Systems & Control Letters 62 (2013) 664–672

Contents lists available at SciVerse ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Model identification of a network as compressing sensing

D. Materassi [a], G. Innocenti [b], L. Giarré [c,*], M. Salapaka [d]

[a] Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, Cambridge (MA), United States

[b] Dipartimento di Ingegneria dell'Informazione, Universitá di Firenze, via S. Marta 3, 50139 Firenze, Italy

[c] DEIM, Universitá di Palermo, viale Delle Scienze, 90128 Palermo, Italy

[d] Department of Electrical and Computer Engineering, University of Minnesota, 200 Union St SE, 55455, Minneapolis (MN), United States

ARTICLE INFO

Article history:
Received 2 March 2011
Received in revised form
6 April 2013
Accepted 16 April 2013
Available online 8 June 2013

Keywords:
Identification
Sparsification
Reduced models
Networks
Compressive sensing

ABSTRACT

In many applications, it is of interest to derive information about the topology and the internal connections of multiple dynamical systems interacting together. Examples can be found in fields as diverse as Economics, Neuroscience and Biochemistry. The paper deals with the problem of deriving a descriptive model of a network with no a-priori knowledge on its topology. It is assumed that the network nodes are passively observed and data are collected in the form of time series. The underlying structure is then determined by the non-zero entries of a "sparse Wiener filter". We cast the problem as the optimization of a quadratic cost function, where a set of parameters are used to operate a trade-off between accuracy and complexity in the final model.

1. Introduction

The interest on networks of dynamical systems is increasing in recent years, especially because of their capability of modeling and describing a large variety of phenomena and behaviors. Remarkably, while networks of dynamical systems are well studied and analyzed in physics [1–3] and engineering [4–6], there are fewer results that address the problem of reconstructing an unknown dynamical network, since it poses formidable theoretical and practical challenges [7]. However, unraveling the interconnectedness and the interdependency of a set of processes is of significant interest in many fields and the necessity for general tools is rapidly emerging (see [8–10] and the bibliography therein for recent results). In the literature, authors have approached this problem in different ways and with various purposes, such as deriving a network topology from just sampled data (see e.g. [11,8,10,12–14]) or determining the presence of substructures in the networked system (see e.g. [3,9]). The Unweighted Pair Group Method with Arithmetic mean (UPGMA) [15] is one of the first techniques proposed to reveal an unknown topology. It has found widespread use in the reconstruction of phylogenetic trees and is widely employed in other areas such as communication systems and resource allocation problems [16]. Another well-known technique for the identification of a tree network is developed in [11] for the analysis of a stock portfolio. The authors identify a tree structure according to the following procedure: (i) a metric based on the correlation index is defined among the nodes; (ii) such a metric is employed to extract the Minimum Spanning Tree [17] which forms the reconstructed topology. However, in [18] a severe limit of this strategy is highlighted, where it is shown that, even though the actual network is a tree, the presence of dynamical connections or delays can lead to the identification of a wrong topology. In [19] a similar strategy, where the correlation metric is replaced by a metric based on the coherence function, is numerically shown to provide an exact reconstruction for tree topologies. Furthermore, in [20] it is analytically proved that the correct reconstruction can be guaranteed for any topology with no cycles.

An approach for the identification of more general topologies is developed in the area of Machine Learning for Bayesian dynamical networks [21,22]. In this case, however, a massive quantity of data needs to be collected in order to accurately evaluate conditional probability distributions.

In [9] different techniques to quantify and evaluate the modular structure of a network are compared and a new one is proposed, trying to combine both the topological and dynamic information of the complex system. However, the network topology is only qualitatively estimated in terms of "clusters" [23].

In [8] a method to identify a network of dynamical systems is described. However, primary assumptions of the technique are the

* Corresponding author. Tel.: +39 3293173773.
E-mail addresses: donnie13@mit.edu (D. Materassi), giacomo.innocenti@unifi.it (G. Innocenti), giarre@unipa.it, lagiarre@gmail.com (L. Giarré), murti@umn.edu (M. Salapaka).

0167-6911/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sysconle.2013.04.004

possibility to manipulate the input of every single node and to conduct as many experiments as needed to detect the link connectivity.

More recently, in [10,24] interesting equivalences between the identification of a dynamical network and a l_0 sparsification problem are highlighted, suggesting the difficulty of the reconstruction procedure [25,26].

In this paper, the main idea is to cast the problem of unveiling an unknown structure as the estimate of a "sparse Wiener filter". Given a set of n stochastic processes $\mathcal{X} = \{x_1, \ldots, x_n\}$, we consider each x_j as the output of an unknown dynamical system, the input of which is given by at most m_j stochastic processes $\{x_{\alpha_{j,1}}, \ldots, x_{\alpha_{j,m_j}}\}$ selected from $\mathcal{X} \setminus \{x_j\}$. The choice of $\{x_{\alpha_{j,1}}, \ldots, x_{\alpha_{j,m_j}}\}$ is realized according to a criterion that takes into account the mean square of the modeling error. The parameters m_j can be a-priori defined, if we intend to impose a certain degree of sparsity on the network, or they can be dynamically defined by introducing a self-tuning strategy that penalizes the introduction of any additional link, when it does not provide a significant reduction of the cost.

For any possible choice of $\{x_{\alpha_{j,1}}, \ldots, x_{\alpha_{j,m_j}}\}$, the computation of the related Wiener Filter leads to the definition of a modeling error, which is a natural way to measure the quality of the description of x_j granted by the time series $\{x_{\alpha_{j,1}}, \ldots, x_{\alpha_{j,m_j}}\}$ in terms of predictive/smoothing capability. Once this step has been performed, each system is represented by the node of a graph and, then, the arcs linking any $x_{\alpha_{j,m_k}}$ to x_j are introduced for each node x_j. At the end of this procedure a graph modeling the network topology has been obtained.

We start introducing a pre-Hilbert space for wide-sense stochastic processes, where the inner product defines the notion of perpendicularity between two stochastic processes. We will show that this formulation of the problem has strong similarities with the l_0-minimization problem, which has been a very active topic of research in Signal Processing during the last few years. Indeed, a standard l_0-minimization problem amounts to finding the "sparsest" solution of a set of linear equations in a finite dimension Hilbert space [26]. With no additional assumptions on the solution, the problem is combinatorially intractable [26]. This has propelled the study of relaxed problems involving, for example, the minimization of the ℓ_1 norm, which is a convex problem and is known to provide solutions with at least a certain order of sparsity [10].

The rest of the paper is organized as follows. In Section 2 the network topology identification problem is formulated. In Section 3 a geometric interpretation and the construction of a pre-Hilbert space needed to define a distance and an inner product for stochastic processes is addressed. In Section 4 the connection with the compressive sensing problem is shown. In Section 5 a greedy algorithm addressing the problem is presented along with an alternative approach based on iterated reweighted least squares. Finally, in Section 6 the results obtained by applying the techniques to numerical data are discussed. In the Appendix most of the definitions, propositions, lemmas and proofs necessary for the construction of the pre-Hilbert space are reported.

Notation:

$\mathbb{N}$: the natural set;
$\mathbb{Z}$: the integer set;
$\mathbb{R}$: the real set;
$\mathbb{C}$: the complex set;
$E[\cdot]$: the mean operator;
$(\cdot)^T$: the transpose operator.

2. Problem formulation

In this section we provide the main definitions to cast the problem of modeling an unknown network structure. We consider n stochastic processes $x_1, \ldots, x_n$ representing the output of n interconnected dynamical systems, and we intend to derive a suitable network structure, describing their links and that can be represented as a directed graph. To this aim, we interpret the processes $x_1, \ldots, x_n$ as nodes of the graph. Furthermore, if the process x_i is "related" to the process x_j we represent this as an edge from x_i to x_j. In order to determine the links connecting the nodes, we follow a procedure based on estimation techniques. Given a process x_j and a parameter $m_j \in \mathbb{N}$, we search for the m_j processes $x_{\alpha_1}, \ldots, x_{\alpha_k}, \ldots, x_{\alpha_{m_j}}$ with $\alpha_k \neq j$, $\forall k = 1, \ldots, m_j$, which provide the best estimate of x_j according to a quadratic criterion. The value m_j is a tuning parameter allowing one to operate a trade-off between the sparsity and the accuracy of the model. Thus, m_j can be a-priori chosen or, conversely, determined using a self-tuning strategy. As a result of this modeling procedure, we obtain a directed graph where, for each node, the entering edges provide information about the nodes that are more useful in estimating its output (see Fig. 1).

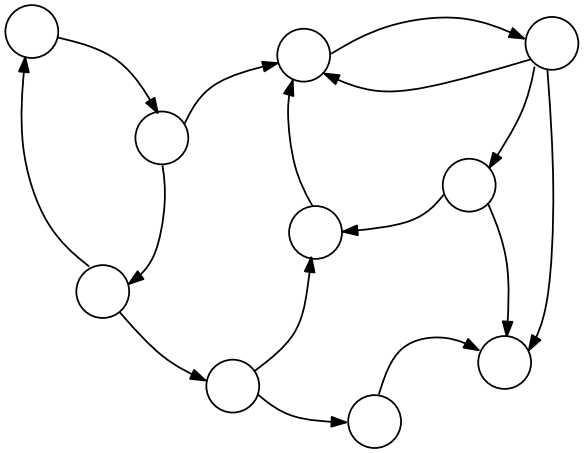

Fig. 1. Example of a possible outcome of the estimation procedure in a 10-node network.

We now introduce some definitions and results, which turn out essential for the rigorous formulation of the problem. For the sake of clarity, we report in the Appendix all the additional definitions and properties.

Definition 1. Let $e_i(t)$, with $i = 1, \ldots, n$ and $t \in \mathbb{Z}$, be n scalar time-discrete, zero-mean, jointly wide-sense stationary random processes in a probability space (Ω, σ, Π), where Ω is the sample space, σ is a sigma algebra on Ω and Π is a probability measure on σ. Then, for any $t \in \mathbb{Z}$ define the vector $e(t) := (e_1(t), .., e_n(t))^T$, describing a n-dimensional time-discrete, zero-mean, wide-sense stationary random process. Moreover, for any $t_1, t_2 \in \mathbb{Z}$ denote the $(n \times n)$ covariance matrix as

$$R_e(t_1, t_2) := E[e(t_1)e^T(t_2)]. \tag{1}$$

The entry (i, j) of $R_e(t_1, t_2)$, with $i, j \in \{1, \ldots, n\}$, is given by

$$R_{e_ie_j}(t_1, t_2) := E[e_i(t_1)e_j(t_2)].$$

Since any two processes e_i and e_j are jointly wide-sense stationary by definition, $R_{e_ie_j}(t_1, t_2)$ only depends on $\tau := t_2 - t_1$:

$$R_{e_ie_j}(0, t_2 - t_1) = R_{e_ie_j}(t_1, t_2), \quad \forall t_1, t_2 \in \mathbb{Z},$$

and, thus, $R_e(t_1, t_2)$ too depends only on $t_2 - t_1$, i.e.

$$R_e(0, t_2 - t_1) = R_e(t_1, t_2), \quad \forall t_1, t_2 \in \mathbb{Z}.$$

Abusing the notation it is possible to write more concisely $R_e(\tau) = R_e(0, \tau)$.

We limit the formulation to real-rational functions and processes with real-rational power spectral density in order to

guarantee the existence of power spectral factors necessary for the computation of the Wiener filter.

Definition 2. Consider a vector-valued sequence $h(k) \in \mathbb{R}^{1\times n}$ with $k \in \mathbb{Z}$. We define its $\mathcal{Z}$-transform as:

$$H(z) := \sum_{k=-\infty}^{\infty} h(k) z^{-k},$$

and we assume that the sum converges for any $z \in \mathbb{C}$ such that $r_1 < |z| < r_2$ with $r_1 < 1 < r_2$. We also assume that any entry of the n-dimensional vector $H(z)$ is a real-rational function of z. By the properties of the $\mathcal{Z}$-transform along with the convergence domain defined by r_1 and r_2, $H(z)$ uniquely identifies the sequence $h(k)$. Moreover, given a vector of rationally related random processes $e(t) := (e_1(t), .., e_n(t))^T$, denote for any $t \in \mathbb{Z}$ the random variable

$$y_t := \sum_{k=-\infty}^{\infty} h(k) e(t-k).$$

Then, we define by $H(z)e$ the related stochastic process such that

$$(H(z)e)(t) = y_t \quad \forall\, t \in \mathbb{Z}.$$

Definition 3. Given a n-dimensional time-discrete, zero-mean, wide-sense stationary random process $e(t) := (e_1(t), .., e_n(t))^T$, we define its power spectral density $\Phi_e(z)$ as:

$$\Phi_e(z) := \sum_{\tau=-\infty}^{\infty} R_e(\tau) z^{-\tau},$$

having a certain domain of convergence $\mathcal{D} \subseteq \mathbb{C}$ in the variable z. Denoting by $\Phi_{e_i e_j}(z)$ the entry (i, j) of $\Phi_e(z)$, it follows that

$$\Phi_{e_i e_j}(z) := \sum_{\tau=-\infty}^{\infty} R_{e_i e_j}(\tau) z^{-\tau}.$$

If for any $i, j \in \{1, \ldots, N\}$ the power spectral density $\Phi_{e_i e_j}(z)$ exists on the unit circle $|z| = 1$ of the complex plane and it is a real-rational function of z, we formally write

$$\Phi_{e_i e_j}(z) = \frac{A(z)}{B(z)} \quad \text{for } i, j = 1, \ldots, N,$$

with $A(z)$, $B(z)$ real coefficient polynomials, such that $B(z) \neq 0$ for any $z \in \mathbb{C}$, $|z| = 1$. In such a case, we say that e is a vector of rationally related random processes.

For the following developments, it is useful to introduce the sets:

$$\mathcal{F} := \{W(z) | W(z) \text{ is a real-rational scalar function of } z \in \mathbb{C} \text{ defined for } |z| = 1\}$$

$$\mathcal{F}^{m\times n} := \{W(z) | W(z) \in \mathbb{C}^{m\times n} \text{ and all of its entries is in } \mathcal{F}\}.$$

Definition 4. Let $e = (e_1, \ldots, e_n)^T$ be a vector of n rationally related random processes. We define the set $\mathcal{F}e$, as

$$\mathcal{F}e := \left\{x = H(z)e \mid H(z) \in \mathcal{F}^{1\times n}\right\}.$$

Problem 5. Consider a set $\mathcal{X} := \{x_1, \ldots, x_n\} \subset \mathcal{F}e$ of n rationally related processes with zero mean, known (cross)-power spectral densities $\Phi_{x_i x_j}(z)$ and let the sparsity indexes $m_j \in \mathbb{N}$, $m_j \leq n\ \forall j = 1, \ldots, n$, be given. Then, in the above framework the mathematical formulation of the considered problem can be stated as follows:

$$\min_{\substack{\alpha_{j,1},\ldots,\alpha_{j,m_j}\neq j \\ W_{j,\alpha_{j,k}}(z)\in\mathcal{F}}} E\left\{\left|x_j - \sum_{k=1}^{m_j} W_{j,\alpha_{j,k}}(z) x_{\alpha_{j,k}}\right|^2\right\}, \tag{2}$$

where every $W_{j,\alpha_{j,k}}(z)$, with $k = 1, \ldots, m_j$, is a possibly non-causal transfer function.

Remark 6. Fixed any set $\{\alpha_{j,k}\}_{k=1}^{m_j}$, Problem 5 is immediately solved by a multiple input Wiener filter. However, the determination of the parameters $\alpha_{j,k}$ makes the problem combinatorial.

3. A geometric interpretation

It is possible to give a geometrical interpretation of (2) by embedding the processes $x_1, \ldots, x_n$ in a suitable vector space. This interpretation has the main advantage of giving to the Wiener filter the meaning of a projective operator in such a space.

Proposition 7. *The ensemble $(\mathcal{F}e, +, \cdot, \mathbb{R})$ is a vector space.*

Proof. See the Appendix. □

Definition 8. For any $x \in \mathcal{F}e$ we denote the norm induced by the inner product $\langle\cdot, \cdot\rangle$ (refer to Definition 15 for a formal introduction of this operator) as

$$\|x\| := \sqrt{\langle x, x\rangle}.$$

As shown in detail in the Appendix, the set $\mathcal{F}e$ along with the operation $\langle\cdot, \cdot\rangle$ is a pre-Hilbert space, under the technical assumption that two different x_1 and x_2 are considered the same process, if they are alike according to a specific equivalence relation "$\sim$", i.e. if $x_1 \sim x_2$ (refer to Definition 11 for a formal introduction of the equivalence relationship necessary for this statement).

We provide an ad-hoc version of the Wiener Filter (guaranteeing that the filter will be real rational) with an interpretation in terms of the Hilbert projection theorem. Indeed, given signals $y, x_1, \ldots, x_n \in \mathcal{F}e$, the Wiener Filter estimating y from $x := (x_1, \ldots, x_n)$ can be interpreted as the operator that determines the projection of y onto the subspace $\mathcal{F}x$.

For the sake of simplicity, in the following we will denote the evaluation of the operator $\Phi_x(z)$ on the unit circle $z = e^{i\omega}$ just as $\Phi_x(\omega)$.

Proposition 9. *Let e be a vector of rationally related processes. Let y and $x_1, \ldots, x_n$ be processes in the space $\mathcal{F}e$. Define $x := (x_1, \ldots, x_n)^T$ and consider the problem*

$$\inf_{W\in\mathcal{F}^{1\times n}} \|y - W(z)x\|^2. \tag{3}$$

If $\Phi_x(\omega) > 0$, for all $\omega \in [-\pi, \pi]$, then the solution exists, is unique and has the form

$$W(z) = \Phi_{yx}(z)\Phi_{xx}(z)^{-1}.$$

Moreover, for any $W'(z) \in \mathcal{F}^{1\times n}x$, it holds that

$$\langle y - W(z)x, W'(z)x\rangle = 0. \tag{4}$$

Proof. Observe that the cost function satisfies

$$\|y - W(z)x\|^2 = \int_{-\pi}^{\pi} \Phi_{yy}(\omega) + W(\omega)\Phi_{xx}(\omega)W^*(\omega) - \Phi_{xy}(\omega)W^*(\omega) - W(\omega)\Phi_{yx}(\omega)d\omega.$$

The integral is minimized by minimizing the integrand for all $\omega \in [-\pi, \pi]$. Then, it is straightforward to find that the minimum is achieved for

$$W(\omega) = \Phi_{yx}(\omega)\Phi_{xx}^{-1}(\omega).$$

Defining the filter $W(z) = \Phi_{yx}(z)\Phi_{xx}(z)^{-1}$, we obtain a real-rational transfer matrix with no poles on the unit circle, that has the specified frequency response. Thus $\hat{x} = W(z)x$ minimizes the cost (3). Eq. (4) is an immediate consequence of the Hilbert projection theorem (for pre-Hilbert spaces) [27]. □

Problem 10. The mathematical formulation of Problem 5 can now be cast as follows:

$$\min_{\substack{\alpha_{j,1},\ldots,\alpha_{j,m_j}\neq j\\ W_{j,\alpha_{j,k}}(z)\in\mathcal{F}}} \left\| x_j - \sum_{k=1}^{m_j} W_{j,\alpha_{j,k}}(z) x_{\alpha_{j,k}} \right\|^2 . \tag{5}$$

4. Links with compressive sensing

In this section we highlight the connection between the problem of modeling a network topology and the compressive sensing problem. Such a connection is possible because of the pre-Hilbert structure constructed in Section 3 and Appendix. Indeed, the concept of inner product defines a notion of "projection" among stochastic processes and makes it possible to seamlessly import tools developed for the compressive sensing problem in order to tackle that of describing a sparsified topology.

In the last few years sparsity problems have attracted the attention of researchers in the area of Signal Processing. This is mainly due to the possibility of representing a signal using only few elements (words) of a redundant base (dictionary). Applications are numerous, ranging from data-compression to "super-resolution", and noise filtering [28,29].

There are many formalizations of the problem, but one of the most common is to cast it as

$$\min_w \|x_0 - \Psi w\|_2 \quad \text{subject to } \|w\|_0 \leq m, \tag{6}$$

where $x_0 \in \mathbb{R}^p$, $\Psi \in \mathbb{R}^{p\times n}$ is a matrix with $n < p$, whose columns represent a redundant base employed to approximate x_0 and the "zero-norm" (it actually is not a norm).

$$\|w\|_0 := |\{i \in \mathbb{N} | w_i \neq 0\}| \tag{7}$$

is defined by the number of non-zero entries of a vector w. It can be said that w is a "simple" way to express x_0 as a linear combination of the columns of Ψ, where the concept of "simplicity" is given by a constraint on the number of non-zero entries of w.

For each $j = 1, \ldots, n$ define the following sets:

$$\mathcal{W}^{(j)} = \{W(z) \in \mathcal{F}^{1\times n} | W_j(z) = 0\}, \tag{8}$$

where $W_j(z)$ denotes the j-th component of $W(z)$. For any $W \in \mathcal{W}^{(j)}$, define the "zero-norm" as

$$\|W\|_0 = \{\#\text{ of entries such that } \exists z \in \mathbb{C},\ W_i(z) \neq 0\}$$

and define the random vector

$$x = (x_1, \ldots, x_n)^T. \tag{9}$$

Then, the problem (2) can be formally cast as

$$\min_{W\in\mathcal{W}_j} \|x_j - Wx\|^2 \quad \text{subject to } \|W\|_0 \leq m, \tag{10}$$

which is, from a formal point of view, equivalent to the standard l_0 problem as defined in (6).

5. Solution via suboptimal algorithms

The problem of "modeling network interconnections/complexity reduction" we have formulated in this paper is equivalent to the problem of determining a sparse Wiener filter, as explained in the previous section, once a notion of orthogonality is introduced. This formal equivalence shows how deriving a suitable topology can immediately inherit a set of practical tools already developed in the area of compressing sensing.

Here we present, as illustrative examples, modifications of algorithms and strategies, well-known in the Signal Processing community, which can be adopted to obtain suboptimal solutions to the problem of modeling the network interconnections.

While formally identical to (6), the problem of a topology reconstruction cast as in (10) still has its own characteristics. Since the "projection" procedure in (10) is given by the estimation of a Wiener filter, it is computationally more expensive than the standard projection in the space of vectors of real numbers. For this reason greedy algorithms offer a good approach to tackle the problem, since the speed becomes a fundamental factor. Moreover, since the complexity of the network model is here one of the final goals, greedy algorithms are a suitable solution, since they allow one to specify explicitly the connection degree m_j of every node x_j. This feature is in general not provided by other algorithms, but it can be easily incorporated usually without increasing their computational complexity. As an alternative approach to greedy algorithms we also describe a strategy based on iterated reweighted optimizations as described in [26].

5.1. A modified Orthogonal Least Squares (Cycling OLS)

Orthogonal Least Squares (OLS) is a greedy algorithm proposed for the first time in [30] and in many ways it resembles the algorithm of Matching Pursuit developed in [31]. It basically consists of iterated orthogonal projections on elements of a (possibly redundant) base in order to approximate a given vector. For the details of this algorithm we remand the reader to [30]. However, for the sake of clarity, we reformulate it in terms of our problem. The initialization occurs at the first step, setting the set of the chosen elements of the dictionary to $\Gamma^{(1)} = \emptyset$. At the l-th iteration step, OLS determines the term $\hat{x}_j^{(l,i)}$ to be added to the reduced dictionary by projecting x_j onto the space generated by $\Gamma^{(l,i)} := \Gamma^{(l-1)} \cup \{x_i\}$ for any $i \neq j$. Then $\Gamma^{(l)}$ is defined as the $\Gamma^{(l,i)}$ for which $\|x_j - \hat{x}_j^{(l,i)}\|$ is the smallest, and the algorithm moves to the next iteration step. The standard OLS goes on at every step introducing a new vector until a stopping condition is met (usually based on the norm of the residual r^k or on the number of iterations).

We propose an algorithm, which derives directly from OLS, but it does not increase the number of vectors $x_{\alpha_{j,k}}$ approximating x_j above m_j. The variation from OLS is very simple. At any iteration, given the set of vectors $\Gamma^{(l-1)}$, if it already contains m_j vectors, the algorithm chooses a vector in $\Gamma^{(l-1)}$ to be removed and tries to replace it with another vector in order to improve the quality of the approximation and updates it. If such an improvement is not possible by removing any of the vectors in the current selection, the algorithm stops. The implementation can be described using the following pseudocode, where we use the notation $=$ for variable assignment and $==$ for variable comparison.

`Cycling Orthogonal Least Squares:`

0. define $x_0 = 0$ (null time series) and $c = 0$.
1. initialize the m_j-tuple $S = (x_0, x_0 \ldots, x_0)$ and $k = 1$
2. while $c \leq m_j$
   - 2a. for $i = 1, \ldots, n$, $i \neq j$
     define S_i as the m_j-tuple where x_i replaces the k-th element of S
     and
     define $\hat{x}_j^{(i)}$ as the projection of x_j on to S_i
   - 2b. $\alpha = \arg\max_i \|x_j - \hat{x}_j^{(i)}\|$
   - 2c. if $x_\alpha == S[k]$ then $c = c + 1$
   - 2d. else $S[k] = x_\alpha$, $c = 1$, $k = k \bmod m_j$, $k = k + 1$
3. return S.

The reason for our modification is here explained. COLS implements a coordinate descent guaranteeing that the number of non-zero components of the solution does not exceed m_j. Once such a limit has been reached, it tries to improve the quality of the approximation without reducing the sparsity of the current solution.

5.2. Solution via Reweighted Least Squares (RWLS)

Another possible approach to "encourage" sparse solutions is provided by reweighted minimization algorithms, as proposed in [26,28]. A comparison between reweighted norm-1 and norm-2 methods is performed in [29]. We consider only reweighted least squares, because such an algorithm is easier to implement, but the intuition behind the two techniques is basically the same.

Using Parseval's theorem, problem (2), can be formulated as

$$\min_{\substack{\alpha_{j,1},\ldots,\alpha_{j,m_j}\neq j\\ W_{j,\alpha_{j,k}}(z)\in\mathcal{F}}} \int_{-\pi}^{\pi} \Phi_{[x_j-\sum_{k=1}^{m_j} W_{j,\alpha_{j,k}}(\omega)x_{\alpha_{j,k}}]}(\omega)d\omega. \tag{11}$$

Consider the following convex variation of the problem

$$\min_{W_{j,k}(z)\in\mathcal{F}} \left\{ \int_{-\pi}^{\pi} \Phi_{[x_j-\sum_{k\neq j} W_{j,k}(\omega)x_k]}(\omega) \right\} \tag{12}$$

subject to

$$\sum_{k=1}^{n} \int_{-\pi}^{\pi} \mu_k W^*_{j,k}(\omega) W_{j,k}(\omega) d\omega \leq 1,$$

where $\mu_k \in \mathbb{R}^n$ is a set of weights for the filters $W_{j,k}(z)$. Using the compact notation introduced in Section 4, we can equivalently write

$$\min_{W\in\mathcal{W}_j} \|x_j - Wx\|^2 \quad \text{subject to } \|W\|^2_\mu \leq 1, \tag{13}$$

where, for a vector $\mu = (\mu_1, \ldots, \mu_n)$, we define

$$\|W\|^2_\mu := \frac{1}{m_j} \sum_{1}^{n} \int_{-\pi}^{\pi} \mu_k W^*_{j,k}(\omega) W_{j,k}(\omega) d\omega.$$

Let us assume that the $\alpha_{j,k}$'s and the relative $W_{\alpha_{j,k}}$ solving (2) are known. Technically, we could set

$$\mu_l := \frac{1}{m_j} \left(\int_{-\pi}^{\pi} W^*_l(\omega) W_l(\omega) d\omega \right)^{-1}, \tag{14}$$

if $l = \alpha_{j,k}$ for some $k = 1, \ldots, m_j$ and $\mu_l = +\infty$ otherwise. With such a choice of weights, the two problems (2) and (13) would be equivalent, since they would provide the same solutions. However, problem (13) has the advantage of being convex. Of course, the values $\alpha_{j,k}$ are not a-priori known, thus it is not possible to evaluate (14). An iterative approach has been proposed making use of the intuition that we have just formulated to estimate the weights (14).

`Reweighted Least Squares:`

0. For all x_j
1. initialize the weight vector $\mu := 0$
2. while a stop criterion is met
   - 2a. solve the convex problem
     $$\min_{W\in\mathcal{W}_j} \|x_j - Wx\|^2 \quad \text{subject to } \|W_j\|^2_\mu \leq 1$$
   - 2b. compute the new weights
     $$\mu_k = \frac{1}{m_j} \int_{-\pi}^{\pi} \|W_j(\omega)\| d\omega$$
3. return all the W_j's.

At any iteration the convex relaxation of the problem is solved, and new weights are computed as a functions of the current solution. When a stopping criterion is met (usually based on the number of iterations), the final solution can be obtained by selecting the m_j largest entries of each W_j.

6. Applications and examples

In this section we report numerical results obtained implementing the algorithms described in the previous section. In order to evaluate the performances provided by the two algorithms (COLS and RWLS) we have considered a network of 20 nodes as represented in Fig. 2 (true). In the graph every node N_j describes a stochastic process x_j, while every directed arc from node N_i to node N_j represents a transfer function $H_{ji}(z) \neq 0$. The absence of such an arc implies that $H_{ji}(z) = 0$. The above transfer functions have been randomly selected from a class of causal FIR filters of order 5. More precisely, a set of parametric fifth-order polynomial rational functions have been built by randomly choosing their coefficients until their roots were found stable. Then, the complete network has been simulated to check for the stability of the overall interconnected system until such a condition has been met. Thus, each process x_j follows the dynamics

$$x_j = e_j + \sum_{i\neq j} H_{ji}(z) x_i. \tag{15}$$

Every node signal is also implicitly considered affected by an additive white Gaussian noise e_j such that the Signal-to-Noise Ratio (SNR) is 4 (a very noisy scenario). All the noise processes are independent from each other. The network has been simulated for 2000 steps obtaining 20 time series. The time series have been employed to estimate a non-causal FIR approximation of the Wiener Filters of order 21 both in the COLS and in the RWLS algorithm. In Fig. 2 we report the results of the identification. Using a global search the global minima of (6) provide the topologies in Fig. 2 (reduced 2) and Fig. 2 (reduced 3) for the case $m_j = 2$ and $m_j = 3$ for each node respectively. In Fig. 2 (no reduction) we report the topology obtained with no constraint on the maximum number of edges: a small threshold has been introduced to remove any edge (N_i, N_j) associated with $W_{ji}(z) \simeq 0$. In the second row of graphs, the results for COLS for the cases $m_j = 1$, $m_j = 2$ and $m_j = 3$ are presented. In Fig. 2 (COLS variable) we report the result given by the implementation of a strategy to automatically determine the number of edges: the number of edges is increased only if it gives a reduction of 20% of the residual error. In the third row of graphs, we report the analogous results for the RWLS algorithm. These simulation results suggest the following conclusions.

Given data generated according to a certain sparse topology, as in Fig. 2 (true), a reconstruction with no reduction techniques tends to determine many more links than in the actual topology. This highlights the necessity of methods to reduce the number of links. A-priori fixing an arbitrary value for m_j produces the results depicted in Fig. 2 (COLS1-3 and RWLS1-3), showing that the choice of m_j is an essential feature in order to obtain a suitable reconstruction: a too small m_j determines a too sparse network. The simple adaptive strategy suggested in this paper provides a good compromise between the accuracy of the estimate and the final structure complexity, leading to a more faithful reconstruction of the actual topology. Both COLS and RWLS with the adaptive strategy provide similar outcomes, even though COLS, which is based on simple projections and not on iterated estimates, tends to have a lower computational cost.

Finally, we also provide a comparison between COLS2 (i.e. COLS with fixed $m_j = 2\ \forall j$) and two other methods, specifically designed for the analysis of network topologies and graphical models. The first one is a correlation based procedure presented in [7] (see p. 216), while the second one is the Reweighted Sparsification algorithm, presented in [32] and based on the usage of Moving Average models, where we use, instead, a generic frequency domain analysis. For the comparison setup we have considered a simple though challenging scenario consisting of the ribbon

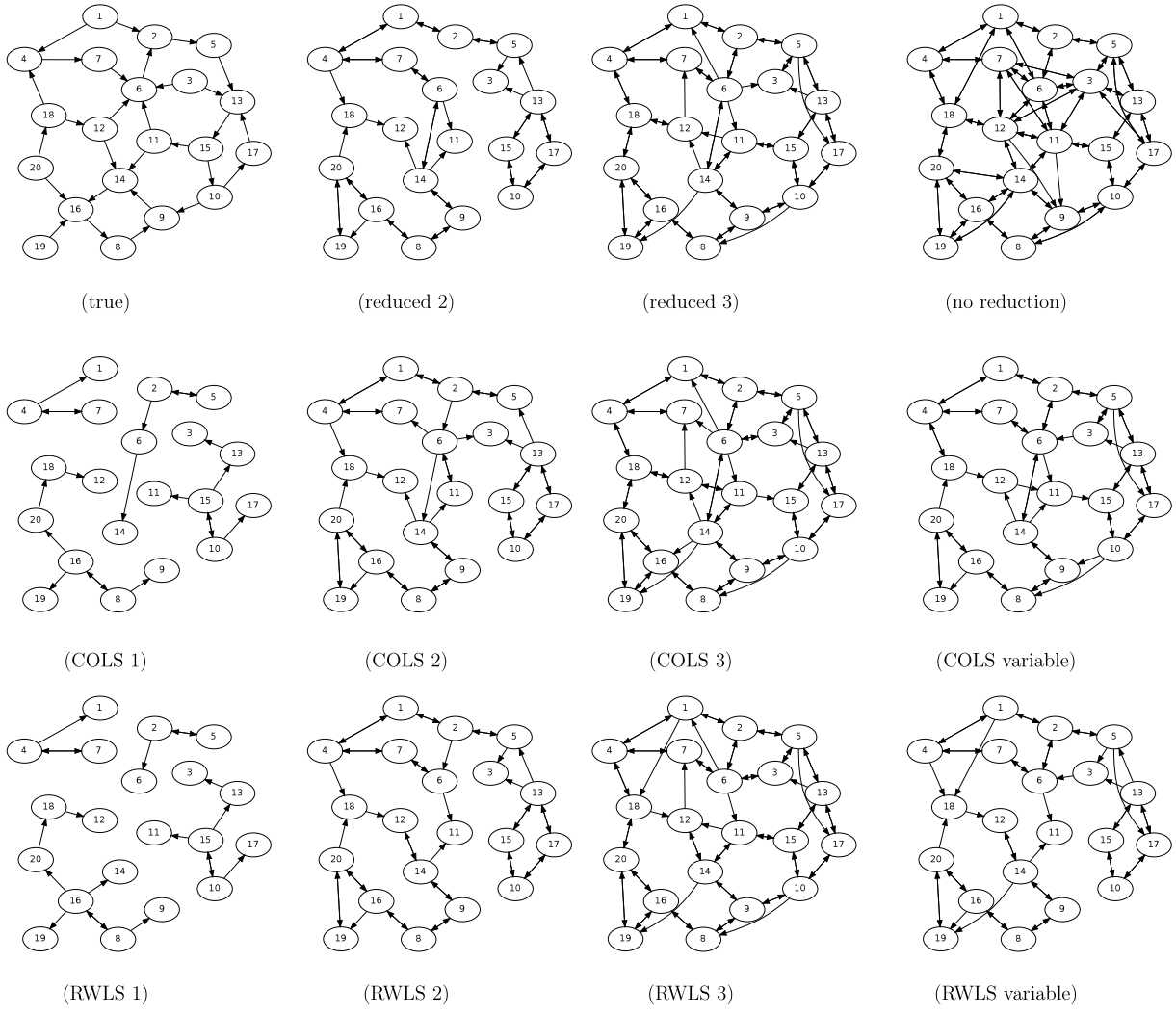

Fig. 2. The actual network topology (true); the topology obtained by a global minimization of (3) with $m_j = 2$ for every node (reduced 2); the topology with $m_j = 3$ for every node (reduced 3); the topology with no constraint on m_j but with a "small" threshold imposed on a norm of the Wiener Filters to avoid a complete graph (no reduction); the topologies obtained used the COLS suboptimal approach with $m_j = 1$ (COLS 1), $m_j = 2$ (COLS 2), $m_j = 3$ (COLS 3), and a self-adjusting strategy for m_j (COLS variable): a link is introduced if it gives at least a reduction of 20% of the residual error; the topologies obtained by RWLS after 10 iterations and keeping only the m_j filters with largest norm (RWLS 1), (RWLS 2), (RWLS 3), or a self-adjusting strategy (RWLS variable).

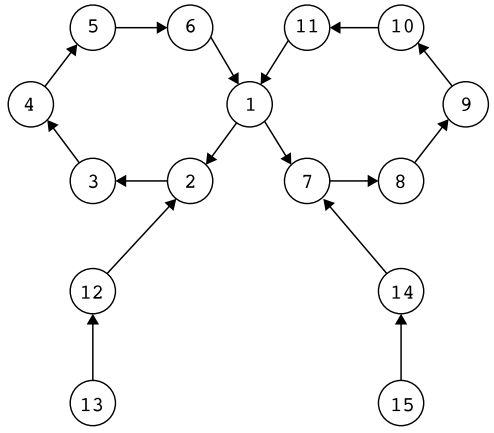

Fig. 3. The twisted loop network used to compare the COLS2 algorithm with the reweighted sparsification in [32] and the standard correlation analysis in [7].

network depicted in Fig. 3, i.e. a twisted loop fed by two different input chains. Each node of the network is given by a randomly generated second order Infinite Impulse Response (IIR) filter with random delay. Moreover, additive white noises affect the systems' outputs, so to provide at regime a quarter of the total energy of the signals. Then, the final network is simulated for different values of the models' parameters until a stable configuration is met, and eventually 5000 samples belonging to the regime condition are taken. In Table 1 the three considered strategies are compared over three different experiments. The sensitivity and specificity indexes are given by

$$\text{Sensitivity} = \frac{\#\text{ correctly identified links}}{\#\text{ correctly identified links} + \#\text{ incorrectly unidentified links}}$$

$$\text{Specificity} = \frac{\#\text{ correctly unidentified links}}{\#\text{ correctly unidentified links} + \#\text{ incorrectly identified links}}.$$

As shown in the table, the main difficulty in identifying the edges of this network is that the IIR models, along with the twisted loop topology, make every signal strongly correlated to each other. Therefore, every attempt aimed to improve the sensitivity has a negative effect on the specificity, and vice versa. Even though a

Table 1
The table reports the sensitivity and specificity indexes of the methods COLS2, reweighted sparsification and simple correlation analysis over three randomly generated experiments related to the ribbon network of Fig. 3. The reweighted sparsification method and the correlation based analysis state the absence of any link between two nodes only, when a specific function is equal to zero. Then, in order to actually apply such procedures we have used a threshold equal to 0.01, that has provided a reasonable compromise between sensitivity and specificity among all the considered experiments.

Method	Sensitivity	Specificity
COLS2	0.6250	0.8365
Reweighted sparsification	0.7500	0.4212
Correlation	0.8750	0.3708
COLS2	0.5625	0.8077
Reweighted sparsification	0.6250	0.4904
Correlation	1	0.3483
COLS2	0.6875	0.8558
Reweighted sparsification	0.4375	0.6154
Correlation	0.9375	0.4831

detailed statistical analysis is out of the scope of this comparison, a number of useful observations are in order.

The correlation method appears as the most biased procedure, preferring sensitivity over specificity, as expected by the loop presence. The Reweighted Sparsification, instead, is the more dependent on the single experiment realization, as expected by the usage of a specific dynamical model. Conversely, COLS seems to provide the more stable behavior, denoting a reasonable compromise between the two performance indexes. Indeed, if we consider the minimum among sensitivity and specificity as a reliability factor, COLS turns out as the most truthful method. It is also worth stressing that the correlation based analysis depends on the inverse of the covariance matrix, and, therefore, it may present remarkable numerical problems just when the network is particularly sparse. In the end, as a final remark, we want to highlight that between COLS and the Reweighted Sparsification algorithm the second one has the heavier computational burden, as it can be directly observed by the fact that it relies on a possibly large number of optimizations subproblems.

7. Final remarks

We have formulated the problem of deriving a link structure from a set of time series, obtained by sampling the output of as many interconnected dynamical systems. Every time series is represented as a node in a graph and their dependencies as connecting edges. The approach we follow in determining the graph arcs relies on (linear) identification techniques based on an ad-hoc version of the Wiener Filter (guaranteeing that the filter will be real rational) with an interpretation in terms of the Hilbert projection theorem. If a time series X_i turns outs "useful" to model the time series X_j, then the directed arc (i, j) is introduced in the graph. In order to modulate the complexity of the final graph, a maximum number m_j of arcs pointing at X_j is assumed, and a cost function is minimized to find the most appropriate arcs. The problem has a similar formulation and strong connections with the problem of compressing sensing, which has been widely studied in recent years. Such a connection is possible because of the pre-Hilbert structure we have constructed. Indeed, the concept of inner product defines a notion of "projection" among stochastic processes and makes it possible to seamlessly import tools developed for the compressive sensing problem in order to tackle the problem of modeling a network topology.

The problem of "topology reconstruction/complexity reduction" is equivalent to the problem of determining a sparse Wiener filter, as explained. However, since an optimization problem must be solved for any single node, we consider the application of suboptimal solutions. In particular, we have introduced a suboptimal greedy algorithm obtained as a modification of the Orthogonal Least Squares (COLS), and an alternative approach based on iterated ReWeighted Least Squares (RWLS). By the comparison of the two algorithms on numerical data we have shown the effectiveness of the proposed solutions. Note that in the present paper no absolute error metric is provided.

Future work will investigate some measure criteria to judge the performance of the algorithm. For instance, as a starting method we could count the correct identified links and the wrong ones, when the underlying topology is known, to define the "most accurate" topology and extend such a measure of accuracy in some norms to the unknown topology case.

Appendix

We provided hereafter the additional definitions and propositions which are needed for the construction of the pre-Hilbert space.

Definition 11. Given two discrete-time scalar, zero-mean, wide-sense jointly stationary random processes $x_1(t)$ and $x_2(t)$, we write that $x_1 \sim x_2$ if and only if, for any $t \in \mathbb{Z}$, $E[(x_2(t) - x_1(t))^2] = 0$, that is $x_1(t) \overset{\text{a.s.}}{=} x_2(t)$ ($x_1(t) = x_2(t)$ almost surely for any $t \in \mathbb{Z}$).

Proposition 12. *The relation $\sim$ is an equivalence relation on any set X of zero-mean time-discrete wide-sense jointly stationary scalar random processes defined on the time domain $\mathbb{Z}$.*

Proposition 13. *Let $x_1 := H^{(1)}(z)e$, $x_2 := H^{(2)}(z)e$ be two elements of $\mathcal{F}e$. Then x_1, x_2 are scalar, zero-mean, wide-sense jointly stationary random processes with rational power cross-spectral densities having no poles in the set $\{z \in \mathbb{C} | \, |z| = 1\}$.*

Proof. The processes x_1 and x_2 are scalar by the definition of $\mathcal{F}e$. Since $H^{(1)}(z), H^{(2)}(z) \in \mathcal{F}^{1\times n}$, they are real-rational and defined on the unit circle, and, as a consequence, they admit a unique representation in terms of the bilateral $\mathcal{Z}$-transform

$$H^{(1)}(z) = \sum_{k=-\infty}^{+\infty} h_k^{(1)} z^{-k} \tag{16}$$

$$H^{(2)}(z) = \sum_{k=-\infty}^{+\infty} h_k^{(2)} z^{-k}, \tag{17}$$

with $h_k^{(1)}, h_k^{(2)} \in \mathbb{R}^{1\times n}$ for $k \in \mathbb{Z}$, such that the convergence is guaranteed on the unit circle $|z| = 1$.

First, let us evaluate the mean of $x_1(t)$, that is

$$E[x_1(t)] = E\left[\sum_{k=-\infty}^{\infty} h_k^{(1)} e(t-k)\right] = \sum_{k=-\infty}^{\infty} h_k^{(1)} E[e(t-k)] = H^{(1)}(1)E[e(0)] = 0.$$

Thus, it does not depend on the time t. Analogously $E[x_2(t)] = 0$. Now, let us evaluate the cross-covariance function

$$\begin{aligned} R_{x_1x_2}(t, t+\tau) &:= E[x_1(t)x_2(t+\tau)^T] \\ &= E\left[\left(\sum_{k=-\infty}^{\infty} h^{(1)}(k)e(t-k)\right)\left(\sum_{l=-\infty}^{\infty} e^T(t+\tau-l)(h^{(2)}(l))^T\right)\right] \\ &= E\left[\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h^{(1)}(k)e(t-k)e^T(t+\tau-l)(h^{(2)}(l))^T\right] \end{aligned}$$

$$= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h^{(1)}(k)E[e(t-k)e^T(t+\tau-l)](h^{(2)}(l))^T$$
$$= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h^{(1)}(k)R_e(\tau-l+k)(h^{(2)}(l))^T = R_{x_1x_2}(0,\tau).$$

Thus, the cross-covariance does not depend on the time t and, abusing notation, it is possible to define

$$R_{x_1x_2}(\tau) := R_{x_1x_2}(0,\tau). \tag{18}$$

Finally, by evaluating the bilateral $\mathcal{Z}$-transform of $R_{x_1x_2}(\tau)$, we have that

$$\begin{aligned}\Phi_{x_1x_2}(z) &:= \sum_{\tau=-\infty}^{\infty} R_{x_1x_2}(\tau)z^{-\tau}\\ &= \sum_{\tau=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h^{(1)}(k)R_e(\tau-l+k)\\ &\quad\times (h^{(2)}(l))^T z^{-\tau-l+k}z^{-k}z^{l}\\ &= H^{(1)}(z)\Phi_e(z)H^{(2)}(z^{-1})\end{aligned}$$

and, thus, one can observe that the above function converges on the unit circle as well. □

Proposition 14. *Given a rationally related vector e, the set $\mathcal{F}e$ is closed with respect to addition, transformation by $H(z) \in \mathcal{F}$ and multiplication by scalar $\alpha \in \mathbb{R}$. Moreover, it holds that, for $x_1 = H^{(1)}(z)e \in \mathcal{F}e$ and $x_2 = H^{(2)}(z)e \in \mathcal{F}e$,*

$$H^{(1)}(z)e + H^{(2)}(z)e = \left[H^{(1)}(z) + H^{(2)}(z)\right]e$$
$$H(z)[H^{(1)}(z)e] = \left[H(z)H^{(1)}(z)\right]e$$
$$\alpha[H^{(1)}(z)e] = \left[\alpha H^{(1)}(z)\right]e.$$

Proof. Let

$$H(z) = \sum_{k=-\infty}^{+\infty} h(k)z^{-k}. \tag{19}$$

- Sum.
  We have
$$\begin{aligned}x_1(t)+x_2(t) &= \left[\sum_{k=-\infty}^{\infty} h_k^{(1)}e(t-k)\right] + \left[\sum_{k=-\infty}^{\infty} h_k^{(2)}e(t-k)\right]\\ &= \sum_{k=-\infty}^{\infty}[h_k^{(1)}+h_k^{(2)}]e(t-k)\\ &= ([H^{(1)}(z)+H^{(2)}(z)]e)(t).\end{aligned}$$
  Since $[H^{(1)}(z)+H^{(2)}(z)]$ has no poles on the set $\{z \in \mathbb{C}\,|\,|z| = 1\}$, $x_1 + x_2 \in \mathcal{F}e$.
- Multiplication by $H(z) \in \mathcal{F}$.
  Since $x_1 \in \mathcal{F}e$, it is a 1-dimensional rationally related vector. Then, it makes sense to compute the random process $H(z)x_1$ for $H(z) \in \mathcal{F} = \mathcal{F}^{1\times 1}$
$$\begin{aligned}(H(z)x_1)(t) &:= \sum_{k=-\infty}^{\infty} h(k)x_1(t-k) &(20)\\ &= \sum_{k=-\infty}^{\infty} h(k)\sum_{l=-\infty}^{\infty} h_l^{(1)}e(t-k-l) &(21)\\ &= \sum_{k=-\infty}^{\infty} h(k)\sum_{l=-\infty}^{\infty} h_{l-k}^{(1)}e(t-l) &(22)\\ &= \sum_{l=-\infty}^{\infty}\left[\sum_{k=-\infty}^{\infty} h(k)h_{l-k}^{(1)}\right]e(t-l) &(23)\\ &= \left(\left[H(z)H^{(1)}(z)\right]e\right)(t), &(24)\end{aligned}$$
  where the last equality comes from the properties of the convolution. Since $H(z)H^{(1)}(z)$ has no poles in the set $\{z \in \mathbb{C}\,|\,|z| = 1\}$, $H(z)x_1 \in \mathcal{F}e$.
- Multiplication by scalar $\alpha \in \mathbb{R}$.
  It is a special case of the previous property. □

Proof of Proposition 7. Let us denote as $\overline{\mathcal{F}e}$ the partition set of $\mathcal{F}e$ induced by the equivalence operator $\sim$. Then, consider $X_1, X_2, X_3 \in \overline{\mathcal{F}e}$, $x_1 \in X_1$, $x_2 \in X_2$, $x_3 \in X_3$, and $\alpha_1, \alpha_2 \in \mathbb{R}$.

- Commutativity for the sum.
  Consider
$$x := x_1 + x_2 \in X_1 + X_2. \tag{25}$$
  Then, since $x = x_2 + x_1$, we also have that $x \in X_2 + X_1$. Since $(\overline{\mathcal{F}e}, +, \cdot, \mathbb{R})$ is a partition, we obtain $X_1 + X_2 = X_2 + X_1$.
- Associativity for the sum.
  Consider
$$x_a := x_1 + (x_2 + x_3) \in X_1 + (X_2 + X_3) \tag{26}$$
$$x_b := (x_1 + x_2) + x_3 \in (X_1 + X_2) + X_3. \tag{27}$$
  Then, since $x_a = x_b$, for the property of a partition set, $X_1 + (X_2 + X_3) = (X_1 + X_2) + X_3$.
- Additive identity.
  The process $x_0(t) = 0$ for any t is in $\mathcal{F}e$, because the zero transfer function is in $\mathcal{F}$. Let the set $X_0 \in \overline{\mathcal{F}e}$ be the set that constraints x_0. Since $x_1 + x_0 = x_1$ for any x_1, we have that X_0 is the identity element for the addition.
- Additive inverse.
  If $x_1 \in \mathcal{F}e$, then also $-x_1 \in \mathcal{F}e$, because the transfer function $-1 \in \mathcal{F}$.
- Scalar multiplication identity.
  Let x_1 be a process in X_1. The scalar 1 is the multiplication identity. Indeed, we have
$$1 \cdot x_1 = x_1 \in X_1. \tag{28}$$
  Thus, it holds that $1 \cdot X_1 = X_1$.
- Associativity of the scalar multiplication.
  Since we have that $x = \alpha_1(\alpha_2 x_1) = (\alpha_1\alpha_2)x_1$, we also have that $\alpha_1(\alpha_2 X_1) = (\alpha_1\alpha_2)X_1$.
- Distributivity of the scalar sum.
  Since we have $(\alpha_1 + \alpha_2)x_1 = \alpha_1 x_1 + \alpha_2 x_1$, we also have $(\alpha_1 + \alpha_2)X_1 = \alpha_1 X_1 + \alpha_2 X_1$.
- Distributivity of the vector sum.
  Since we have $\alpha_1(x_1 + x_2) = \alpha_1 x_1 + \alpha_1 x_2$, we also have $\alpha_1(X_1 + X_2) = \alpha_1 X_1 + \alpha_1 X_2$. □

Definition 15. We define a scalar binary operation $\langle\cdot,\cdot\rangle$ on $\mathcal{F}e$ in the following way

$$\langle x_1, x_2\rangle := R_{x_1x_2}(0).$$

Proposition 16. *The set $\mathcal{F}e$, along with the operation $\langle\cdot,\cdot\rangle$ is a pre-Hilbert space (with the technical assumption that x_1 and x_2 are the same processes if $x_1 \sim x_2$).*

Proof. According to the above results the statement is proven if $\langle\cdot,\cdot\rangle$ is an internal product for $\mathcal{F}e$. To this aim, let us consider the partition set of $\mathcal{F}e$ induced by the equivalence operator $\sim$ and denote it as $\overline{\mathcal{F}e}$. Then, consider $x_1 \in X_1 \in \overline{\mathcal{F}e}$ and $x_2 \in X_2 \in \overline{\mathcal{F}e}$ and evaluate the internal product properties by inspection.

- Symmetry: $\langle X_1, X_2\rangle = \langle X_2, X_1\rangle$.
  it is a direct consequence of the analogous property of the correlation, i.e. $R_{x_1x_2}(0) = R_{x_2x_1}(0)$.

- Linearity in the first argument with respect to a scalar value $\alpha \in \mathbb{R}$: $\langle \alpha X_1, X_2\rangle = \alpha\langle X_1, X_2\rangle$.
  it is sufficient to observe that $R_{\alpha x_1, x_2}(0) = E[\alpha x_1(0) x_2^T(0)] = \alpha E[x_1(0)x_2^T(0)] = \alpha R_{x_1,x_2}(0)$.
- Positive definiteness: $\langle X_1, X_1\rangle \geq 0$.
  by the definition it follows that $\langle x_1, x_1\rangle = E[x_1(0)x_1^T(0)] \geq 0$. Observe that $E[x_1(0)x_1^T(0)] = 0$ implies $x_1(t) = 0$ almost everywhere. Moreover, also the process $X_0(t) = 0, \forall t$, belongs to X_1 and, thus, X_1 turns out the identity element for the sum property among the set of the partition $\overline{\mathcal{F}e}$. □

References

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Complex networks: structure and dynamics, Physics Reports 424 (4–5) (2006) 175–308.
[2] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences 99 (12) (2002).
[3] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical Review E 69 (2) (2004).
[4] H. Zhang, Z. Liu, M. Tang, P.M. Hui, An adaptive routing strategy for packet delivery in complex networks, Physics Letters A 364 (2007) 177–182.
[5] R. Olfati-Saber, Distributed Kalman filtering for sensor networks, in: Proc. of IEEE CDC, New Orleans, 2007, pp. 5492–5498.
[6] I. Schizas, A. Ribeiro, G. Giannakis, Consensus in ad hoc WSNs with noisy links-part i: distributed estimation of deterministic signals, IEEE Transactions on Signal Processing 56 (1) (2008).
[7] E.D. Kolaczyk, Statistical Analysis of Network Data: Methods and Models, Springer-Verlag, Berlin, Germany, 2009.
[8] M. Timme, Revealing network connectivity from response dynamics, Physical Review Letters 98 (22) (2007) 224101.
[9] S. Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, A. Rapisarda, Detecting complex network modularity by dynamical clustering, Physical Review E 75 (2007) 045102.
[10] D. Napoletani, T. Sauer, Reconstructing the topology of sparsely connected dynamical networks, Physical Review E 77 (2008) 026103.
[11] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, UK, 2000.
[12] M. Ozer, M. Uzuntarla, Effects of the network structure and coupling strength on the noise-induced response delay of a neuronal network, Physics Letters A 375 (2008) 4603–4609.
[13] B.M. Sanandaji, T.L. Vincent, M.B. Wakin, Exact topology identification of large-scale interconnected dynamical systems from compressive observations, in: American Control Conference, ACC, 2011, IEEE, 2011, pp. 649–656.
[14] A. Chiuso, G. Pillonetto, A Bayesian approach to sparse dynamic network identification, Automatica 48 (8) (2012) 1553–1565. http://dx.doi.org/10.1016/j.automatica.2012.05.054.
[15] C.D. Michener, R.R. Sokal, A quantitative approach to a problem of classification, Evolution 11 (1957) 490–499.
[16] R.B. Freckleton, P.H. Harvey, M. Pagel, Phylogenetic analysis and comparative data: a test and review of evidence, American Naturalist 160 (2002) 712–726.
[17] R. Diestel, Graph Theory, Springer-Verlag, Berlin, Germany, 2006.
[18] G. Innocenti, D. Materassi, A modeling approach to multivariate analysis and clusterization theory, Journal of Physics A 41 (20) (2008) 205101.
[19] Donatello Materassi, Giacomo Innocenti, Unveiling the connectivity structure of financial networks via high-frequency analysis, Physica A: Statistical Mechanics and its Applications 388 (18) (2009) 3866–3878.
[20] G. Innocenti, D. Materassi, Modeling the topology of a dynamical network via Wiener filtering approach, Automatica 48 (5) (2012) 936–946.
[21] L. Getoor, N. Friedman, B. Taskar, D. Koller, Learning probabilistic models of relational structure, Journal of Machine Learning Research 3 (2002) 679–707.
[22] Nir Friedman, D. Koller, Being Bayesian about network structure: a Bayesian approach to structure discovery in Bayesian networks, Machine Learning 50 (2003) 95–126.
[23] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008 (10) (2008) P10008.
[24] D. Materassi, G. Innocenti, L. Giarré, Reduced complexity models in the identification of dynamical networks: links with sparsification problems, in: Proceedings of the 48th IEEE Conference on Decision and Control, 2009 Held Jointly with the 2009 28th Chinese Control Conference, CDC/CCC 2009, 2009, pp. 4796–4801.
[25] E.J. Candès, Terence Tao, Decoding by linear programming, IEEE Transactions on Information Theory 51 (12) (2005) 4203–4215.
[26] Emmanuel Candès, Michael Wakin, Stephen Boyd, Enhancing sparsity by reweighted l1 minimization, Journal of Fourier Analysis and Applications 14 (2008) 877–905.
[27] David G. Luenberger, Optimization by Vector Space Methods, John Wiley & Sons Inc., Hoboken, New Jersey, 1969.
[28] I. Daubechies, R. DeVore, M. Fornasier, C.Si. Güntürk, et al., Iteratively reweighted least squares minimization for sparse recovery, Communications on Pure and Applied Mathematics 63 (1) (2009) 1–38.
[29] D.P. Wipf, S. Nagarajan, Iterative reweighted l1 and l2 methods for finding sparse solutions, in: Signal Processing with Adaptive Sparse Structured Representations, Rennes, France, April 2009.
[30] S. Chen, S.A. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification, International Journal of Control 50 (5) (1989) 1873–1896.
[31] S. Mallat, Z. Zhang, Matching pursuits with time–frequency dictionaries, IEEE Transactions on Signal Processing 41 (12) (1993).
[32] M. Ayazoglu, M. Sznaier, N. Ozay, Blind identification of sparse dynamic networks and applications, in: Proceedings of 50th IEEE Conference on Decision and Control and European Control Conference, 2011, pp. 2944–2950.