fields of science.

Considering the simplicity of this method, scientists have become creative to improve its ability to detect physical interactions. The ARCACNE algorithm [29], for example, tries to fix the problem of false positive connections by deleting the weakest connection in a circle of size 3 using information theoretic arguments.

2.4.2. Pearson Correlation

The simplest way to estimate stochastic dependencies of two random variables X, Y is the Pearson correlation coefficient C_{xy}. The measure is traditionally used in almost all disciplines that are accustomed to statistics and probabilistic statements, like medicine or social science. In many cases it is even the obligatory standard measure for scientific journals.

Pearson correlation is defined as the normalized cross-covariance Σ_{xy} of two variables

$$C_{xy} := \frac{\Sigma_{xy}}{\sqrt{\Sigma_{xx} \cdot \Sigma_{yy}}} \, . \tag{2.48}$$

In order to generalize the analysis for time series of processes in a stationary state it is convenient to introduce a time delay τ that makes it possible to track the changes of C_{xy} regarding future and past values of X and Y.

$$C_{xy}(\tau) := \frac{\Sigma_{xy}(\tau)}{\sqrt{\Sigma_{xx} \cdot \Sigma_{yy}}} \tag{2.49}$$

Here $\Sigma(\tau)$ is the delayed covariance matrix of the joint probability density $p(x, y)$. And Σ_{xx} and Σ_{yy} are the diagonal elements at time delay $\tau = 0$.

If the joint probability density $p(x, y)$ is Gaussian and X and Y are non-correlated, i.e. $C_{xy} = 0$, the joint probability distribution factorizes $p(x, y) = p(x) \cdot p(y)$ and both variables are statistically independent by definition.
Since higher order momenta of the joint probability distribution are not considered, Pearson correlation does only reflect statistical dependencies for Gaussian probability densities. These distributions are mostly generated by linearly related random variables, so that Pearson correlation does not reflect statistical dependencies in nonlinear systems.
C_{xy} can take values in the interval $[-1, 1]$. If $C_{xy} = 1$ the trajectories of $X(t)$ and $Y(t)$ are related by a linear deterministic equation $X = m \cdot Y + c$ with $m > 0$. For $C_{xy} = -1$ the same is true for negative m.

We define conditional correlation $C_{x|y}$ of two Gaussian distributed random variables X, Y as the normalized covariance $\Sigma_{x|y}$ of the conditional joint probability distribution $p(x|y)$.
$\Sigma_{x|y}$ can be computed via

$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \, . \tag{2.50}$$

This formula can be derived by $p(x|y) = p(x, y)/p(y)$ (see Appendix A.2).
Keep in mind, that X and Y can be multivariate processes, so that $C_{x|y}$ is a matrix.

2.4.3. Mutual Information

Mutual information and other information theoretic measures of statistical dependence in nonlinear systems have become popular lately. Here, we introduce delayed mutual information, derive mutual information for the OU process and show that for this process Pearson correlation is actually sufficient to measure statistical dependencies.

In the late 1940's, the American mathematician Claude E. Shannon (*1916, †2001) proposed a way to quantify and measure information. His idea was mathematically so convenient and general that it is now used as a solution for many problems in other fields.
The most convincing feature of the introduced measure is the possibility to quantity statistic dependencies in any kind of relationship. As such, it is a helpful tool to analyze the dependencies in complex systems which are often nonlinear and of unknown nature.

Shannon's information $H(X)$ of a random variable X with probability distribution $p(x)$ is basically an entropy. For a continuous random variable it is defined as the integral

$$H(X) := -\int p(x)\log\left(p(x)\right)\ \mathrm{d}x\ . \tag{2.51}$$

For two random processes X, Y, this entropy can be displayed as a Venn diagram (Fig. 2.4). The entropy of the combined process $H(X,Y)$ is the sum of the individual entropies $H(X)$, $H(Y)$ minus the entropy they have in common $I(X;Y)$

$$\begin{aligned}
H(X,Y) &:= -\iint p(x,y)\log\left(p(x,y)\right)\ \mathrm{d}x\mathrm{d}y \\
&= -\iint p(x,y)\log\left(p(x,y)\right)\ \mathrm{d}x\mathrm{d}y + H(X) - H(X) + H(Y) - H(Y) \\
&= H(X) + H(Y) - \iint p(x,y)\log\left(p(x,y)\right)\ \mathrm{d}x\mathrm{d}y \\
&\quad + \iint p(x,y)\log\left(p(x)\right)\ \mathrm{d}x\mathrm{d}y + \iint p(y)\log\left(p(x,y)\right)\ \mathrm{d}x\mathrm{d}y \\
&= H(X) + H(Y) - \underbrace{\iint p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)\ \mathrm{d}x\mathrm{d}y}_{:=I(X;Y)}\ .
\end{aligned} \tag{2.52}$$

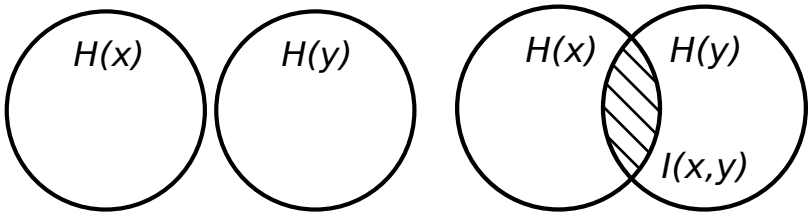


Figure 2.4: Venn Diagramm for Mutual Information, Left side: If X and Y are statistically independent they share no information. Hence $H(X,Y) = H(X) + H(Y)$ and $I(X;Y) = 0$. Right side: If the variables are statistically dependent they have information in common. Thus $I(X;Y) \neq 0$

$I(X;Y)$ is called mutual information of X and Y.

Mutual information has convincing properties: It is always positive and only zero if X and Y are statistically independent [9, Eq. 2.90]. Since this holds for all probability distributions, mutual information $I(X;Y)$ is a terrific pairwise measure of statistical dependence.

Mutual information of two variables X, Y is symmetric, i.e. $I(X;Y) = I(Y;X)$. For stationary random processes with asymmetric physical interactions, e.g., $X \to Y$, it is convenient to break the symmetry of mutual information $I(X;Y)$ by introducing a time delay τ.

The delayed mutual information is defined by

$$I_\tau(X;Y) = \int\int p(x,\tau;y,0) \log\left(\frac{p(x,\tau;y,0)}{p(x)p(y)}\right) \mathrm{d}x\mathrm{d}y \,, \tag{2.53}$$

where $p(x,\tau;y,0)$ is the joint probability density of X being in state x are time $t_0 + \tau$ and Y being in state y at time t_0. Since the process (X,Y) is in a stationary state, the joint distribution only depends on the time delay τ and the marginal distributions are time independent.

Delayed mutual information of negative time delays is defined by $I_{-\tau}(X;Y) := I_\tau(Y;X)$.

2.4.4. Mutual Information of the Ornstein-Uhlenbeck Process

Based on the probability distributions of the Ornstein-Uhlenbeck process (OU process, Sec. 2.3.7), we compute the delayed mutual information between two state space variables and show that it solely depends on their Pearson correlation coefficient. The calculation can be found in [26].

The delayed mutual information of two variables $X_i, X_j \in \{X_1, \ldots X_N\}$ of a stationary mul-

tivariate process is given by

$$I_\tau(X_i; X_j) := \iint p(x_i, \tau; x_{0j}, 0) \log\left(\frac{p(x_i, \tau; x_{0j}, 0)}{p(x_i)p(x_{0j})}\right) \mathrm{d}x_i \mathrm{d}x_{0j} \ , \tag{2.54}$$

$p(x_i, \tau; x_{0j}, 0)$ is the joint probability distribution of the time shifted state between two variables, $p(x_i)$ is the marginal distribution of one variable.

For the OU process, we already calculated the joint probability distribution of the global time shifted state $p(\boldsymbol{x}, \tau; \boldsymbol{x}_0, 0)$ and the probability distribution of the stationary state $p(\boldsymbol{x}_0)$. The low-dimensional densities $p(x_i, \tau; x_{0j}, 0)$ and $p(x_i)$ are computed by integrating the high-dimensional densities $p(\boldsymbol{x}, \tau; \boldsymbol{x}_0, 0)$ and $p(\boldsymbol{x}_0)$ over all but the required variables. We get

$$\begin{aligned} p(x_i) &= \int \cdots \int p(\boldsymbol{x}_0) \, \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \ldots \mathrm{d}x_N \\ &= \sqrt{2\pi\Sigma_{0,ii}}^{-1} \exp\left\{-\frac{x_i^2}{2\Sigma_{0,ii}}\right\} \end{aligned} \tag{2.55}$$

$$\begin{aligned} p(x_i, \tau; x_{0j}, 0) &= \int \cdots \int p(\boldsymbol{x}, \tau; \boldsymbol{x}_0, 0) \, \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \ldots \mathrm{d}x_N \mathrm{d}x_{0,1} \ldots x_{0,j-1} \mathrm{d}x_{0,j+1} \ldots \mathrm{d}x_{0,N} \\ &= \sqrt{(2\pi)^2 \det \Sigma'}^{-1} \exp\left\{-\frac{1}{2}(x_j, x_i)\Sigma'^{-1}(x_j, x_i)^\intercal\right\} \end{aligned} \tag{2.56}$$

where Σ_0 is the covariance matrix of the stationary state (Eq. 2.39), Σ_τ is the time shifted covariance of the stationary state calculated (Eq. 2.41) and

$$\Sigma' = \begin{pmatrix} \Sigma_{0,jj} & \Sigma_{\tau,j,i}^\intercal \\ \Sigma_{\tau,i,j} & \Sigma_{0,i,i} \end{pmatrix} \ . \tag{2.57}$$

Hence,

$$
\begin{aligned}
I_\tau(X_i; X_j) = &\iint p(x_i, \tau; x_{0j}, 0) \log \left\{ \sqrt{\frac{\Sigma_{0,ii}\Sigma_{0,jj}}{\det(\Sigma')}} \right\} \, \mathrm{d}x_i \mathrm{d}x_{0j} \\
&- \frac{1}{2} \iint p(x_i, \tau; x_{0j}, 0)(x_{0j}, x_i) \Sigma'^{-1} (x_{0j}, x_i)^\intercal \, \mathrm{d}x_i \mathrm{d}x_{0j} \\
&+ \frac{1}{2} \iint p(x_i, \tau; x_{0j}, 0) \left(\frac{x_i^2}{\Sigma_{0,ii}} + \frac{x_{0j}^2}{\Sigma_{0,jj}} \right) \, \mathrm{d}x_i \mathrm{d}x_{0j} \\
&= \log \left\{ \sqrt{\frac{\Sigma_{0,ii}\Sigma_{0,jj}}{\det(\Sigma')}} \right\} - \frac{2\Sigma_{0,ii}\Sigma_{0,jj} - 2\Sigma_{\tau,ij}^2}{2 \det \Sigma'} + \frac{1}{2}\frac{\Sigma_{0,ii}}{\Sigma_{0,ii}} + \frac{1}{2}\frac{\Sigma_{0,jj}}{\Sigma_{0,jj}} \\
&= \log \left\{ \sqrt{\frac{\Sigma_{0,ii}\Sigma_{0,jj}}{\Sigma_{0,ii}\Sigma_{0,jj} - \Sigma_{\tau,ij}^2}} \right\} \\
&= -\frac{1}{2} \log \left\{ 1 - \frac{\Sigma_{\tau,ij}^2}{\Sigma_{0,ii}\Sigma_{0,jj}} \right\} = -\frac{1}{2} \log \left\{ 1 - C_{ij}^2(\tau) \right\} \ , \qquad (2.58)
\end{aligned}
$$

where $C_{ij}(\tau)$ is the Pearson correlation coefficient.

Since the logarithm is monotone, this proves that Pearson correlation correctly captures the statistical dependencies between variables of the OU process.

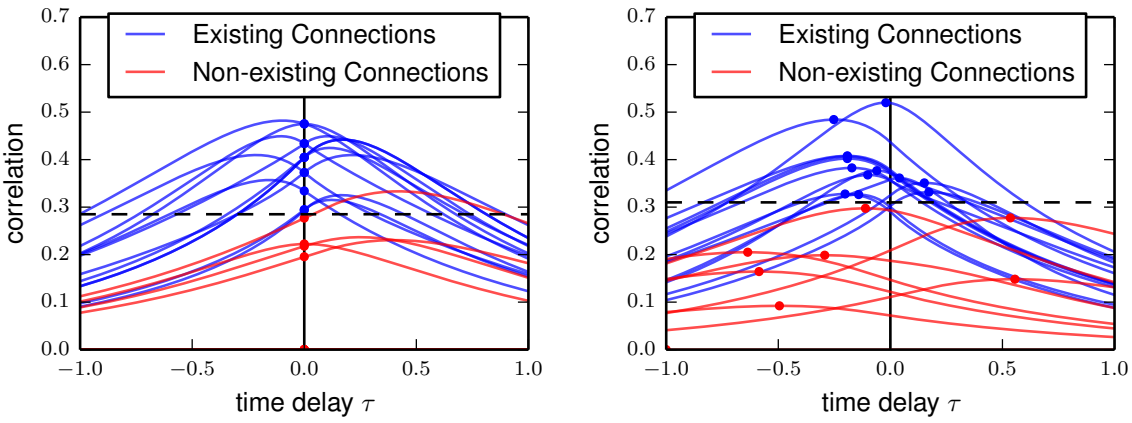


Figure 2.5: Different types of thresholding return different results. Two stationary systems each with statistically differently depending units. (a) is reconstructible by instantaneous thresholding: at $\tau = 0$ the curves are separable by the dashed threshold. (b) is reconstructible by maximum thresholding: regarding only the maximal values the curves for existing and non-existing connections are separable.

Instantaneous Thresholding and Maximum Thresholding

Two members of the family of thresholding methods are often used in the literature: instantaneous thresholding and maximum thresholding.

Let $\{i_n\}_{n\in\mathbb{N}}$ and $\{j_n\}_{n\in\mathbb{N}}$ be two time series and

$$c_{ij}(\tau) = f(\{i_{n+\tau}\}_{n\in\mathbb{N}}, \{j_n\}_{n\in\mathbb{N}}) \tag{2.59}$$

be a measure of statistical dependence including a delay τ, so that $c_{ij}(-\tau) := c_{ji}(\tau)$

Instantaneous thresholding compares only values at time delay zero $c_{ij}(0)$. Because the measure is symmetric, i.e. $c_{ij}(0) = c_{ji}(0)$, the directions of the dependencies can not be deduced. Hence, only undirected graphs are reconstructed.

Maximum thresholding considers the maximum correlation value $c_{ij,\max} = \max_{\tau} c_{ij}(\tau)$. This way, an asymmetry is introduced that can be used to make statements about the direction of the connection. By intuition causality implies that the cause always precedes its effect [23], so that $c_{ij,\max}$ with a positive time delay suggests that changes in j cause changes in i, hence $j \to i$. However, this immediately raises the issue of reconstructing asymmetric bidirectional connections so that the reconstruction of actual directed networks remains problematic.

In the literature, instantaneous thresholding is the more popular method, most likely due to its simplicity. When comparing both methods regarding their ability to reconstruct networks of physical interaction, we find examples of processes which are only fully reconstructible by one of the two methods (Fig. 2.5).

2.5. Bayesian Approach

In this section, we present the Bayesian network approach to network reconstruction. While thresholding measures of statistical dependence focuses on the reconstruction of effective connections, the Bayesian approach aims at the reconstruction of conditional statistical dependencies.

In the first part of this section, we introduce the concept of a Bayesian network. In the second part, we introduce dynamical Bayesian networks which extend the concept of Bayesian networks to random processes and dynamical systems. In the last part, we introduce the

powerful inductive causation algorithm (IC* algorithm) proposed by Pearl [33, Sec. 2.6], an algorithm that reconstructs all Bayesian networks for a given probability distribution. It was used by Runge [35] to reconstruct physical dependencies in random processes.

2.5.1. Bayesian Network

A Bayesian network is a graphical representation of a set of random variables and their conditional dependencies.

We have already seen that the joint probability density of a set of random variables can be decomposed in conditional probability densities by the chain rule (Eq. 2.17), but not all conditional dependencies have to depend on all remaining variables.
In a Bayesian network illustrating such a decomposition, every variable X_i is represented by a vertex. Directed edges are drawn from X_j to X_i if the conditional probability of X_i depends explicitly on X_j. This defines directed acyclic graphs (DAGs).
Fig.2.6 shows the Bayesian network for the decomposed joint probability

$$P(X, Y, Z, V) = P(X|Y, Z, V) \cdot P(Y|Z) \cdot P(Z|V) \cdot P(V) \ , \tag{2.60}$$

which is not unique.
For a given probability distribution the corresponding Bayesian network is unique up to edge orientations that do not create or destroy so called v-structures, i.e. converging edges, whose source vertices are not connected by an edge [33, p. 19].

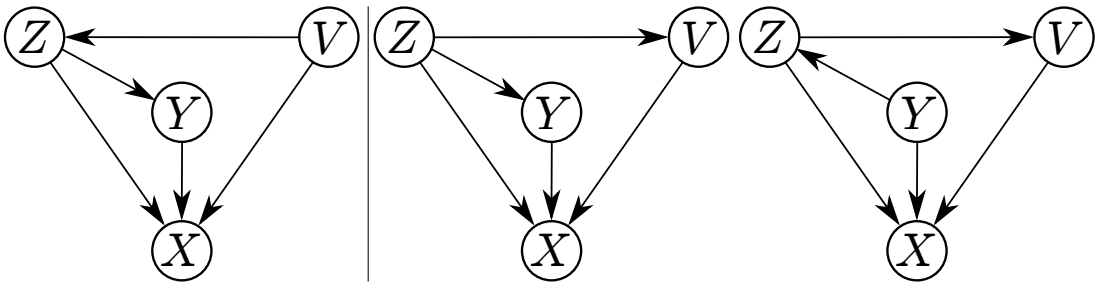


Figure 2.6: Example of Three Bayesian network Representations for the Same Joint Probability Ditribution, The network on the left represents the joint probability decomposition $P(X, Y, Z, V) = P(X|Y, Z, V)P(Y|Z)P(Z|V)P(V)$. The networks on the right constitute equivalent representations. The graph is not unique because the connections $V \to Z$ and $Z \to Y$ can be reversed without destroying or creating a v-structure.

2.5.2. Dynamical Bayesian Networks

Dynamical Bayesian networks have been developed to study the relationships between state space variables in multivariate random processes.

It should be made clear that the network of physical interactions is not the Bayesian network of the system. Networks of physical interactions are defined by a stochastic dynamical systems, for instance a Langevin equation or another multivariate random process. In these systems the concept of time is crucial. Information about state changes diffuse through these systems from one variable to another with an intrinsic velocity. For Bayesian networks, the concept of time is of no importance. It only reflects conditional dependencies in joint probability distributions.
For this reason, let us consider the joint probability distribution $p(\boldsymbol{x}_2, t_2; \boldsymbol{x}_1, t_1; \boldsymbol{x}_0, t_0)$ of three successive states $\boldsymbol{X}_0$, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ at times $t_0 < t_1 < t_2$ of a multivariate first order Markov process in a stationary state. Because of its Markov property, it can certainly be written as

$$p(\boldsymbol{x}_2, t_2; \boldsymbol{x}_1, t_1; \boldsymbol{x}_0, t_0) = p(\boldsymbol{x}_2, t_2|\boldsymbol{x}_1, t_1) \cdot p(\boldsymbol{x}_1, t_1|\boldsymbol{x}_0, t_0) \cdot p(\boldsymbol{x}_0, t_0) \tag{2.61}$$

and because of its stationarity it only depends on time differences

$$p(\boldsymbol{x}_2, t; \boldsymbol{x}_1, t-\tau; \boldsymbol{x}_0, t-2\tau) = p(\boldsymbol{x}_2, t|\boldsymbol{x}_1, t-\tau) \cdot p(\boldsymbol{x}_1, t-\tau|\boldsymbol{x}_0, t-2\tau) \cdot p(\boldsymbol{x}_0, t-2\tau) \ . \tag{2.62}$$

This already defines a Bayesian network between the successive states $\boldsymbol{x}_0$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ called a dynamical Bayesian network.
If discrete-time systems like

$$x_{i,t} = \underbrace{f_i(x_{i,t-\tau}) + \sum_{j \neq i} A_{ij} g_{ij}(x_{i,t-\tau}, x_{j,t-\tau})}_{\text{discrete-time network}} + \underbrace{\sum_j B_{ij} \eta_j(t)}_{\text{noise term}} \ , \tag{2.63}$$

with time step τ and white noise $\eta(t)$ are considered, the conditional probability density

between successive states can be decomposed even more:

$$p(x_{i,t}|\ldots,x_{j,t-\tau},\ldots) \quad \Leftrightarrow \quad A_{ij} \neq 0 \tag{2.64}$$

Since A is the adjacency matrix of the network of physical interactions, a link in the dynamical Bayesian network reflects a physical connection. Hence, reconstructing the Bayesian network structure between successive time steps leads to a reconstruction of the network of physical interactions.

Fig.2.7 illustrates an example for $N = 3$ variables. The network of physical interactions in Fig.2.7a has the dynamical Bayesian network representation displayed in Fig.2.7b.

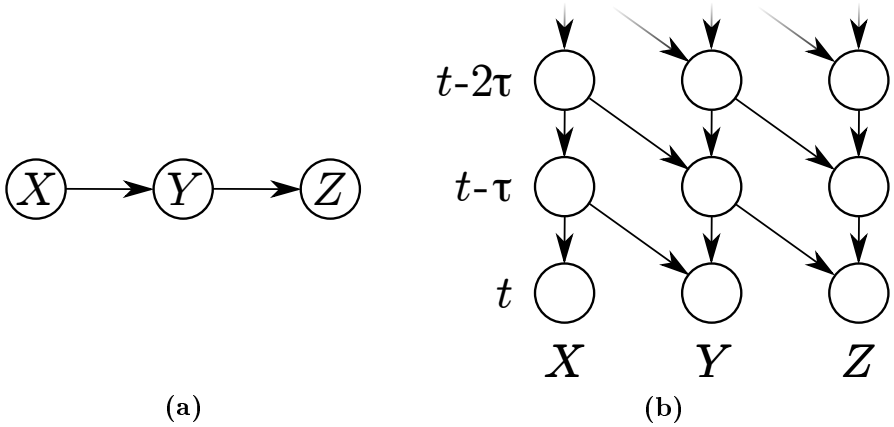


Figure 2.7: Chain topology and its dynamical Bayesian network representation. (a) shows a network of physical interactions, (b) shows is Bayesian representation in case of a discrete-time random process with time step τ

2.5.3. Inductive Causation Algorithm

Here, we explain the inductive causation algorithm (IC* algorithm). An algorithm that was proposed for the reconstruction of Bayesian networks by Pearl [33, Sec. 2.6].

Bayesian network reconstruction approaches have become a benchmark. They split into two groups: direct methods and scoring methods. Direct methods try to reconstruct the Bayesian network using conditional statistical dependencies and machine learning algorithms. Scoring methods search the solution space of all possible Bayesian networks trying to find the one structure that explains the measured statistic the best. Often, scoring methods involve max-

imization a likelihood function or include the assumption of a particular model.
Both approaches have to deal with a solution space that grows super-exponentially.
Here, we introduce the IC* algorithm because it can be treated analytically later on.

In a Bayesian network a v-structure is fundamentally distinguishable from other patterns This is the basic idea of the IC* algorithm. Consider the following three structures:

$$1\text{: } X \to Y \to Z\ , \quad 2\text{: } X \leftarrow Y \to Z\ , \quad 3\text{: } X \to Y \leftarrow Z$$

Then for the v-structure (3) it is true that X and Z are statistically independent (short: $X \perp\!\!\!\perp Z$) and that they are not statistically independent given Y (short: $(X \not\perp\!\!\!\perp Z) \mid Y$). For the others the other two structures, the opposite is true: $X \not\perp\!\!\!\perp Z$ and $(X \perp\!\!\!\perp Z) \mid Y$.
In case of the Bayesian networks (1) and (2), the variable Y is a so called separating context S under which X and Z become independent. A context can consist of a set of variables, but is of course not allowed to include the considered variables X and Z themselves.

Let us consider a Bayesian network of variables in which not all variables have been observed. The unobserved variables are so called hidden variables or latent variables and denoted by U. When taking hidden variables into account, we can define three different connections types between two variables X, Y (modulo the direction):

1. $X \to Y$, a direct causal influence of X on Y, which can not be distinguished from ...

2. ... $X \to U \to Y$, where U acts as a mediator for the connection. And ...

3. ... $X \leftarrow U \to Y$, where U serves as a common input for X and Y.

The knowledge about an edge can be summarized in the following notation (see also Tab. 2.1): If we know that there is a connection between X and Y but we do not know its nature then we write $X - Y$. We write $X \to Y$, if we know that there has to be a direct or meditated causal influence of X on Y or that both share a common input. For a common input we write $X \leftrightarrow Y$ and for a known directed influence we write $X \xrightarrow{*} Y$.

Abbreviation	Possible Connections
$X - Y$	$X \leftarrow Y$, $X \leftarrow U \leftarrow Y$, $X \rightarrow Y$, $X \rightarrow U \rightarrow Y$ and $X \leftarrow U \rightarrow Y$
$X \rightarrow Y$	$X \rightarrow Y$, $X \rightarrow U \rightarrow Y$ and $X \leftarrow U \rightarrow Y$
$X \leftrightarrow Y$	$X \leftarrow U \rightarrow Y$
$X \overset{*}{\rightarrow} Y$	$X \rightarrow Y$ and $X \rightarrow U \rightarrow Y$

Table 2.1: The IC* classification table. Four different abbreviations summarize four different degrees of knowledge.

The knowledge about a connection can be gradually increased in three steps:

First, infer whether $X - Y$: Following the definition of conditional dependence, interaction can only be present if X and Y are statistically dependent in every context S,
hence $(X \not\perp\!\!\!\perp Y) \mid S$ for all S. If this is the case, we are allowed two write $X - Y$.

Secondly, if $X - Y$, check whether $X \rightarrow Y$: This is more complicated. According to Pearl we have to find a variable Z and a context S for which $(X \perp\!\!\!\perp Z) \mid S$ and $(Y \not\perp\!\!\!\perp Z) \mid S$. These conditions rule out that X and Y are only interacting via $X \overset{*}{\leftarrow} Y$. The directed connection from Y to X can not be present because if $(Y \not\perp\!\!\!\perp Z) \mid S$, either (a) Y and Z are directly connected $Y - Z$ or (b) we found a v-structure $Y \rightarrow S \leftarrow Z$. So, let as assume $X \overset{*}{\leftarrow} Y$ was true:

In case (a) there would be an "information flow" between Y and Z in which S is not involved. Hence, $(X \not\perp\!\!\!\perp Z) \mid S$, contradicting the assumption.

Case (b) leads to the same contradiction: Knowing something about S and Z leads to knowledge about Y due to $Z \rightarrow S \leftarrow Y$. Knowing something about Y leads to knowledge about X due to $X \overset{*}{\leftarrow} Y$. Hence, knowing something about Z and S leads to knowledge about X, so that X and Z can not be statistically independent given S, i.e. $(X \not\perp\!\!\!\perp Z) \mid S$.

Both cases (a) and (b) lead to contradictions, so that $X \overset{*}{\leftarrow} Y$ can not be true.

And last, verify that $X \overset{*}{\rightarrow} Y$. This is done by finding a variable Z that is a potential cause for X, and a context S, such that $(Y \not\perp\!\!\!\perp Z) \mid S$ and $(Y \not\perp\!\!\!\perp Z) \mid S \cup X$. This guarantees that some of the information transfered from Z to Y has to be passed through X, hence $X \overset{*}{\rightarrow} Y$.

The question remains in which order the variables should be compared. This is answered

by an algorithm proposed by Peter Spirtes and Clark Glymour, the PC algorithm [36]. The main problems are the sheer number of combinatorial possibilities and the estimation of high dimensional probability densities. Spirtes and Glymour suggest beginning with two dimensional probability densities to infere $X_i \not\perp\!\!\!\perp X_j \mid \emptyset \; \forall i \neq j$ and then increasing the number of variables in the context set gradually.

By using the IC* algorithm, it is possible to infer all possible conditional decompositions of a given joint probability or joint probability density.

2.6. ROC curves

Receiver operator characteristic (short: ROC or ROC curve) provide a method to visualize and evaluate the quality of binary classifiers.

A binary classifier is a functions h which classifies whether a sample $v \in \mathcal{M}$ belongs to a certain class ($h(v) = \text{True}$) or not ($h(v) = \text{False}$). $\mathcal{M}$ is called sample space.

$$h: \quad \mathcal{M} \to \{\text{False}, \text{True}\} \tag{2.65}$$

Let $\mathcal{M}^+ \subseteq \mathcal{M}$ be the set of samples actually belonging to class and let $\mathcal{M}^- \subseteq \mathcal{M}$ be a set of samples not belonging to that class. Let them have cardinalities $N^+ := |\mathcal{M}^+|$ and $N^- := |\mathcal{M}^-|$, so that $\mathcal{M} = \mathcal{M}^+ \cup \mathcal{M}^-$ and $N := |\mathcal{M}| = N^+ + N^-$. Then a perfect classifier has to fulfill the conditions

$$v \in \mathcal{M}^+ \;\Leftrightarrow\; h(v) = \text{True} \tag{2.66}$$

$$v \in \mathcal{M}^- \;\Leftrightarrow\; h(v) = \text{False}\,. \tag{2.67}$$

However, real classifiers are usually imperfect; they produce false classifications.
These failures can either be false positive, if a sample is incorrectly classified as a member of the class, or false negative, if a member of the class is not identified as such. Correctly categorized samples constitute true positive or true negative classifications accordingly.
Let $\mathcal{T}^+, \mathcal{T}^-, \mathcal{F}^+, \mathcal{F}^- \subseteq \mathcal{M}$ be the subsets of true positive, true negative, false positive and

false negative classifications. Hence,

$$\mathcal{T}^+ \cup \mathcal{F}^- = \mathcal{M}^+ \tag{2.68}$$

$$\mathcal{T}^- \cup \mathcal{F}^+ = \mathcal{M}^- \ . \tag{2.69}$$

The fraction of true positive classifications with respect to the overall numbers of positive samples is called true positive rate $t^+ = \frac{|\mathcal{T}^+|}{|\mathcal{M}^+|}$ or *sensitivity* and $f^- = \frac{|\mathcal{F}^-|}{|\mathcal{M}^+|}$ is called false negative rate. True negative rate or *specificity* t^- and the false positive rate f^+ are defined analogously.

Every non-trivial classifier depends on parameters which determine its output. In this thesis, all classifiers depend on a one dimensional criterion: the threshold. By varying this threshold and measuring sensitivity and specificity, a finger print of performance in f^+-t^+ space is obtained. This finger print is called ROC curve.
Depending on the shape of the curve the quality of the classifier can be extracted visually.
For example, consider the witless random classifier which decides at random with a probability p if a sample is classified positively. For large N^+ the true positive rate is then $t^+ \approx \frac{p \cdot N^+}{N^+} = p$. Same holds for the false positive rate in case of large N^- since $f^+ \approx \frac{p \cdot N^-}{N^-} = p$. Hence, $t^+ = f^+$. This is why the ROC of every random classifier lies on the identity in f^+-t^+ space.
The ROC curve of ideal classifier has to intersect the point $(0, 1)$ in f^+-t^+ space because no false positives and false negatives are produced for some criterion value. These two extreme cases and an intermediate case are shown in Fig. 2.8 which illustrates the generation of the ROC curve for a one-parameter classifier.

When separating two classes by thresholding of a criterion value, the curve start at $(0, 0)$ and end at $(1, 1)$. If both sets can be separated, the classifier is perfect and the ROC has a rectangular shape. The area under the curve will be exactly one. Otherwise the integral will lead to smaller values. As a result, the area under the curve (AUC) is often used as a quality measure for classifiers.
Our measure of choice will instead be derived from the minimal sum of false negatives and

false positives. We define the quality measure ϑ as follows:

$$\vartheta := 1 - \min_{\Theta} \left(f^+(\Theta) + f^-(\Theta) \right) \tag{2.70}$$

Graphically, it is the minimal ℓ^1-distance from the point of perfect reconstruction $(0, 1)$ to the ROC curve minus one. ϑ is one for perfect reconstruction and zero for any classification worse than random guessing.

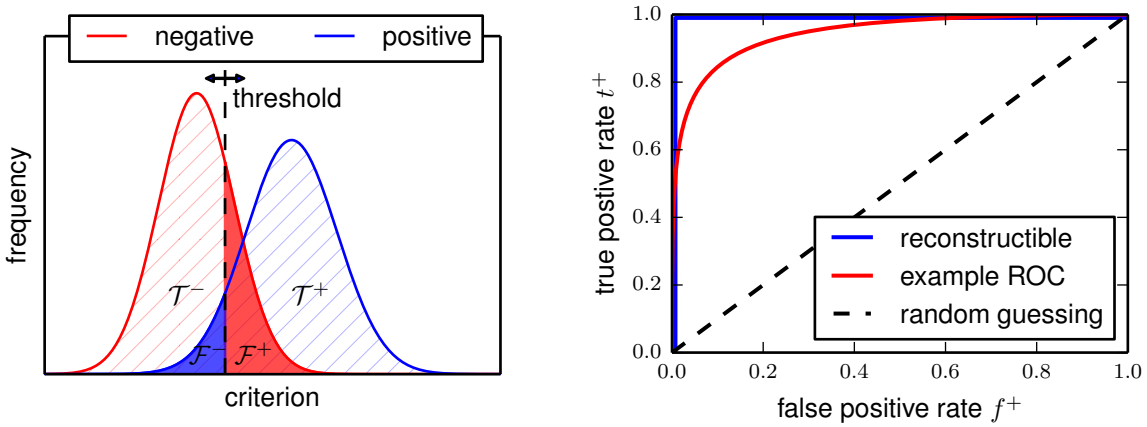


Figure 2.8: Generating and interpreting ROC curves. (a) shows how ROC curves are generated. The number of false positives $\mathcal{F}^+$ and true positives $\mathcal{T}^+$ depend on the threshold. (b) displays the corresponding ROC curve in red. For each threshold (f^+, t^+) is plotted to get a curve in f^+-t^+ space.

2.7. Simulating Random Processes

In this paragraph, we introduce the Euler-Maruyama algorithm, a numerical integration scheme for stochastic differential equations.

For our analysis it is necessary to generate trajectories of the Ornstein-Uhlenbeck process. There are two possible ways to generate the required data sets:
Since the Ornstein-Uhlenbeck process has an analytic solution (see Eq. 2.35) and the probability density of the conditional probability distribution $p(\boldsymbol{x_{n+1}}, t_n + \Delta t | \boldsymbol{x_n}, t_n)$ is known (see Eq. 2.37), the first option is to compute the mean $\langle \boldsymbol{X}_{n+1} \rangle_x$ and the covariance $\Sigma_{n+1,n+1}$ analytically for the next time step based on the current state of the process $\boldsymbol{x}_n$. The next

state $\boldsymbol{x}_{n+1}$ is then drawn from this multivariate normal distribution. The respective equations are

$$\Sigma_{n+1,n+1} = \int_0^{\Delta t} e^{J(\Delta t-\tau)} B B^{\intercal} e^{J^{\intercal}(\Delta t-\tau)} \, \mathrm{d}\tau \tag{2.71}$$

$$\bar{\boldsymbol{x}}_{n+1} = e^{J\Delta t} \boldsymbol{x}_n \ , \tag{2.72}$$

where Δt it the time step between the data points.
If no further assumptions are made these equations have to be solved analytically, either with pen and paper or the help of computational power, or numerically. This gives rise to new problems since analytic solutions are only feasible for small or regular networks and numerical solution are computationally extremely expensive for large networks.
Still, once $e^{J\Delta t}$ and $\Sigma_{\Delta t,\Delta t}$ are computed data sets for any step size can be generated.

The simpler solution is to integrate the Langevin equation numerically using an iterative numerical integration scheme. The classical integration method for the Ornstein-Uhlenbeck process is the Euler–Maruyama method.
Consider a multivariate stochastic differential equation

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{\eta}(t), \tag{2.73}$$

then the sequence of successive states $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ separated by a time step Δt is described by

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \int_{t_n}^{t_n+\Delta t} \boldsymbol{f}(\boldsymbol{x}(t)) \, \mathrm{d}t + \int_{t_n}^{t_n+\Delta t} g(\boldsymbol{x}(t))\boldsymbol{\eta}(t) \, \mathrm{d}t \ . \tag{2.74}$$

A zero order Taylor expansion of $\boldsymbol{f}(\boldsymbol{x}(t))$ and $g(\boldsymbol{x}(t))$ at t_n yields

$$\boldsymbol{x}_{n+1} \approx \boldsymbol{x}_n + \boldsymbol{f}(\boldsymbol{x}_n)\Delta t + g(\boldsymbol{x}_n) \int_{t_n}^{t_n+\Delta t} \mathrm{d}\boldsymbol{W}(t), \tag{2.75}$$

where $\boldsymbol{W}(t)$ is the multivariate Wiener process discussed in Sec. 2.3.6.
Since $\boldsymbol{W}(t)$ is a Gaussian distributed random process, the difference

$$\int_{t_n}^{t_n+\Delta t} \mathrm{d}\boldsymbol{W}(t) = \boldsymbol{W}_{n+1} - \boldsymbol{W}_n = \boldsymbol{W}_{\Delta t} \tag{2.76}$$

is also a Gaussian distributed random process. We have shown that $\boldsymbol{W}_{\Delta t}$ has zero mean and and a covariance matrix of $\Sigma_{\boldsymbol{W}_{\Delta t}} = \sqrt{\Delta t}\mathbb{1}$. Thus, a zero order Taylor expansion leads to the iterative integration scheme

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{f}(\boldsymbol{x}_n)\Delta t + g(\boldsymbol{x}_n)\sqrt{\Delta t}\boldsymbol{\Psi}, \tag{2.77}$$

in which $\boldsymbol{\Psi}$ is a multivariate Gaussian random variable with zero mean and
a covariance matrix $\Sigma_{\boldsymbol{\Psi}} = \mathbb{1}$. This scheme is known as the Euler-Maruyama method.

In general the Euler-Maruyama method has deterministic order one (i.e. the deterministic truncation error is of order $\mathcal{O}(\Delta t^2)$), weak order of convergence 1 and strong order of convergence $\frac{1}{2}$. In the special case of the Ornstein-Uhlenbeck process the Euler-Maruyama method strongly converges of order one because it coincides with the Milstein method [17, Sec. 4.7.3.2]. Weak and strong order of numerical SDE integration schemes give information about the accuracy of the statistical properties of the simulated processes [17, Sec. 4.7.3.1].

3. Results

After establishing a theoretic basis, we now present the insights and results of our own research. This part of the thesis is divided in three chapters:

In the first chapter (Sec. 3.1), we show that the IC* algorithm (Sec. 2.5.3), a reconstruction method for Bayesian dependency networks, does not apply to continuous-time processes in principle.
In the second chapter (Sec. 3.2), we turn towards the easier, but scientifically very relevant, thresholding of pairwise measures of statistical dependence (Sec. 2.4.1).
We discuss general performance issues and examine the performance of correlation thresholding under idealized conditions. It is shown that the reconstruction performance of correlation thresholding depends on the topology itself. Also, we find analytically that correlation thresholding is best for weakly coupled systems and networks with regular indegree.
In the last chapter (Sec. 3.3), we introduce covariance inversion. A new method for the reconstruction of physical dependency networks of linear homogeneous systems driven by additive noise. For infinite time series this method is always successful. A numerical analysis shows promissing results for time series with reasonable length. More importantly, we demonstrate that this method does not depend on the topology and outperforms correlation thresholding when it comes to the reconstruction of non-regular topologies.

3.1. Inductive Causation and Temporal Continuity

In this section, we discuss how time-continuity influences the Bayesian network approach in respect to its ability to reconstruct networks of physical interaction. For this purpose, we use the Ornstein-Uhlenbeck process (Sec. 2.3.7) and see if a simple chain structur can be reconstructed by means of the IC* algorithm (Sec. 2.5.3). This algorithm returns all suitable Bayesian networks for a set of random variables given their joint probability distribution.

We have shown in Sec. 2.5 that the dynamical Bayesian network between successive states of a discrete-time random process coincides with the network of physical interactions. To what extent does this hold for continuous-time processes like the Ornstein-Uhlenbeck process (OU process)? The question is: Are we able to reconstruct a dependency network such as

the one shown in Fig. 2.7 using the IC* algorithm for homogeneous OU processes from time series?

This process is given by the stochastic differential equation

$$\dot{x}_i = -x_i + \beta \sum_j A_{ij} \cdot (x_j - x_i) + \eta_i(t) \,, \tag{3.1}$$

where β is the coupling coefficient and A is the adjacency matrix. For the vector $\boldsymbol{X} = (X_1, X_2, X_3) = (X, Y, Z)$, the adjacency matrix corresponding to the network of Fig.2.7a is given by

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \,. \tag{3.2}$$

Let us assume that we measure the process time in intervals of τ. The joint probability density $p(\boldsymbol{x}_t, \boldsymbol{x}_{t-\tau}, \boldsymbol{x}_{t-2\tau})$ in the stationary state is a multivariate Gaussian with zero mean and covariance matrix given by

$$\Sigma = \begin{pmatrix} \Sigma_0 & \Sigma_\tau & \Sigma_{2\tau} \\ \Sigma_\tau^\intercal & \Sigma_0 & \Sigma_\tau \\ \Sigma_{2\tau}^\intercal & \Sigma_\tau^\intercal & \Sigma_0 \end{pmatrix} \,, \tag{3.3}$$

where Σ_0 is the stationary covariance matrix given by Eq. 2.39) and Σ_t is the delayed covariance matrix of the stationary state given by Eq. 2.41 derived in Sec. 2.3.7. Using Eq. 2.40 and Eq. 2.41 Σ can be computed analytically.

In order to reconstruct the chain structure of the given network, it is necessary to show that $(X_{t-2\tau} \perp\!\!\!\perp Z_t) \mid \{Y_{t-\tau}, Z_{t-\tau}\}$. Since the probability distribution is Gaussian, statistical dependence is captured by Pearson correlation, so that it is sufficient to show that $X_{t-2\tau}$ and Z_t are uncorrelated given $Y_{t-\tau}$ and $Z_{t-\tau}$. For this purpose the corresponding conditional correlation matrix is calculated. First, we integrate the Gaussian joint probability distribution over all variables except $X_{t-2\tau}$, Z_t, $Y_{t-\tau}$ and $Z_{t-\tau}$ and get a Gaussian with zero mean and

covariance matrix

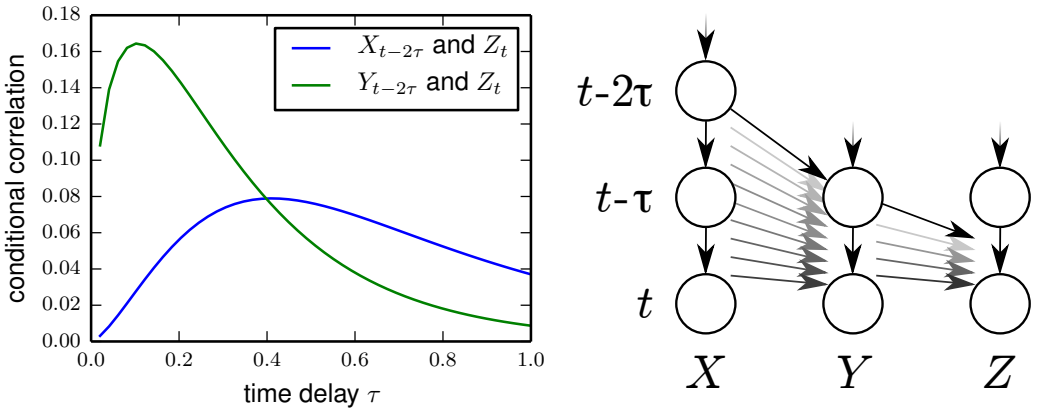
$$\Sigma' = \begin{pmatrix} \Sigma_{0,zz} & \Sigma_{\tau,zy} & \Sigma_{\tau,zz} & \Sigma_{2\tau,zx} \\ \Sigma_{\tau,zy} & \Sigma_{0,yy} & \Sigma_{\tau,yz} & \Sigma_{\tau,yx} \\ \Sigma_{\tau,zz} & \Sigma_{\tau,yz} & \Sigma_{0,zz} & \Sigma_{\tau,zx} \\ \Sigma_{2\tau,zx} & \Sigma_{\tau,xy} & \Sigma_{\tau,zx} & \Sigma_{0,xx} \end{pmatrix} \tag{3.4}$$

for the vector $\boldsymbol{X}' = (Z_t, Y_{t-\tau}, Z_{t-\tau}, X_{t-2\tau})^\intercal$. Then, the conditional covariance matrix is computed by means of Eq. 2.50:

$$\Sigma_{X_{t-2\tau},Z_t \mid Y_{t-\tau},Z_{t-\tau}} = \begin{pmatrix} \Sigma_{0,zz} & \Sigma_{2\tau,zx} \\ \Sigma_{2\tau,zx} & \Sigma_{0,zz} \end{pmatrix} - \begin{pmatrix} \Sigma_{\tau,zy} & \Sigma_{\tau,zz} \\ \Sigma_{\tau,yx} & \Sigma_{\tau,zx} \end{pmatrix} \begin{pmatrix} \Sigma_{0,yy} & \Sigma_{0,yz} \\ \Sigma_{0,yz} & \Sigma_{0,zz} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{\tau,zy} & \Sigma_{\tau,yx} \\ \Sigma_{\tau,zz} & \Sigma_{\tau,zx} \end{pmatrix} \tag{3.5}$$

Finally, this matrix is normalized according to $C_{ij|Y} = \frac{\Sigma_{ij|Y}}{\sqrt{\Sigma_{ii|Y}\Sigma_{jj|Y}}}$ to obtain the conditional correlation matrix $C(X_{t-2\tau}, Z_t \mid Y_{t-\tau}, Z_{t-\tau})$.

We use the same procedure for $C(Y_{t-2\tau}, Z_t \mid X_{t-\tau}, Z_{t-\tau})$. Both conditional correlations are compared to test whether the process is inferred as $X \to Y \to Z$ or $Y \to X \to Z$.



(a) **(b)**

Figure 3.1: Continuous-time leads to spurious correlations. (a) The conditional correlations $C(X_{t-2\tau}, Z_t|Y_{t-\tau}, Z_{t-\tau})$ and $C(Y_{t-2\tau}, Z_t|X_{t-\tau}, Z_{t-\tau})$ are non-zero for every finite $\tau \neq 0$. Hence, the unmodified IC* algorithm would consider X and Z to be connected. The graph was generated for $\beta = 2$. (b) A sketch of the actual Bayesian network representation of the continuous-time stochastic process in Fig.2.7a. Conditioning on nodes separated by finite time steps does not stop the flow of information because countless other nodes exist illustrated by arrows. These nodes transfer information around the context.

It is shown in Fig.3.1a that the conditional correlation between Z_t and $X_{t-\tau}$ only vanishes for time delay $\tau = 0$. As a result, the vertices are statistically independent only for $\tau = 0$ which constitutes only a trivial case because $(X \perp\!\!\!\perp Y) \mid X$ for all Y. Following the argumentation of the IC* algorithm the considered variables should have been statistically independent instead. This means, that the IC* algorithm fails to reconstruct the Bayesian network.

Why is this so? Because the the OU process is continuous in time. The process does not work as a finite Bayesian network (Fig.2.7b) but rather as the Bayesian network with infinitely many nodes (sketched in Fig.3.1b). The vertex $X_{t-2\tau}$ does not communicate with $X_{t-\tau}$ and $Y_{t-\tau}$ but instead with vertices located an infinitesimal step into the future $X_{t-2\tau+\mathrm{d}\tau}$ and $Y_{t-2\tau+\mathrm{d}\tau}$. These are, for their part, connected to vertices another infinitesimal time step into the future and so on. The context $S = \{Y_{t-\tau}, Z_{t-\tau}\}$ we choose to separate $X_{t-2\tau}$ and Z_t did not prevent the information flow indicated by shaded arrows (Fig.3.1b). The only context separating $X_{t-2\tau}$ and Z_t is $S' = \{Y_{t-\mathrm{d}t}, Z_{t-\mathrm{d}t}\}$.

In consequence, even for infinite time series this method is not able to identify non-existing links. Instead, a thresholding procedure would have to be introduced to decide which variables are unconnected.

Moreover, even a thresholding procedure does not solve this problem in general because for larger time delayes τ (bad time resolution) $C(X_{t-2\tau}, Z_t|Y_{t-\tau}, Z_{t-\tau}) < C(Y_{t-2\tau}, Z_t|X_{t-\tau}, Z_{t-\tau})$ so that the chain structure $Y \to X \to Z$ would be preferred.

Thus, in theory, the powerful IC* is not applicable for continuous-time processes without further modifications. However, in many cases finite Bayesian networks might remain a good approximation for the correct infinite representations of continuous-time random processes, so that the IC*-algorithm might still be helpful.

3.2. Performance of Thresholding of Pairwise Measures of Statistical Dependence

In this chapter we examine the performance of thresholding of pairwise measures of statistical dependence (Sec. 2.4) under ideal conditions. We investigate how good the method can ultimatley become and obtain qualitative rules for its usage.

To exclude estimation errors and computationally intensive simulations, both, the analyzed

system and the measure of statistical dependence, are kept simple.
The performance of this method is tested using the example of correlation thresholding (CT). Since Pearson correlation reflects statistical dependence only in the case of Gaussian distributed random variables (Sec. 2.4.2), we choose the stationary Ornstein-Uhlenbeck process (OU process, Sec. 2.3.7) as the underlying system. Since the correlation matrix of this process can be obtained analytically, we refrain from numerical simulations and conduct a fully analytic analysis.

This chapter contains five sections:
In the first section we define and explain the conditions that we believe to be ideal for correlation thresholding (CT). In the process we show that the correlation curves of the OU process depend on the topology of the network as well as its dynamic properties.
The second chapter contains a list of common reconstruction issues caused by the topology regarding thresholding of pairwise measures of statistical dependence.
In the third chapter we investigate the influence of the dynamical parameters on CT. We show that strongly coupled systems are not reconstructible by CT in general and that the method works best for weakly coupled systems.
The last two chapters are dedicated to topologically induced reconstruction errors of CT.
Here, the analysis is done half analytically and half numerically. We reveal that CT works well for networks with regular indegree (e.g. regular ring structures are always reconstructible). However, analyzing topologies with more complex indegree distribution leads to a dramatic decrease of the reconstruction performance.

3.2.1. Ideal Conditions and Parameters of Correlation Thresholding

In the following section, we describe the model used for the analysis of correlation thresholding (CT). The model stresses the arguments why CT should lead to a reconstruction of the dependency network (Sec. 2.4.1) and reflects conditions that we believe to be ideal for this method.

It was already mentioned that the dynamical system of our choice is the Ornstein-Uhlenbeck process (OU process) because the probability distribution of its stationary state is a Gaussian.

Hence, Pearson correlation correctly reflects statistical dependencies in this system. For the OU process, these correlations can be obtained analytically. This prevents statistical fluctuations and estimation errors from compromising our results.
Since a stationary state is required, the Jacobi matrix J has to be negative definite, otherwise the trajectories simply diverge.
Furthermore, arbitrary connection weights are a general hindrance for the reconstruction by thresholding because thresholding only classifies between connections and non-connections. Weakly connected variables, however, are intuitively statistically less dependent than strongly connected variables. This implies that even simple networks could become non-reconstructible if the weights are chosen accordingly. We choose the connection weights of our model to be globally constant to avoid this particular scenario.
For the same reason, arbitrary auto-dependencies and noise strengths are not ideal because these parameters define the time scale of each vertex and thereby rescale the connection weights for each node individually.
In addition, we assume independent noise sources for each vertex. Allowing vertices to share noise sources results in correlations which are not related to the connections between the vertices and thus infer with the reconstruction based on correlations.
A summary of these considerations yields

- stationary OU process
- constant auto-dependency $\alpha \geq 0$
- independent noise sources for each vertex
- constant connection weights $\beta \geq 0$
- globally constant noise strength $\gamma \geq 0$
- perfect correlation estimation, infinitely long time series .

It is important to mention that these assumptions do not lead to a network model that it easier to reconstruct in all cases, but instead stresses the argument why thresholding of measures of statistical dependence should lead to the actual network behind the dynamics in the first place.

This fully specifies the model given by

$$\dot{x_i} = -\alpha x_i + \beta \sum_{j=1}^{N} A_{ij} \cdot (x_j - x_i) + \gamma \eta_i(t) \ , \tag{3.6}$$

where N is the number of nodes, α is the auto-dependency, β is the coupling constant, γ is the noise strength and A is the adjacency matrix of the system.
Fortunately, the covariance of the OU process has an analytic solution so that we neglect errors due to false estimation of correlation.
In vector notation Eq. 3.6 reads

$$\dot{\boldsymbol{x}} = \underbrace{-(\alpha \mathbb{1} + \beta L)}_{=J} \boldsymbol{x} + \underbrace{\gamma \mathbb{1}}_{=B} \boldsymbol{\eta}(t) \tag{3.7}$$

with Laplace matrix $L_{ij} = -A_{ij} + \delta_{ij} \sum_j A_{ij}$. Gershgorin's Theorem guarantees that the Jacobi Matrix J is indeed negative definite (see Appendix A.1). The delayed covariance of the stationary state is given by Eq. 2.41 which states

$$\begin{aligned} \Sigma_\tau &= e^{J\tau}\Sigma_0 = e^{J\tau} \int_0^\infty e^{Jt} B B^\intercal e^{J^\intercal t} \, \mathrm{d}t \\ &= \gamma^2 e^{-(\alpha\mathbb{1}+\beta L)\tau} \int_0^\infty e^{-2\alpha t} e^{-\beta L t} e^{-\beta L^\intercal t} \, \mathrm{d}t \, . \end{aligned} \tag{3.8}$$

Rescaling time to $t' = \beta t$ and defining the dynamical constant $\delta := \frac{2\alpha}{\beta}$ yields

$$\Sigma_\tau = \underbrace{\frac{\gamma^2}{\beta} e^{-\frac{\delta}{2}\tau}}_{\text{independent of } i,j} \quad e^{-L\tau} \int_0^\infty e^{-\delta t} e^{-Lt} e^{-L^\intercal t} \, \mathrm{d}t \, . \tag{3.9}$$

The correlation curves C_τ are generated by normalizing Σ_τ via $C_{\tau,ij} = \frac{\Sigma_{\tau,ij}}{\sqrt{\Sigma_{0,ii}\Sigma_{0,jj}}}$. Thus, the factor $\frac{\gamma^2}{\beta}$ cancels. Since the Laplace matrix L depends only on the adjacency matrix A, the Pearson correlation curves C_τ are functions of the time delay τ, the dynamical constant δ and the topology A.

$$\boxed{C_{\tau,ij} = f(\tau, \delta, A)} \tag{3.10}$$

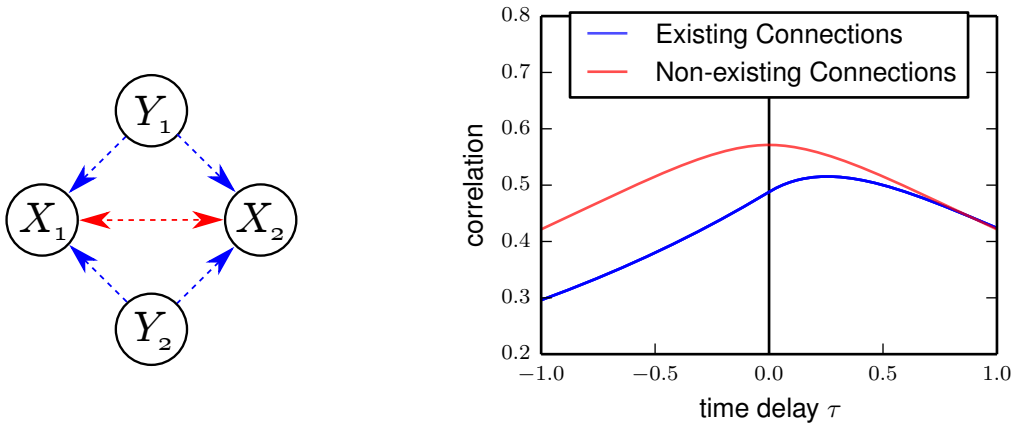
3.2.2. Main Reconstruction Issues

In this section, we illustrate the main challenges that occur when trying to reconstruct graphs by correlation thresholding. Each class of challenges is accompanied by an illustrative example

generated with the linear homogeneous model introduced in the last section.

(1) The Common Cause Problem

If two nodes have sources in common it is possible that they are more correlated with each other than with one of their input nodes. This issue, in the literature referred to as common cause problem, is illustrated by Fig. 3.2. In this example the problem will not be solved by any method that thresholds statistical dependence globally.



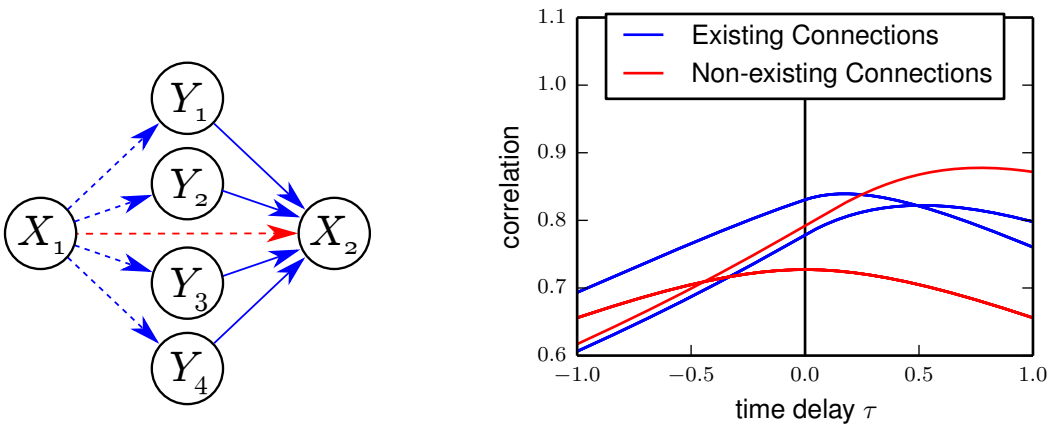
(a) **(b)**

Figure 3.2: Common Cause Problem, (a) Blue arrows denote existing connections, red arrows denote non-existing connections. Dashed arrows mark possible false positive or false negative edges. (b) The plot shows, that the non-existing connection between X_1 and X_2 is more correlated than the existing connection between each X-Y pair.

(2) Relay Problem

If the connection from one variable to another is directed, indirect and relayed via several intermediate nodes it can happen that the indirectly connected nodes are statistically more dependent than nodes that are actually connected. We call this the relay problem.

Correlation thresholding then reconstructs a structure in which the relay function of the intermediate nodes is not visible. For some networks, like gene regulatory networks, this might be problematic since the important role of the mediating structure (the relaying nodes) can not be inferred. In Fig. 3.3 we show a simple structure of six nodes which displays this problem.



(a) **(b)**

Figure 3.3: Relay Destruction, (a) Blue arrows denote existing connections, red arrows denote non-existing connections. Dashed arrows mark possible false positive or false negative edges. (b) Second order effects render the reconstruction challenging: Either none of the edges from X_1 to the mediating nodes Y_i is inferred or a non-existing direct dependency from X_1 to X_2 is deduced as well.

(3) Subgraph Dilemma

Often the network of physical interactions is not reconstructible by global thresholding while for a nontrivial partition each subgraph itself would be fully reconstructible by an individual threshold. Especially networks in which the local connection density varies strongly possess this feature. We refer to this problem as subgraph dilemma.

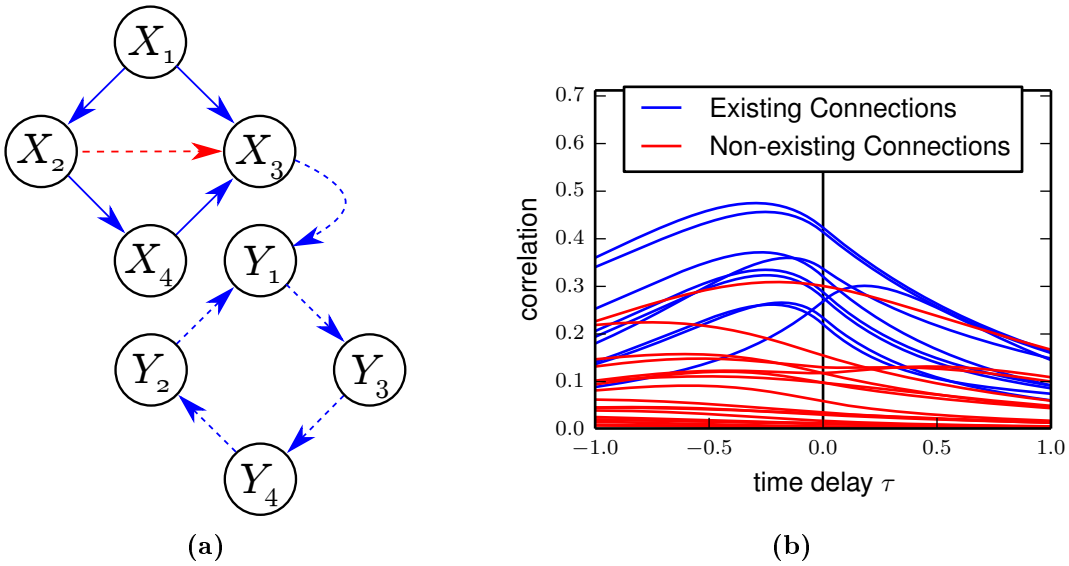
Fig. 3.4 illustrates this problem. The subgraphs, each containing rings of four nodes, are fully reconstructible themselves while the complete network is not.

Even though it might seem trivial, the existence of the subgraph dilemma has far reaching consequences.

First, the larger the system the more different local topological variations are possible. As a consequence, larger networks are in principle more difficult to reconstruct than smaller networks.

Also, this problem illustrates in a simple way that the search for problematic patterns within the graph, called pattern matching, does not lead to insights about the reconstructibility of the network. Even though structures are known to be problematic, their inner correlations change when exposed to external signals. The network is more than the sum of its parts. Hence, to decide whether a certain topology is reconstructible by global thresholding or not,

the whole graph has to be taken into account.



(a) **(b)**

Figure 3.4: Subgraph Dilemma. (a) Blue arrows denote existing connections, red arrows denote non-existing connections. Dashed arrows mark possible false positive or false negative edges. (b) Each subgraph consisting only of X- or Y vertices would be reconstructible while the figure proves that the unification is not.

3.2.3. The Influence of the dynamical constant δ

Here, we discuss the influence of the dynamical constant δ on the reconstruction of physical interactions for the homogeneous Ornstein-Uhlenbeck process (OU process) by correlation thresholding (CT). We show that common cause and relay structures are reconstructible for weak coupling and not reconstructible for strong coupling.
Assuming that the qualitative behavior of integrated and modified structures of this kind is the same, we conclude that every weakly coupled system is reconstructible by CT.

We have shown in Sec. 3.2.1 that the correlation curves C_τ of the OU process depend on the time shift τ, the dynamical constant δ and the adjacency matrix A (see Eq. 3.10). The dynamical constant is defined as $\delta := \frac{2\alpha}{\beta}$, i.e. the fraction of the auto-dependency α and the coupling coefficient β. Hence, small δ corresponds to strong coupling and large δ corresponds to weak coupling.
How do the correlation curves depend on δ?

To answer this question, we analyze simple typical model structures that have proven to induce problematic higher order correlation effects: The common cause problem and the relay structure, which were introduced in the previous section. For a more general analysis we consider an arbitrary number of intermediate vertices.
Even though these structures are unlikely to be embedded in larger networks in their pure form without incoming connections and incoming connections will always alter correlations significantly, they still model the qualitative behavior of higher order correlation effects correctly. Hence, learning how the statistical dependencies in these two structures change when changing δ is an helpful indicator for the behavior of more complex networks.

We proceed as follows:
First, we compute the instantaneous covariance matrix of the OU process by solving the integral given by Eq. 3.9 for $\tau = 0$. Hence,

$$\Sigma_0 = \frac{\gamma^2}{\beta} \int_0^\infty e^{-\delta t} \underbrace{e^{-Lt} e^{-L^\intercal t}}_{=:\Lambda(t)} \, \mathrm{d}t \; . \tag{3.11}$$

For this purpose, we calculate the matrix Λ, which is determined by the topology, and integrate element-wise.
Then, we compute the Pearson correlation matrix C_0 using $C_{0,ij} = \frac{\Sigma_{0,ij}}{\sqrt{\Sigma_{0,ii}\Sigma_{0,jj}}}$.

Common Cause

We show that common cause structures with more than one intermediate node are always reconstructible in the weak coupling limit and never reconstructible in the strong coupling limit. The network structure of the high order common cause problem is shown in Fig. 3.5a.

Let $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})^\intercal$ $\boldsymbol{X} \in \mathbb{R}^2$, $\boldsymbol{Y} \in \mathbb{R}^m$ be the vector of random variables and let each element of $\boldsymbol{Y}$ be a source node of each element of $\boldsymbol{X}$. Then, the adjacency and the resulting

Laplacian for $\boldsymbol{Z}$ have the form

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad \Rightarrow \quad L = \begin{pmatrix} m & 0 & -1 & -1 & \cdots \\ 0 & m & -1 & -1 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} . \tag{3.12}$$

Here, m is the number of source nodes. The matrix power of L yields

$$L^n = \begin{cases} m^{n-1}L & n \neq 0 \\ \mathbb{1} & n = 0 \end{cases} \quad n \in \mathbb{N} . \tag{3.13}$$

Thus, the matrix exponential is given by

$$\begin{aligned} e^{-Lt} &= \sum_{n=0}^{\infty} \frac{(-t)^n}{n!} L^n \\ &= \mathbb{1} + \sum_{n=1}^{\infty} \frac{(-t)^n m^{n-1}}{n!} L \\ &= \mathbb{1} + \frac{e^{-mt} - 1}{m} L . \end{aligned} \tag{3.14}$$

Hence,

$$\Lambda := e^{-Lt} \cdot e^{-L^\intercal t} = \mathbb{1} + \frac{e^{-mt} - 1}{m}(L + L^\intercal) + \left(\frac{e^{-mt} - 1}{m} \right)^2 LL^\intercal \tag{3.15}$$

with

$$LL^T = \begin{pmatrix} m^2 + m & m & 0 & \cdots \\ m & m^2 + m & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} , \tag{3.16}$$

so that the entries of Λ are given by

$$\Lambda_{11} = \Lambda_{22} = 1 + 2(e^{-mt} - 1) + \frac{m^2 + m}{m^2}(e^{-mt} - 1)^2 \tag{3.17}$$

$$\Lambda_{33} = \ldots = \Lambda_{NN} = 1 \tag{3.18}$$

$$\Lambda_{12} = \frac{(e^{-mt} - 1)^2}{m} \tag{3.19}$$

$$\Lambda_{13} = \ldots = \Lambda_{1N} = \Lambda_{23} = \ldots = \Lambda_{2N} = \Lambda_{13} = -\frac{e^{-mt} - 1}{m} \, . \tag{3.20}$$

All remaining entries not defined by $\Lambda = \Lambda^\intercal$ are zero.

Integration of

$$\Sigma_{0,ij} = \frac{\gamma^2}{\beta} \int_0^\infty e^{-\delta t} \Lambda(t) \, \mathrm{d}t \tag{3.21}$$

yields

$$\Sigma_{0,11} = \Sigma_{0,22} = \frac{\gamma^2}{\beta} \frac{1}{m} \left\{ \frac{1}{\delta} - \frac{2}{m + \delta} + \frac{m + 1}{2m + \delta} \right\} \tag{3.22}$$

$$\Sigma_{0,33} = \ldots = \Sigma_{0,NN} = \frac{\gamma^2}{\beta} \frac{1}{\delta} \tag{3.23}$$

$$\Sigma_{0,12} = \frac{\gamma^2}{\beta} \frac{2m}{\delta(m + \delta)(2m + \delta)} \tag{3.24}$$

$$\Sigma_{0,13} = \ldots = \Sigma_{0,1N} = \Sigma_{0,23} = \ldots = \Sigma_{0,2N} = \Sigma_{0,13} = \frac{\gamma^2}{\beta} \frac{m}{\delta(m + \delta)} \, . \tag{3.25}$$

Normalizing yields two different correlation values: The correlation

$$C_{xx} = \frac{2m}{\delta^2 + m\delta + 2m} \tag{3.26}$$

between the non-connected nodes X_1 and X_2. And the correlation

$$C_{xy} = \sqrt{\left(1 + \frac{m}{(\delta + m)}\right) \frac{1}{(\delta^2 + m\delta + 2m)}} \tag{3.27}$$

between one element of $\boldsymbol{X}$ and one element of $\boldsymbol{Y}$.

These correlations depend on δ and m and have interesting properties:

First of all, C_{xx} and C_{xy} both decrease monotonically with increasing dynamical constant δ.

(for C_{xy} see Appendix B.1.1). This confirms our intuition that weaker coupling causes weaker correlations.

In the limit of $\delta \to 0$, i.e. infinitely strong connections, the correlations become

$$C_{xx}\Big|_{\delta=0} = 1 \quad \text{and} \quad C_{xy}\Big|_{\delta=0} = \frac{1}{\sqrt{m}} \,. \tag{3.28}$$

Hence, in this case the trajectories of X_1 and X_2 are identical. This is reasonable, because X_1 and X_2 both get the same input from within the network (the source nodes $\boldsymbol{Y}$) and the additional influences of the noise sources on each element of $\boldsymbol{X}$ are negligible in the strong coupling limit.

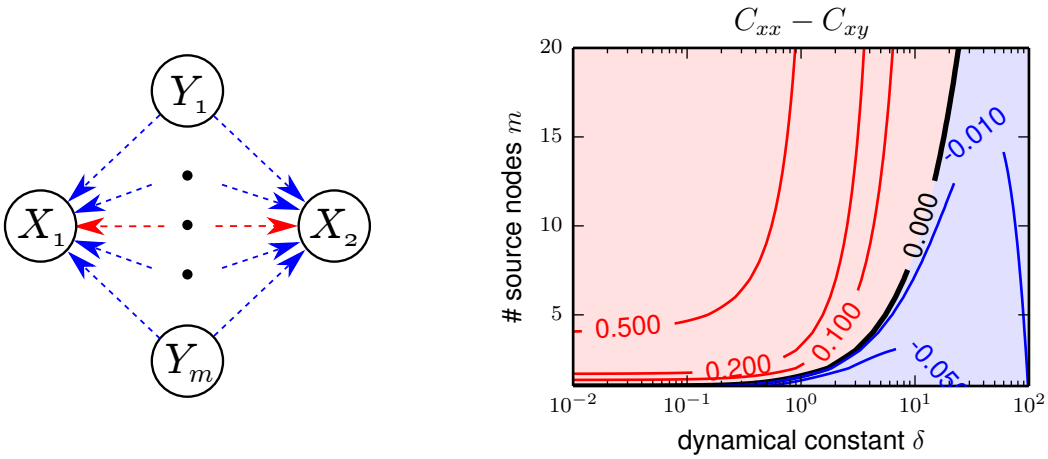
Elements of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are less correlated since each source node in $\boldsymbol{Y}$ has to share the influence on a node in $\boldsymbol{X}$ in equal proportions with the other source nodes. Apparently, the system is not reconstructible for strongly coupled nodes (and $m \geq 2$) because $C_{xx} > C_{xy}$.

For the weak coupling limit we substitute $\Delta = \frac{1}{\delta}$ and take the limit $\Delta \to 0$. Here both correlation values are zero. This is expected because for $\Delta = 0$ nodes are not connected at all. However, the functions increase differently with Δ. A second order Taylor expansion around $\Delta = 0$ yields

$$C_{xx}\Big|_{\Delta=0} \approx 2m\Delta^2 + \mathcal{O}(\Delta^3) \quad \text{and} \quad C_{xy}\Big|_{\Delta=0} \approx \Delta + \mathcal{O}(\Delta^3) \,. \tag{3.29}$$

So, while C_{xx} increases quadratically for small Δ, C_{xy} increases linearly. Thus, C_{xy} increases faster. We conclude that common cause structures are reconstructible in weakly coupled systems.

Fig. 3.5b shows the regime in which the common cause problem is reconstructible in blue. For each m there is a finite δ so that the network becomes reconstructible. $C_{xx} - C_{xy} = 0$ defines a boundary for the non-reconstructible and reconstructible regime.



(a) **(b)**

Figure 3.5: Reconstructability in common cause structures. (a) Common cause problem for m intermediate nodes. Blue/Red arrows denote existing/non-existing connections, dashed lines correspond to possible false positive or false negative connections. (b) The difference $C_{xx} - C_{xy}$ as defined in Eq. 3.26 and Eq. 3.27. The reconstructible regime is shaded blue, the non-reconstructible regime shaded red.

Relay Structure

We perform the same analysis that was done for the common cause structure (see above) for the relay structure. It is revealed that relay structures are also always reconstructible in the weak coupling limit, and never reconstructible in the strong coupling limit if the number of transmitting nodes exceeds three. All calculation refer to the network shown in Fig. 3.6a.

Here, we define $\boldsymbol{Z} = (X_2, \boldsymbol{Y}, X_1)^\intercal$. Each element of $\boldsymbol{Y}$ gets inputs from X_1 and each element of $\boldsymbol{Y}$ is a source node of X_2.

The adjacency matrix and the Laplacian of the network for $\boldsymbol{Z}$ are

$$A = \underbrace{\begin{pmatrix} 0 & 1 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}}_{m+2} \quad \Rightarrow \quad L = \begin{pmatrix} m & -1 & \cdots & -1 & 0 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{3.30}$$

The matrix power of the Laplacian yields

$$L^n = \begin{pmatrix} m^n & -\frac{1-m^n}{1-m} & \cdots & -\frac{1-m^n}{1-m} & \frac{m-m^n}{1-m} \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{3.31}$$

where used the geometric series.

Hence, the matrix exponential is given by

$$e^{-Lt} = \begin{pmatrix} e^{-mt} & \frac{e^{-t}-e^{-mt}}{m-1} & \cdots & \frac{e^{-t}-e^{-mt}}{m-1} & \frac{m(1-e^{-t})-1+e^{-mt}}{m-1} \\ 0 & e^{-t} & \cdots & 0 & 1-e^{-t} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & e^{-t} & 1-e^{-t} \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{3.32}$$

The matrix Λ and the covariance matrix Σ_0 are computed following the same ideas as in the previous paragraph (see Appendix B.1.2). We find four correlation values:

Two for the existing connections $X_1 \to Y_i$

$$C_{xy} = \sqrt{\left(1 + \frac{1}{\delta+1}\right) \frac{1}{\delta^2+\delta+2}} \tag{3.33}$$

and $Y_i \to X_1$

$$\begin{aligned} C_{yx} = {} & \sqrt{2m+\delta}\left(2m^2+4m\delta+2m+\delta^3+\delta^2(m+1)\right) \cdot \\ & \sqrt{(m+\delta)(m+\delta+1)(\delta^2+\delta+2)\left(4m^3+4m^2+\delta^5+\delta^4(2m+4)+\right.} \\ & \left.\overline{\delta^3(m^2+9m+5)+\delta^2(5m^2+9m+2)+\delta(10m^2+2m)}\right)^{-1} \end{aligned} \tag{3.34}$$

and two for the non-existing connections $Y_i \leftrightarrow Y_j$

$$C_{yy} = \frac{2}{\delta^2+\delta+2} \tag{3.35}$$

and $X_1 \to X_2$

$$C_{xx} = m\sqrt{(2m+\delta)(\delta+2)(m+\delta+1)}\cdot \\ \sqrt{(m+\delta)(\delta+1)(4m^3+4m^2+\delta^5+\delta^4(2m+4)+} \\ \overline{\delta^3(m^2+9m+5)+\delta^2(5m^2+9m+2)+\delta(10m^2+2m)}^{-1}. \tag{3.36}$$

Like in the previous example the correlation curves decrease monotonically with δ, confirming our intuition that weaker coupling causes weaker correlations (see Appendix B.1.2).

Taking the limits $\delta \to 0$ reveals that all correlation values become $C_{xy} = C_{yx} = C_{yy} = C_{xx} \to 1$. Each elements of $\boldsymbol{Y}$ copies the trajectory of X_1 perfectly because – as mentioned above – the stochastic effects of noise sources are negligible compared to the influences of coupled nodes. Thus, the trajectories of all elements of $\boldsymbol{Y}$ are identical. X_2 in turn is coupled infinitely strong to each element of $\boldsymbol{Y}$ and copies their trajectories. Ergo: All variables follow the trajectory of X_1.

Even though all correlation values are similar in the strong coupling limit $\delta \to 0$, they decrease differently with increasing δ.

First and second order Taylor expansions yield

$$C_{xy}\Big|_{\delta=0} \approx 1 - \frac{1}{2}\delta + \frac{1}{8}\delta^2 + \mathcal{O}(\delta^3) \tag{3.37}$$

$$C_{yx}\Big|_{\delta=0} \approx 1 - \left(\frac{1}{4} + \frac{1}{2m(m+1)}\right)\delta + \mathcal{O}(\delta^2) \tag{3.38}$$

$$C_{yy}\Big|_{\delta=0} \approx 1 - \frac{1}{2}\delta - \frac{1}{4}\delta^2 + \mathcal{O}(\delta^3) \tag{3.39}$$

$$C_{xx}\Big|_{\delta=0} \approx 1 - \left(\frac{1}{4} + \frac{1}{m+1} + \frac{1}{2m(m+1)}\right)\delta + \mathcal{O}(\delta^2)\,. \tag{3.40}$$

Since $\left(\frac{1}{4} + \frac{1}{2m(m+1)}\right) < \frac{1}{2}$ for all $m \geq 3$, the fastest decreasing correlation value among existing connections is not C_{yx} but C_{xy} which decreases linearly with $-\frac{1}{2}\delta$. Evidently, C_{yy} decreases even faster than C_{xy}.

The problematic correlation value among non-existing connections is C_{xx}: C_{xx} decreases slower than C_{xy} for four or more transmitting nodes ($m \geq 4$) because $\left(\frac{1}{4} + \frac{1}{m+1} + \frac{1}{2m(m+1)}\right)\Big|_{m\geq 4} < \frac{1}{2}$. Hence, $C_{xx} > C_{xy}$ at some point, so that the structure is not reconstructible for more than three transmitting nodes $m \geq 4$ and strong coupling $0 < \delta \ll 1$.

For the weak coupling limit $\frac{1}{\delta} =: \Delta \to 0$, Taylor expansions around $\Delta = 0$ yield

$$C_{xy}\Big|_{\Delta=0} \approx \Delta + \mathcal{O}(\Delta^2) \tag{3.41}$$

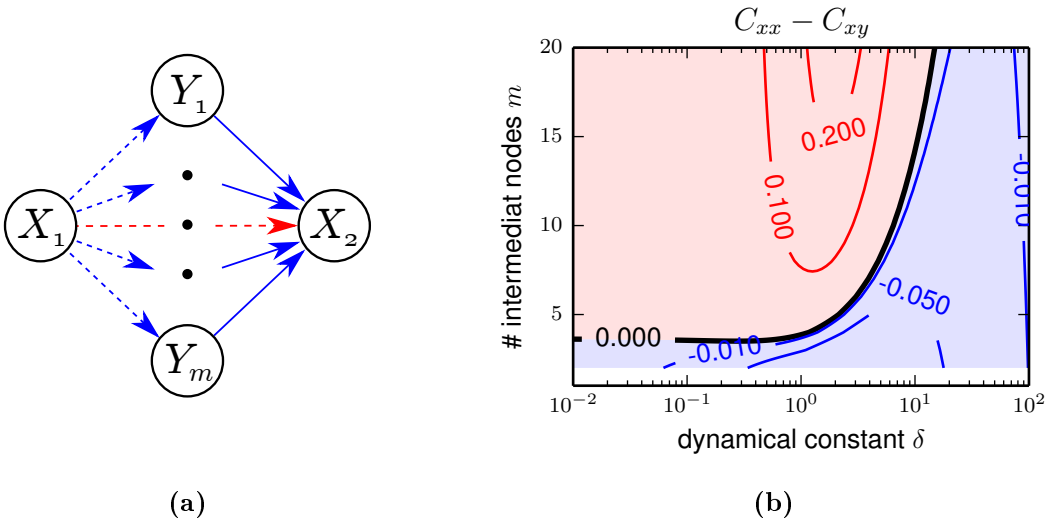
$$C_{yx}\Big|_{\Delta=0} \approx \Delta + \mathcal{O}(\Delta^2) \tag{3.42}$$

$$C_{yy}\Big|_{\Delta=0} \approx 2\Delta^2 + \mathcal{O}(\Delta^3) \tag{3.43}$$

$$C_{xx}\Big|_{\Delta=0} \approx m\Delta^2 + \mathcal{O}(\Delta^3)\ . \tag{3.44}$$

We conclude that relay structures are reconstructible if weakly coupled $1 \ll \delta < \infty$ since the statistical dependencies between connected vertices increase linearly and the dependencies between unconnected vertices increase only quadratically with Δ.

In Fig. 3.6b the reconstructible regime is shaded blue. Relay structures with less than four intermediate nodes are always reconstructible.



(a) **(b)**

Figure 3.6: Reconstructability of relay structures (a) Typical relay structure for m intermediate nodes. Blue/Red arrows denote existing/non-existing connections, dashed lines correspond to possible false positive or false negative connections. (b) Contour plot of $C_{xx} - C_{yx}$ as defined in Eq. 3.36 and Eq. 3.33. Networks in red shaded regions are not reconstructible, those in blue shaded areas are reconstructible by global thresholding.

Since common cause structures and relay structures are reconstructible in the weak coupling limit, we conclude that weakly coupled systems are in general reconstructible by correlation thresholding for infinite time series. However, later numerical analysis reveal that stochastic fluctuations due to finite time series renders precise estimation of correlation values impos-

sible for weakly coupled systems (Fig. 3.17). Hence, correlation thresholding has conceptual limitations.

3.2.4. Influence of the Topology

Here, we analyze the influence of the topology on the reconstruction quality of thresholding of instantaneous correlations (CT). As before, the underlying system model is the homogeneous Ornstein-Uhlenbeck process (OU process, Sec. 3.2.1) .

In the first paragraph, we show analytically that regular k-rings are always reconstructible by thresholding of instantaneous correlation.
In the second paragraph, we analyze more complex topologies using different random graph algorithms and find that the reconstruction error mostly depends on three parameters: the number of nodes N, the connection density ρ and the shape of the indegree distribution.
The reconstruction quality of CT decreases with increasing network size and increasing connection density in the interval $\rho \in [0, 0.5]$.
Networks with unimodal indegree distribution are usually harder to reconstruct with increasing distribution width. For instance, CT performs extremely well for networks with constant indegree, i.e. $k_{\text{in},i} = k$ for all $i \in 1, \ldots, N$ while Erdös-Renyí networks (ER networks) are much harder to reconstruct.

Thresholding Homogeneous k-Rings

In this paragraph, we prove that homogeneous Ornstein-Uhlenbeck processes (OU processes) with k-ring topology can be completely reconstructed by thresholding of instantaneous correlations (CT).
We do this by demonstrating that the correlation between nodes decreases monotonically with their distance in the ring. In addition, we show that the correlation with the farthermost connected node is always larger than the correlation with closest unconnected node. Since connected nodes are closer than unconnected nodes, the combination of both propositions proves the claim.
The proof contains four steps and holds independent of network size N, connection density $\rho = \frac{k}{N-1}$ and dynamical constant δ.

(1) *General Component Equation for Homogeneous OU Processes*

The homogeneous OU process is described by the Langevin equation

$$\dot{\boldsymbol{x}} = \underbrace{-(\alpha\mathbb{1} + \beta L)}_{J}\boldsymbol{x} + \underbrace{\gamma\mathbb{1}}_{B}\boldsymbol{\eta}\,, \tag{3.45}$$

where L the Laplace matrix of the network, $\alpha > 0$ is the auto-dependency, $\beta > 0$ is the coupling coefficient, γ is the noise strength and $\boldsymbol{\eta}(t)$ is a white noise vector.
The instantaneous covariance matrix Σ_0 has to fulfill Eq. 2.40

$$J\Sigma_0 + \Sigma_0 J^\intercal = -BB^\intercal\,. \tag{3.46}$$

Inserting $J = -(\alpha\mathbb{1} + \beta L)$ and $B = \gamma\mathbb{1}$ into the equation above yields

$$\Sigma_0 = \frac{1}{2\alpha}\left\{\gamma^2\mathbb{1} - \beta\left(L\Sigma_0 + \Sigma_0 L^\intercal\right)\right\}\,. \tag{3.47}$$

Thus, the components of Σ_0 satisfy

$$\Sigma_{0,ij} = \frac{1}{2\alpha}\left\{\gamma^2\delta_{ij} - (k_{\text{in},i} + k_{\text{in},j})\beta\Sigma_{0,ij} + \beta\left(\sum_{\{l:i\leftarrow l\}}\Sigma_{0,jl} + \sum_{\{l:j\leftarrow l\}}\Sigma_{0,li}\right)\right\} \tag{3.48}$$

$$\Rightarrow\quad \Sigma_{0,ij} = \frac{1}{2\alpha + (k_{\text{in},i} + k_{\text{in},j})\beta}\left\{\gamma^2\delta_{ij} + \beta\left(\sum_{\{l:i\leftarrow l\}}\Sigma_{0,jl} + \sum_{\{l:j\leftarrow l\}}\Sigma_{0,li}\right)\right\}\,, \tag{3.49}$$

where δ_{ij} is the Kronecker delta, $k_{\text{in},i}$ is the in-degree of vertex i and $\sum_{\{l:i\leftarrow l\}}$ denotes the sum over all nodes l that are in-neighbors of i.

(2) *Symmetries of* Σ_0 *Yield Equation for Components* $\sigma_n := \Sigma_{0,i,i+n}$

So far, only the homogeneity of auto-dependency α, coupling strength β and noise strength γ was used. The topology of the k-ring determines how to resolve the two sums:

$$\Sigma_{0,ij} = \frac{1}{2(\alpha + k\beta)}\left\{\gamma^2\delta_{ij} + \beta\left(\sum_{l=1}^{k}\Sigma_{0,j,i+l} + \sum_{l=1}^{k}\Sigma_{0,j+l,i}\right)\right\} \tag{3.50}$$

In a k-ring $k = k_{\text{in},i} = k_{\text{in},j}$ is the in- and out-degree of each node and also the maximum distance between connected nodes. For this reason $2k + 1 < N$. Equality already denotes a network in which all nodes are connected either by incoming or outgoing connections, so that a reconstruction is trivial because no unconnected pairs exist.

The topological features of a k-ring yield further conditions:

Due to the fact that such a graph is rotationally symmetric, the covariance between two nodes only depends on their distance in the ring. Thus, Σ_0 is a circulant matrix, i.e. $\Sigma_{0,i+n,j+n} = \Sigma_{0,ij}$ for all $n \in \mathbb{Z}$. Σ_0 is fully determined by the sequence $(\Sigma_{0,i,i+n})_{n=0}^{N-1}$. Also, the correlation values $C_{ij} := \frac{\Sigma_{0,ij}}{\sqrt{\Sigma_{0,ii}\Sigma_{0,jj}}} = \frac{\Sigma_{0,i,i+n}}{\Sigma_{0,ii}}$ are just proportional to the covariance values. Hence, thresholding covariance is fully equivalent to thresholding correlation.

For computational reasons, we define the periodic sequence $\sigma \widehat{=} (\sigma_n)_{n=-\infty}^{\infty}$ with period N and $\sigma_n := \Sigma_{0,i,i+n}$. The sequence has to fulfill $\sigma_{n+N} = \sigma_n$.

In addition, Σ_0 is symmetric i.e. $\Sigma_{0,ij} = \Sigma_{0,ji}$, simply because it is an instantaneous covariance matrix. Hence, the periodic sequence σ also has to fulfill the condition $\sigma_n = \sigma_{-n}$ for all $n \in \mathbb{Z}$. By use of both symmetries Eq. 3.50 becomes

$$\sum_{l=1}^{k} \Sigma_{0,i,i+n-l} - (\delta + 2k)\Sigma_{0,i,i+n} \sum_{l=1}^{k} \Sigma_{0,i,i+n+l} = -\frac{\gamma^2}{\beta}\delta_{i,i+n} \tag{3.51}$$

$$\Rightarrow \quad \sum_{l=1}^{k} \sigma_{n-l} - (\delta + 2k)\sigma_n + \sum_{l=1}^{k} \sigma_{n+l} = -\frac{\gamma^2}{\beta}\delta_{0,n}\,. \tag{3.52}$$

where $\delta := \frac{2\alpha}{\beta} > 0$ is the dynamical constant.

(3) *Use Fourier Transform to Compute σ*

We make use of the periodicity of σ by applying a Fourier transform $\mathcal{F}[\,\cdot\,]$ (see Appendix B.2.1, B.2.2, B.2.3). $\mathcal{F}[\,\sigma\,]$ can be computed using Eq. 3.52. The inverse Fourier transform of $\mathcal{F}[\,\sigma\,]$ yields

$$\sigma_n = \frac{\gamma^2}{\beta(\delta + 2k + 1)} \left(\delta_{0n} + \sum_{l=1}^{\infty} \frac{\zeta_{k,n}^{*l}}{(\delta + 2k + 1)^l} \right)\,, \tag{3.53}$$

where ζ_k is the periodic step sequence

$$\zeta_{k,n} = \begin{cases} 1 & \text{if} \quad n \mod N \leq k \quad \text{or} \quad n \mod N \geq N-k \\ 0 & \text{otherwise} \end{cases} \tag{3.54}$$

and ζ_k^{*l} is the lth convolution of ζ_k with itself

$$\zeta_{k,n}^{*l} = \sum_{m=0}^{N-1} \zeta_{k,n} \cdot \zeta_{k,m-n}^{*(l-1)} \qquad \zeta_k^{*1} := \zeta_k \; . \tag{3.55}$$

All convolutions ζ_k^{*l} are monotonically decreasing for $|n| < \frac{N}{2}$ (see Appendix B.2.4), so that σ is a sum of monotonically decreasing sequences. Hence, σ is monotonically decreasing for $|n| < \frac{N}{2}$.

(4) *The Farthermost Connected Node is Always more Correlated Than Closest Unconnected Node*

The difference between σ_k (i.e the covariance of i and the farthermost connected node $(i+k)$) and σ_{k+1} (i.e the covariance of i and the closest unconnected node $(i+k+1)$) is computed via Eq. 3.52 (Appendix B.2.5). It yields

$$\sigma_k - \sigma_{k+1} = \frac{1}{\delta + 2k + 1} \left(\sigma_0 - \sigma_{2k+1}\right) \; . \tag{3.56}$$

Eq. 3.53 shows that $\sigma_0 > \sigma_n$ for all $n \mod N \neq 0$ because the convolution sequences ζ_k^{*l} are monotonically decreasing for $|n| < \frac{N}{2}$ for all l and the Kronecker delta in the sum increases the value of σ_0 even more.
Moreover, $\sigma_{2k+1} \neq \sigma_0$ since we chose k such that it fulfills $2k+1 < N$. Hence,

$$\sigma_0 - \sigma_{2k+1} > 0 \quad \Rightarrow \quad \sigma_k - \sigma_{k+1} > 0 \; . \tag{3.57}$$

σ_n is monotonically decreasing for $|n| < \frac{N}{2}$ and the farthermost connected node is more correlated than the closest connected node. Hence, connected nodes are strictly more correlated than unconnected nodes. Thus, k-ring topologies of this model are always reconstructible.

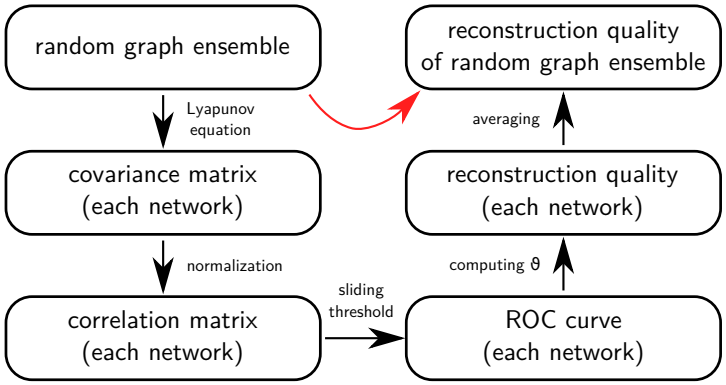


Figure 3.7: Analysis of the Influence of the Topology on Correlation Thresholding, To access the reconstruction quality of an ensemble of networks, we compute the reconstruction quality of each network individually and average over the ensemble. Investigating ensembles with different topological features leads to insights how the topology influences the reconstruction quality of correlation thresholding (CT).

Thresholding Correlations in Random Graphs

In this paragraph, we use random graph algorithms to investigate the influence of topological features on correlation thresholding (CT). The chosen model is again the homogeneous Ornstein-Uhlenbeck processes (OU process).

We find that the reconstruction performance of CT depends strongly on the the indegree distribution of the network itself. Networks with constant indegree are far more reconstructible than networks with broad indegree distribution. For unimodal indegree distributions the reconstruction quality decreases with the standard deviation of the indegree distribution and the connection density ρ in the regime $\rho \in [0, 0.5]$. Surprisingly, correlation thresholding of networks with unimodal indegree distribution does not explicitly depend on the network size N.

It was mentioned in Sec. 3.2.2 that patterns in the topology of networks of physical interaction can not be treated separately when discussing CT because incoming connections reconfigure correlation values. These changes can render integrated patterns reconstructible that are not reconstructible outside the network (and vice versa). Hence, it is not sufficient to search for patterns that are known to be not reconstructible by CT in a structural network, to decide whether the it is reconstructible or not. For this reason each network has to be checked in its entirety.

Here, we assume that similar topologies generate by the same random graph algorithm lead to similar reconstruction qualities, and that the appearance of special topologies (like k-rings)

is so unlikely that they have no influence in the average reconstruction quality of an ensemble.

To analyze the influence of the topology on the reconstruction quality of CT, we proceeded as follows (Fig. 3.7):

We generated random graph ensembles with network size N. We fixed the dynamical constant δ and computed the instantaneous covariance matrix Σ_0 for each network individually by solving the LYAPUNOV equation (Eq. 2.40) numerically by means of the Bartels-Stewart algorithm [3]. The instantaneous covariance matrix Σ_0 yields the instantaneous correlation matrix C_0 by normalization. The correlation matrix C_0 was then used to reconstruct the network by CT. A sliding threshold generated ROC curves (Sec. 2.6) which we evaluated with the reconstruction quality measure ϑ (Eq. 2.70). The quality values ϑ were then averaged over the random graph ensemble to assess its mean reconstruction quality.

Using this procedure no lengthy simulations have to be made in order to compute the correlations between variables. Instead, correlations can be accessed directly for time series of infinite length. The Bartels-Stewart algorithm computes correlations with machine precision (Fig. 3.8), so that no significant errors are expected in this respect.

Note, that not the actual adjacency matrix was taken as reference to count false positives and false negatives, but its undirected representation, because CT is only able to reconstruct connections without their direction.

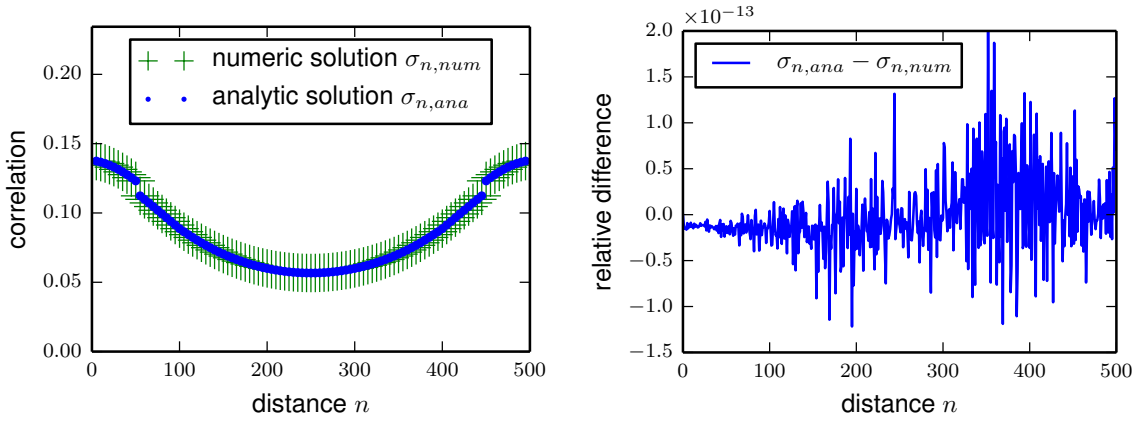


Figure 3.8: The Bartels-Stewart algorithm computes correlations with machine precision, (a) We compare the analytic solution $\sigma_{n,\mathrm{ana}}$ of Eq. 3.53 and numerical solution $\sigma_{n,\mathrm{num}}$ obtained by means of the Bartels-Stewart algorithm for a directed k-ring with $N = 500$ and $k = 50$. Every 5th point is plotted. (b) The relative error $(\sigma_{n,\mathrm{ana}} - \sigma_{n,\mathrm{num}})/\sigma_{n,\mathrm{ana}}$ lies within machine precision.

We used this procedure to analyze CT for networks in the domain between regular and random generated by the WS1, WS2 and WS3 algorithm (see Watts-Strogatz model Sec. 2.1.2). In the Watts-Strogatz model each random graph ensemble depends on three parameters: the network size N, the rewiring probability q and the connection density ρ.
We kept the network size at $N = 250$, varied q and ρ and generate for each combination ensembles of 100 networks. The LYAPUNOV equations were solved for a dynamical constant of $\delta = 2$.

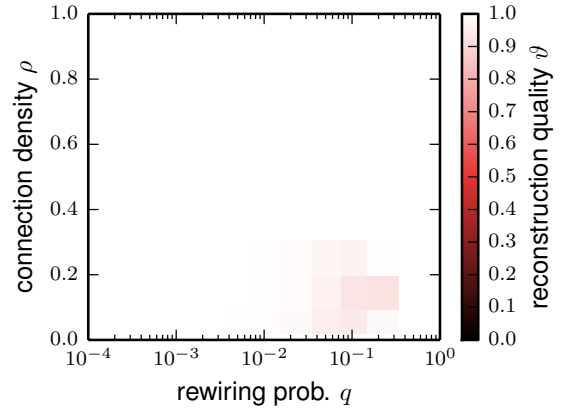
The results of this analysis is shown on the left side of Fig. 3.9 (red figures).
Four general observations can be made:
First, ensembles for $q = 0$ and $\rho = 1$ are reconstructible, i.e. the reconstruction quality measure ϑ is one. This is always true because these networks are k-rings which were shown to be reconstructible in the last paragraph.
Secondly, networks with constant indegree (generated by the WS1 algorithm) are mostly reconstructible except those in the regime of highly randomized connections $q \approx 1$ and low connection density ρ (Fig. 3.9a).
Thirdly, keeping the outdegree constant and randomizing the indegree instead decreases the reconstruction quality significantly (Fig. 3.9c). In this case, the reconstruction quality exhibits a global minimum at $(q, \rho) = (1, 0.5)$, i.e. for dense Erdös-Renyí networks. In fact, reconstructing networks in this regime is not much better than random guessing. Keep in mind that $\rho = 0.5$ corresponds to networks in which almost every pair of nodes is connected either by an incoming or an outgoing connection. Also, the indegree distribution of this ensembles have the largest variance among networks in this setting.
And last, randomizing both, indegree as well as outdegree, does not lead to further changes in the reconstruction quality (Fig. 3.9e) when compared to only randomizing the indegree.

The almost perfect reconstruction of graphs with constant indegree is a consequence of the almost discretely distributed correlation values among connected variables: If every node has the same number of incoming connections k_{in} (and additional second order effects are neglected), every source node has the same influence on this target node as the other $k_{\text{in}} - 1$ source nodes. Hence, the statistical dependencies between connected nodes are almost everywhere in the graph the same. Also, second order effects become less dominant for acyclic structures

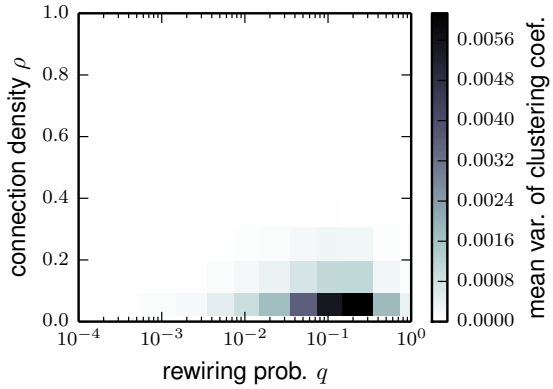
with increasing k_{in}, because statistical dependencies are blurred over long distances by the k_{in} source nodes of each node on the path. For constant indegree the decrease in statistical dependence over one connection is the same for all connections in the network.
The slight decrease in reconstruction quality for WS1 graphs in case of sparse highly randomized graphs can be explained by the mean variance of the local clustering coefficient (Fig. 3.9b). This measure reflects the differences in the local connection density and has its maximum ruffly in the same region where the reconstruction quality has its minimum. The low connection density combined with a heterogeneous local connection density results in networks that consist of weakly connected clusters so that second order effects have a strong influence.

The qualitative similarities of the reconstruction quality and the standard deviation of the indegree distribution (compare Fig. 3.9c and Fig. 3.9d) in combination with the simple fact that an additional randomization of the outdegree distribution has an negligible effect on the reconstruction quality (compare Fig. 3.9c and Fig. 3.9e) lead to the conjecture that the shape of the indegree distribution is the best indicator for reconstruction errors.
Since the degree distributions of Watts-Strogatz graphs are dominated by the second momentum, i.e. the variance, there might even be a direct relation between the reconstruction quality and the standard deviation of Gaussian-like indegree distributions.

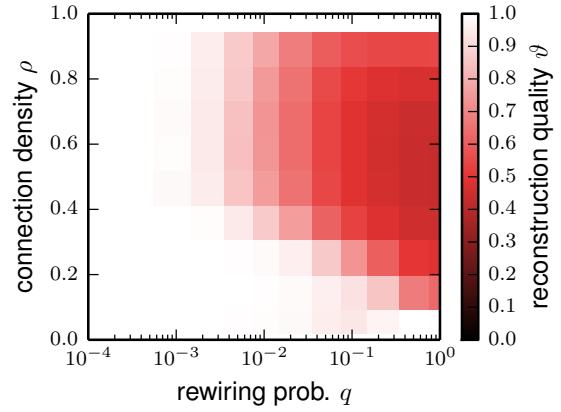
We used the WS4 algorithm to investigate the behavior of the reconstruction quality in dependence of the variances of the in- and outdegree distribution individually (Fig. 3.10).
The same data is used to generate the scatter plots in which the reconstruction quality is plotted against the standard deviation of the respective distributions (Fig. 3.11).
Again it is clearly shown that manipulations of the outdegree distribution lead to minor changes in the reconstruction quality, while small changes of the indegree distribution have an enormous impact.
This definitely confirms our conjecture: The indegree distribution mostly determines the reconstruction quality of the network.



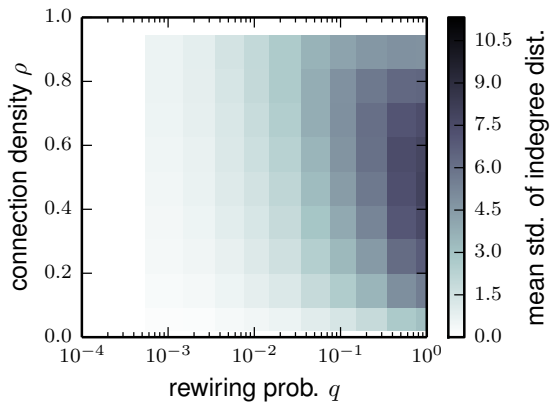
(a) randomized outgoing connections
reconstruction quality



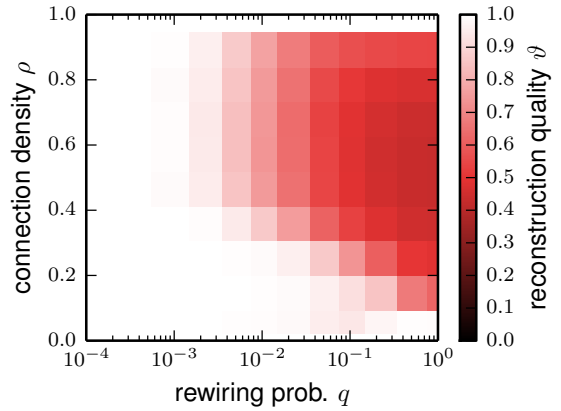
(b) randomized outgoing connections
mean variance of clustering coefficient



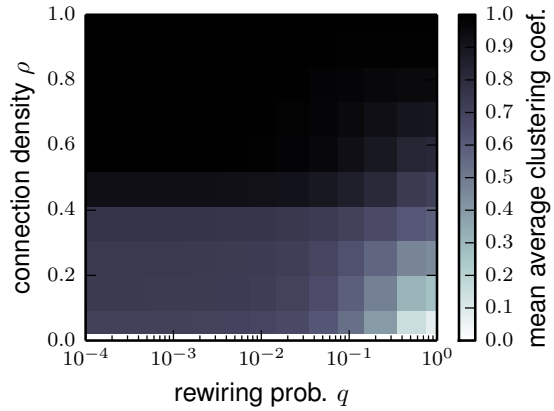
(c) randomized incoming connections
reconstruction quality



(d) randomized incoming connections
standard deviation of indegree distribution



(e) randomized directions
reconstruction quality



(f) in all three cases
mean average clustering coefficient

Figure 3.9: Reconstruction quality and graph properties for WS models.
Mean reconstruction quality and mean graph properties for the WS1, WS2 and WS3 model. Each data point is generated by an ensemble of 100 graphs with $N = 250$ nodes and $\delta = 2$. (a) Keeping the indegree distribution constant results in reconstructible graphs. (b) Variance of the clustering coefficient with maximum in the same region as (a). (c) Randomizing incoming connections leads to an strong increase of errors. (d) Standard deviation of indegree distribution shows qualitatively similar behavior. (e) Additional randomization of outgoing connections has no significant effect. (f) Mean clustering coefficient does not explain the error curves.

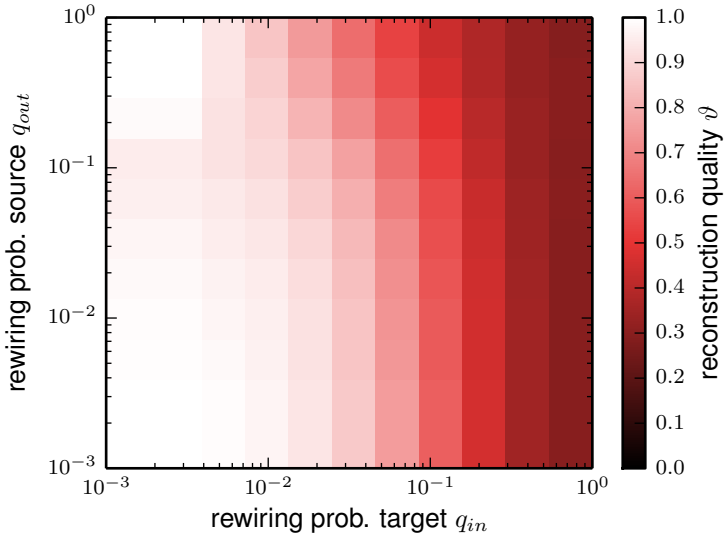
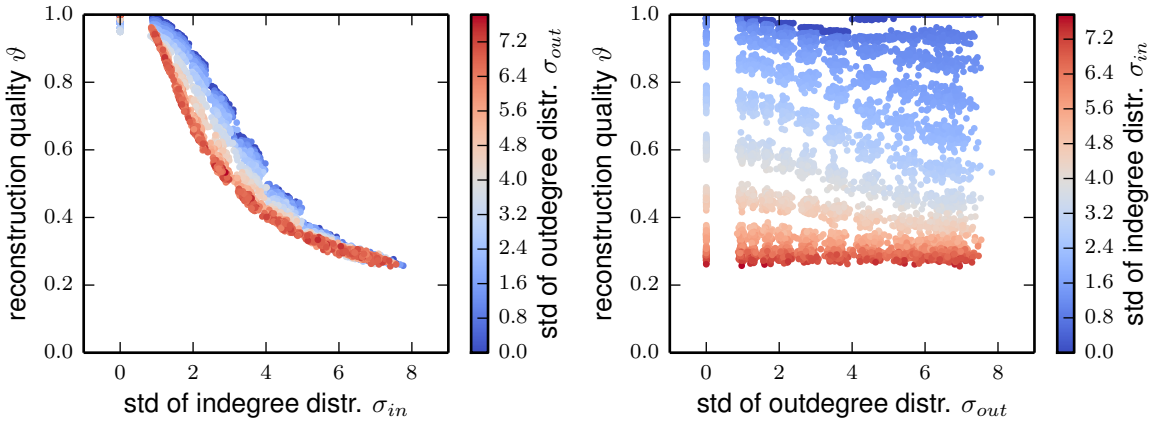


Figure 3.10: Changes in the Indegree Distribution have a Strong Impact on the Reconstruction Quality. The WS4 algorithm is used to examine the influence of the in- and outdegree distribution separately. Randomizing the indegree distribution (q_{in}) leads to greater changes than randomizing the outdegree distribution (q_{out}). For each data point the reconstruction quality of 50 networks with $N = 250$ nodes, a density of $\rho = 0.25$ and the respective number of randomized source and target nodes has been averaged. The dynamical constant was kept constant at $\delta = 2$.



(a) **(b)**

Figure 3.11: There is no functional connection between the shape of the outdegree distribution and the reconstruction quality. (a) The reconstruction quality is almost fully determined by the STD of the indegree distribution, while (b) the STD of the outdegree distribution contains little information about the reconstruction quality. The data of Fig. 3.10 was used to generate these plots.

The reconstruction quality of Watts-Strogatz networks also depends on the the connection density ρ (see Fig. 3.12).

Minimal reconstruction quality is achieved for $\rho \approx 0.5$. The reason for this is possibly the large number of one-directional connections and the relatively small number of non-connected node pairs: Since second order effects have a strong influence (as a result of high connection density) and since the correlations among existing connections are comparatively low (because most of the connections are only one-directional), relatively many non-existing connections and a significant amount of existing connections are not classified correctly.

It is also observable that denser graphs with a connection density of $\rho > 0.5$ are usually more difficult to reconstruct than their sparser counterpart with a connection density of $\rho' = 1 - \rho < 0.5$. Hence, the reconstruction quality ϑ is not fully symmetric in dependence of ρ even though its minimum is at $\rho = 0.5$. Also, statistical fluctuations in the topology have a stronger impact on the reconstruction quality of denser networks, since the curve becomes noisier with increasing ρ.

The asymmetry in shape and sensitivity is only natural since ϑ is based on the minimal weighted sum of false positives and false negatives: While the number of non-connections decreases with the connection density, the probability of an incorrect classification of non-connections increases because of second order effects. Since the false positive rate is the percentage of incorrectly classified non-connections, the impact of one false positive on the quality measure ϑ is amplified in denser graphs. This also increases the sensitivity of ϑ to topological changes in denser graphs. In contrast, sparse networks are naturally easy to reconstruct since second order effects play only a minor role.

Fig. 3.12 also shows that the network size N has only minor influences on the reconstruction quality ϑ if the density is known. Of course, the network size bounds the standard deviation of the indegree distribution, so that larger networks are more likely to be difficult to reconstruct than small networks if the connection density is kept constant.

Hence, the reconstruction quality of networks with unimodal indegree distribution only depend on the standard deviation of the indegree distribution $\sqrt{\langle k_{\text{in}} \rangle}$ and the connection density ρ.

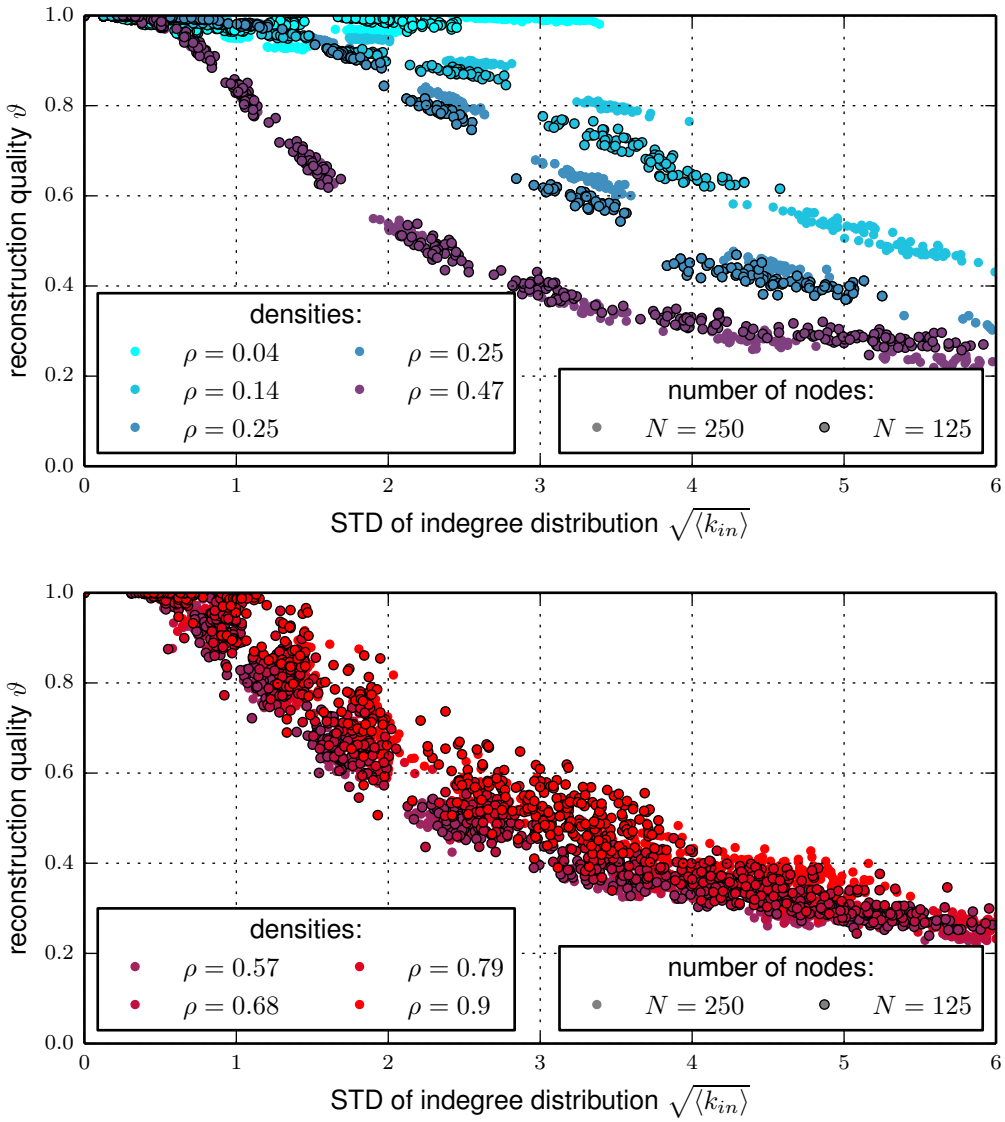
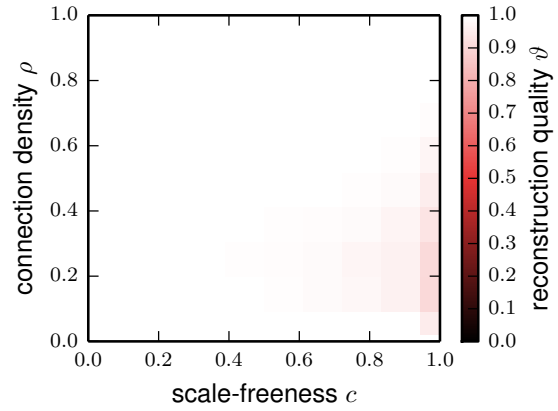


Figure 3.12: The connection density also determines the reconstruction quality. Plotting the reconstruction quality ϑ versus the standard deviation of the indegree distribution $\sqrt{\langle k_{\text{in}} \rangle}$ for different density values ρ reveals that the connection density has an impact on the reconstruction quality independent of standard deviation. Networks with $\rho \approx 0.5$ are hardest to reconstruct. Networks with $\rho > 0.5$ are harder to reconstruct than networks with $\rho < 0.5$: the relationship is not symmetric in connection and non-connection. If the network density is known, the network N has only minor influences on the reconstruction quality since the scatter plots for $N = 250$ and $N = 125$ mostly coincide.

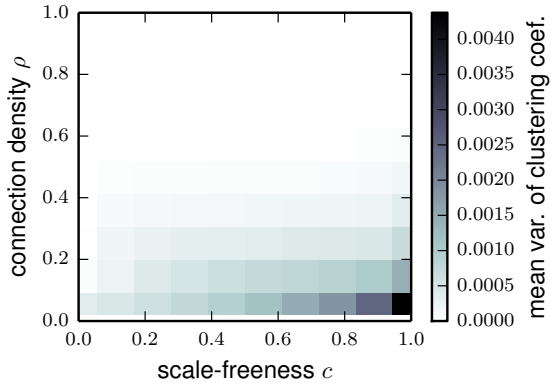
In addition to the Watts-Strogatz model, we used the analysis scheme of Fig. 3.7 to examine the performance of correlation thresholding in scale-free graphs using the modified Barabasí-Albert model (see Sec. 2.1.2). The corresponding algorithms BA1, BA2 and BA3 depend on three parameters: the network size N, the connection density ρ and the scale-freeness s. We kept the network size at $N = 250$ and conducted the analysis scheme for $\delta = 2$ (see left graphs of Fig. 3.13).

Again, the reconstruction quality mainly depends on the indegree distribution: A constant indegree mostly results in reconstructible graphs (Fig. 3.13a). Randomizing the indegree decreases the reconstruction error dramatically (Fig. 3.13c) and additional randomization of the outdegree distribution has no significant effect (Fig. 3.13e). Hence, the impact of the indegree distribution is much stronger than the impact of the outdegree distribution.

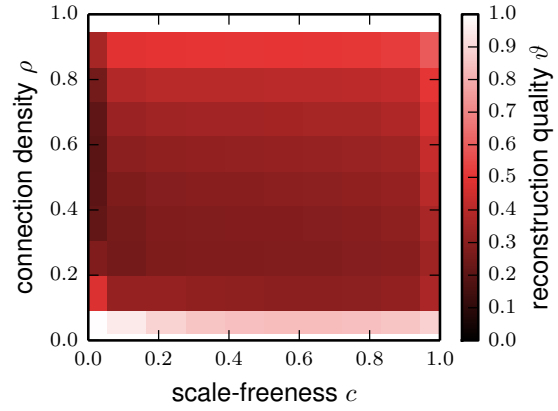
However, for scale-free graphs we do not find qualitative similarities between the reconstruction quality and the observed quantities (right side of Fig.3.13). Especially, when comparing the reconstruction quality (Fig. 3.13c) and the standard deviation of the indegree distribution in (Fig. 3.13d) no obvious statement can be made. This is not very surprising since the standard deviation is not the shape determining factor of scale-free distributions. In fact, the variance of scale-free distributions is often not finite.



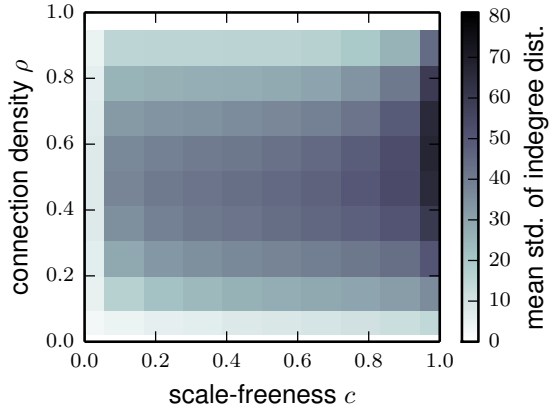
(a) randomized outgoing connections
reconstruction distance



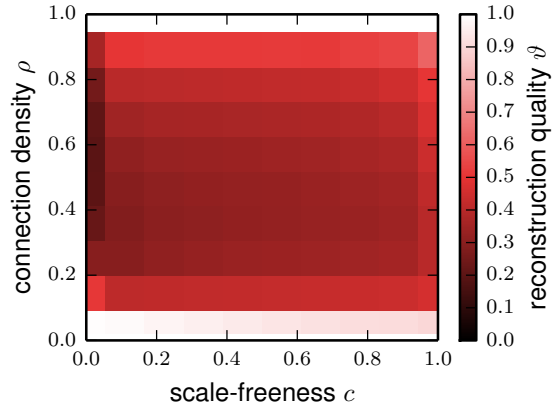
(b) randomized outgoing connections
variance of clustering coefficient



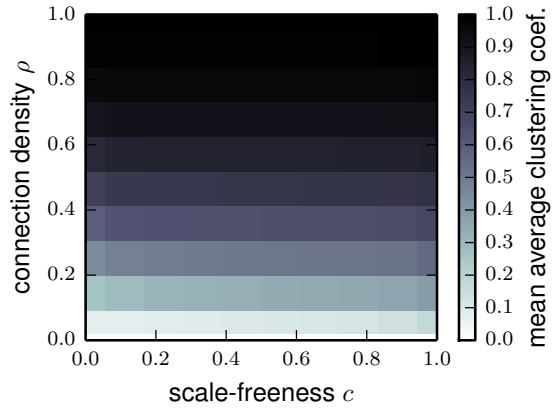
(c) randomized incoming connections
reconstruction distance



(d) randomized incoming connections
variance of indegree distribution



(e) randomized directions
reconstruction distance



(f) in all three cases
mean clustering coefficient

Figure 3.13: Reconstruction quality and graph properties for scale-free networks. For three different modifications of the Barabasí-Albert model the mean reconstruction quality and other global graph properties are computed over an ensemble of 100 random graphs with $n = 250$ nodes. (a) Keeping the indegree distribution constant results in reconstructable graphs. (b) Variance of the clustering coefficient shows minor similarities. (c) Randomizing incoming connections leads to an explosion of errors. (d) Variance of indegree distribution does captures the error's behavior. (e) Additional randomization of outgoing connections has no significant effect. (f) Mean clustering coefficient does not explain the error curves.

3.3. A New Method: Covariance Inversion

In this section, we introduce covariance inversion (CI), a method for the reconstruction of physical dependencies in linear dynamical systems driven by additive noise.

This section is divided in five parts:
In the first part, we introduce CI and discuss possible challenges and conceptual advantages compared to other methods of network reconstruction.
Additionally, an algorithm to reconstruct the matrix of physical dependencies from time series of homogeneous Ornstein-Uhlenbeck processes (OU processes) is defined (Sec. 3.2.1). In the second part, we discuss the implications of using the Euler-Maruyama scheme (Sec. 2.7) to simulate the OU process. The last three parts are dedicated to the performance of CI:
First, we examine the influence of parameters related to the dynamical system. It is shown that, in practice, CI displays the best results when dealing with strongly coupled systems.
Then, we investigate if the performance of CI depends on the topology of the dependency network itself. We find that the performance of CI does not depend on the structure of the network but on its average degree. In the last section we compare covariance inversion (CI) and correlation thresholding (CT) regarding their performance.

3.3.1. Definition of Covariance Inversion

Covariance inversion (CI) is an inference method for physical interactions in linear dynamical systems driven by additive noise. Here, we explain the method as well as its advantages and possible limitations.

The approach is based on the relationship between the covariance matrix Σ_0,
the delayed covariance matrix Σ_τ and the Jacobian matrix J of a stationary OU process.
In Sec. 2.3.7, we defined the stationary multivariate OU process as the solution of the SDE

$$\dot{\boldsymbol{x}} = J\boldsymbol{x} + B\boldsymbol{\eta}(t) \tag{3.58}$$

with negative definite Jacobi matrix J, noise input matrix B and white noise $\boldsymbol{\eta}(t)$ for $t \to \infty$. Independent of B the covariance matrices Σ_0 and Σ_τ are connected via

$$\Sigma_\tau = e^{J\tau}\Sigma_0 \tag{3.59}$$

for a time delay τ (for details see derivation of Eq. 2.41).

A simple conversion yields

$$\boxed{J = \frac{1}{\tau}\log(\Sigma_\tau \Sigma_0^{-1})\ ,} \tag{3.60}$$

where $\log(\cdot)$ is the matrix logarithm.

Hence, a reconstruction of the Jacobi matrix J is theoretically possible just by measuring the instantaneous covariance matrix Σ_0 and the delay covariance matrix Σ_τ for a specific time delay τ.

For the homogeneous OU process (Eq. 3.6)

$$\dot{x}_i = -\alpha x_i + \beta \sum_{j=1}^{N} A_{ij}(x_j - x_i) + \gamma \eta_i(t) \tag{3.61}$$

with white noise $\eta_i(t)$, adjacency matrix A, auto-dependency $\alpha \geq 0$, coupling coefficient $\beta \geq 0$ and noise strength γ, the Jacobian J is given by

$$J_{ij} = \beta A_{ij} - \delta_{ij}\left(\alpha + \beta \sum_{j=1}^{N} A_{ij}\right)\ . \tag{3.62}$$

Thus, the adjacency matrix A can be computed element-wise from the Jacobian J via

$$A_{ij} = \begin{cases} \frac{1}{\beta} J_{ij} & i \neq j \\ 0 & \text{otherwise} \end{cases}\ . \tag{3.63}$$

This method always returns the correct matrix of directed physical dependencies for infinite time series. Unlike correlation thresholding, the approach is independent of topology, dynamical parameters as well as time delay τ and even works for correlated noise sources.

The existence of hidden variables, however, depicts a limitation of this approach, since it has

been shown that hidden variables result in several possible solutions [16,20].
In the following, we call this approach covariance inversion (CI) because it requires an inversion of the instantaneous covariance matrix.

In practice, only finite time series are available which only provide imperfect estimates of the matrices required for CI, Σ_0 and Σ_τ. Let $\{\boldsymbol{x}_k\}_{k=1}^{M}$ be a time series of length M of an N-dimensional random process $\boldsymbol{X} = \boldsymbol{x} \in \mathbb{R}^N$, then the unbiased estimator of the general time shifted covariance matrix $\Sigma_{\Delta t \cdot m}$ is

$$\widehat{\Sigma}_{\Delta t \cdot m, ij} := \frac{1}{M-m-1} \sum_{k=1}^{M-m} (x_{i,k+m} - \bar{x_i})(x_{j,k} - \bar{x_j}) \tag{3.64}$$

$$\text{with mean} \quad \bar{x_i} := \frac{1}{M} \sum_{k=1}^{M} x_{i,k} \,, \tag{3.65}$$

where Δt is the time step between the measurements and m is the delay (or time shift) measured in Δt. An estimator $\widehat{\Omega}$ is unbiased if its expected value $\langle\widehat{\Omega}\rangle$ corresponds to Ω, that is $\langle\widehat{\Omega}\rangle = \Omega$.
The naive estimator of the Jacobian matrix given by CI is

$$\boxed{\widehat{J} := \frac{1}{\Delta t \cdot m} \log\left(\widehat{\Sigma}_{\Delta t \cdot m} \widehat{\Sigma}_0^{-1}\right)} \tag{3.66}$$

It should be clear, that the estimator $\widehat{J}$ might be problematic for at least three reasons:
First, the estimator could be biased, meaning that the expected value of $\widehat{J}$ is different from the actual value of J, i.e. $\langle\widehat{J}\rangle \neq J$. Secondly, the inversion of an estimated matrix is often troublesome and, finally, the matrix logarithm may cause problems since it is neither real nor unique in general. Here, it has to suffice that the real part of $\widehat{J}$, the part of interest, is unique [22].

We propose the following scheme to reconstruct the adjacency matrix of the homogeneous OU processes from time series:

1. Estimate the covariance matrices Σ_0 and Σ_τ, for instance, by means of $\widehat{\Sigma}_{\Delta t \cdot m}$.
2. Compute the estimated Jacobi matrix $\widehat{J}$ and take the real part $\mathfrak{R}(\widehat{J})$.
3. Use a threshold Θ to classify every connection $j \to i$ whose entry $\mathfrak{R}(\widehat{J})_{ij}$ is above Θ as existing, and every connection $j \to i$ whose entry $\mathfrak{R}(\widehat{J})_{ij}$ is below Θ as non-existing.

This scheme promises good results because entries $\mathfrak{R}(\widehat{J})_{ij}$ of existing connections $j \to i$ are distributed around the coupling coefficient β, while entries $\mathfrak{R}(\widehat{J})_{ij}$ that reflect the absence of a connection $j \nrightarrow i$ are distributed around zero (see Eq. 3.62.
For reasons of simplicity, we denote the range of values of the entries of $\mathfrak{R}(\widehat{J})$ with r. If the time series is long enough, estimation errors are small and the conditional distribution $p(r|\text{connection})$, i.e. the distribution of the values of the reconstructed Jacobian $\mathfrak{R}(\widehat{J})$ that correspond to actual connections, and the conditional distribution $p(r|\text{non-connection})$ are separable by a threshold, so that the value of $\mathfrak{R}(\widehat{J})_{ij}$ becomes a useful criterion to determine whether a connection from j to i exist or if the variables are unconnected.

The performance of this method can be evaluated using the same procedure as in the previous chapter, i.e. by creating ROC curves and measuring the reconstruction quality ϑ (Sec. 2.6).
It is obvious that the reconstruction quality of CI depends decisively on the distribution of the entry values r of the reconstructed Jacobian matrix $\mathfrak{R}(\widehat{J})$.
For infinite time series and perfect correlation estimation, the conditional distributions $p(r|\text{connection})$ and $p(r|\text{non-connection})$ are sharply peaked at β and zero respectively. We approximate these distribution by Gaussian bell curves.
Dividing the entry values r by the connection strength β makes it possible to compare distributions of reconstructions for different coupling strengths. For this reason we measure the variances of the distributions $p(\frac{r}{\beta}|\cdot)$ instead of $p(r|\cdot)$.
The variances of these unimodal conditional distributions are crucial for the reconstruction quality. For small variances the entry values of connections and non-connections are well separable, while for larger variances both distributions overlap and become indistinguishable (Fig.3.14).

Therefore, the scaling behavior of the reconstruction quality of covariance inversion is best quantified by the scaling of these variances.

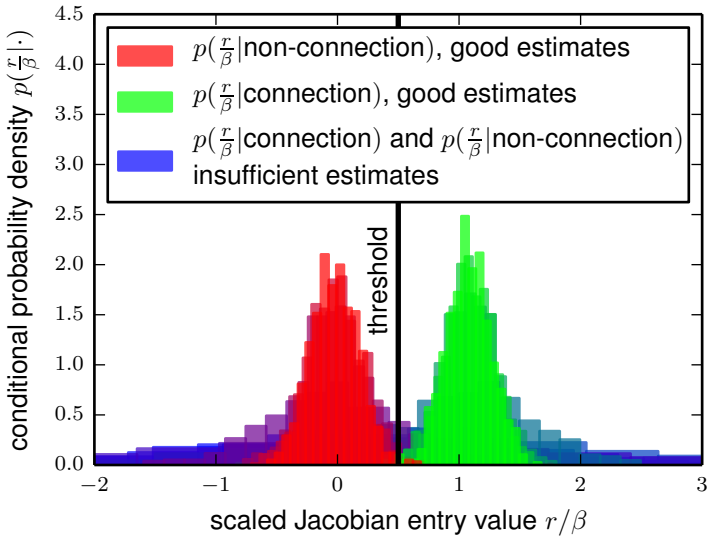


Figure 3.14: For sufficient estimates $\Re(\widehat{J})$ the conditional distributions are well separable. The normalized histograms for reconstructed Jacobian entry values r divided by the coupling strength β. The width of the distributions for connections (green) and non-connections (red) changes with the quality of the estimation of the covariance matrices. Lowering the quality (blurred colors) renders the distributions indistinguishable (blue).

3.3.2. Simulating the Homogeneous Ornstein-Uhlenbeck Processes

To simulate the homogeneous OU process we use the Euler-Maruyama integration scheme (Sec. 2.7). The chosen model has implications for the choice of the integration time step Δt.

The following analyses are based on the homogeneous OU process

$$\dot{\boldsymbol{x}} = J\boldsymbol{x} + \gamma\boldsymbol{\eta}(t) \tag{3.67}$$

$$= -(\alpha\mathbb{1} + \beta L)\boldsymbol{x} + \gamma\boldsymbol{\eta}(t) \tag{3.68}$$

with Jacobi matrix $J = -(\alpha\mathbb{1} + \beta L)$, Laplace matrix L, auto-dependency $\alpha \geq 0$, coupling coefficient $\beta \geq 0$ and noise strength γ. $\boldsymbol{\eta}(t)$ is white noise.
Measuring the time in unites of the intrinsic time scale $\frac{1}{\alpha}$ and defining the dynamical constant

$\delta := \frac{2\alpha}{\beta}$ as well as the scaled noise strength $\gamma' := \frac{\gamma}{\alpha}$ yields

$$\dot{\boldsymbol{x}} = -(\mathbb{1} + \frac{2}{\delta}L)\boldsymbol{x} + \gamma'\boldsymbol{\eta}(t) \; . \tag{3.69}$$

We generate artificial time series by solving the SDE using the Euler-Maruyama scheme (see Eq. 2.77 and Sec. 2.7 for details). For this SDE the integration scheme yields

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \underbrace{\overbrace{\left(-\mathbb{1} - \frac{2}{\delta}L\right)}^{:=J'}\boldsymbol{x}_n\Delta t}_{\text{Euler step}} + \gamma'\sqrt{\Delta t}\boldsymbol{\Psi} \; . \tag{3.70}$$

Since the deterministic Euler-step in the Euler-Maruyama method has stability issues, one should choose the time step Δt very carefully. To make sure that the simulations do not diverge, Δt should be chosen such that

$$|\lambda \cdot \Delta t + 1| \leq 1 \tag{3.71}$$

for all eigenvalues λ of the propagation matrix $J' := -(\mathbb{1} + \frac{2}{\delta}L)$. In general, the time step Δt should be chosen as small as possible.

For the chosen model, a rule of thumb can be formulated that enables estimation of Δt based on global system parameters (see Appendix C.1). According to this rule, Δt is a reasonable choice if

$$\left(1 + \frac{2k}{\delta}\right)\Delta t \leq 1 \; , \tag{3.72}$$

where δ is the dynamical constant and k is the average degree. This condition is easier to check than the actual stability condition Eq. 3.71 since the average degree is directly accessible through the connection density ρ. Even though this rule does not imply stability under all circumstances, it is still be an useful indicator for which choice of Δt simulations become unstable.

It is helpful to realize that the matrix J' always has an eigenvalue $\lambda' = -1$ because the

smallest eigenvalue of any Laplace matrix L is zero.

$$J'(1,1,\ldots,1)^\intercal = -\mathbb{1}(1,1,\ldots,1)^\intercal - \underbrace{\frac{2}{\delta}L(1,1,\ldots,1)^\intercal}_{=0} = -(1,1,\ldots,1)^\intercal \,. \tag{3.73}$$

Applying Gershgorin's theorem (see Appendix A.1), we see that λ' is the eigenvalue of J' with the largest real part. Hence, there is always an eigenvalue for which the stability condition Eq. 3.71 just holds.

Fig.3.15a shows the trajectories of the three dimensional system with connections $X_2 \to X_1$, $X_3 \to X_1$ and $X_3 \to X_2$. It is shown that the measured correlation curves approximate the analytical curves for infinite time series (Fig. 3.15b).

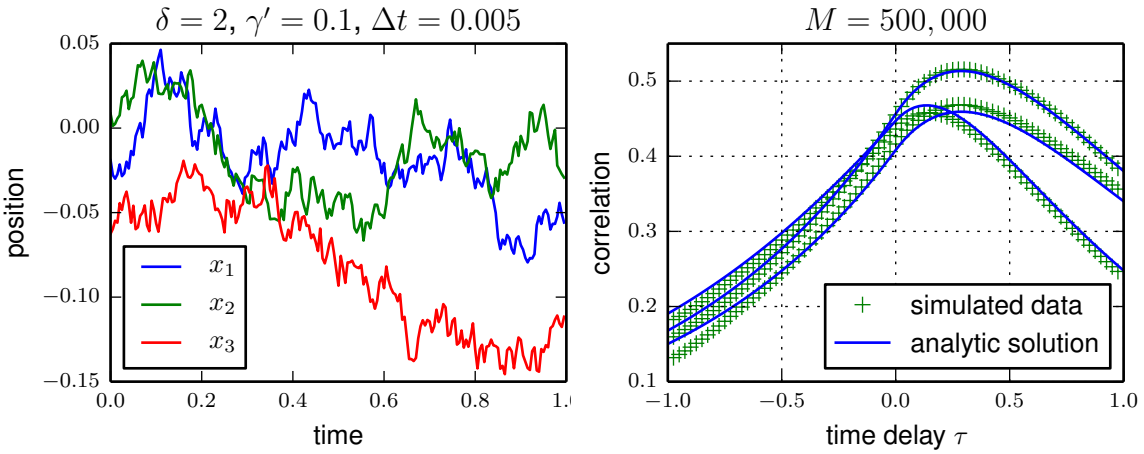


Figure 3.15: Simulated data yields the expected correlation curves. (a) The trajectory of a three dimensional system with connections $X_2 \to X_1$, $X_3 \to X_1$ and $X_3 \to X_2$. The dynamical constant was set to $\delta = 2$, the noise strength was set to $\gamma' = 0.1$. The time step was chosen to be $\Delta t = 0.005$.(b) The correlation curves were computed using Eq. (3.65) for a time series of length $M = 500,000$.

3.3.3. Performance of CI: Influence of the Dynamical Parameters

We examine the performance of covariance inversion (CI) depending on the dynamical parameters of the homogeneous OU process. The numerical analysis shows that CI is best for strongly coupled variables and does not depend the noise strength γ'.

The dynamic behavior of the homogeneous OU process

$$\dot{\boldsymbol{x}} = -\left(\mathbb{1} + \frac{2}{\delta}L\right)\boldsymbol{x} + \gamma'\boldsymbol{\eta}(t) \tag{3.74}$$

as defined in the previous chapter is determined by the dynamical constant $\delta := \frac{2\alpha}{\beta}$, the ratio of auto-dependency α and coupling strength β, as well as the reduced noise term $\gamma' = \frac{\gamma}{\alpha}$.

In order to investigate the influence of these parameters, we conducted a purely numerical analysis. We generated different data sets for varying δ and γ' but fixed topology A_{fix} and measured the reconstruction quality ϑ.
For each choice of δ and γ' we produced five datasets with same length M for two different time steps that guarantee stability of the Euler integration scheme in this setting.
The time delay τ needed for CI was chosen as small as possible, i.e. $\tau = \Delta t$. We only considered the undirected representation of the network A_{fix} for the reconstruction to produce results that are comparable to correlation thresholding (CT) for later use.

For the analysis we chose dynamical constants in the interval $\delta \in [0.01, 10.24]$ and a fixed noise strength of $\gamma' = 0.1$. The integration was performed using time steps $\Delta t = 0.0001$ and $\Delta t = 0.0005$ for $M = 20,000$ integration steps. A_{fix} is an adjacency matrix of a network with 50 nodes generated using the WS2 model with connection density $\rho_{\text{fix}} = 0.2$ and rewiring probability $q_{\text{fix}} = 0.25$.

Fig.3.16a shows the outcome of this analysis. If the sudden drop in reconstruction quality for the larger time step is disregarded, both curves show the same qualitative relation. The performance of covariance inversion is best for small δ, i.e. strongly coupled networks. Surprisingly, the reconstruction quality is better for the larger time step. This behavior is also visible in Fig.3.16b which shows the variance of the distributions of of the entry values r of the reconstructed Jacobian $\Re(\widehat{J})_{ij}$ corresponding to connections $p(r\frac{\delta}{2}|\text{conn.})$ and non-connections $p(r\frac{\delta}{2}|\text{non-conn.})$. To be able to compare different connection strengths, we normalized the values of $\Re(\widehat{J})_{ij}$ by dividing by the connection strength $\frac{2}{\delta}$. This way the conditional distributions are peaked at zero and one instead of zero and $\frac{2}{\delta}$.

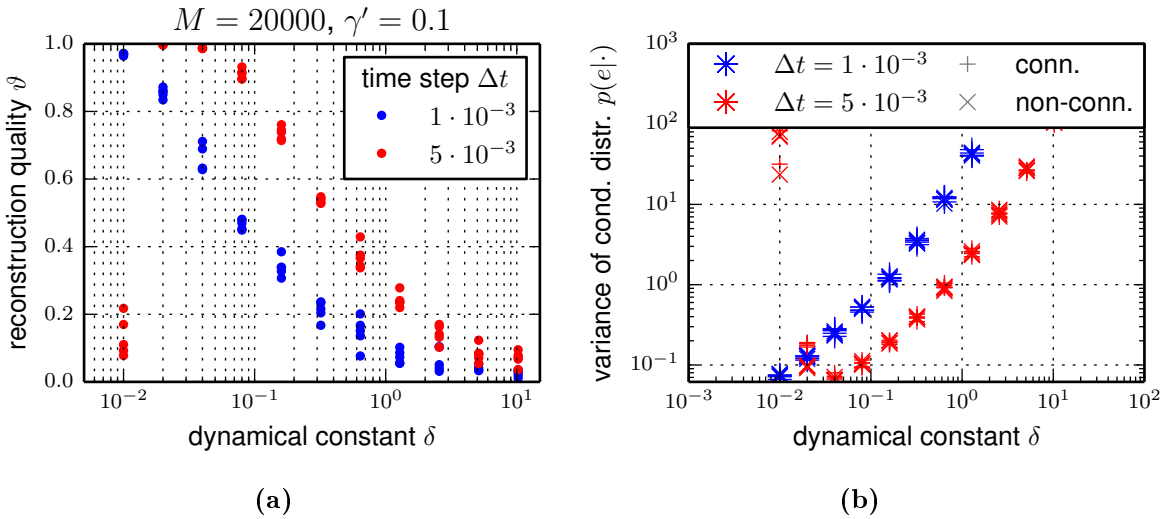
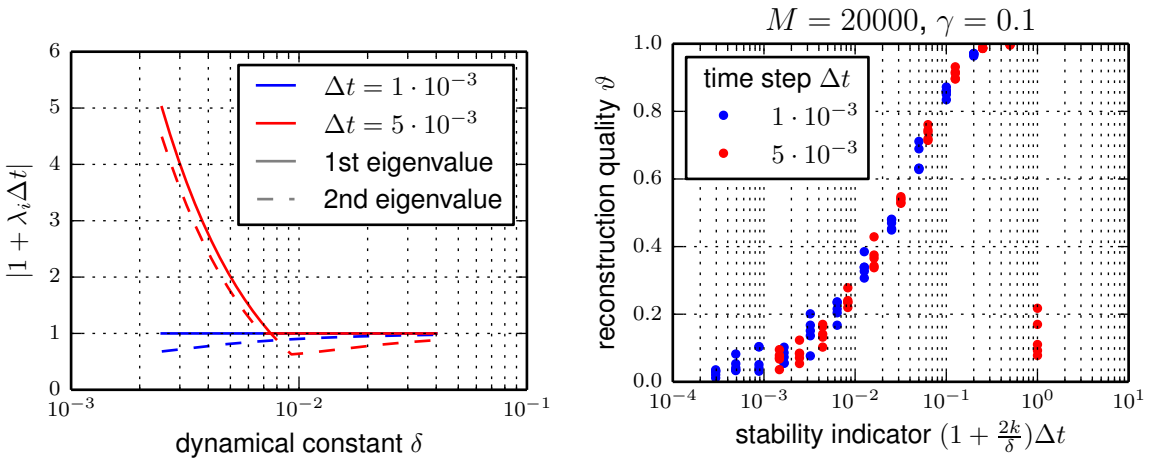


Figure 3.16: Reconstruction quality decreases with increasing dynamical constant δ. (a) Reconstruction quality of CI for different dynamical constants δ and time steps Δt. The reconstruction quality for the larger time step is better for all $\delta > 0.01$ than for the smaller time step. In general, CI is better for strongly coupled networks ($\delta \ll 1$). (b) The variances of the conditional distributions $p(\frac{r}{\beta}|\text{connection})$ and $p(\frac{r}{\beta}|\text{non-connection})$ increase sub-exponentially and almost power-law-like with increasing δ. Both distributions have almost identical shape. All datasets have the same size $M = 20,000$ and simulate the same network A_{fix} generated with the WS2 algorithm for $N = 50$ nodes, a connection density of $\rho = 0.2$ and a rewiring probability of $q = 0.25$. The noise strength was set to $\gamma' = 0.1$

The variances scale sub-exponentially and almost power-law-like with the dynamical constant δ. Also, the variances for the two different conditional distributions are almost equal for all δ. The sudden drop of the reconstruction quality for $\Delta t = 0.0005$ and $\delta = 0.01$ is an artifact of the Euler-Maruyama integration scheme which becomes unstable for these parameters (Fig.3.17a). It is shown that for $\Delta t = 0.0005$ in the vicinity of $\delta = 0.01$ the eigenvalues of the Jacobian matrix J' become problematic. In fact, the stability condition $|1 + \lambda \Delta t| \leq 1$ does not hold for $\delta \approx 0.007$, i.e. for $\delta = 0.01$ the integration scheme is at the brink of instability. The rule of thumb (Eq. 3.72) is also not fulfilled for $\delta = 0.01$ (Fig.3.17b). Thus, inaccurate datasets due to integration errors prevent the reconstruction.



(a) **(b)**

Figure 3.17: Stability of the Euler-Maruyama integration scheme.(a) The two largest values of the stability condition $|1 + \lambda \Delta t| \leq 1$ for the adjacency matrix A_{fix} used in Fig.3.16 and different time steps Δt dependent on the dynamical constant δ. For $\Delta t = 0.0005$ and $\delta = 0.007$ the integration becomes unstable, i.e. the condition is not fulfilled. (b) The reconstruction quality ϑ dependent on the fraction $(1 + 2k/\delta)\Delta t$ which serves as the rule of thumb to determine the stability of the numerical integration scheme. $2k\Delta t/\delta$ should be smaller then one. The reconstruction quality increases as the fraction approaches this boundary, where it decreases dramatically.

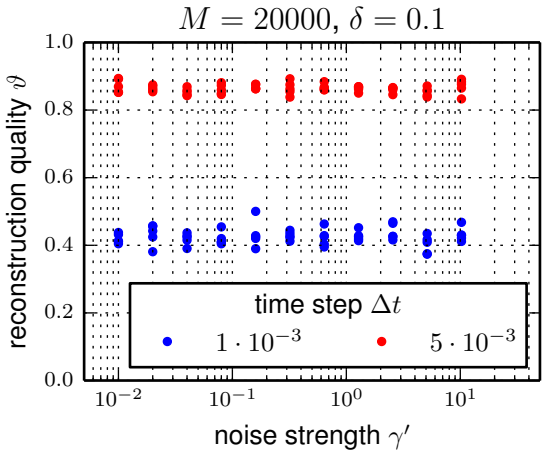


Figure 3.18: The reconstruction quality does not depend on the noise strength. The reconstruction quality ϑ plotted against the noise strength δ for different time steps Δt. Changes in γ' have no influence on the reconstruction quality. The dynamical constant was set to $\delta = 0.1$, the size of the time series is $M = 20,000$, the adjacency matrix of the system is again A_{fix} used in Fig.3.16.

The same analysis is done for a fixed dynamical constant δ and varying noise strength γ' (Fig.3.18). The reconstruction quality shows no particular dependency on the noise strength at all, since the plot shows a horizontal line.

We conclude that covariance inversion is more reliable for the reconstruction of strongly coupled systems.
The unintuitive outcome that datasets generated with larger time steps are easier to reconstruct is not fully understood yet. We speculate that either bigger time steps result in simulation errors that lead to a virtual increase of the coupling strength or that bigger changes in the covariance curves help to estimate the Jacobi matrix.

3.3.4. Performance of CI: Influence of the Topology

In this section, we investigate if topological differences interfere with the network reconstruction by CI. Our analysis is only numerical and reveals that CI has no topological preferences but that its reconstruction performance scales exponentially with the mean degree $k = (N-1)\rho$ and the time shift Δt.

In order to investigate which topological parameters covariance inversion depends on, we produced datasets of systems with different topology and size while keeping the dynamical parameters constant. We generated different networks by means of the WS2 algorithm for varying network size N, connection density ρ and rewiring probability q. For each network the Langevin equation of the Ornstein Uhlenbeck process (Eq. 3.74) was integrated for $M = 200,000$ time steps of size $\Delta t = 0.0001$. Afterwards, we reconstructed the dependency networks from the datasets for different time shifts $\tau = m\Delta t$ using the covariance inversion method and evaluated the reconstructions utilizing the reconstruction quality measure ϑ.

The dynamical system was set to $\delta = 50$ and $\gamma' = 0.1$. We generated $400 \cdot 11 = 4400$ reconstructed Jacobian matrices in total using eight different network sizes, ten different connection densities, five different rewiring probabilities and eleven different time shifts.
The size of the datasets $M = 200,000$ was chosen such, that the choice of the time shift and the resulting reduction of the dataset has an negligible effect on the covariance estimator. For

instance, choosing $m = 90$ instead of $m = 1$ results in a difference of less than 0.05% of the data points for the estimator of $\Sigma_{m\Delta t}$.

In Fig.3.19, we plotted the reconstruction quality ϑ as a function of the time shift $m = \frac{\tau}{\Delta t} = \#\Delta t$ ordered by connection density ρ for three network sizes. The plots reveal that the reconstruction quality decreases for increasing network size, density and chosen time shift. The five different topologies generated by five different rewiring probabilities q are not labeled. However, curves with the same color corresponding to different values of q coincide and show the same behavior. Hence, it is reasonable to assume that the reconstruction quality does not primarily depend on the topology of the system.

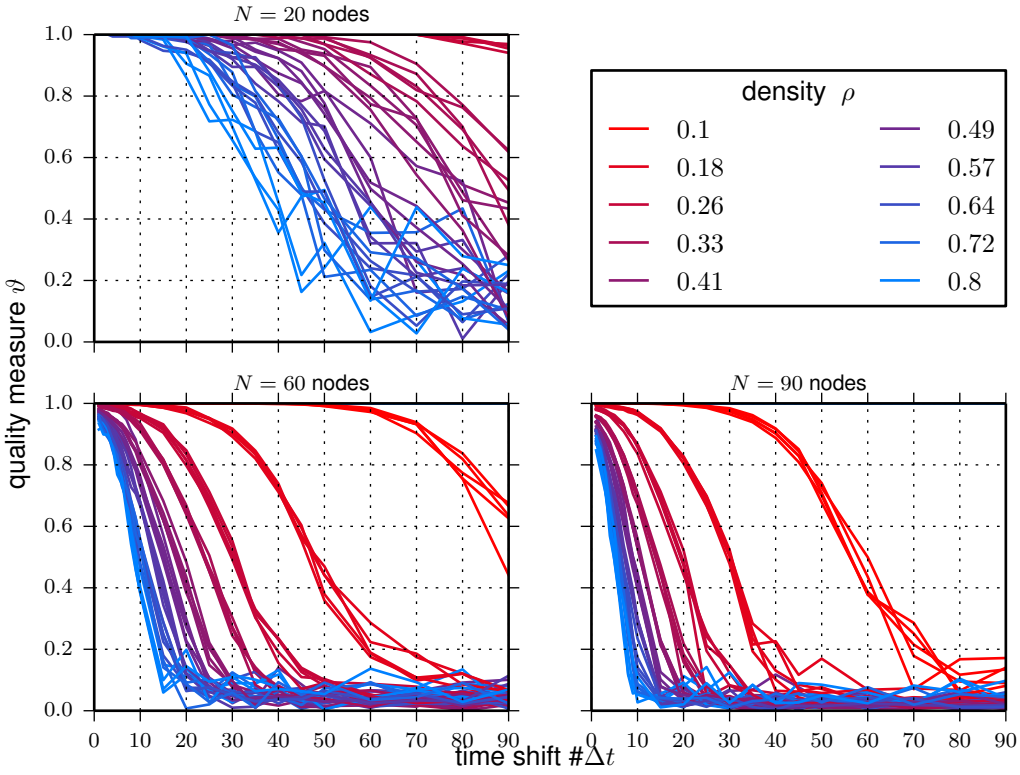


Figure 3.19: Performance of CI decreases with increasing network size N, connection density ρ and time shift $m = \#\Delta t$. The reconstruction quality ϑ of covariance inversion is shown in dependence of the used number of time shifts $\#\Delta t$. The graphs show the performance of CI for different network sizes. For each network size 50 networks were constructed using the WS2 model for 10 different connection densities ρ and 5 different rewiring probabilities $q \in \{0, 0.25, 0.5, 0.75, 1\}$. The density is color-coded.

Fig.3.20 (left plot) verifies this conjecture: CI is independent of q. It is illustrated how the variance of the conditional distributions depends on the topology determining parameter q for different system sizes N and constant connection densities ρ. The plot shows noisy horizontal lines which seem to be randomly distributed. However, for each specific combination of system size and connection density, the variances of connection and non-connection distributions follow the same curve. A general relationship between the variances and the network topology cannot be deduced. The straight lines suggest that CI is independent of the topology.
The right plot in Fig.3.20 shows the dependency between the variances and the connection density ρ for different network sizes and constant rewiring probability q. Here, the plots for constant N and different q coincide. Again, it is observable that variances for non-connection entry distribution and connection entry distribution show the same behavior. The variances increase for increasing connection density ρ and they increase faster for larger network sizes N. The plots confirm that the reconstruction quality is mostly independent of the system topology. They also show that the reconstruction primarily depends on the network size and the connection density.

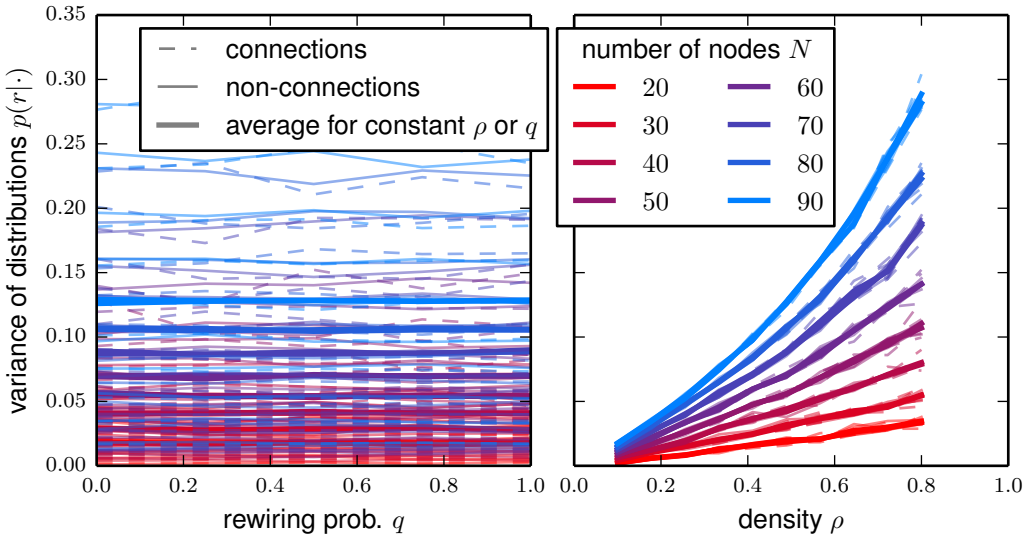


Figure 3.20: CI does not depend on the topology of the network q, but on the connection density ρ. (Left) The variances of the conditional distributions $p(r\frac{\delta}{2}|\text{connection})$ and $p(r\frac{\delta}{2}|\text{non-connection})$ as a function of the rewiring probability q for constant connection density ρ and different network sizes N. The straight curves indicate independence. (Right) Same variances as functions of ρ for constant q. Curves for different q values and the same network size coincide and reveal a functional connection. The time shift was kept $m = 1$.

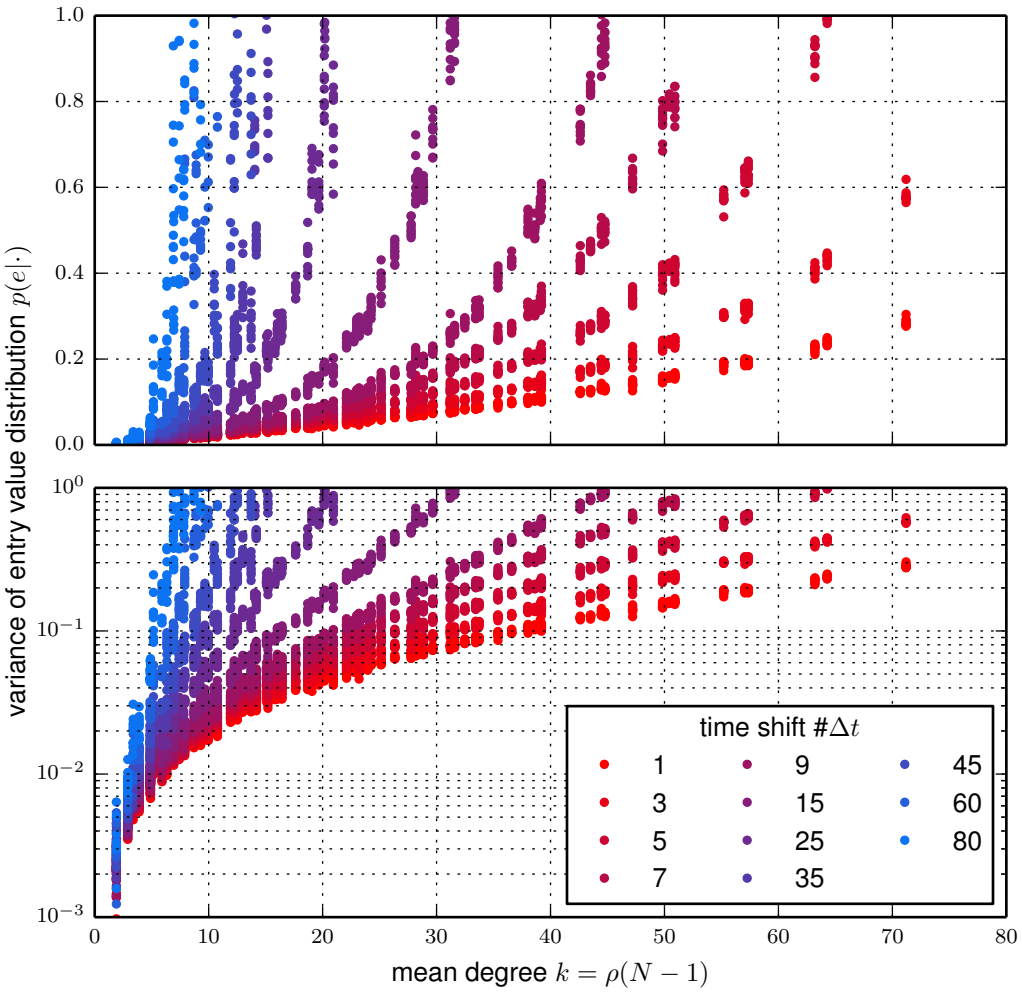


Figure 3.21: The variance of the distributions is a function of mean degree k and time shift $\#\Delta t$. The variance of the distribution of reconstructed Jacobian entry values corresponding to connection and non-connection as a function of the mean degree $k == \rho(N-1)$ for different time shifts $\#\Delta t$.The plots suggest a functional connection between the mean variance and the mean degree dependent on the time shift. (Top) Data plotted in linear scale reveals an almost linear relation between k and the variance of the distributions. (Bottom) Logarithmic scale shows that larger time shifts result in exponential scaling for medium mean degree.

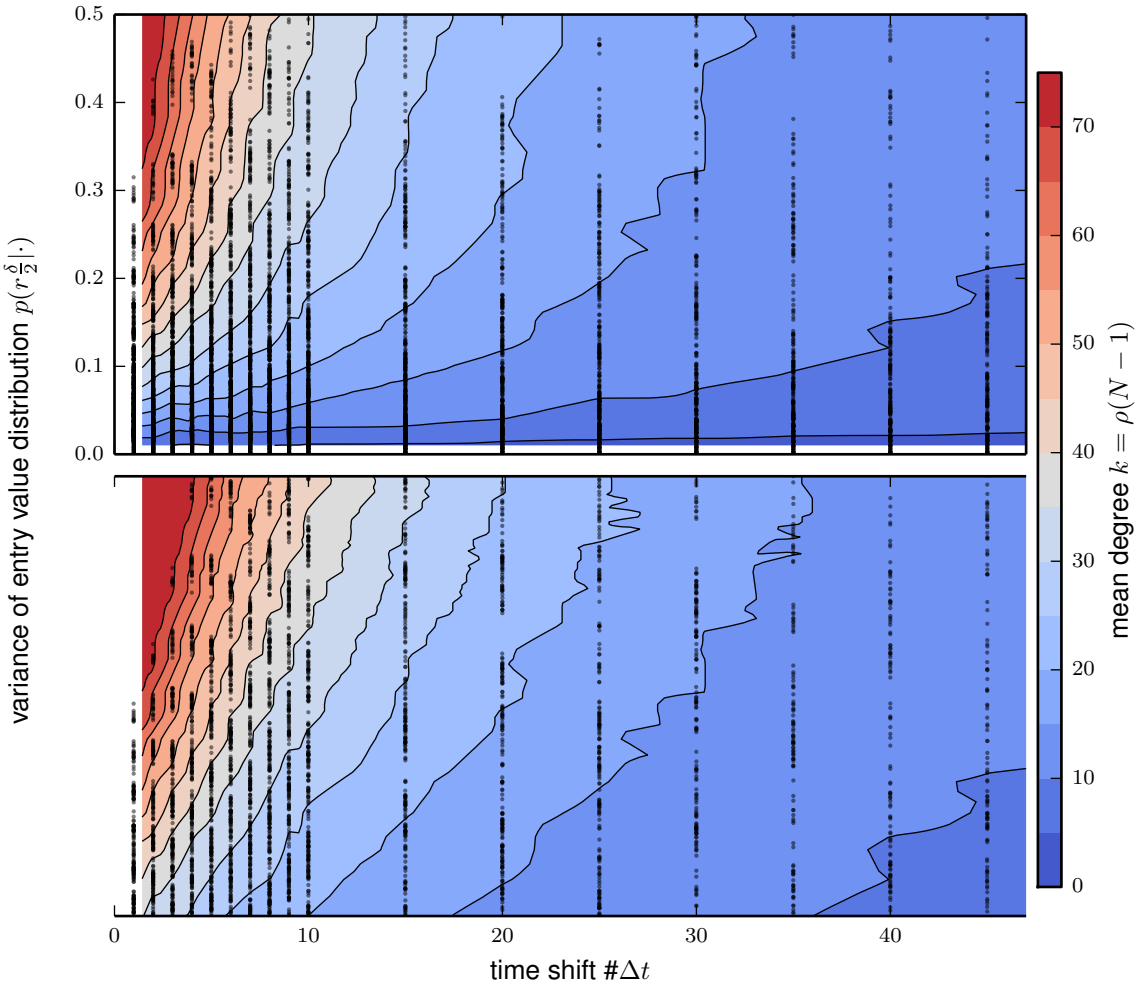


Figure 3.22: Variance scales exponentially with the mean degree k. The data of Fig.3.21 is used to generate an interpolated plot of the variance of the conditional distributions $p(r\frac{\delta}{2}|\text{connection})$ and $p(r\frac{\delta}{2}|\text{non-connection})$ dependent on the time shift m for constant mean degree $k = \rho(N-1)$. The data points utilized for the interpolation are plotted as black dots. (Top) Interpolated heat map with linear scaling. (Bottom) Logarithmic scaling shows horizontal lines indicating an exponential dependency.

After excluding the topology of the network as a quality determining parameter, three network parameters are left: The network size N, the connection density ρ and the time shift m which is used to reconstruct the Jacobian.
Fig.3.21 shows the variances for different time shifts depending on the mean degree $k = (N-1)\rho$.
Two observations can be made. First, all points for the same time shift lie on a smooth curve. This indicates that covariance inversion only depends on the product $\rho(N-1)$ and not on both parameters individually. Secondly, the logarithmic plot shows curves that become asymptoticly linear for large mean degree. Thus, the scaling for all variances becomes eventually exponential. Fig.3.22 shows an interpolation of the variance for different mean degrees dependent on the time shift. Again, the logarithmic plot reveals an exponential dependency.

3.3.5. Comparison of Covariance Inversion and Correlation Thresholding

In this section, we prove the usefulness of covariance inversion (CI), by comparing it with correlation thresholding (CT). The latter is a scientific standard method and intensively discussed in Sec.2.4.1. For this purpose, we applied CI and CT to reconstruct networks from surrogate data and used the quality measure ϑ to evaluate their performance.

We have shown numerically that CI works better for strongly coupled systems (Sec. 3.3.3, Fig.3.16a). By contrast, we found analytically that CT is always applicable to weakly coupled systems for infinite time series (Sec.3.2). But how does CT perform on real data, i.e. finite time series?
We compared the reconstruction quality of CT to that of CI for different dynamical constants δ (Fig.3.23). The analysis was based on the dataset of Fig.3.16a. It is shown that even though correlation thresholding is theoretically useful for weakly coupled systems ($1 \ll \delta$) the practical limitations of time series analysis prevent a reconstruction for large δ for both reconstruction methods. This limitation is caused by analytical cross-covariance values that are small compared to the statistical fluctuations of the covariance estimation. The relative estimation error increases, so that covariance values cannot be inferred reliably. For stronger coupled systems ($\delta \ll 1$), the relative estimation error in the entries of the covariance matrix are smaller, but correlation thresholding fails because second order effects render the systems

non-reconstructible by thresholding of measures of statistical dependence (Sec. 3.2.3).
This is the reason why the maximal performance of correlation thresholding is somewhere in between strongly coupled and weakly coupled systems (depending on the topology).
In contrast to CT, CI benefits from the improved estimation of strongly coupled systems without drawbacks and therefore shows better results in this dynamical regime. For weakly coupled systems both methods perform equally unsatisfactory.

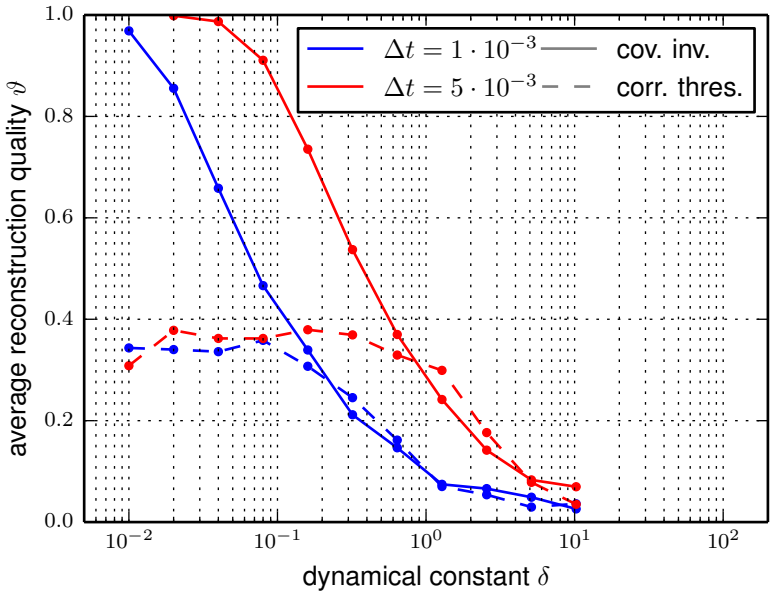


Figure 3.23: Reconstruction methods are best for strong coupling, CI outperforms CT. The reconstruction quality measure ϑ plotted versus the dynamical constant δ for different integration time steps and reconstruction methods. Full lines: covariance inversion (CI), dashed lines: correlation thresholding (CT). CI reaches maximal performance for $\delta \to 0$ and outperforms CT. Number of nodes $N = 50$ with adjacency matrix A_{fix} and dataset as in Fig.3.16 and Fig.3.17

Since the performance of CT depends on the topology, the differences in performance between CT and CI also do.
The question is: Which reconstruction method should be used to reconstruct which kind of topology? It was shown above, that strongly coupled systems are better reconstructed by CI, while weakly coupled systems are problematic for both methods. Thus, to answer this question, we keep the dynamical constant δ at an intermediate value and study how the reconstruction quality scales with increasing network size and increasing number of data points. Now, the question for each reconstruction method is: How many data points are needed to

faithfully reconstruct the network of physical interactions?

We chose three different parameter settings for the WS2 random graph algorithm for three different performance regimes of CT. We picked a setup that is beneficial for CT $(q, \rho) = (0.01, 0.14)$, an intermediate setup with $(q, \rho) = (0.02, 0.36)$ and a parameter combination that lies in the problematic regime of CT $(q, \rho) = (0.05, 0.47)$. For each setting, we constructed 25 networks for different network sizes from $N = 30$ to $N = 200$ nodes and generated trajectories using a time step of $\Delta t = 0.005$. These time series were then passed to both reconstruction methods using different amounts of data points M. The performance of each method was evaluated utilizing the reconstruction quality measure ϑ. We averaged the reconstruction quality ϑ over the ensemble of 25 networks to generate heat maps that reveal the scaling in N and M.

The result of this analysis is shown in Fig.3.24 and in Fig.3.25.

Fig.3.24 shows the reconstruction quality of both methods in comparison. Double logarithmic plots reveal that the contour lines follow a power law for both reconstruction methods. Also, the reconstruction quality of CT changes dramatically with the choice of topological parameters (q, ρ) while the reconstruction quality of CI is less effected by changes in (q, ρ). This confirms our expectations (see Sec. 3.2.4).

In Fig.3.25 we display the differences in reconstruction quality $\vartheta_{\text{ci}} - \vartheta_{\text{ct}}$ which show that CI is superior to CT in most cases. In fact, CT is only more successful than covariance inversion for $(q, \rho) = (0.01, 0.14)$, i.e. for more or less regular graphs.

For more complex network structures this analysis indicates that CI is the better choice. It should be noted, that this comparison only involves the reconstruction of the undirected representation of the network of physical interactions. Yet, CI is even capable of reconstructing the directed version. Hence, the performance of CT compared to CI is even worse if the actual network of physical interactions is considered.

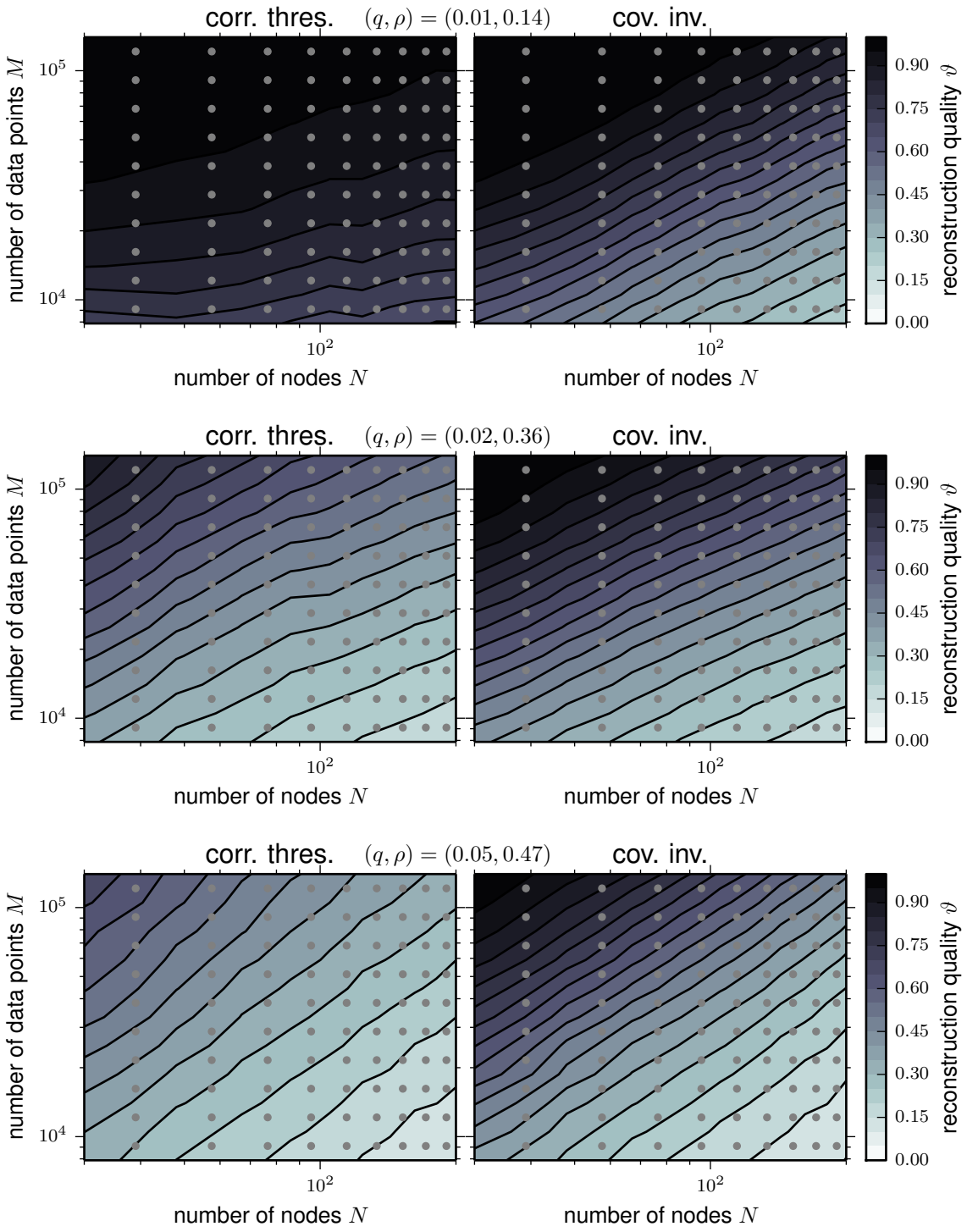


Figure 3.24: The performances of CT and CI follow power laws depending on Network Size N and length of time series M. Average reconstruction quality ϑ of correlation thresholding (left) and covariance inversion (right) depending on the number of nodes N and the number of data points M. (top) Rewiring probability $q = 0.01$ and connection density $\rho = 0.14$, (center) $q = 0.02$ and $\rho = 0.36$, (bottom) $q = 0.05$ and $\rho = 0.47$. Data points are drawn as gray dots. Networks are generated by means of the WS2 random graph algorithm. Each data point was averaged over 25 representations. Datasets are generated by numerical integration of OU processes with $\delta = 2$ and $\gamma = 0.1$ using a time step of $\Delta t = 0.005$. Covariance inversion was performed utilizing a minimal time shift.

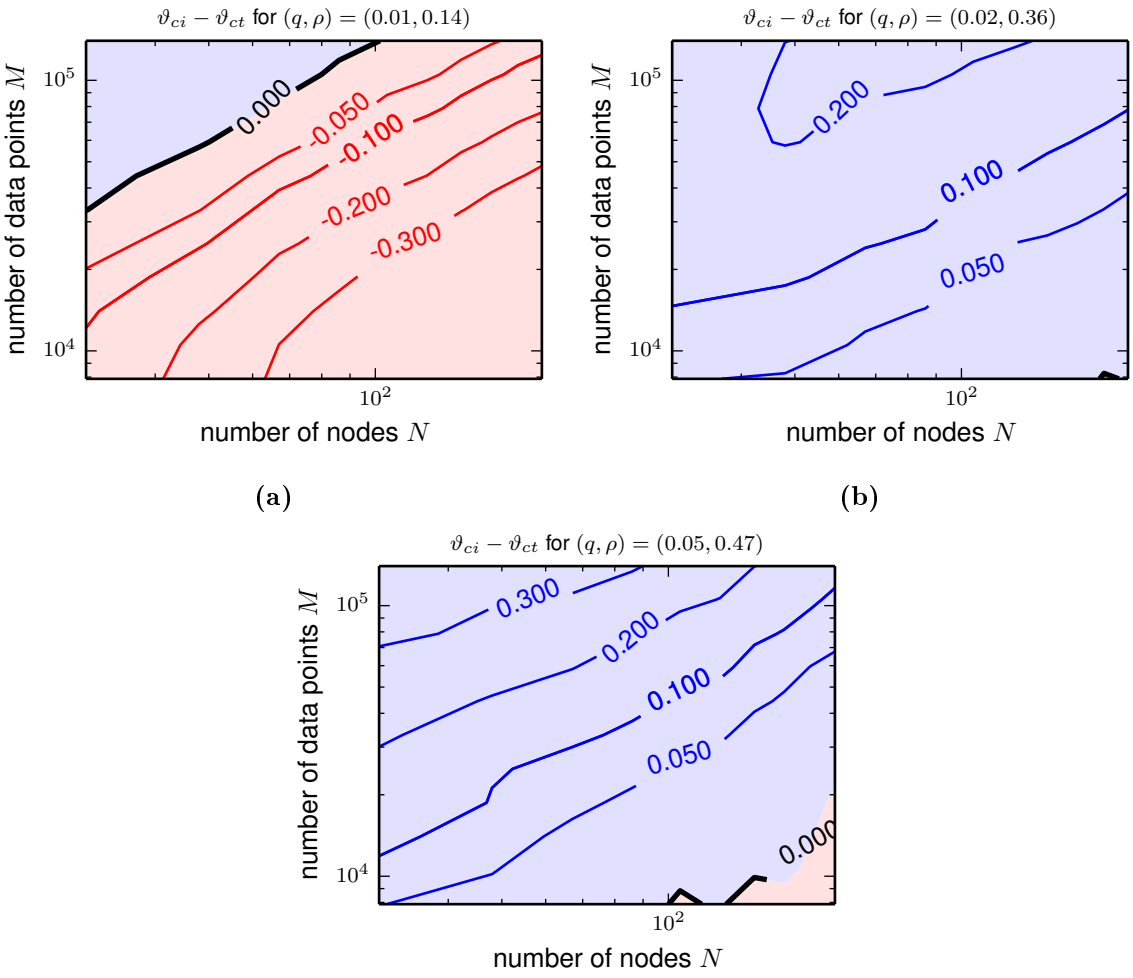


Figure 3.25: CI outperforms CT for irregular structures. The difference $\vartheta_{\mathrm{ci}} - \vartheta_{\mathrm{ct}}$ of the performance of reconstruction quality ϑ_{ci} and correlation thresholding ϑ_{ct}. (a) rewiring probability $q = 0.01$ and connection density $\rho = 0.14$, (b) $q = 0.02$ and $\rho = 0.36$, (c) $q = 0.05$ and $\rho = 0.47$. We used the same data as for Fig.3.24. Positive regions are shaded blue, negative regions are shaded red.

4. Conclusion

In this thesis, we studied the reconstruction of networks of physical interactions from stationary random processes. To investigate principal limits of reconstructibility in a minimal but non-trivial model, we chose to study homogeneous Ornstein-Uhlenbeck processes. This choice allows an exact analysis, in particular analytic computations of pairwise correlations. The homogeneity of the model refers to homogeneous auto-dependency, coupling strength and noise strength. Using this model, we focused on three basic questions addressing the very foundations of present network reconstruction approaches:

(1) *Does time-continuity interfere with the concept of Bayesian reconstruction methods?*

(2) *Which networks are reconstructible by thresholding of measures of statistical dependence?*

(3) *Is it possible to estimate the reconstruction error of thresholding approaches?*

In addition, we introduced our own inference approach, covariance inversion, for linear systems with homogeneous coupling strength and compared our method to correlation thresholding regarding their performance to reconstruct physical interactions.

In the first part of this thesis (Sec. 3.1), we analyzed the promising dynamical Bayesian network approach regarding its capability to cope with continuous-time random processes. We tested analytically if the powerful inductive causation algorithm (IC* algorithm [33, Sec. 2.6]) succeeds in reconstructing a chain topology without further adjustments.
We showed, that already this most basic network structure cannot be easily inferred. Time continuity causes additional correlations between all unconnected nodes which have at least one indirectly connected source in common. These spurious correlations cannot be reduced by any statistical conditioning and for any time lag. Thus, for time-continuous systems, there is a fundamental difference between the network of statistical dependencies and the network of physical interactions. The Bayesian approach only returns an approximation of the actual network of physical dependencies even in the ideal case of infinite time series.
This finding may have consequences for Bayesian reconstruction approaches which deal with time-continuous systems [14, 35]. Future usage of Bayesian network approaches for the reconstruction of physical interactions should thus include not only a statistical but also a

conceptual error analysis.

In the second part (Sec. 3.2), we investigated the performance of thresholding of pairwise measures of statistical dependence under idealized conditions. In particular, we analyzed the influence of dynamical and topological parameters on thresholding of linear correlations of Ornstein-Uhlenbeck dynamics.
We showed analytically that weakly coupled systems are in principle reconstructible by this method because second order effects can be neglected. However, numerical analysis of finite time series revealed that statistical fluctuations prevent the reconstruction in this regime. In strongly coupled systems, second order effects are dominant and complicate the reconstruction further.
We proved that regular k-rings are always reconstructible. Moreover, we illustrated for irregular topologies that the reconstruction quality crucially depends on the indegree distribution while the outdegrees distribution has only a minor influence. For networks with unimodal indegree distribution, we found that the variance of the indegree distribution and the connection density are the leading factors controlling the reconstruction quality.
Accordingly, even under idealized conditions, correlation thresholding does not perform well neither in the weak coupling limit nor in the strong coupling limit. Also, the reconstruction quality depends crucially on the topology that is ought to be reconstructed.
Hence, our findings clearly show that thresholding of measures of statistical dependence has conceptual limitations and that the topologies returned by these methods often have hardly any physical meaning.
Since thresholding methods are applied in various fields, the outcome of many studies may have to be seen in a different light (e.g., [4, 12, 28, 41] and certainly much more). Still, the found dependencies help to estimate reconstruction errors if the indegree distribution and the dynamical properties of the network are known or can be estimated beforehand.

In the third part of this thesis (Sec. 3.3), we introduced covariance inversion as a reconstruction method for linear systems. For infinite time series, this method always reconstructs directed physical interactions perfectly, independent of dynamical parameters, topology or correlated noise sources. Even noise strengths can be reconstructed. Comparable methods for less gen-

eral unidirectional networks have been proposed in [8,34].
We compared covariance inversion to correlation thresholding using surrogate data and found that our method outperforms correlation thresholding in most reasonable settings. Moreover, covariance inversion is capable of reconstructing directed networks whereas correlation thresholding is not. These results demonstrate that covariance inversion is applicable in general because it easily competes against a well excepted benchmark. However, this method has one downside. The trajectories of all involved units have to be known. This is rarely the case in real world scenarios.

Taken together, our findings provide parts of answers to the questions raised above.

(1) *Time-continuity weakens Bayesian network approaches because correlated but unconnected nodes can in general not be statistically separated in continuous-time systems.*

(2) *Homogeneous systems with constant indegree, including regular topologies, are well reconstructible. The best reconstruction performance is achieved for systems with moderate coupling strengths.*

(3) *If the dynamical properties and the indegree distribution of the system are known, the reconstruction error can be estimated by computationally analyzing smaller networks with the same properties.*

Overall we were able to shed some light on the interrelation of physical interactions, conditional dependencies and pairwise statistical dependencies. Our findings reveal that current network reconstruction methods have conceptual limitations when trying to infer physical interactions. In addition, we developed a new promising approach that focuses on eliminating exactly these limitations.
We hope that this thesis contributes new and helpful insights in the field of network reconstruction and gives fruitful impulses for future research.

Outlook

We were able to answer some questions about the interrelation of structural, Bayesian and effective networks and even introduced a novel inference method. However, our research raises new questions that may be answered in future projects.

For instance, we showed that time-continuous processes generate spurious correlations that represent a challenge for Bayesian reconstruction methods. This raises the question how the gained knowledge could be used to overcome this challenge and to improve existing inference algorithms.
Moreover, we have investigated the performance of correlation thresholding which considers traditionally only correlation values. It seems reasonable that it is possible to conceive an inference method that takes not only the correlation values but also the correlation curves as a whole into account.
And most importantly, we developed, covariance inversion, an inference method for linear systems based on covariance. In the light of the direct relationship between covariance and mutual information for linear systems (see Appendix C.2), we have to ask ourselves if covariance inversion can be used to infer structural connectivities in non-linear systems. Even if covariance inversion only yields an approximation of the system, this would still yield a new approach to the reconstruction of networks.

We believe that answering these questions would lead to tremendous improvements of existing reconstruction methods and thus would amount to novel and deep insights in the structure and inner workings of networks that define and determine us and the world we are living in.

A. Appendix: Theorems

A.1. Gershgorin's Theorem

Theorem 1 (Gershgorin's Theorem)**.** *Let $A \in \mathbb{K}^{n\times n}$ and λ be an eigenvalue of A, then*

$$\exists i \in \{1, \dots n\} : \quad \|\lambda - A_{ij}\| \leq \sum_{j\neq i}^{n} \|A_{ij}\| \tag{A.1}$$

In order to prove Theorem 1 we make use of strictly diagonally dominant matrices.

Definition A.1.1 (strictly diagonally dominant)**.** Let A be a matrix $A \in \mathbb{K}^{n\times n}$, then it is strictly diagonally dominant if and only if

$$\|A_{ii}\| > \sum_{j\neq i}^{n} \|A_{ij}\| \quad \forall i \in \{1, \dots n\} \tag{A.2}$$

Theorem 2. *A strictly diagonally dominant matrix is not singular.*

Proof. Let us assume $A \in \mathbb{K}^{n\times n}$ is strictly diagonally dominant and singular, then

$$\exists \boldsymbol{u} \in \mathbb{K}^n,\ \boldsymbol{u} \neq \mathbf{0} \ : \quad A\boldsymbol{u} = \mathbf{0} \tag{A.3}$$

Let us denote the maximal absolute value of $\boldsymbol{u}$ as u_i, then

$$A_{ii}u_i + \sum_{j\neq i}^{n} A_{ij}u_j = 0 \tag{A.4}$$

$$\Leftrightarrow \quad \|A_{ii}\|\|u_i\| \leq \sum_{i\neq j}^{n} \|A_{ij}\|\|u_j\| \leq \sum_{i\neq j}^{n} \|A_{ij}\|\|u_i\| \tag{A.5}$$

$$\Rightarrow \quad \|A_{ii}\| \leq \sum_{i\neq j}^{n} \|A_{ij}\| \quad \lightning \tag{A.6}$$

Hence, there is no singular strictly diagonally dominant matrix. □

Proof, Gershgorin's Theorem. Let us assume $A \in \mathbb{K}^{n\times n}$ does not fulfill Theorem 1. Hence, there exists an eigenvalue λ for which

$$\|\lambda - A_{ij}\| > \sum_{j\neq i}^{n} \|A_{ij}\| \quad \forall i \in \{1, \dots, n\} \tag{A.7}$$

Therefore, the matrix $M := \lambda \mathbb{1} - A$ is strictly diagonally dominant and for this reason non-singular. This contradicts that λ is eigenvalue of A. ↯

Hence, Gershgorin's Theorem holds. □

A.2. Conditional Covariance

Let $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$ be a Gaussian distributed multivariate random variable with

$$\text{mean} \qquad \boldsymbol{z_0} = (\boldsymbol{x_0}, \boldsymbol{y_0}) \quad \text{and} \tag{A.8}$$

$$\text{covariance matrix} \quad \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^\intercal & \Sigma_{yy} \end{pmatrix} \quad \text{with} \quad \Sigma^{-1} = \begin{pmatrix} A & B \\ B^\intercal & C \end{pmatrix} \tag{A.9}$$

where $A = A^\intercal$ and $C = C^\intercal$. For reasons of simplicity, all vectors are row vectors.

We define $U := \Sigma_{xy}\Sigma_{yy}^{-1}$. The inverse covariance matrix is given by

$$A = \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^\intercal\right)^{-1} \tag{A.10}$$

$$B = -AU \qquad B^\intercal = -U^\intercal A \tag{A.11}$$

$$C = \Sigma_{yy}^{-1} + U^\intercal A U \ . \tag{A.12}$$

The joint probability distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by

$$\begin{aligned}
p(\boldsymbol{x}, \boldsymbol{y}) &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0, \boldsymbol{y} - \boldsymbol{y}_0)\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{x}_0, \boldsymbol{y} - \boldsymbol{y}_0)^\intercal\right\} \\
&= \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x} - \boldsymbol{x}_0)A(\boldsymbol{x} - \boldsymbol{x}_0)^\intercal + (\boldsymbol{x} - \boldsymbol{x}_0)B(\boldsymbol{y} - \boldsymbol{y}_0)^\intercal\right.\right. \\
&\qquad \left.\left. + \ (\boldsymbol{y} - \boldsymbol{y}_0)B^\intercal(\boldsymbol{x} - \boldsymbol{x}_0)^\intercal + (\boldsymbol{y} - \boldsymbol{y}_0)C(\boldsymbol{y} - \boldsymbol{y}_0)^\intercal\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x} - \boldsymbol{x}_0)A(\boldsymbol{x} - \boldsymbol{x}_0)^\intercal - (\boldsymbol{x} - \boldsymbol{x}_0)AU(\boldsymbol{y} - \boldsymbol{y}_0)^\intercal\right.\right. \\
&\qquad \left.\left. - \ (\boldsymbol{y} - \boldsymbol{y}_0)U^\intercal A(\boldsymbol{x} - \boldsymbol{x}_0)^\intercal + (\boldsymbol{y} - \boldsymbol{y}_0)\left(\Sigma_{yy}^{-1} + U^\intercal A U\right)(\boldsymbol{y} - \boldsymbol{y}_0)^\intercal\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x} - \boldsymbol{x}_0) - (\boldsymbol{y} - \boldsymbol{y}_0)U^\intercal\right]A\left[(\boldsymbol{x} - \boldsymbol{x}_0) - (\boldsymbol{y} - \boldsymbol{y}_0)U^\intercal\right]^\intercal\right. \\
&\qquad \left. -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{y}_0)\Sigma_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{y}_0)^\intercal\right\} \ .
\end{aligned} \tag{A.13}$$

The marginal probability distribution of $\boldsymbol{Y}$ is given by

$$p(y) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{y_0})\Sigma_{yy}^{-1}(\boldsymbol{y}-\boldsymbol{y_0})^{\intercal}\right\} . \tag{A.14}$$

Hence, the conditional probability distribution is given by

$$\begin{aligned} p(x\,|\,y) &= \frac{p(x,y)}{p(y)} \\ &\propto \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{x}_0)-(\boldsymbol{y}-\boldsymbol{y}_0)U^{\intercal}\right]A\left[(\boldsymbol{x}-\boldsymbol{x}_0)-(\boldsymbol{y}-\boldsymbol{y}_0)U^{\intercal}\right]^{\intercal}\right\} . \end{aligned} \tag{A.15}$$

This defines a normal distribution with

$$\text{mean} \qquad \bar{\boldsymbol{x}}_0 = \boldsymbol{x}_0 + (\boldsymbol{y}-\boldsymbol{y}_0)U^{\intercal} = \boldsymbol{x}_0 + (\boldsymbol{y}-\boldsymbol{y}_0)\Sigma_{yy}^{-1}\Sigma_{xy}^{\intercal} \tag{A.16}$$

$$\text{covariance matrix} \qquad \Sigma_{x|y} = A^{-1} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^{\intercal} . \tag{A.17}$$

For column vectors this reads

$$\text{mean} \qquad \bar{\boldsymbol{x}}_0 == \boldsymbol{x}_0 + \Sigma_{xy}\Sigma_{yy}^{-1}(\boldsymbol{y}-\boldsymbol{y}_0) \tag{A.18}$$

$$\text{covariance matrix} \qquad \Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^{\intercal} . \tag{A.19}$$

B. Appendix: Correlation Thresholding

B.1. Dependence of Dynamical Constant δ

B.1.1. Common Cause Structure: Calculations

The correlation

$$C_{xy} = \sqrt{\left(1 + \frac{m}{(\delta + m)}\right) \frac{1}{(\delta^2 + m\delta + 2m)}} \tag{B.1}$$

of Eq. 3.27 is monotonically decreasing in δ for $\delta \geq 0$. To see this, we prove that products of positive monotonically decreasing functions are also positive and monotonically decreasing: Let $f(x), g(x)\mathbb{R}^+$ be two positive monotonically decreasing functions in the interval $x \in [a, b] \subseteq \mathbb{R}$, i.e. $f'(x), g'(x) \leq 0$. Then the function $h(x) = f(x) \cdot g(x)$ is also positive and its derivative

$$h'(x) = \underbrace{f'(x)g(x)}_{<0} + \underbrace{f(x)g'(x)}_{<0} \tag{B.2}$$

negative in interval $x \in [a, b] \subseteq \mathbb{R}$.
Since C_{xy} is a product of two functions that are clearly positive and monotonically decreasing for $\delta \geq 0$, the correlation C_{xy} is also monotonically decreasing for $\delta \geq 0$.

B.1.2. Relay Structure: Calculations

The matrix Λ is given by

$$\Lambda_{11} = e^{-2mt} + \frac{m\left(e^{-t} - e^{-mt}\right)^2}{(m-1)^2} + \frac{\left(m(e^{-t} - 1) - e^{-mt} + 1\right)^2}{(m-1)^2} \tag{B.3}$$

$$\Lambda_{22} = \ldots = \Lambda_{N-1,N-1} = e^{-2t} + \left(e^{-t-1}\right)^2 \tag{B.4}$$

$$\Lambda_{12} = \ldots = \Lambda_{1,N-1} = \frac{e^{-t} - e^{-mt}}{m-1} e^{-t} + \frac{m(e^{-t} - 1) - 1 + e^{-mt}}{m-1}(1 - e^{-t}) \tag{B.5}$$

$$\Lambda_{1N} = \frac{m(e^{-t} - 1) - 1 + e^{-mt}}{m-1} \quad \Lambda_{2N} = \ldots = \Lambda_{N-1,N} = (1 - e^{-t}) \quad \Lambda_{NN} = 1\,. \tag{B.6}$$

All entries that are not given by $\Lambda = \Lambda^\intercal$ are found to be

$$\Lambda_{ij} = (1 - e^{-t})^2\,. \tag{B.7}$$

The covariance matrix Σ_0 is given by

$$\Sigma_{11} = \frac{4m^3 + 4m^2 + \delta^5 + \delta^4(2m+4) + \delta^3(m^2+9m+5) + \delta^2(5m^2+9m+2) + \delta(10m^2+2m)}{\delta(m+\delta)(2m+\delta)(\delta+1)(\delta+2)(m+\delta+1)} \tag{B.8}$$

$$\Sigma_{22} = \ldots \Sigma_{N-1,N-1} = \frac{\delta^2+\delta+2}{\delta(\delta+1)(\delta+2)} \tag{B.9}$$

$$\Sigma_{12} = \ldots = \Sigma_{1,N-1} = \frac{2m^2+4m\delta+2m+\delta^3+\delta^2(m+1)}{\delta(m+\delta)(\delta+1)(\delta+2)(m+\delta+1)} \tag{B.10}$$

$$\Sigma_{1N} = \frac{m}{\delta(m+\delta)(\delta+1)} \quad \Sigma_{2N} = \ldots = \Sigma_{N-1,N} = \frac{1}{\delta(\delta+1)} \quad \Sigma_{NN} = \frac{1}{\delta}\,. \tag{B.11}$$

All entries that are not given by $\Sigma = \Sigma^\intercal$ are found to be

$$\Sigma_{ij} = \frac{2}{\delta(\delta+1)(\delta+2)}\,. \tag{B.12}$$

The correlation

$$C_{xy} = \sqrt{\left(1+\frac{1}{\delta+1}\right)\frac{1}{\delta^2+\delta+2}} \tag{B.13}$$

of Eq. 3.33 is monotonically decreasing because it is a product of positive monotonically decreasing functions (see above).

The correlation

$$\begin{aligned} C_{yx} &= \sqrt{2m+\delta}\left(2m^2+4m\delta+2m+\delta^3+\delta^2(m+1)\right)\cdot \\ &\quad \sqrt{(m+\delta)(m+\delta+1)(\delta^2+\delta+2)\left(4m^3+4m^2+\delta^5+\delta^4(2m+4)+\right.} \\ &\quad \left.\overline{\delta^3(m^2+9m+5)+\delta^2(5m^2+9m+2)+\delta(10m^2+2m)}\right)^{-1} \\ &= \sqrt{\frac{f(\delta,m)}{g(\delta,m)}} \end{aligned} \tag{B.14}$$

of Eq. 3.34 is monotonously decreasing in $\delta \in \mathbb{R}^+$. We prove this by differentiating $\frac{f(\delta,m)}{g(\delta,m)}$ with respect to δ.

$$\partial_\delta \frac{f(\delta,m)}{g(\delta,m)} = \frac{g(\delta,m)\cdot\partial_\delta f(\delta,m) - f(\delta,m)\cdot\partial_\delta g(\delta,m)}{g^2(\delta,m)} \tag{B.15}$$

This derivative is alway negative for $\delta \geq 0$ because

$$\begin{aligned} g(\delta,m)\cdot\partial_\delta f(\delta,m) - f(\delta,m)\cdot\partial_\delta g(\delta,m) = \\ &- 5\delta^{12} - \delta^{11}(34m+30) - \delta^{10}(92m^2+224m+78) \\ &- \delta^{9}(+128m^3+680m^2+640m+116) \\ &- \delta^{8}(97m^4+1096m^3+2129m^2+1070m+105) \\ &- \delta^{7}(38m^5+1018m^4+3758m^3+3818m^2+1162m+54) \\ &- \delta^{6}(6m^6+548m^5+3825m^4+7006m^3+4437m^2+838m+12) \\ &- \delta^{5}(160m^6+2256m^5+7320m^4+8232m^3+3432m^2+392m) \\ &- \delta^{4}(20m^7+718m^6+4506m^5+8486m^4+6246m^3+1696m^2+112m) \\ &- \delta^{3}(96m^7+1628m^6+5068m^5+6108m^4+2852m^3+488m^2+16m) \\ &- \delta^{2}(332m^7+1708m^6+3348m^5+2468m^4+688m^3+64m^2) \\ &- \delta(32m^8+288m^7+960m^6+1120m^5+480m^4+64m^3) \\ &- (16m^8+112m^7+208m^6+144m^5+32m^4) \leq 0\,. \end{aligned} \tag{B.16}$$

Hence, C_{yx} is monotonically decreasing for $\delta \geq 0$.

The correlation

$$\begin{aligned} C_{xx} &= m\sqrt{(2m+\delta)(\delta+2)(m+\delta+1)}\cdot \\ &\quad \sqrt{(m+\delta)(\delta+1)(4m^3+4m^2+\delta^5+\delta^4(2m+4)+} \\ &\quad \overline{\delta^3(m^2+9m+5)+\delta^2(5m^2+9m+2)+\delta(10m^2+2m)}^{-1} \\ &= \sqrt{\frac{f(\delta,m)}{g(\delta,m)}} \end{aligned} \tag{B.17}$$

of Eq. 3.36 is also monotonously decreasing in $\delta \in \mathbb{R}^+$.
We prove this by differentiating $\frac{f(\delta,m)}{g(\delta,m)}$ with respect to δ

$$\partial_\delta \frac{f(\delta,m)}{g(\delta,m)} = \frac{g(\delta,m)\partial_\delta f(\delta,m) - f(\delta,m)\partial_\delta g(\delta,m)}{g^2(\delta,m)}\,, \tag{B.18}$$

the nominator of which

$$\begin{aligned} g(\delta,m)\partial_\delta f(\delta,m) - f(\delta,m)\partial_\delta g(\delta,m) = \\ &- 4\delta^9 m^2 - \delta^8(24m^3+30m^2) - \delta^7(54m^4+176m^3+90m^2) \\ &- \delta^6(58m^5+386m^4+510m^3+138m^2) \\ &- \delta^5(30m^6+404m^5+1064m^4+740m^3+114m^2) \\ &- \delta^4(6m^7+204m^6+1054m^5+1420m^4+566m^3+48m^2) \\ &- \delta^3(40m^7+508m^6+1292m^5+960m^4+216m^3+8m^2) \\ &- \delta^2(98m^7+592m^6+774m^5+304m^4+32m^3) \\ &- \delta(128m^7+304m^6+208m^5+32m^4) - (8m^8+48m^7+56m^6+16m^5) \end{aligned} \tag{B.19}$$

is always negative for $\delta \geq 0$.

Hence, C_{xx} is monotonically decreasing for $\delta \geq 0$.

B.2. Correlation Thresholding in k-rings

B.2.1. Fourier Transform

The Fourier Transform $\mathcal{F}[\,\cdot\,]$ and its inverse $\mathcal{F}^{-1}[\,\cdot\,]$ of a sequence $a \widehat{=} (a)_{n=-\infty}^{\infty}$ with periodicity N is given by

$$\begin{aligned} \alpha = \mathcal{F}[\,a\,] \qquad & \alpha_m = \mathcal{F}[\,a\,]_m := \sum_{n=0}^{N-1} a_n e^{-2\pi \frac{nm}{N}} \\ a = \mathcal{F}^{-1}[\,\alpha\,] \qquad & a_n = \mathcal{F}^{-1}[\,\alpha\,]_n := \frac{1}{N}\sum_{n=0}^{N-1} \alpha_m e^{2\pi \frac{nm}{N}} \ . \end{aligned} \tag{B.20}$$

Inverse Fourier transforms of scalar products of sequences are convolutions of their individual inverse Fourier transforms: $\mathcal{F}^{-1}[\,\alpha \cdot \beta\,] = \mathcal{F}^{-1}[\,\alpha\,] * \mathcal{F}^{-1}[\,\beta\,]$.

Proof. Let α, β be two sequences with periodicity N and inverse Fourier transforms $a = \mathcal{F}^{-1}[\,\alpha\,]$, $b = \mathcal{F}^{-1}[\,\beta\,]$, then

$$\mathcal{F}^{-1}[\,\alpha \cdot \beta\,]_n = \frac{1}{N}\sum_{m=0}^{N-1} \alpha_m \beta_m e^{2\pi i \frac{nm}{N}} \tag{B.21}$$

$$= \frac{1}{N}\sum_{m=0}^{N-1}\left(\sum_{l=0}^{N-1} a_l e^{-2\pi i \frac{lm}{N}}\right)\beta_m e^{2\pi \frac{nm}{N}} \tag{B.22}$$

$$= \sum_{l=0}^{N-1} a_l \left(\frac{1}{N}\sum_{m=0}^{N-1}\beta_m e^{2\pi \frac{(n-l)m}{N}}\right) \tag{B.23}$$

$$= \sum_{l=0}^{N-1} a_l b_{n-l} = \mathcal{F}^{-1}[\,\alpha\,] * \mathcal{F}^{-1}[\,\beta\,] \tag{B.24}$$

□

B.2.2. Fourier Transform $s := \mathcal{F}[\,\sigma\,]$

We define the Fourier transform $s := \mathcal{F}[\,\sigma\,]$ of σ.

The sequence σ has to fulfill Eq. 3.52

$$\sum_{l=1}^{k} \sigma_{n-l} - (\delta + 2k)\sigma_n + \sum_{l=1}^{k} \sigma_{n+l} = -\frac{\gamma^2}{\beta}\delta_{0,n}\,. \tag{B.25}$$

Multiplying by $e^{-2\pi i \frac{nm}{N}}$ and summing the resulting equation over all $m \in [0, N-1]$ yields

$$\sum_{m=0}^{N-1}\left\{\sum_{l=1}^{k}\sigma_{n-l}e^{-2\pi i\frac{nm}{N}} - (\delta+2k)\sigma_n e^{-2\pi i\frac{nm}{N}} + \sum_{l=1}^{k}\sigma_{n+l}e^{-2\pi i\frac{nm}{N}}\right\} = -\frac{\gamma^2}{\beta} \tag{B.26}$$

$$\Rightarrow \quad \sum_{l=1}^{k}\left\{\sum_{m=0}^{N-1}\sigma_{n-l}e^{-2\pi i\frac{nm}{N}} - \frac{\delta+2k}{k}\sum_{m=0}^{N-1}\sigma_n e^{-2\pi i\frac{nm}{M}} + \sum_{m=0}^{N-1}\sigma_{n+l}e^{-2\pi i\frac{nm}{N}}\right\} = -\frac{\gamma^2}{\beta} \tag{B.27}$$

$$\Rightarrow \quad \sum_{l=1}^{k} e^{-2\pi i\frac{lm}{N}} s_m - (\delta+2k)s_m + \sum_{l=1}^{k} e^{2\pi i\frac{lm}{N}} s_m = -\frac{\gamma^2}{\beta} \tag{B.28}$$

$$\Rightarrow \quad s_m = \frac{\gamma^2}{\beta}\frac{1}{\delta + 2k - 2\sum_{l=1}^{k}\cos\left(2\pi\frac{lm}{N}\right)}\,. \tag{B.29}$$

B.2.3. Inverse Fourier Transform $\sigma = \mathcal{F}^{-1}[\, s\,]$

We rewrite s_m from above:

$$\begin{aligned} s_m &= \mathcal{F}[\,\sigma\,]_m \\ &= \frac{\gamma^2}{\beta} \frac{1}{(\delta + 2k + 1) - \underbrace{\left(2\sum_{l=1}^{k} \cos\left(2\pi \frac{lm}{N}\right) + 1\right)}_{:= z_{k,m}}} \\ &= \frac{\gamma^2}{\beta(\delta + 2k + 1)} \left(1 - \frac{z_{k,m}}{\delta + 2k + 1}\right)^{-1} \\ &= \frac{\gamma^2}{\beta(\delta + 2k + 1)} \sum_{l=0}^{\infty} \left(\frac{z_{k,m}}{\delta + 2k + 1}\right)^{l} \end{aligned} \tag{B.30}$$

Here, we used the geometric series and the fact that $|z_{k,m}| < \delta + 2k + 1$ for all $\delta > 0$.
$z_k \widehat{=} (z_{k,m})_{m=-\infty}^{\infty}$ is a periodic sequence the inverse Fourier transform of which $\zeta_k := \mathcal{F}^{-1}[\, z_k\,]$ yields

$$\begin{aligned} \zeta_{k,n} &= \mathcal{F}^{-1}[\, z_k\,]_n \\ &= \frac{1}{N} \sum_{m=0}^{N-1} z_{k,m} e^{2\pi i \frac{nm}{N}} \\ &= \frac{1}{N} \sum_{m=0}^{N-1} \left\{2\sum_{l=1}^{k} \cos\left(2\pi \frac{lm}{N}\right) + 1\right\} e^{2\pi i \frac{nm}{N}} \\ &= \sum_{l=-k}^{k} \frac{1}{N} \sum_{m=0}^{N-1} e^{2\pi i \frac{(n-l)m}{N}} = \sum_{l=-k}^{k} \delta_{nl}\ , \end{aligned} \tag{B.31}$$

which is the periodic step sequence

$$\zeta_{k,n} = \begin{cases} 1 & \text{if} \quad n \mod N \leq k \quad \text{or} \quad n \mod N \geq N - k \\ 0 & \text{otherwise} \end{cases} . \tag{B.32}$$

We iteratively define the sequence ζ_k^{*l} of sequences

$$\zeta_k^{*l} := (\zeta_k * \zeta_k^{*(l-1)})\ , \quad \zeta_k^{*1} = \zeta_k\ . \tag{B.33}$$

Thus, the inverse Fourier transform $\sigma = \mathcal{F}^{-1}[\,s\,]$ yields

$$\begin{aligned}\sigma_n = \mathcal{F}^{-1}[\,s\,]_n &= \frac{1}{N}\sum_{m=0}^{N-1} s_m e^{2\pi i \frac{nm}{N}} \\ &= \frac{\gamma^2}{\beta(\delta+2k+1)}\sum_{l=0}^{\infty}\frac{1}{N}\sum_{m=0}^{N-1}\left(\frac{z_{k,m}}{\delta+2k+1}\right)^l e^{2\pi i \frac{nm}{N}} \\ &= \frac{\gamma^2}{\beta(\delta+2k+1)}\left\{\delta_{0n} + \sum_{l=1}^{\infty}\frac{\mathcal{F}^{-1}[\,z_k^l\,]_n}{(\delta+2k+1)^l}\right\} \\ &= \frac{\gamma^2}{\beta(\delta+2k+1)}\left\{\delta_{0n} + \sum_{l=1}^{\infty}\frac{\zeta_{k,n}^{*l}}{(\delta+2k+1)^l}\right\} \end{aligned} \tag{B.34}$$

Hence, the covariance σ_n between two nodes i and $(i+n)$ is an infinite weighted sum of simple sequences.

B.2.4. Monotony of ζ_k^{*l}

Let ζ_k be the periodic step sequence

$$\zeta_{k,n} = \begin{cases} 1 & \text{if} \quad n \mod N \le k \quad \text{or} \quad n \mod N \ge N-k \\ 0 & \text{otherwise} \end{cases} . \tag{B.35}$$

and let the sequence of sequences ζ_k^{*l} be defined by

$$\zeta_k^{*l} := (\zeta_k * \zeta_k^{*(l-1)}) \,, \quad \zeta_k^{*1} = \zeta_k \,. \tag{B.36}$$

Furthermore, let $k, N \in \mathbb{N}$ and $\delta > 0$ with $2k+1 < N$.

Then, for all l, ζ_k^{*l} is symmetric (i.e. invariant under $n \mapsto -n$).

Proof. Mathematical induction.

(Basis)

We note that $\zeta_k^{*1} \mathrel{\widehat{=}} \zeta_k$ is obviously symmetric. *(Inductive Step)*

Let $\zeta_k^{*(l-1)}$ be symmetric. Then, ζ_k^{*l} is also symmetric because

$$\zeta_{k,-n'}^{*l} = \sum_{m=0}^{N-1} \zeta_{k,m}\zeta_{k,-n'-m}^{*(l-1)} = \sum_{m=0}^{N-1} \zeta_{k,N-m}\zeta_{k,-n'-(N-m)}^{*(l-1)} = \sum_{m=0}^{N-1} \zeta_{k,m}\zeta_{k,n'-m}^{*(l-1)} = \zeta_{k,n'}^{*l} \tag{B.37}$$

□

More importantly, for all l, ζ_k^{*l} is monotonically decreasing in the interval $n \in [0, \frac{N}{2})$.

Proof. Mathematical induction.

(Basis)

We note that $\zeta_k^{*1} \mathrel{\widehat{=}} \zeta_k$ is monotonically decreasing in the interval $n \in [0, \frac{N}{2})$, that is

$$\zeta_{k,n}^{*(l-1)} - \zeta_{k,n+1}^{*(l-1)} \geq 0\,. \tag{B.38}$$

(Induction Step)

Let $\zeta_k^{*(l-1)}$ be monotonically decreasing in the interval $n \in [0, \frac{N}{2})$. Then, the difference of two subsequent elements of ζ_k^{*l} with $0 \leq n < \frac{N}{2}$ is

$$\begin{aligned}
\zeta_{k,n}^{*l} - \zeta_{k,n+1}^{*l} &= \sum_{m=0}^{N-1} \zeta_{k,m}(\zeta_{k,n-m}^{*(l-1)} - \zeta_{k,n+1-m}^{*(l-1)}) \\
&= \sum_{m=0}^{k} \zeta_{k,n-m}^{*(l-1)} - \zeta_{k,n+1-m}^{*(l-1)} + \sum_{m=N-k}^{N-1} \zeta_{k,n-m}^{*(l-1)} - \zeta_{k,n+1-m}^{*(l-1)} \\
&= \sum_{m=-k}^{k} \zeta_{k,n-m}^{*(l-1)} - \zeta_{k,n+1-m}^{*(l-1)} = \zeta_{k,n-k}^{*(l-1)} - \zeta_{k,n+k+1}^{*(l-1)}\,.
\end{aligned} \tag{B.39}$$

We prove that the r.h.s. of Eq. B.39 is larger than zero by checking all four different cases.

Case I: If $n < k$ and $n + k + 1 < \frac{N}{2}$

$$\zeta_{k,n-k}^{*(l-1)} - \zeta_{k,n+k+1}^{*(l-1)} = \zeta_{k,k-n}^{*(l-1)} - \zeta_{k,n+k+1}^{*(l-1)} \geq 0 \tag{B.40}$$

because $k-n, n+k+1 < \frac{N}{2}$ and $k-n < k-n+2n+1 = n+k+1$ and $\zeta_k^{*(l-1)}$ is monotonically decreasing in this interval.

Case (II: If $n < k$ and $n + k + 1 > \frac{N}{2}$

$$\zeta^{*(l-1)}_{k,n-k} - \zeta^{*(l-1)}_{k,n+k+1} = \zeta^{*(l-1)}_{k,k-n} - \zeta^{*(l-1)}_{k,N-n-k-1} \geq 0 \tag{B.41}$$

because $k-n, N-n-k-1 < \frac{N}{2}$ and $k-n < k-n+N-2k-1 = N-n-k-1$ and $\zeta^{*(l-1)}_k$ is monotonically decreasing in this interval.

Case III: If $n > k$ and $n + k + 1 < \frac{N}{2}$

$$\zeta^{*(l-1)}_{k,n-k} - \zeta^{*(l-1)}_{k,n+k+1} \geq 0 \tag{B.42}$$

because $n-k, n+k+1 < \frac{N}{2}$ and $n-k < n-k+2k+1 = n+k+1$ and $\zeta^{*(l-1)}_k$ is monotonically decreasing in this interval.

Case IV: If $n > k$ and $n + k + 1 > \frac{N}{2}$

$$\zeta^{*(l-1)}_{k,n-k} - \zeta^{*(l-1)}_{k,n+k+1}\zeta^{*(l-1)}_{k,n-k} - \zeta^{*(l-1)}_{k,N-n-k-1} \geq 0 \tag{B.43}$$

because $n-k, N-n-k-1 < \frac{N}{2}$ and $n-k \leq n-k+N-2n-1 = N-n-k-1$ and $\zeta^{*(l-1)}_k$ is monotonically decreasing in this interval.

Ergo, the r.h.s. of Eq. B.39 is always positive.

Thus, the sequences ζ^{*l}_k are symmetric and monotonically decreasing in the interval $n \in [0, \frac{N}{2})$. □

B.2.5. The Difference $\sigma_k - \sigma_{k+1}$

Eq 3.52 yields the difference $\sigma_k - \sigma_{k+1}$:

$$\sum_{l=1}^{k} (\sigma_{k-l} - \sigma_{k+1-l}) - (\delta + 2k)(\sigma_k - \sigma_{k+1}) + \sum_{l=1}^{k} (\sigma_{k+l} - \sigma_{k+1+l}) = 0 \tag{B.44}$$

$$\Rightarrow \quad \sigma_0 - \sigma_k - (\delta + 2k)(\sigma_k - \sigma_{k+1}) + \sigma_{k+1} - \sigma_{2k+1} = 0 \tag{B.45}$$

$$\Rightarrow \quad \sigma_k - \sigma_{k+1} = \frac{1}{\delta + 2k + 1}(\sigma_0 - \sigma_{2k+1}) \tag{B.46}$$

C. Appendix: Covariance Inversion

C.1. Rule of Thumb

For the homogeneous Ornstein-Uhlenbeck process, the deterministic part of the Langevin equation reads

$$\dot{x} = \underbrace{\mathbb{1} + \frac{2}{\delta} L}_{J'} \boldsymbol{x} \tag{C.1}$$

where δ is the dynamical constant, L is the Laplace matrix and J is the probation matrix. The stability condition for the Euler intergration scheme yields

$$|1 + \lambda \Delta t| \leq 1 \tag{C.2}$$

for all eigenvalues λ of J'.

Additionally, Gershgorin's theorem states that each eigenvalue λ of J' lies in the union of discs defined by

$$D_i := \left\{ \lambda' \ : \ |\lambda' - J_{ii}| \leq \sum_{j \neq i} |J'_{ij}| \right\} \ , \quad \lambda \in \bigcup_{i=1}^{N} D_i \ . \tag{C.3}$$

Hence,

$$\lambda \in \bigcup_{i=1}^{N} \left\{ \lambda' \ : \ |\lambda' + 1 + \frac{2k_i}{\delta}| \leq \frac{2k_i}{\delta} \right\} \ , \tag{C.4}$$

where k_i is the indegree of vertex i.

These discs include each other

$$D_i \subseteq D_j \quad \Leftrightarrow \quad k_i \leq k_j \ . \tag{C.5}$$

Thus, there is a disc

$$D(k') := \left\{ \lambda' \ : \ |\lambda' + 1 + \frac{2k'}{\delta}| \leq \frac{2k'}{\delta} \right\} \tag{C.6}$$

defined by k' which contains all eigenvalues.

Now, we expand the stability condition and to pin down a reasonable regime for Δt:

$$|1+\lambda\Delta t| = |1+\lambda\Delta t - \left(1+\frac{2k'}{\delta}\right)\Delta t + \left(1+\frac{2k'}{\delta}\right)\Delta t| \tag{C.7}$$

$$\leq |1-\left(1+\frac{2k'}{\delta}\right)\Delta t| + |\lambda+1+\frac{2k'}{\delta}|\Delta t \tag{C.8}$$

$$\leq |1-\left(1+\frac{2k'}{\delta}\right)\Delta t| + \frac{2k'}{\delta}\Delta t \overset{!}{\leq} 1 \tag{C.9}$$

$$\Rightarrow |1-\frac{2k'}{\delta}\Delta t - \Delta t| \leq 1-\frac{2k'}{\delta}\Delta t \tag{C.10}$$

Hence, if we chose Δt such that

$$\left(1+\frac{2k'}{\delta}\right)\Delta t \leq 1 \tag{C.11}$$

the stability condition $|1+\lambda\Delta t| \leq 1$ is fulfilled.

Now we assume k' is smaller or equal the average degree k. Thus, we get the rule of thumb

$$\boxed{\left(1+\frac{2k}{\delta}\right)\Delta t \leq 1\,.} \tag{C.12}$$

C.2. Mutual Information and Covariance Inversion

We showed in Sec. 2.4.4 that the mutual information of state variables of the Ornstein-Uhlenbeck process (OU process) only depends on the correlation, or covariance, between both variables (Eq. 2.58). Moreover, covariance inversion returns the adjacency matrix of the random process if the covariance matrices of the process are given (Eq. 3.60). Hence, it is possible to reconstruct the system if the delayed mutual information is given:

$$I_\tau(X_i, X_j) = -\frac{1}{2}\log\left(1-\frac{\Sigma^2_{\tau,ij}}{\Sigma_{ii}\Sigma_{jj}}\right) \tag{C.13}$$

$$\Rightarrow \Sigma_{\tau,ij} = \sqrt{\Sigma_{ii}\Sigma_{jj}}e^{-I_\tau(X_i,X_j)} := \Upsilon_{\tau,ij} \tag{C.14}$$

Also,

$$J = \frac{1}{\tau} \log\left(\Sigma_\tau \Sigma_0^{-1}\right) \tag{C.15}$$

$$\Rightarrow J = \frac{1}{\tau} \log\left(\Upsilon_\tau \Sigma_0^{-1}\right) \ . \tag{C.16}$$

Thus, by measuring the stationary covariance matrix and the delayed mutual information, a reconstruction of the Jacobin matrix is challenging but possible. It remains unclear, if this method is useful to estimate the Jacobian of nonlinear systems.

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.

[2] A.-L. Barabási and E. Bonabeau. Scale-free. *Scientific American*, 2003.

[3] R. H. Bartels and G. Stewart. Solution of the matrix equation ax+ xb= c [f4]. *Communications of the ACM*, 15(9):820–826, 1972.

[4] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

[5] A. J. Butte and I. S. Kohane. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. In *Pac Symp Biocomput*, volume 5, pages 418–429, 2000.

[6] C. F. Cadieu and K. Koepsell. Phase coupling estimation from multivariate phase statistics. *Neural Computation*, 22(12):3107–3126, 2010.

[7] J. Casadiego and Ti. Network dynamics as an inverse problem mathematical technology of networks. *PROMS*, 2015.

[8] E. S. Ching, P.-Y. Lai, and C. Leung. Extracting connectivity from dynamics of networks with uniform bidirectional coupling. *Physical Review E*, 88(4):042817, 2013.

[9] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[10] E. Davidson and M. Levin. Gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14):4935–4935, 2005.

[11] G. Deco, V. K. Jirsa, and A. R. McIntosh. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12(1):43–56, 2011.

[12] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. The backbone of the climate network. *EPL (Europhysics Letters)*, 87(4):48007+, Feb. 2010.

[13] M. J. Dunlop, R. S. Cox, J. H. Levine, R. M. Murray, and M. B. Elowitz. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature Genetics*, 40(12):1493–1498, 2008.

[14] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4):601–620, 2000.

[15] C. Gardiner. *Stochastic methods*. Springer-Verlag, Berlin–Heidelberg–New York–Tokyo, 1985.

[16] J. Gonçalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of lti networks. *Automatic Control, IEEE Transactions on*, 53(7):1670–1674, 2008.

[17] M. Grigoriu. *Stochastic calculus: applications in science and engineering*. Springer Science & Business Media, 2002.

[18] C. Gros. *Complex and adaptive dynamical systems: A primer*. Springer Science & Business Media, 2013.

[19] A. G. Haldane and R. M. May. Systemic risk in banking ecosystems. *Nature*, 469(7330):351–355, 2011.

[20] D. Hayden, Y. Yuan, and J. Goncalves. Network reconstruction from intrinsic noise. *arXiv preprint arXiv:1310.0375*, 2013.

[21] M. Hecker, S. Lambeck, S. Toepfer, E. Van Someren, and R. Guthke. Gene regulatory network inference: data integration in dynamic models—a review. *Biosystems*, 96(1):86–103, 2009.

[22] N. J. Higham. *Functions of matrices: theory and computation*. Siam, 2008.

[23] K. Karimi. A brief introduction to temporality and causality. *arXiv preprint arXiv:1007.2449*, 2010.

[24] A. Karr. *Probability*. Springer Texts in Statistics. Springer New York, 1993.

[25] M. Keeling. The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, 67(1):1–8, 2005.

[26] C. Kirst. *Dynamics of Pulsed-Coupled Neuronal Oscillators with Partial Reset*. PhD thesis, University of Göttingen, 2011.

[27] D. C. Knill and A. Pouget. The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719, 2004.

[28] L. Kuo-Ching and W. Xiaodong. Gene regulatory network reconstruction using conditional mutual information. *EURASIP Journal on Bioinformatics and Systems Biology*, 2008, 2008.

[29] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, and A. Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7(Suppl 1):S7, 2006.

[30] C. J. Needham, J. R. Bradford, A. J. Bulpitt, and D. R. Westhead. A primer on learning in bayesian networks for computational biology. *PLoS computational Biology*, 3(8):e129, 2007.

[31] E. Nicolato and E. Venardos. Option pricing in stochastic volatility models of the ornstein-uhlenbeck type. *Mathematical Finance*, 13(4):445–466, 2003.

[32] M. Paluš, D. Hartman, J. Hlinka, and M. Vejmelka. Discerning connectivity from dynamics in climate networks. *Nonlinear Processes in Geophysics*, 18(5):751–763, 2011.

[33] J. Pearl. *Causality: models, reasoning and inference*, volume 29. Cambridge Univ Press, 2000.

[34] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai. Noise bridges dynamical correlation and topology in coupled oscillator networks. *Physical Review Letters*, 104(5):058701, 2010.

[35] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical Review Letters*, 108(25):258701, 2012.

[36] P. Spirtes and C. Glymour. An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, 9(1):62–72, 1991.

[37] S. H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.

[38] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(7083):545–550, 2006.

[39] M. Thattai and A. Van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001.

[40] M. Timme and J. Casadiego. Revealing networks from dynamics: an introduction. *Journal of Physics A: Mathematical and Theoretical*, 47(34):343001, 2014.

[41] A. Tsonis and P. Roebber. The architecture of the climate network. *Physica A: Statistical and Theoretical Physics*, 333:497–504, Feb. 2004.

[42] A. A. Tsonis, K. L. Swanson, and P. J. Roebber. What do networks have to do with climate? *Bulletin of the American Meteorological Society*, 87(5):585–595, May 2006.

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

[44] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440–442, 1998.

[45] D. Witthaut and M. Timme. Braess's paradox in oscillator networks, desynchronization and power outage. *New Journal of Physics*, 14(8):083036, 2012.

[46] J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis. Advances to bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, 20(18):3594–3603, 2004.

Acknowledgments / Danksagung

An dieser Stelle würde ich gerne all jenen von tiefstem Herzen meinen Dank aussprechen, die mir mit wissenschaftlichem Rat und persönlicher Unterstützung während der Enstehung dieser Arbeit zur Seite gestanden haben.

Mein besonderer Dank gilt meinem Betreuer und Erstprüfer Prof. Dr. Marc Timme, dessen hilfreichen Rat und fachliche Expertise ich bei etlichen Gelegenheiten in Anspruch nehmen durfte. Sein Optimismus und sein Vertrauen in meine Fähigkeiten haben mich dazu ermutigt mein Bestes zu geben und diese Arbeit zu dem zu machen was sie ist.
Ich danke seiner Arbeitsgruppe für Netzwerkdynamik für die wunderbare Arbeitsatmosphäre und die Erfahrungen, welche ich während meiner Zeit am Insitut machen konnte. Ihr habt starke Eindrücke hinterlassen, die meine persönliche und berufliche Zukunft prägen werden.
Auch danke ich Prof. Dr. Ulrich Parlitz, welcher so freundlich war zuzustimmen der Bewertung dieser Arbeit als Zweitprüfer beizuwohnen.

Ferner bedanke ich mich bei José Luis Casadiego Bastidas für all die fruchtbaren Diskussionen, nützlichen Anmerkungen und Korrekturen. Ich danke Dr. Sarah Hallerberg für ihre hilfreichen Kommentare bezüglich Klassifikatorevaluation und ROC-Kurven sowie für die Durchsicht einiger Kapitel dieser Arbeit. Des Weiteren danke ich Dr. Martin Rohden, Matthias Wendland und Hauke Hähne. Auch sie nahmen sich Zeit, um sich einige Passagen meiner Arbeit genauer anzusehen. Vielen Dank für Eure Unterstützung!

Nicht zuletzt danke ich allen Bacheloranten und Masteranten, mit denen ich die Freude hatte ein Großraumbüro teilen zu dürfen. Unter anderem sind das Burooj Ghani, Hauke Hähne, Arke Vogelt, Gunter Weber, Francesca Schönberg, Benjamin Schäfer und Johannes Kassel.
Ich danke Euch für die freundschaftliche Arbeitsatmosphäre und all die anregenden Gespräche.

Weiteren Dank schulde ich den Technikern ohne die meine Arbeit nicht möglich gewesen wäre. Sven Jahnke, Yorck-Fabian Beensen und Denny Flieger waren immer erreichbar und schnell zur Stelle, um technische Hindernisse aus dem Weg zu räumen.

Mein Dank gilt meinem Freund und Kollegen Alexander Schmidt, dessen scharfes Auge für Details die Ursache für die letzten kleinen Änderungen an dieser Arbeit darstellt. Du hast mich durch mein Studium und die halbe Welt begleitet. Ich könnte mir keinen besseren Reisegefährten wünschen. Ich danke auch Arne Rohlfing, einem Freund, der es wie kein anderer versteht eine Atmosphäre der Konzentration zu schaffen. Danke Euch beiden!

Zu guter Letzt möchte ich - wie schon in anderen Arbeiten zuvor - meiner Mutter, meinem Vater und meinem Bruder danken, deren uneingeschränkte Liebe, Vertrauen und Unterstützung ich mir jeden Tag meines Studiums gewiss sein durfte.

Erklärung nach §17(9) der Prüfungsordnung für den Masterstudiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüber hinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nicht bestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 18. März 2015

(Benedict Johannes Lünsmann)