2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Causal and non-causal filtering for network reconstruction

Donatello Materassi   Murti V. Salapaka

Abstract— **In many diverse areas, determining the connectivity of various entities in a network is of significant interest. This article's main focus is on a network (graph) of nodes (vertices) that are linked via filters that represent the edges of a graph. Both cases of the links being non-causal and causal are considered. Output of each node of the graph represents a scalar stochastic process driven by an independent noise source and by a sum of filtered outputs of nodes linked to the node of interest. It is shown that the method provided will identify all true links in the network with some spurious links added. The spurious links remain local in the sense that they are added within a hop of a true link. In particular, it is proven that the method determines a link to be present only between the kins of a node where kins of a node consist of parents, children and co-parents (other parents of all of its children) in the graph. Main tools for determining the network topology is based on Wiener filtering. Another significant insight provided by the article is that the Wiener filter estimating a stochastic process, represented by a node, based on other processes in a network configuration remains local in the sense that the Wiener filter utilizes only measurements local to the node being estimated.**

I. 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. While networks of dynamical systems are well studied and analyzed in physics [1], [2] and engineering [3], [4], there are fewer results that address the problem of reconstructing the topology of a network. Unravelling the interconnectedness of a set of processes is of significant interest in many fields, with the necessity for general tools rapidly increasing (see [5], [6] and [7] and the bibliography therein for recent results). However, such a problem poses formidable theoretical as well as practical challenges (see [8]). Existing results derive a network topology from sampled data (see e.g. [9], [5], [7], [10]) or to determine the presence of substructures (see e.g. [2], [6]). A well-known technique for the identification of a tree network is developed in [9] for the analysis of a stock portfolio. However, in [11] a severe limitation 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 [12] 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. In [12] it is shown that a correct reconstruction can be guaranteed for a topology with no cycles.

In [6] different techniques for quantifying and evaluating 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 [5] a method to identify a network of dynamical systems is described. However, primary assumptions of the technique are the possibility to manipulate the input of every single node and the possibility of conducting experiments to detect the link connectivity.

In [13] an interesting and novel approach based on auto-regressive models and Granger-causality [14] is proposed for reconstructing a network of dynamical systems. This technique relies on multivariate identification procedure to detect the presence of a link, but still no theoretical sufficient or necessary conditions are derived to check the correctness of the results.

In this paper the problem of reconstructing a network of dynamical systems where every node represents an observable scalar signal and the dynamics is linear and represented by the connecting links is addressed. The problem, when analyzed from a systems theory point of view, provides a method for correctly identifying a topology that belongs to the pre-specified class of self-kin networks. Moreover, if the network does not belong to such a class, conditions about the optimality of the identified topology according to a defined criterion is estabilished. From this perspective, sufficient conditions for the exact reconstruction of a large class of networks, which we name self-kin, are derived. In the case the network is not self-kin, the reconstructed topology is guaranteed to be the smallest self-kin network containing the actual one. The theory developed is not bayesian and relies directly on Wiener filtering theory. Conditions derived for the detection of links are based on sparsity properties of the (non-causal) Wiener filter modeling the network. Indeed, conditions under which the Wiener filter smoothing a signal of the network is “local” are derived. From a different perspective, another important contribution of the paper is given by providing conditions for a local and distributed implementation of the Wiener filter.

The results obtained bear a striking similarity to the ones developed in the area of machine learning for Bayesian Networks (BNs) [15], [16] where the topology of a network of nodes that represent random variables is sought. The main result obtained in the BNs literature (see [17]) is that the probability distribution of a random variable conditioned on the rest of the random variables of the network is equal to the probability distribution of the random variable conditioned only on the random variables within the kin set of the random variable. It is assumed that the network has no loops. The problem considered in this article is for a network of random processes and is not restricted to random variables as is the case for BNs. Evidently issues concerning causality and stability do not arise for BNs which have to be addressed for a network of random processes. Moreover, in this article no assumption on the absence of loops is made as is the case in [17].

The paper is organized as follows. In Section II definitions are provided based on standard notions of graph theory; in Section III the main problem is formulated; in Section IV the main results are provided for non-causal Wiener filtering; in Section V the results are extended to causal Wiener filtering and Granger causality; in Section VI the implementation of algorithms for the detection of network topologies are discussed for different scenarios.

Notation:
The symbol $:=$ denotes a definition
$\|x\|$: 2-norm of a vector x
W^T: the transpose of a matrix or vector W
W^*: the conjugate transpose of a matrix or vector W
x_i or $\{x\}_i$: the i-th element of a vector x
W_{ji}: the entry (j,i) of a matrix W
W_{j*}: j-th row of a matrix W
W_{*i}: i-th column of a matrix W
x_V: when $V=(v_1,...,v_n)$ is a n-tuple of natural numbers denotes the vector $(x_{v_1}\ ...\ x_{v_n})^T$
$|A|$: cardinality (number of elements) of a set A
$E[\cdot]$: mean operator;
$R_{XY}(\tau):=E[X(t)Y^T(t+\tau)]$: cross-covariance function of wide-sense stationary vector processes X and Y;
$R_X(\tau):=R_{XX}(\tau)$: autocovariance;
$\mathcal{Z}(\cdot)$: Zeta-transform of a signal;
$\Phi_{XY}(z):=\mathcal{Z}(R_{XY}(\tau))$: cross-power spectral density;
$\Phi_X(z):=\Phi_{XX}(z)$: power spectral density;
b_i : i-th element of the canonical base of $\mathbb{R}^n$.

II. Preliminary definitions

In this section, basic notions of graph theory, which are functional to the subsequent developments, will be recalled. For an extensive overview see [18]. First, the standard definition of undirected and oriented graphs is provided.

Definition 1 (Directed and Undirected Graphs): An undirected graph G is a pair (V,A) where V is a set of vertices or nodes and A is a set of edges or arcs, which are unordered subsets of two distinct elements of V.
A directed (or oriented) graph G is a pair (V,A) where V is a set of vertices or nodes and A is a set of edges or arcs, which are ordered pairs of elements of V.
In the following, if not specified, oriented graphs are considered.

Definition 2 (Topology of a graph): Given an oriented graph $G=(V,A)$, its topology is defined as the undirected graph $G'=(V,A')$ such that $\{N_i,N_j\}\in A'$ if and only if $(N_i,N_j)\in A$ or $(N_j,N_i)\in A$, and $top(G):=G'$.
By removing the orientation on any edge of an oriented graph G, an undirected graph G' is obtained that is its topology. An example of a directed graph and its topology is represented in Figure 1.

Definition 3 (Children and Parents): Given a graph $G=(V,A)$ and a node $N_j\in V$, the children of N_j are defined as $\mathcal{C}_G(N_j):=\{N_i|(N_j,N_i)\in A\}$ and the parents of N_j as $\mathcal{P}_G(N_j):=\{N_i|(N_i,N_j)\in A\}$.
Extending the notation, children and the parents of a set of nodes are denoted as follows

$$\mathcal{C}_G(\{N_{j_1},...,N_{j_m}\}):=\cup_{k=1}^m\mathcal{C}_G(N_{j_k})$$
$$\mathcal{P}_G(\{N_{j_1},...,N_{j_m}\}):=\cup_{k=1}^m\mathcal{P}_G(N_{j_k}).$$

Definition 4 (Kins): Given an oriented graph $G=(V,A)$ and a node $N_j\in V$, kins of N_j are defined as

$$\mathcal{K}_G(N_j):=\{N_i|N_i\neq N_j \text{ and } N_i\in\mathcal{C}_G(N_j)\cup \cup\,\mathcal{P}_G(N_j)\ \cup\ \mathcal{P}_G(\mathcal{C}_G(N_j))\}.$$

Kins of a set of nodes are defined in the following way

$$\mathcal{K}_G(\{N_{j_1},...,N_{j_m}\}):=\cup_{k=1}^m\mathcal{K}_G(N_{j_k}).$$

Definition 5 (Proper Parents and Proper Children): Given an oriented graph $G=(V,A)$ and a node N_j, N_i is a proper parent (child) of N_j if it is a parent (child) of N_j and $N_i\notin\mathcal{P}_G(\mathcal{C}_G(N_j))$. N_i is a proper kin if it is a kin and $N_i\notin\mathcal{P}_G(N_j)\cup\mathcal{C}_G(N_j)$.
Note that the kin relation is symmetric, in the sense that $N_i\in\mathcal{K}_\mathcal{G}(N_j)$ if and only if $N_j\in\mathcal{K}_\mathcal{G}(N_i)$.

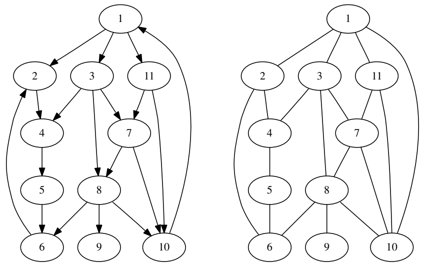
Definition 6 (Kin-graph): Given an oriented graph $G=(V,A)$, its kin-graph is the undirected graph $\tilde{G}=(V,\tilde{A})$ where

$$\tilde{A}:=\{\{N_i,N_j\}|N_i\in\mathcal{K}_G(N_j) \text{ for all } j\}.$$

and it is denoted as $kin(G)=\tilde{G}$.
A directed graph and its kin-graph are represented in Figure 2. Note that the kin-graph of G is an undirected graph. It could be defined as a directed graph, but, because of the symmetry of the kin relation, a directed graph contains exactly the same information. Moreover such a choice is motivated by the following definition

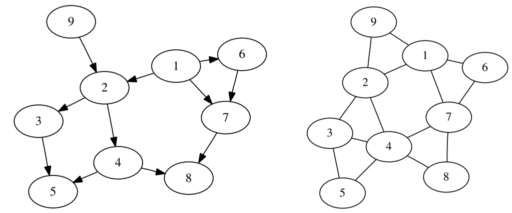
Definition 7 (Self-kin Graph): An oriented graph G is self-kin if $top(G)=kin(G)$.
Many graphs are self-kin, such as rooted trees, rings and triangular lattices [18].

Definition 8: Let $\mathcal{E}$ be a set containing time-discrete scalar, zero-mean, jointly wide-sense stationary random processes such that, for any $e_i,e_j\in\mathcal{E}$, the power spectral density $\Phi_{e_ie_j}(z)$ exists, is real rational with no poles on the unit circle and given by $\Phi_{e_ie_j}(z)=\frac{A(z)}{B(z)}$, where $A(z)$ and $B(z)$ are polynomials with real coefficients such that



(a) (b)

Fig. 1. A directed graph (a) and its topology (b).



(a) (b)

Fig. 2. An oriented graph (a) and its kin topology (b).

$B(z) \neq 0$ for any $z \in \mathbf{C}$, with $|z| = 1$. Then, $\mathcal{E}$ is a set of rationally related random processes.

Definition 9: The set $\mathcal{F}$ is defined as the set of real-rational single-input single-output (SISO) transfer functions that are analytic on the unit circle $\{z \in \mathbf{C} | \ |z| = 1\}$.

Definition 10: Given a SISO transfer function $H(z) \in \mathcal{F}$, represented as

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

the causal truncation operator is defined as

$$\{H(z)\}_C := \sum_{k=0}^{\infty} h_k z^{-k}. \tag{2}$$

Lemma 11: For every $H(z) \in \mathcal{F}$, it holds that $\{H(z)\}_C \in \mathcal{F}$.

Definition 12: The set $\mathcal{F}^+$ is defined as the set of real-rational SISO transfer functions in $\mathcal{F}$ such that

$$\{H(z)\}_C = H(z). \tag{3}$$

Definition 13: Let $\mathcal{E}$ be a set of rationally related random processes. The set $\mathcal{FE}$ is defined as

$$\mathcal{FE} := \left\{ x = \sum_{k=1}^{m} H_k(z) e_k \mid e_k \in \mathcal{E}, H_k(z) \in \mathcal{F}, m \in \mathbb{N} \right\}.$$

Lemma 14: The set $\mathcal{FE}$ is a vector space with the field of real numbers. Let

$$< x_1, x_2 >:= R_{x_1 x_2}(0) = \int_{-\pi}^{\pi} \Phi_{x_1 x_2}(e^{i\omega}),$$

which defines an inner product on $\mathcal{FE}$ with the assumption that two processes x_1 and x_2 are considered identical if $x_1(t) = x_2(t)$, almost always for any t.

Proof: The proof is left to the reader ■

For any $x \in \mathcal{FE}$, the norm induced by the inner product is defined as $\|x\| := \sqrt{< x, x >}$.

Definition 15: For a finite number of elements $x_1, ..., x_m \in \mathcal{FE}$, tf-span is defined as

$$\text{tf-span}\{x_1, ..., x_m\} := \left\{ x = \sum_{i=1}^{m} \alpha_i(z) x_i \mid \alpha_i(z) \in \mathcal{F} \right\}.$$

Definition 16: For a finite number of elements $x_1, ..., x_m \in \mathcal{FE}$, c-tf-span is defined as

$$\text{c-tf-span}\{x_1, ..., x_m\} := \left\{ x = \sum_{i=1}^{m} \alpha_i(z) x_i \mid \alpha_i(z) \in \mathcal{F}^+ \right\}.$$

Lemma 17: The tf-span and c-tf-span operators define a subspace of $\mathcal{FE}$.

Proof: The proof is left to the reader. ■

The following definition provides a class of models for a network of dynamical systems. It is assumed that the dynamics of each agent (node) in the network is represented by a scalar random process $\{x_j\}_{j=1}^n$ that is given by the superposition of a noise component e_j and the "influences" of some other "parent nodes" through dynamic links. The noise acting on each node is assumed not related with the other noise components. If a certain agent "influences" another one a directed edge can be drawn and a directed graph can be obtained.

Definition 18 (Linear Dynamic Graph): A Linear Dynamic Graph $\mathcal{G}$ is defined as a pair $(H(z), e)$ where

- $e = (e_1, .., e_n)^T$ is a vector of n rationally related random processes such that $\Phi_e(z)$ is diagonal
- $H(z)$ is a $n \times n$ matrix of transfer functions in $\mathcal{F}$ such that $H_{jj}(z) = 0$, for $j = 1, ..., n$.

The output $x := (x_1, ..., x_c)$ of the LDG is defined as

$$x(t) = e(t) + H(z)x(t). \tag{4}$$

Let $V := \{x_1, ..., x_n\}$ and let $A := \{(x_i, x_j) | H_{ji}(z) \neq 0\}$. The pair $G = (V, A)$ is the associated directed graph of the LDG. Nodes and edges of a LDG will mean nodes and edges of the graph associated with the LDG.

Observe that Equation (4) defines a map from a vector of rationally related processes x to a vector of rationally related processes e. Indeed, $e = (\mathcal{I} - H(z))x$ and each entry of $(\mathcal{I} - H(z))$ has no poles on the unit circle. If the operator $(\mathcal{I} - H(z))$ is invertible on the space of rationally related processes it can be guaranteed that, for any vector of rationally related processes e, a vector x of processes that are still rationally related will be obtained. For this reason, the following definition is introduced.

Definition 19: A LDG $(H(z), e)$ is well-posed if each entry of $(\mathcal{I} - H(z))^{-1}$ belongs to $\mathcal{F}$. Thus,

$$x = (\mathcal{I} - H(z))^{-1} e.$$

can be written. A LDG $(H(z), e)$ is causally well-posed if all the entries of $(\mathcal{I} - H(z))$ and $(\mathcal{I} - H(z))^{-1}$ belong to $\mathcal{F}^+$.

A LDG is a complex interconnection of linear transfer functions $H_{ji}(z)$ connected according to a graph G and forced by stationary additive mutually uncorrelated noise. The following definition will be useful for determining sufficient conditions for detection of links in a network.

Definition 20: A LDG $\mathcal{G} = (H(z), e)$ is topologically detectable if $\Phi_{e_i}(e^{i\omega}) > 0$ for any $\omega \in [-\pi, \pi]$ and for any $i = 1, ..., n$.

III. Problem Formulation

Problem 21: Consider a well-posed LDG $\mathcal{G} = (H, e)$ where its associated graph G is unknown. Given the Power (Cross-) Spectral Densities of $\{x_j\}_{j=1,...,n}$, reconstruct the unknown topology of G.

IV. Sparsity of the non-causal Wiener Filter

First, a lemma is provided that guarantees that any element in tf-span$\{x_i\}_{i=1,...,n}$ admits a unique representation if the cross-spectral density matrix of its generating processes has full normal rank.

Lemma 22: Let q and $x_1, ..., x_n$ be processes in the space $\mathcal{FE}$. Define $x = (x_1, ..., x_n)^T$. Suppose that $q \in$ tf-span$\{x_i\}_{i=1,...,n}$ and that $\Phi_x(e^{i\omega}) > 0$ almost for any $\omega \in [-\pi, \pi]$. Then there exists a unique transfer matrix $H(z)$ such that $q = H(z)x$.

Proof: Note that if $H(z)$ is such that $q = H(z)x = 0$, then $\Phi_{qq}(e^{i\omega}) = 0 = H(e^{i\omega})\Phi_x(e^{i\omega})H^*(e^{i\omega})$. Since $\Phi_x(e^{i\omega}) > 0$ for any $\omega \in [-\pi, \pi]$, it holds that $H(e^{i\omega}) = 0$ almost everywhere which implies that $H(z) = 0$. Now, by contradiction assume that $q = H_1(z)x = H_2(z)x$, with $H_1(z) \neq H_2(z)$. Then $0 = [H_2(e^{i\omega}) - H_1(e^{i\omega})]\Phi_x(e^{i\omega})[H_2(e^{i\omega}) - H_1(e^{i\omega})]^*$ implying that $H_1(z) = H_2(z)$. ■

A specific formulation of the non-causal Wiener filter is introduced for the defined spaces.

Proposition 23: Let v and $x_1, ..., x_n$ be processes in the space $\mathcal{FE}$. Define $x := (x_1, ..., x_n)^T$ and $X :=$ tf-span$\{x_1, ..., x_n\}$. Consider the problem

$$\inf_{q \in X} \|v - q\|^2. \tag{5}$$

If $\Phi_x(e^{i\omega}) > 0$, for $\omega \in [-\pi, \pi]$, the solution $\hat{v} \in X$ exists, is unique and with $\hat{v} = W(z)x$ where

$$W(z) = \Phi_{vx}(z)\Phi_x(z)^{-1}.$$

Moreover, $\hat{v}$ is the unique element in X such that, for any $q \in X$,

$$< v - \hat{v}, q >= 0. \tag{6}$$

Proof: The proof is left to the reader, but it can be obtained following the standard arguments reported, for example, in [19]. ■

In the following definition a notion of conditional non-causal Wiener-uncorrelation is given.

Definition 24: Let v, $x_1, ..., x_n$ be processes in the space $\mathcal{FE}$. Define $x := (x_1, ..., x_n)^T$ and $X :=$ tf-span$\{x_1, ..., x_n\}$. For any $i \in \{1, ..., n\}$, the process v is conditionally non-causally Wiener-uncorrelated with x_i given the processes $\{x_k\}_{k \neq i}$ if the i-th entry of the Wiener filter to estimate v from x is zero, that is

$$\Phi_{vx}\Phi_x^{-1} b_i = 0. \tag{7}$$

where b_i is the vector of $\mathbb{R}^n$ that has 1 as the $i - th$ entry and 0 in all other entries.

The following lemma provides an immediate relationship between non-causal Wiener-uncorrelation and the inverse of the cross-spectral density matrix. This result presents strong similarities with the property of the inverse of the covariance matrix for jointly Gaussian random-variables. Indeed, it is well-known that the entry (i, j) of inverse of the covariance matrix of n random variables $x_1, ..., x_n$ is zero if and only if x_i and x_j are conditionally independent given other variables.

Lemma 25: Let $x_1, ..., x_n$ be processes in the space $\mathcal{FE}$. Define $x = (x_1, ..., x_n)^T$. Assume that Φ_x has full normal rank. The process x_i is non-causally Wiener-uncorrelated with x_j given the processes $\{x_k\}_{k \neq i,j}$, if and only if the entry (i, j), or equivalently the entry (j, i), of $\Phi_x^{-1}(z)$ is zero, that is, for $i \neq j$,

$$b_j^T \Phi_x^{-1} b_i = b_i^T \Phi_x^{-1} b_j = 0. \tag{8}$$

Proof: Without any loss of generality, let $j = n$ and define $x_{\overline{n}} := (x_1, ..., x_{n-1})^T$. Suppose the non-causal Wiener filter estimating x_n from $x_{\overline{n}}$ is $W_{n\overline{n}}$. Then

$$x_n = \varepsilon_n + W_{n\overline{n}}(z) x_{\overline{n}} \tag{9}$$

where, from (6), the error ε_n has the property that $\Phi_{\varepsilon_n x_{\overline{n}}}(z) = 0$. Define $r := (x_{\overline{n}}^T, \varepsilon_n)^T$ and observe that

$$r = \begin{pmatrix} \mathcal{I} & 0 \\ -W_{n\overline{n}}(z) & 1 \end{pmatrix} x; \qquad x = \begin{pmatrix} \mathcal{I} & 0 \\ W_{n\overline{n}}(z) & 1 \end{pmatrix} r.$$

It follows that

$$\Phi_x^{-1} = \begin{pmatrix} \mathcal{I} & W_{n\overline{n}}(z)^* \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Phi_{x_{\overline{n}}}^{-1} & 0 \\ 0 & \Phi_{\varepsilon_n}^{-1} \end{pmatrix} \begin{pmatrix} \mathcal{I} & 0 \\ W_{n\overline{n}}(z) & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \Phi_{x_{\overline{n}}} + \frac{W_{n\overline{n}}^* W_{n\overline{n}}}{\Phi_{\varepsilon_n}} & W_{n\overline{n}}^* \Phi_{\varepsilon_n}^{-1} \\ \Phi_{\varepsilon_n}^{-1} W_{n\overline{n}} & \Phi_{\varepsilon_n}^{-1} \end{pmatrix}.$$

The assertion is proven by premultiplying by b_n^T and post-multiplying by b_i ■

The following theorem provides a sufficient condition to determine if two nodes in a LDG are kins.

Theorem 26: Consider a well-posed and topologically detectable LDG $(H(z), e)$ with associated graph G. Let $x = (x_1, ..., x_n)^T$ be its output. Define the space $X_j =$ tf-span$\{x_i\}_{i \neq j}$. Consider the problem of approximating the signal x_j with an element $\hat{x}_j \in X_j$, as defined below

$$\inf_{\hat{x}_j \in X_j} \|x_j - \hat{x}_j\|^2. \tag{10}$$

Then the optimal solution $\hat{x}_j$ exists, is unique and

$$\hat{x}_j = \sum_{i \neq j} W_{ji}(z) x_i \tag{11}$$

where $W_{ji}(z) \neq 0$ implies $\{x_i, x_j\} \in kin(G)$.

Proof: The LDG dynamics is given by $x = (\mathcal{I} - H(z))^{-1} e$ implying that

$$\Phi_x^{-1} = (\mathcal{I} - H)^* \Phi_e^{-1} (\mathcal{I} - H). \tag{12}$$

Consider the j-th row of Φ_x^{-1}. We have

$$b_j^T \Phi_x^{-1} = (b_j^T - H_{*j}^*) \Phi_e^{-1} (\mathcal{I} - H). \tag{13}$$

The $k-$th row element of the vector $(b_j^T - H_{*j}^*)$ is zero if $k \neq j$ and x_k is not a parent of x_j. Since Φ_e is diagonal the i-th column of $\Phi_e^{-1}(\mathcal{I} - H)$ has zero entries for any $k \neq i$ that is not a parent of i. Given $i \neq j$, if i is not a parent of j and i is not a child of j and i and j have no common children (they are not coparents), it follows that the entry (j, i) of $\Phi_{xx}^{-1}(z)$ is zero. Using Lemma (25) the assertion is proven. ■

The following result provides a sufficient condition for the reconstruction of a link in a LDG.

Corollary 27: Consider a well-posed and topologically detectable LDG $\mathcal{G}$ with associated graph G. Let $x = (x_1, ..., x_n)^T$ be its output. Let $W_{ji}(z)$ be the entry of the non-causal Wiener filter estimating x_j from $\{x_k\}_{k \neq j}$ corresponding to the process x_i. If $\mathcal{G}$ is self-kin, then $W_{ji}(z) \neq 0$ implies $(x_j, x_i) \in top(G)$.

Proof: Since $\mathcal{G}$ is self-kin, $\mathcal{P}_G(x_j) \cup \mathcal{C}_G(x_j) \cup \mathcal{P}_G(\mathcal{C}_G(x_j)) = \mathcal{C}_G(x_j) \cup \mathcal{P}_G((x_j))$. Thus, from the previous theorem the assertion follows immediately. ■

The following lemma is a key result to explicitly determine the expression of the Wiener filter for a LDG in the non-causal and in the causal scenarios.

Lemma 28: Consider a well-posed LDG $\mathcal{G} = (H(z), e)$ with associated graph G and output $x = (x_1, ..., x_n)^T$. Fix $j \in \{1, ..., n\}$ and define the set

$$C := \{c | x_c \in \mathcal{C}_G(x_j)\} = \{c_1, ... c_{n_c}\}$$

containing the indexes of the n_c children of x_j. Then, for $i \neq j$,

$$x_i \in \text{tf-span}\left\{ \left\{ \bigcup_{k \in C} (e_k + H_{kj}(z) e_j) \right\} \cup \left\{ \bigcup_{k \notin C \cup \{j\}} \{e_k\} \right\} \right\}.$$

Furthermore, if $\mathcal{G}$ is causal,

$$x_i \in \text{c-tf-span}\left\{ \left\{ \bigcup_{k \in C} (e_k + H_{kj}(z) e_j) \right\} \cup \left\{ \bigcup_{k \notin C \cup \{j\}} \{e_k\} \right\} \right\}.$$

Proof: Define

$$\begin{aligned}
&\varepsilon_j := 0 \\
&\varepsilon_k := e_k + H_{kj}(z)e_j && \text{if } k \in C \\
&\varepsilon_k := e_k && \text{if } k \notin \{C\} \cup \{j\} \\
&\xi_k := \sum H_{ki}(z)x_i && \text{if } k = j \\
&\xi_k := x_k && \text{if } k \neq j
\end{aligned} \quad (14)$$

and, by inspection, observe that$[\mathcal{I} - H(z)]\xi = \varepsilon$. Since $\mathcal{G}$ is well posed, $[\mathcal{I} - H(z)]$ is invertible implying that the signals $\{\xi_i\}_{i=1,...,n}$ are a linear transformation of the signals $\{\varepsilon_i\}_{i=1,...,n}$. For $i \neq j$, we have

$$x_i = \xi_i \in \text{tf-span}\{\varepsilon_k\}_{k=1,...,n} = \text{tf-span}\{\varepsilon_k\}_{k\neq j}$$

where the first equality follows from (14) and last equality follows from the fact that $\varepsilon_j = 0$. The causality of $\mathcal{G}$ also implies that $x_i = \text{c-tf-span}\{\varepsilon_k\}_{k\neq j}$. This proves the assertion. ■

V. Sparsity of causal filtering operators

First, we need to introduce the following lemma.

Lemma 29: Let $\mathcal{E}$ be a space of rationally related processes and let v and $x_1, ..., x_n$ be processes in $\mathcal{FE}$. Define $x := (x_1, ..., x_n)^T$. Assume that $\Phi_{vx}(e^{i\omega}) = 0$ for all $\omega \in [-\pi, \pi]$. Then $< v, q >= 0$ for all $q \in \text{tf-span}\{x_i\}_{i=1,...,n}$.

Proof: As $q \in \text{tf-span}\{x_i\}_{i=1,...,n}$, it follows that there exist $\alpha_i(z) \in \mathcal{F}$ such that

$$q = \sum_{i=1}^{n} \alpha_i(z)x_i =: \alpha(z)x,$$

where $\alpha(z) = (\alpha_1(z), ..., \alpha_n(z))$ is a row vector of real-rational transfer functions. Then it follows that

$$< v, q >= \int_{-\pi}^{\pi} \Phi_{vx}(e^{i\omega})\alpha(e^{i\omega})^* = 0.$$

■

Now, a specific formulation of the standard causal Wiener filter (see [19]) is introduced for the defined spaces.

Proposition 30: Let v and $x_1, ..., x_n$ be processes in the space $\mathcal{FE}$. Define $x := (x_1, ..., x_n)^T$ and $X := \text{c-tf-span}\{x_1, ..., x_n\}$. Consider the problem

$$\inf_{q \in X} \|v - q\|^2. \quad (15)$$

Let $S(z)$ be the spectral factorization of $\Phi_x(e^{i\omega}) = S(e^{i\omega})S^*(e^{i\omega})$. If $\Phi_x(e^{i\omega}) > 0$, for $\omega \in [-\pi, \pi]$, the solution $\hat{v}^{(c)} \in X$ exists, is unique and has the form

$$\hat{v}^{(c)} = W^{(c)}(z)x$$

where $W^{(c)}(z) = \{\Phi_{vx}(z)\Phi_x(z)^{-1}S(z)\}_C S^{-1}(z)$. Moreover $\hat{v}^{(c)}$ is the unique element in X such that, for any $q \in X$, satisfies

$$< v - \hat{v}^{(c)}, q >= 0. \quad (16)$$

Proof: The proof follows the standard derivation of the Wiener-Hopf filter (see [19]). ■

The following theorem proves the sparsity of the causal Wiener filter stating that the causal Wiener filter estimating x_j from the signals x_i, $i \neq j$, has non-zero entries corresponding to the kin signals of x_j.

Theorem 31: Consider a well-posed, causal and topologically detectable LDG. Let $x_1, ..., x_n \in \mathcal{FE}$ be the signals associated with the n nodes of its graph. Define $X_j = \text{c-tf-span}\{x_i\}_{i\neq j}$. Consider the problem of approximating the signal x_j with an element $\hat{x}_j \in X_j$, as defined below

$$\min_{\hat{x}_j \in X_j} \|x_j - \hat{x}_j\|^2 .$$

Then the optimal solution $\hat{x}_j$ exists, is unique and

$$\hat{x}_j = \sum_{i\neq j} W_{ji}(z)x_i$$

where $W_{ji}(z) \neq 0$ implies $(x_i, x_j) \in kin(G)$.

Proof: For any $i \neq j$, define ε_i as in (14). Also note that

$$e_j := x_j - \sum_i H_{ji}(z)x_i. \quad (17)$$

Consider $\hat{e}_j$ defined as

$$\hat{e}_j := \arg \min_{q \in \text{c-tf-span}\{\varepsilon_i\}_{i\neq j}} \|e_j - q\| = \sum_{i\neq j} C^{(c)}_{ji}(z)\varepsilon_i$$

where the transfer functions $C^{(c)}_{ji}(z)$ are given by the causal Wiener filter estimating e_j from $\{\varepsilon_i\}_{i\neq j}$. Notice that $C^{(c)}_{ji}(z)$ is equal to zero if x_i is not a child of x_j. Now, let us consider the optimization problem

$$\hat{x}_j := \arg \min_{q \in \text{c-tf-span}\{x_i\}_{i\neq j}} \|x_j - q\| = \sum_{i\neq j} W_{ji}(z)x_i$$

where $W_{ji}(z)$ are the entries of the associated causal Wiener filter. Its solution $\hat{x}_j$ satisfies

$$\begin{aligned}
\hat{x}_j &= \sum_{i\neq j} H_{ji}(z)x_i + \arg \min_{q \in \text{c-tf-span}\{x_i\}_{i\neq j}} \|e_j - q\| = \\
&= \sum_{i} H_{ji}(z)x_i + \arg \min_{q \in \text{c-tf-span}\{\varepsilon_i\}_{i\neq j}} \|e_j - q\|
\end{aligned}$$

where the first equality derives from (17) and the last one has been obtained by using Lemma 28. Thus, we have

$$\hat{x}_j = \sum_{i\neq j} W_{ji}x_i = \sum_i H_{ji}(z)x_i + \sum_{i\neq j} C_{ji}\varepsilon_i.$$

Substituting the espression of ε_i, $i \neq j$, as a function of x_i, $i \neq j$, the assertion is proven. ■

The following theorem proves the sparsity of the one step prediction operator (or Granger-causal operator). Granger-causality is a wide-spread technique in econometrics to test the causal dependence of time series. If the stronger hypothesis of strictly causal transfer functions $H_{ji}(z)$ is met, one-step predictor provides an exact reconstruction of parent-child links in a LDG.

Theorem 32: Consider a well-posed, strictly causal and topologically detectable LDG. Let $x_1, ..., x_n \in \mathcal{FE}$ be the signals associated with the n nodes of its graph. Define $X_j = \text{c-tf-span}\{x_1, ..., x_n\}$. Consider the problem of approximating the signal zx_j with an element $\hat{x}_j \in X_j$, as defined below

$$\min_{\hat{x}_j \in X_j} \|zx_j - \hat{x}_j\|^2 . \quad (18)$$

Then the optimal solution $\hat{x}_j$ exists, is unique and

$$\hat{x}_j = \sum_{i=1}^{n} W_{ji}(z) x_i \tag{19}$$

where $W_{ji}(z) \neq 0$ implies $i = j$ or x_i is a parent of x_j.

Proof: For any $i \neq j$, define ε_i as in (14). Also define $\varepsilon_j := e_j$. Note that

$$e_j := x_j - \sum_i H_{ji}(z) x_i. \tag{20}$$

Consider the minimization problem

$$\hat{e}_j := \arg \min_{q \in \text{c-tf-span}\{\varepsilon_i\}_i} \|z e_j - q\| = \sum_{i \neq j} C_{ji}^{(g)}(z) \varepsilon_i$$

where the transfer functions $C_{ji}^{(g)}(z)$ are elements of $\mathcal{F}^+$. We have that $C_{ji}^{(g)}(z) = 0$ for any $i \neq j$. Indeed, since $\Phi_{e_i e_j}(e^{i\omega}) = 0$ for $i \neq j$, it holds that

$$\arg \min_{q \in \text{c-tf-span}\{\varepsilon_i\}_{i=1}^n} \|z e_j - q\| = \arg \min_{q \in \text{c-tf-span}\{e_i\}_{i=1}^n} \|z e_j - q\| = \\ = \arg \min_{q \in \text{c-tf-span}\{e_j\}} \|z e_j - q\|.$$

Conversely, from (30), we find $C_{jj}^{(g)}(z) = \{z S_j(z)\}_C z^{-1} S_j(z)$ where $S_j(z)$ is the spectral factor of e_j. Now, let us consider the problem

$$\arg \min_{q \in \text{c-tf-span}\{x_i\}_{i \neq j}} \|z x_j - q\|.$$

Its solution $\hat{x}_j$ is

$$\begin{aligned}\hat{x}_j &= \sum_k z H_{jk}(z) x_k + \arg \min_{q \in \text{c-tf-span}\{x_i\}_i} \|z e_j - q\| = \\ &= \sum_k z H_{jk}(z) x_k + C_{jj}^{(g)}(z) e_j \\ &= C_{jj}^{(g)}(z) x_j + \sum_{k \neq j} [z H_{jk}(z) - C_{jj}^{(g)}(z) H_{jk}(z)] x_k.\end{aligned}$$

This proves the assertion. ■

VI. A Reconstruction Algorithm

The previous section provides theoretical results allowing for the reconstruction of a topology via Wiener filtering. It needs to be stressed that even in the case of sparse graphs, the reconstruction of the kinship topology can be considered a practical solution. The following algorithm is a pseudocode implementation of the reconstruction technique that was developed in the previous section.

`Reconstruction algorithm`

0. Initialize the set of edges $A = \{\}$
1. For any signal x_j
2. Determine the Optimal filter entries $W_{ji}(z)$ (non-causal Wiener, causal Wiener or one-step predictor)
3. For any $W_{ji}(z) \not\equiv 0$
4. add $\{N_i, N_j\}$ to A
5. end
7. end
8. return A

Under the assumption of ergodicity of the signals, there are a variety of techniques to perform step 2 using data. Most of them rely on estimating the spectral densities of the signal involved. For example an efficient technique based on Gram-Schmidt orthogonalization is described in [20].

VII. Conclusions

This work has illustrated a simple but effective procedure to identify the general structure of a network of linear dynamical systems. The approach followed is based on Wiener Filtering in order to detect the existing links of a network. When the topology of the original graph is described by a self-kin network, the method developed guarantees an exact reconstruction. Self-kin networks provide a non-trivial class of networks since they allow the presence of loops, nodes with multiple inputs and lack of connectivity. Moreover, the paper also provides results about general networks. It is shown that, for a general graph, the developed procedure reconstructs the topology of the smallest self-kin graph containing the original one. Thus, the method is optimal in this sense.

References

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, February 2006.

[2] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, 2004.

[3] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *Proc. of IEEE CDC*, New Orleans, 2007, pp. 5492–5498.

[4] I. Schizas, A. Ribeiro, and G. Giannakis, "Consensus in ad hoc WSNs with noisy links-part i: Distributed estimation of deterministic signals," *IEEE Trans. on Signal Processing*, vol. 56, no. 1, 2008.

[5] M. Timme, "Revealing network connectivity from response dynamics," *Phys. Rev. Lett.*, vol. 98, no. 22, p. 224101, 2007.

[6] S. Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, and A. Rapisarda, "Detecting complex network modularity by dynamical clustering," *Phys. Rev. E*, vol. 75, p. 045102, 2007.

[7] D. Napoletani and T. Sauer, "Reconstructing the topology of sparsely connected dynamical networks," *Phys. Rev. E*, vol. 77, p. 026103, 2008.

[8] E. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. Berlin, Germany: Springer-Verlag, 2009.

[9] R. Mantegna and H. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge UK: Cambridge University Press, 2000.

[10] M. Ozer and M. Uzuntarla, "Effects of the network structure and coupling strength on the noise-induced response delay of a neuronal network," *Physics Letters A*, vol. 375, pp. 4603–4609, 2008.

[11] G. Innocenti and D. Materassi, "A modeling approach to multivariate analysis and clusterization theory," *Journal of Physics A*, vol. 41, no. 20, p. 205101, 2008.

[12] D. Materassi and G. Innocenti, "Topological identification in networks of dynamical systems," *IEEE Trans. Aut. Control*, vol. 55, no. 8, pp. 1860–1871, August 2010.

[13] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Phys. Rev. Lett.*, vol. 100, p. 144103, 2008.

[14] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.

[15] L. Getoor, N. Friedman, B. Taskar, and D. Koller, "Learning probabilistic models of relational structure," *Journal of Machine Learning Research*, vol. 3, pp. 679–707, 2002.

[16] N. Friedman and D. Koller, "Being bayesian about network structure: A bayesian approach to structure discovery in bayesian networks," *Machine Learning*, vol. 50, pp. 95–126, 2003.

[17] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[18] R. Diestel, *Graph Theory*. Berlin, Germany: Springer-Verlag, 2006.

[19] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2000.

[20] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Transactions on Information Theory*, vol. 44, pp. 2943–2959, 1998.