

# Appendix A

## PAPER A

---

**Title:** Online Non-linear Topology Identification from Graph-connected Time Series

**Authors:** **R. Money**, J. Krishnan, B. Beferull-Lozano

**Conference:** IEEE Data Science and Learning Workshop 2021

---

# Online Non-linear Topology Identification from Graph-connected Time Series

---

R. Money, J. Krishnan, B. Beferull-Lozano

---

**Abstract:** Estimating the unknown causal dependencies among graph-connected time series plays an important role in many applications, such as sensor network analysis, signal processing over cyber-physical systems, and finance engineering. Inference of such causal dependencies, often known as topology identification, is not well studied for non-linear non-stationary systems, and most of the existing methods are batch-based which are not capable of handling streaming sensor signals. In this paper, we propose an online kernel-based algorithm for topology estimation of non-linear vector autoregressive time series by solving a sparse online optimization framework using the composite objective mirror descent method. Experiments conducted on real and synthetic data sets show that the proposed algorithm outperforms the state-of-the-art methods for topology estimation.

## A.1 Introduction

Recent advancements in cyber-physical systems (CPS) and sensor networks call for advanced research on data analysis of structured or inter-linked spatio-temporal signals. Such structured signals can be meaningfully represented using graph-connected time series. Graph representation is a prevalent tool to model the inter-dependency of data [52], and it plays a vital role in countless practical applications such as time series prediction [53], change point detection [54], data compression [55], etc. Many of the functional dependencies in real-world time series are causal [56], and inferring the causal dependencies, which we term as *topology identification*, generates a more informative representation of the multivariate data. These dependencies may not be physically observable in some cases; instead, there can be logic connections between data nodes that are not physically connected due to control mechanisms, and inferring such typologies is a challenging task. Linear models, such as structural equation models (SEM), vector auto-regressive (VAR) models, and structural vector auto-regressive (SVAR) models [57] are widely used to study the causal dependencies among the graph-connected time series. SEM being a memory-less model, does not accommodate the temporal dependencies among the data, whereas the VAR is an

ideal choice for modeling the time-lagged interactions; however, it fails to capture the instantaneous causal relations. SVAR is a slightly modified model that unifies both SEM and VAR. The choice of the model depends on the physical nature of the system; for instance, SVAR is a useful model for brain connectivity analyses. However, VAR deserves special attention since the nodal dependencies on many practical sensor networks (e.g., water networks, oil and gas networks) involve mainly time-lagged interactions.

A significant challenge connected to topology identification is that the real-world systems are usually non-stationary, meaning that the statistical properties of dependencies vary over time. The commonly used batch-based off-line methods [11] have two major drawbacks: **i)** they are not effective in tracking the topology of non-stationary systems and **ii)** from a pure computational point of view, they suffer from processing large batch of data; hence, it is necessary to develop online estimation algorithms [48]. Online topology estimation algorithms have been developed for linear models, meaning that the causal dependencies among the data time-series hold a linear relation. For instance, in [48], a novel online linear topology identification algorithm have been proposed by minimizing a group-lasso-regularized [58] objective function.

Although the linear topology identification is a well-studied problem, many practical systems have non-linear dependencies [59]. As an example, in a smart water network, the causal dependencies are non-linear due to various control systems, saturation in valves or pumps, and non-linear physical equations governing the system. Similarly, essential non-linear dependencies are present in most of the real-world systems such as brain networks and finance networks. The ability of nonparametric techniques [60] and deep neural networks to learn non-linear functions is well studied, which has been exploited also in topology identification [9], [11], [14]. However, once again most of these algorithms are batch-based. Kernel-based representations are powerful tools to model the non-linear dependencies [15], which can be exploited to develop algorithms for online non-linear topology identification. For instance, in [61], authors have proposed an online algorithm based on functional gradient descent by considering a SVAR model. In [12], authors used a more general non-additive model for topology identification and a dictionary-based approach to solve the computational complexity imposed by the kernels. But [12] restricts the choice of the kernel functions to be twice differentiable to learn a sparse topology.

This paper proposes an online topology identification algorithm based on a non-linear VAR model using kernels. The proposed algorithm learns sparse and time-varying non-linear typology by solving an online optimization framework using composite objective iterations [39]. We provide strong empirical evidence using real and synthetic data sets, which show that the proposed algorithm outperforms its state-of-the-art counterpart.

## A.2 Problem Formulation

Consider a collection of  $N$  time series, connected by a directed graph and let  $y_n[t]$  be the value of time series at time  $t = 0, 1, \dots, T-1$  measured at node  $1 \leq n \leq N$ . A  $P$ -th order non-linear VAR model of the time series can be formulated as

$$y_n[t] = \sum_{n'=1}^N \sum_{p=1}^P a_{n,n'}^{(p)} f_{n,n'}^{(p)}(y_{n'}[t-p]) + u_n[t], \quad (\text{A.1})$$

where  $f_{n,n'}^{(p)}$  is a non-linear function that captures the causal influence of the  $p$ -lagged data at node  $n'$  on the node  $n$ ,  $a_{n,n'}^{(p)}$  is the corresponding entry of the graph adjacency matrix, and  $u_n[t]$  is the measurement noise. Referring to (A.1), topology identification can be defined as the estimation of the non-linear dependencies expressed by  $\left\{ a_{n,n'}^{(p)} f_{n,n'}^{(p)}(\cdot) \right\}_{p=1}^P$  for  $n = 1, 2, \dots, N$  from the observed time series  $\{y_n[t]\}_{n=1}^N$ .

To circumvent the challenges in topology identification, imposed by the non-linear dependencies, we assume that the functions  $f_{n,n'}^{(p)}(\cdot)$  in (A.1) belong to a reproducing kernel Hilbert space (RKHS):

$$\mathcal{H}_{n'}^{(p)} := \left\{ f_{n,n'}^{(p)} | f_{n,n'}^{(p)}(y) = \sum_{t=0}^{\infty} \beta_{n,n',t}^{(p)} \kappa_{n'}^{(p)}(y, y_{n'}[t-p]) \right\}, \quad (\text{A.2})$$

where  $\kappa_{n'}^{(p)} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is the Hilbert space basis function, often known as the kernel, which measures the similarities between the arguments of the basis function. Using (A.2), a function  $f_{n,n'}^{(p)}$  evaluated at  $y$  can be represented as the linear weighted sum of the similarities between  $y$  and the data samples  $\{y_{n'}[t-p]\}_{t=0}^{t=\infty}$ , where the weights are denoted by  $\beta_{n,n',t}^{(p)}$ . We assume that the Hilbert space is characterized by the inner product  $\langle \kappa_{n'}^{(p)}(y, x_1), \kappa_{n'}^{(p)}(y, x_2) \rangle := \sum_{t=0}^{\infty} \kappa_{n'}^{(p)}(y[t], x_1) \kappa_{n'}^{(p)}(y[t], x_2)$ , with the kernel having the reproducible property  $\langle \kappa_{n'}^{(p)}(y, x_1), \kappa_{n'}^{(p)}(y, x_2) \rangle = \kappa_{n'}^{(p)}(x_1, x_2)$ . Such a Hilbert space with the reproducing kernel constitutes an RKHS with norm  $\|f_{n,n'}^{(p)}\|_{\mathcal{H}_{n'}^{(p)}}^2 = \sum_{t=0}^{\infty} \sum_{t'=0}^{\infty} \beta_{n,n',t}^{(p)} \beta_{n,n',t'}^{(p)} \kappa_{n'}^{(p)}(y_n[t], y_n[t'])$ . We refer to [37] for further reading on RKHS.

The least-squares (LS) estimate  $\{f_{n,n'}^{(p)} \in \mathcal{H}_{n'}^{(p)}; n' = 1, \dots, N, p = 1, \dots, P\}$  for a particular node is obtained by solving the following non-parametric optimization problem:

$$\left\{ \hat{f}_{n,n'}^{(p)} \right\}_{n',p} = \arg \min_{\{f_{n,n'}^{(p)} \in \mathcal{H}_{n'}^{(p)}\}} \frac{1}{2} \sum_{\tau=P}^{T-1} \left[ y_n[\tau] - \sum_{n'=1}^N \sum_{p=1}^P a_{n,n'}^{(p)} f_{n,n'}^{(p)}(y_{n'}[\tau-p]) \right]^2. \quad (\text{A.3})$$

It is to be noted that, in (A.3), the functions  $\{f_{n,n'}^{(p)}\}$  belongs to the RKHS, defined in (A.2), which is an infinite dimensional space. However, by resorting to the Representer Theorem [38], the solution of (A.3) can be written using a finite number of data samples:

$$\hat{f}_{n,n'}^{(p)}(y_{n'}[\tau-p]) = \sum_{t=p}^{p+T-1} \beta_{n,n',(t-p)}^{(p)} \kappa_{n'}^{(p)}(y_{n'}[\tau-p], y_{n'}[t-p]). \quad (\text{A.4})$$

Using (A.4), (A.3) can be reformulated as a parametric optimization problem involving the available data samples, as follows:

$$\left\{ \widehat{\alpha}_{n,n',t}^{(p)} \right\}_{n',p,t} = \arg \min_{\left\{ \alpha_{n,n',t}^{(p)} \right\}} \mathcal{L}^n \left( \alpha_{n,n',t}^{(p)} \right), \quad (\text{A.5})$$

where

$$\mathcal{L}^n \left( \alpha_{n,n',t}^{(p)} \right) := \frac{1}{2} \sum_{\tau=P}^{T-1} \left[ y_n[\tau] - \sum_{n'=1}^N \sum_{p=1}^P \sum_{t=p}^{p+T-1} \alpha_{n,n',t}^{(p)} \kappa_{n'}^{(p)}(\tau, t) \right]^2, \quad (\text{A.6})$$

$$\alpha_{n,n',t}^{(p)} := a_{n,n'}^{(p)} \beta_{n,n',(t-p)}^{(p)}, \quad (\text{A.7})$$

and

$$\kappa_{n'}^{(p)}(\tau, t) := \kappa_{n'}^{(p)}(y_{n'}[\tau - p], y_{n'}[t - p]). \quad (\text{A.8})$$

We stack the entries of  $\left\{ \alpha_{n,n',t}^{(p)} \right\}$  and  $\left\{ \kappa_{n'}^{(p)}(\tau, t) \right\}$  in the lexicographic order of the indices  $p$ ,  $n'$ , and  $t$  to obtain the vectors  $\boldsymbol{\alpha}_n \in \mathbb{R}^{PNT}$  and  $\boldsymbol{\kappa}_\tau \in \mathbb{R}^{PNT}$ , respectively, and rewrite (A.5) as

$$\widehat{\boldsymbol{\alpha}}_n = \arg \min_{\boldsymbol{\alpha}_n} \mathcal{L}^n(\boldsymbol{\alpha}_n), \quad (\text{A.9})$$

where

$$\mathcal{L}^n(\boldsymbol{\alpha}_n) = \frac{1}{2} \sum_{\tau=P}^{T-1} \left[ y_n[\tau] - \boldsymbol{\alpha}_n^\top \boldsymbol{\kappa}_\tau \right]^2 \quad (\text{A.10})$$

Further, to avoid overfitting and to enforce group sparsity, we propose a regularized optimization framework:

$$\widehat{\boldsymbol{\alpha}}_n = \arg \min_{\boldsymbol{\alpha}_n} \mathcal{L}^n(\boldsymbol{\alpha}_n) + \lambda \sum_{n'=1}^N \sum_{p=1}^P \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2, \quad (\text{A.11})$$

where  $\lambda \geq 0$  is the regularization parameter and  $\boldsymbol{\alpha}_{n,n'}^{(p)} = (\alpha_{n,n',0}^{(p)}, \alpha_{n,n',1}^{(p)}, \dots, \alpha_{n,n',T}^{(p)}) \in \mathbb{R}^T$ . The second term in (A.11) is a *group-lasso* regularizer, which promote a *group-sparse structure* in  $\boldsymbol{\alpha}_{n,n'}^{(p)}$ , thereby exploiting the prior information that the number of causal dependencies are typically small for real-world graph-connected time series.

The parametric optimization given by (A.11) is a batch (offline) solver meaning that to solve (A.11), we require all data samples  $\{y_n[\tau]\}_{\tau=P}^{T-1}$  to be available. Such an offline approach has two major drawbacks: **i)** it is not suitable for real-time applications since the solver has to wait for the entire batch of data and **ii)** it suffers from high computation complexity and memory requirements which grows super linearly with the batch size. In the following section, we propose an online algorithm to estimate the coefficients  $\boldsymbol{\alpha}_n$  in (A.11).

### A.3 Online topology estimation

First replace the original loss function  $\mathcal{L}^n(\boldsymbol{\alpha}_n)$  in (A.11) with the instantaneous loss function  $l_\tau^n(\boldsymbol{\alpha}_n) = \frac{1}{2} [y_n[\tau] - \boldsymbol{\alpha}_n^\top \boldsymbol{\kappa}_\tau]^2$ :

$$\widehat{\boldsymbol{\alpha}}_n = \arg \min_{\boldsymbol{\alpha}_n} l_\tau^n(\boldsymbol{\alpha}_n) + \lambda \sum_{n'=1}^N \sum_{p=1}^P \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2. \quad (\text{A.12})$$

A straightforward way to solve (A.12) is by applying the online subgradient descent (OSGD). However, it is to be remarked that the regularizer in (A.12) is non-

differentiable and OSGD fails to provide sparse  $\boldsymbol{\alpha}_{n,n'}^{(p)}$  since it linearizes the entire instantaneous objective function in (A.12) [48]. To mitigate this issue, we use the composite objective mirror descent (COMID) [39] algorithm. The online COMID update can be written as

$$\boldsymbol{\alpha}_n[t+1] = \arg \min_{\boldsymbol{\alpha}_n} J_t^{(n)}(\boldsymbol{\alpha}_n), \quad (\text{A.13})$$

where

$$\begin{aligned} J_t^{(n)}(\boldsymbol{\alpha}_n) &\triangleq \nabla \ell_t^n(\tilde{\boldsymbol{\alpha}}_n[t])^\top (\boldsymbol{\alpha}_n - \tilde{\boldsymbol{\alpha}}_n[t]) \\ &+ \frac{1}{2\gamma_t} \|\boldsymbol{\alpha}_n - \tilde{\boldsymbol{\alpha}}_n[t]\|_2^2 + \lambda \sum_{n'=1}^N \sum_{p=1}^P \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2. \end{aligned} \quad (\text{A.14})$$

In (A.14),  $\tilde{\boldsymbol{\alpha}}_n[t] \in \mathbb{R}^{PN(t+1)}$  is defined as  $[\boldsymbol{\alpha}_n[t]; \mathbf{0}]$ , where  $\boldsymbol{\alpha}_n[t] \in \mathbb{R}^{PNt}$  is the value of  $\boldsymbol{\alpha}_n$  estimated by processing the samples up to time  $t$ . The zero vector  $\mathbf{0} \in \mathbb{R}^{PN}$  is appended as an initialization for the coefficients of the new elements of the kernel vector corresponding to the  $(t+1)^{\text{th}}$  data sample. In (A.14), the first term is the gradient of the loss function and the second term is the Bregman divergence  $B(\boldsymbol{\alpha}_n, \tilde{\boldsymbol{\alpha}}_n[t]) = \frac{1}{2} \|\boldsymbol{\alpha}_n - \tilde{\boldsymbol{\alpha}}_n[t]\|_2^2$ , chosen in such a way that the COMID update has a closed form solution [40] and  $\gamma_t$  is the corresponding step size. Bregman divergence ensures that  $\boldsymbol{\alpha}_n[t+1]$  is close to  $\tilde{\boldsymbol{\alpha}}_n[t]$ , in line with the assumption that the topology changes smoothly. The third term is a sparsity enforcing regularizer, in order to promote sparsity in the updates. The gradient in (A.14) is evaluated as

$$\mathbf{v}_n[t] := \nabla \ell_t^n(\tilde{\boldsymbol{\alpha}}_n[t]) = \boldsymbol{\kappa}_\tau (\boldsymbol{\alpha}_n^\top \boldsymbol{\kappa}_\tau - y_n[\tau]) \quad (\text{A.15})$$

Expanding the objective function in (A.14) by omitting the constants leads to the following formulation:

$$\begin{aligned} J_t^{(n)}(\boldsymbol{\alpha}_n) &\propto \frac{\boldsymbol{\alpha}_n^\top \boldsymbol{\alpha}_n}{2\gamma_t} + \boldsymbol{\alpha}_n^\top \left( \mathbf{v}_n[t] - \frac{1}{\gamma_t} \tilde{\boldsymbol{\alpha}}_n[t] \right) + \lambda \sum_{n'=1}^N \sum_{p=1}^P \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2 \\ &= \sum_{n'=1}^N \sum_{p=1}^P \left[ \frac{\boldsymbol{\alpha}_{n,n'}^{(p)\top} \boldsymbol{\alpha}_{n,n'}^{(p)}}{2\gamma_t} + \boldsymbol{\alpha}_{n,n'}^{(p)\top} \left( \mathbf{v}_{n,n'}^{(p)}[t] - \frac{1}{\gamma_t} \tilde{\boldsymbol{\alpha}}_{n,n'}^{(p)}[t] \right) \right. \\ &\quad \left. + \lambda \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2 \right]. \end{aligned} \quad (\text{A.16})$$

Note that (A.16) is separable in  $n'$ ,  $m$  and  $p$ . Using (A.16), a closed form solution of (A.13) can be obtained in terms of multidimensional shrinkage-thresholding operator [41] as

$$\boldsymbol{\alpha}_{n,n'}^{(p)}[t+1] = \left( \tilde{\boldsymbol{\alpha}}_{n,n'}^{(p)}[t] - \gamma_t \mathbf{v}_{n,n'}^{(p)}[t] \right) \times \left[ 1 - \frac{\gamma_t \lambda \mathbb{1}\{n \neq n'\}}{\|\tilde{\boldsymbol{\alpha}}_{n,n'}^{(p)}[t] - \gamma_t \mathbf{v}_{n,n'}^{(p)}[t]\|_2} \right]_+, \quad (\text{A.17})$$

where  $[x]_+ = \max\{0, x\}$  and

$$\mathbb{1}\{n \neq n'\} = \begin{cases} 1, & \text{if } n \neq n' \\ 0, & n = n'. \end{cases}$$

The term  $\tilde{\boldsymbol{\alpha}}_{n,n'}^{(p)}[t] - \gamma_t \mathbf{v}_{n,n'}^{(p)}[t]$  in (A.17) performs a stochastic gradient update of  $\boldsymbol{\alpha}_{n,n'}^{(p)}$  in a direction that decreases the instantaneous loss function  $\ell_t^n(\boldsymbol{\alpha}_n)$  and the second term in (A.17) promotes group sparsity of  $\boldsymbol{\alpha}_{n,n'}^{(p)}$ . The function  $\mathbb{1}\{n \neq n'\}$  in the second term prevents the enforcement of sparsity of self-connections of the

graph. One major issue with (A.17) is that the size of  $\mathbf{v}_{n,n'}^{(p)}[t]$  becomes prohibitive as  $t$  increases. To mitigate this issue we select the recent  $T_w$  data points to calculate (A.17). For the experiments presented in this paper, we heuristically fix the value of  $T_w$  to 2000. Although this sub-optimal approach affects the performance of the algorithm, we are getting quite competitive empirical performance as shown later in the experiment section.

The proposed algorithm, termed as *Nonlinear Topology Identification via Sparse Online learning* (NL-TISO), is summarized in **Algorithm 5**.

---

**Algorithm 5:** NL-TISO Algorithm

---

**Result:**  $\alpha_{n,n'}^{(p)}$ , for  $n, n' = 1, \dots, N$  and  $p = 1, \dots, P$   
**Store**  $\{\mathbf{y}_n[t]\}_{t=1}^P$ ,  
**Initialize**  $\lambda, \gamma$  (heuristically chosen) and kernel parameters depending on the type of the kernel.  
**for**  $t = P, P + 1, \dots$  **do**  
    Get data samples  $y_n[t], \forall n$  and compute  $\kappa_\tau$   
    **for**  $n = 1, \dots, N$  **do**  
        compute  $\mathbf{v}_n[t]$  using (A.15)  
        **for**  $n' = 1, \dots, N$  **do**  
            compute  $\alpha_{n,n'}^{(p)}[t + 1]$  using (A.17)  
        **end**  
    **end**  
**end**

---

## A.4 Experiments

In this section, we illustrate the effectiveness of the proposed NL-TISO algorithm using synthetic and real data. We compare our results with two state-of-the-art topology estimation algorithms: **i)** TIRSO [48]- a recent online topology estimation algorithm based on COMID update developed for linear causal dependencies and **ii)** functional gradient descent (FGD) algorithm [61]- an online kernel based topology estimation algorithm based on functional gradient descent updates

### A.4.1 Experiments using synthetic data

#### A.4.1.1 Identifying causal dependencies

We generated graph connected time series, based on the non-linear VAR model (C.1) with parameter values  $N = 5, T = 3000$ , and  $P = 2$ . The entries of the graph adjacency matrix  $\{a_{n,n'}^{(p)}\}$  are drawn from a Gaussian distribution  $\mathcal{N}(8, 3)$  with an edge probability  $p_e = 0.1$ . The initial  $P$  samples of the time series are drawn randomly from a Gaussian distribution  $\mathcal{N}(0, 0.1)$  and the remaining samples are generated using model (C.1). A Gaussian kernel centered at the dependent data points and

having variance 0.03 is used to model the non-linear dependencies in (C.1), where the kernel coefficients  $\beta_{n,n'}^{(p)}$  are drawn from a zero mean Gaussian distribution with variance 0.03. The noise  $u_n[t]$  is generated from a zero mean Gaussian distribution with variance 0.01. The causal dependencies  $\{\alpha_{n,n'}^{(p)}[t]\}$  are estimated using the

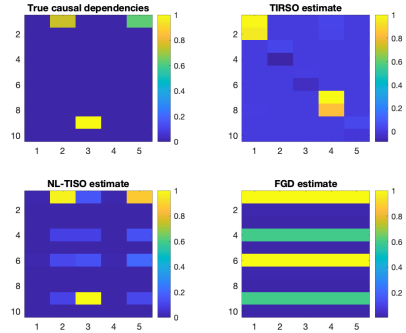


Figure A.1: Causal dependencies (normalized) estimated using different algorithms compared with the true dependency.

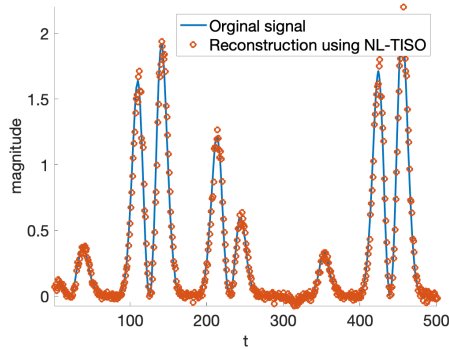


Figure A.2: Reconstruction of true signal in node 1 using estimated coefficients.

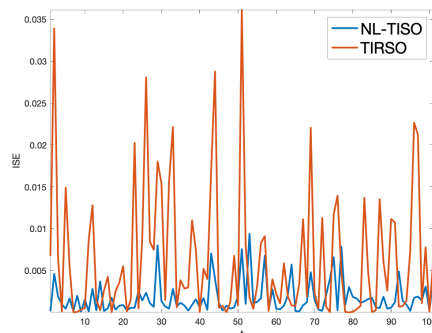


Figure A.3: ISE comparison of NL-TISO and TIRSO when the signal to be reconstructed is rapidly varying.

proposed NL-TISO algorithm using Gaussian kernel having variance 0.1 and with hyper-parameters  $\lambda = 0.1$  and  $\gamma = 10$ . Since a stationary topology is considered in



having variance 0.03 is used to model the non-linear dependencies in (C.1), where the kernel coefficients  $\beta_{n,n'}^{(p)}$  are drawn from a zero mean Gaussian distribution with variance 0.03. The noise  $u_n[t]$  is generated from a zero mean Gaussian distribution with variance 0.01. The causal dependencies  $\{\alpha_{n,n'}^{(p)}[t]\}$  are estimated using the

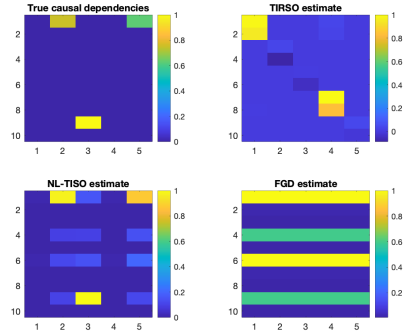


Figure A.1: Causal dependencies (normalized) estimated using different algorithms compared with the true dependency.

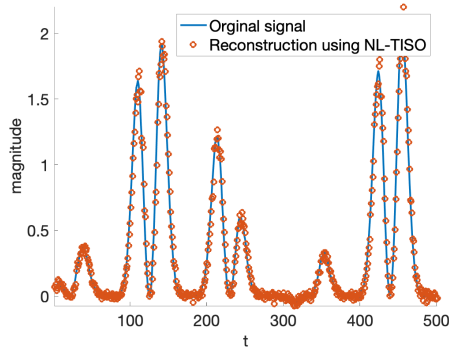


Figure A.2: Reconstruction of true signal in node 1 using estimated coefficients.

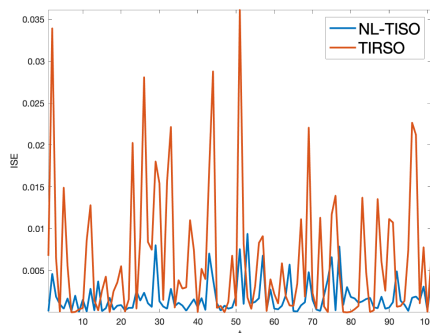


Figure A.3: ISE comparison of NL-TISO and TIRSO when the signal to be reconstructed is rapidly varying.

proposed NL-TISO algorithm using Gaussian kernel having variance 0.1 and with hyper-parameters  $\lambda = 0.1$  and  $\gamma = 10$ . Since a stationary topology is considered in

this experiment, we compute the  $\ell_2$  norms  $\widehat{b}_{n,n'}^{(p)} = \|\boldsymbol{\alpha}_{n,n'}^{(p)}[t]\|_2$  at  $t = T$  and arrange them in a matrix structure similar to the graph adjacency matrix to visualize the causal dependencies. A similar strategy is adopted for the FGD and the TIRSO algorithms, and the estimated adjacency matrix is used to visualize the dependencies. The true and the estimated dependencies are shown in Fig. B.1, in which for each subplot, the  $5 \times 5$  dependency matrices corresponding to  $p = 1$  and 2 are concatenated, resulting in a size  $10 \times 5$  size matrix. From Fig. B.1, it is clear that the NL-TISO algorithm outperforms others in identifying the causal relationship.

#### A.4.1.2 Signal Reconstruction Experiment

In this experiment, using the inferred causal dependencies, we reconstruct the time series and compare it with the true signals. In contrary to the previous experiment, a dynamic graph-topology is considered here using a time varying adjacency matrix

$$a_{n,n'}^{(p)}[t+1] = a_{n,n'}^{(p)}[t] + 0.01 \sin(0.03 * t) \quad (\text{A.18})$$

with a random initialization. We use a different non-linear dependency compared to the previous experiment to generate data:

$$f_{n,n'}^{(p)}(x) = 0.4 \sin(\pi x^2) + 0.3 \sin(2\pi x) + 0.3 \sin(3\pi x). \quad (\text{A.19})$$

Graph-connected time series ( $N = 5$ ) are generated using (C.1), (A.18), and (A.19) in a similar manner as described in A.4.1.1.

The causal dependencies  $\{\boldsymbol{\alpha}_{n,n'}^{(p)}\}$  are estimated from the time series using NL-TISO with a Gaussian kernel having variance 0.02 and with hyper-parameters  $\lambda = 10^{-6}$  and  $\gamma = 10$ . Using the same Gaussian kernel and the estimated dependencies, the time series are reconstructed. In Fig. A.2, a visual comparison of both the true and reconstructed time series at one of the five nodes is shown. We observed that the reconstructed signal is very close to the true one, although a Gaussian-based kernel is used to infer the non-linearity imposed by (A.19), which in turn indicates that kernel-based representations are a powerful tool in handling the non-linear causal dependencies. Further, the signal reconstruction quality of the state-of-the-art algorithms TIRSO [48] and FGD [61] are compared using *instantaneous squared error*, which is defined as  $ISE(t) = (y_n(t) - \hat{y}_n(t))^2$  and is plotted in Fig. A.3, which concludes that NL-TISO outperforms TIRSO by a considerable margin for the non-linear signal models. We have also observed that the ISE of the FGD algorithm is much worse than NL-TISO and TIRSO and is not shown in the figure.

### A.4.2 Experiments using Real Data

In this section, we present experiments using real data collected from Lundin’s offshore oil and gas (O&G) platform Edvard-Grieg<sup>1</sup>. We consider a directed graph with 24 nodes; each node corresponds to temperature (T), pressure (P), or oil-level (L) sensors. These sensors are placed in the separators of decantation tank that separates oil, gas, and water. The time series are obtained by uniformly sampling the sensor readings and applying normalization to have zero mean and unit sample

---

<sup>1</sup><https://www.lundin-energy.com/>

variance. These time series are expected to exhibit causal dependencies due to the underlying physical coupling arising from the pipeline connections and the control systems.

The causal dependencies are learned using NL-TISO with a Gaussian kernel having a variance of 0.1 and with hyper parameter values  $\lambda = 0.1$  and  $\gamma = 10$ . In Fig. A.4, we show one portion of the reconstructed signal corresponding to sensor-1, which is a pressure sensor, and it can be observed that the reconstructed signal is very close to the true sensor reading. Further, in Fig.B.6, we compare the reconstruction error of NL-TISO with TIRSO in terms of ISE for sensor-1 signal samples. We observe that NL-TISO outperforms TIRSO by a considerable margin, which supports the effectiveness of proposed algorithm in learning real world topology. The causal dependencies among the 24 time series obtained by averaging the NL-TISO estimates for one hour is shown in Fig. B.7.

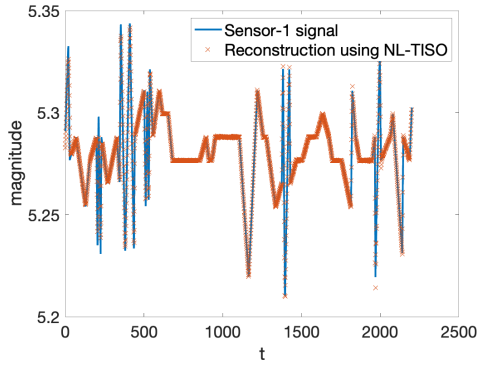


Figure A.4: Reconstruction of sensor-1 signal with sampling time 5s from Lundin data using estimated causal dependencies.

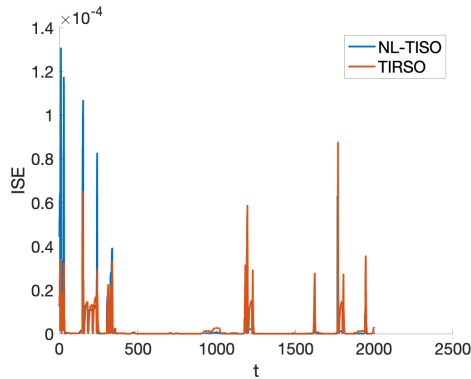


Figure A.5: ISE comparison of NL-TISO and TIRSO using real data.

## A.5 Conclusion

An online algorithm for non-linear topology identification from graph-connected time-series was proposed in this paper. Most of the state-of-the-art algorithms

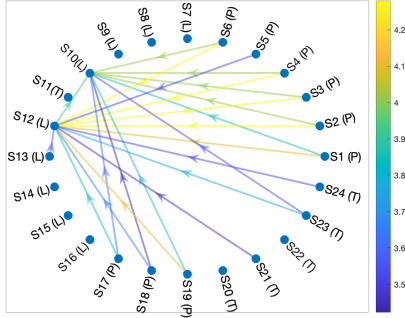


Figure A.6: Causality graph in oil and gas plant estimated by NL-TISO.

solve the topology estimation problem by assuming a linear and stationary topology. However, many real-world networks are highly dynamic and non-linear. The proposed algorithm, NL-TISO, is devised based on kernel representation to handle the non-linearities of the real-world sensor networks. Further, using a composite objective mirror descent method, NL-TISO estimates sparse topology in an online fashion aiming at dynamic system models. Qualitative and quantitative empirical evidence provided in the paper using real and synthetic data show that NL-TISO is an effective algorithm to infer the causal dependencies of real-world sensor networks. We identify two major limitations of the proposed framework: **i)** the computational complexity and memory requirements of kernel-based representations increases considerably with number of data points which is handled in NL-TISO by considering a time window to select recent samples and **ii)** the variance of the Gaussian kernels used in NL-TISO are heuristically chosen. These limitations could be handle by further research on dictionary-based multi-kernel representations, which will be devoted to our future work.