2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Relations between structure and estimators in networks of dynamical systems

Donatello Materassi Murti V. Salapaka Laura Giarrè

***Abstract*— The article main focus is on the identification of a graphical model from time series data associated with different interconnected entities. The time series are modeled as realizations of stochastic processes (representing nodes of a graph) linked together via transfer functions (representing the edges of the graph). Both the cases of non-causal and causal links are considered. By using only the measurements of the node outputs and without assuming any prior knowledge of the network topology, a method is provided to estimate the graph connectivity. In particular, it is proven that the method determines links to be present only between a node and its "kins", where kins of a node consist of parents, children and co-parents (other parents of all of its children) in the graph. With the additional hypothesis of strictly casual links, a similar method is provided that allows one to exactly reconstruct the original graph. Main tools for determining the network topology are based on Wiener, Wiener-Hopf and Granger filtering. Analogies with the problem of Compressing Sensing are drawn and two greedy algorithms to address the problem of reducing the complexity of the network structure are also suggested.**

I. INTRODUCTION

In many diverse areas, determining cause-effect relationships among various entities in a network is of significant interest. Interconnections of simple systems are used to understand the emergence of complicated phenomena (see, for example, [1]) and have provided novel modeling approaches in many fields, such as Economics (see e.g. [2]), Biology (see e.g. [3]), Cognitive Sciences (see e.g. [4]), Ecology (see e.g. [5]) and Geology (see e.g. [6]), especially when the investigated phenomena are characterized by spatial distributions where a multivariate analysis is involved.

Given the widespread interest in the problem of unraveling the interconnectedness of complex networks, the necessity for general tools has been rapidly increasing (see [7] and [8] and the bibliography therein for recent results). Indeed, even though there is considerable work in this area (see [7], [8], [9], [10]), deriving information about a network topology remains a formidable task and such a goal poses many theoretical and practical challenges [11].

Most techniques offer methods to identify a network structure that rely only on heuristic considerations. Theoretical guarantees about the correctness of the reconstruction are usually not provided. For example, in [7] 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 this paper we address the problem of reconstructing a network of dynamical systems using only passive observations. Since we are also aiming to obtain precise theoretical guarantees and characterizations about the reconstructed network, it will be necessary to define a class of network models from which the data have been generated (a *generative class of models*). The class of models that we will consider (Linear Dynamic Graphs) is quite general since it takes into account the presence of loops and of multiple signals influencing the same one. Conditions derived for the detection of links are based on "sparsity" and "local" properties of the estimators defined on the network signals. Indeed, from a different perspective, another important contribution of the paper is given by providing conditions for a local and distributed implementation of the mean square estimators. The results obtained bear a striking similarity to the ones developed in the area of machine learning for Bayesian Networks (BNs) [12] where the topology of a network of nodes that represent random variables is sought. The main result obtained in the BNs literature (see [13]) 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, though, 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. The paper is organized as follows. We start introducing a specific class of networks named Linear Dynamic Graphs and recalling basic notions of mean least square estimation. Then, we state the main objective of the paper: determining the relations between the topology of a Linear Dynamic Graph and the algebraic structure of mean least estimators defined on such a network. In Section III we show that these relations translate in precisely characterized properties of locality and sparsity of variations of the Wiener filter. In Section IV the locality and sparsity properties are used to define an algorithm capable of inferring (and in some case exactly determining) the connectivity structure of an unknown network. The sparsity of the mean least square estimators motivates the discussion developed in Section V where we draw a comparison between the problem of reconstructing a network and the problem of Compressive Sensing. The formal analogies between the two problems can be exploited in order to use Compressive Sensing techniques in the identification of the structure of a network. While we consider only two greedy technique, we stress that the analogies of the two problem are a promising starting point for establishing deeper connections.

NOTATION

- $E[\cdot]$: expectation operator
- $\mathcal{Z}\{\cdot\}$: z-transform operator
- $\Phi_{xy}(z)$: cross power spectral density of two jointly stationary stochastic vectors $\Phi_{xy}(z) := \mathcal{Z}\{E[x(0)y^T(\tau)]\}$
- $\Phi_x(z) := \Phi_{xx}(z)$
- $b_j := (0, 0...0, 1, 0, ..., 0)^T$, vector with j-th element equal to 1 and others equal to zero.

Donatello Materassi is with Laboratory for Informations and Decision Systems, Massachusetts Institute of Technology

Murti Salapaka is with Department of Electrical and Computer Engineering, University of Minnesota.

Laura Giarrè is with Dipartimento di Ingegneria Elettrica, Elettronica e delle Telecomunicazioni, Università di Palermo.

978-1-61284-801-3/11/$26.00 ©2011 IEEE

- H_{ji}: entry (j, i) of the matrix H
- H_{j*}: j-th row of the matrix H
- H_{*i}: i-th column of the matrix H

II. Preliminary definitions and notions

In this section we intend to provide the basic intuition behind the underlying generative class of models that we will study and at the same time to recall fundamental results in linear estimation. We begin with definitions about operators described by rational transfer functions.

Definition 1: 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 $\{z \in \mathbf{C} | \ |z| = 1\}$.
What follows is a standard notion of causality.

Definition 2: Given a SISO transfer function $H(z) \in \mathcal{F}$, represented as $H(z) = \sum_{k=-\infty}^{\infty} h_k z^{-k}$, the causal truncation operator is defined as

$$\{H(z)\}_C := \sum_{k=0}^{\infty} h_k z^{-k}. \tag{1}$$

A transfer function $H(z)$ is causal if $\{H(z)\}_C = H(z)$. The set $\mathcal{F}^+$ is defined as the set of causal real-rational SISO transfer functions in $\mathcal{F}$.
The following definition provides us with two vector spaces of jointly stationary stochastic processes.

Definition 3: For a finite number of wide-sense jointly stationary processes $x_1, ..., x_n$, their tfspan is defined as

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

Analogously, their $\mathrm{ctfspan}$ is defined as

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

An inner product can be defined for $\mathrm{tfspan}\{x_1, ..., x_m\}$ (and $\mathrm{ctfspan}\{x_1, ..., x_m\}$).

Definition 4: For two elements $a, b \in \mathrm{tfspan}\{x_1, ..., x_m\}$, we define the operation $< a, b >= E[a(0)b(0)]$.
We leave to the reader to prove that $< \cdot, \cdot >$ is actually an inner product on the vector space $\mathrm{tfspan}\{x_1, ..., x_m\}$, making it a pre-Hilbert space (with the technical assumption that two processes are the same if they are equal almost always). Also observe that $\mathrm{ctfspan} x_1, ..., x_n$ is pre-Hilbert space as well, since it is a subspace of $\mathrm{tfspan}\{x_1, ..., x_m\}$.

Definition 5: We define the norm induced by the inner product in $\mathrm{tfspan}\{x_1, ..., x_m\}$ as $\|a\| :=< a, a >$.

The generative class of models: Linear Dynamic Graphs

In order to obtain theoretical guarantees about the reconstruction of a network of agents, we are going to define a class of models for the network. We provide the motivations behind our assumptions. In many applications data are collected in the form of time series where each time series represents a distinct agent. We consider a network where each of these agents is observed through a scalar output that is modeled as the realization of a stochastic process x_j, for $j = 1, ..., n$. In the absence of the processes x_i, $i \neq j$, the processes x_j is assumed to have an "independent behavior". This "independent behavior is described by the stochastic process e_j. At the same time, x_j can also be linearly and additively "influenced" by none, one or more of the processes x_i, $i \neq j$. By this we mean that, by removing from x_j its "independent behavior" e_j, we find a superposition of influences determined by the processes x_i, $i \neq j$, through some dynamical transfer functions $H_{ji}(z)$, with $i \neq j$,

$$x_j = e_j + \sum_{i \neq j} H_{ji}(z) x_i. \tag{2}$$

A block diagram representation of Equation (2) is reported in Figure 1. This basic intuition leads to a more formal

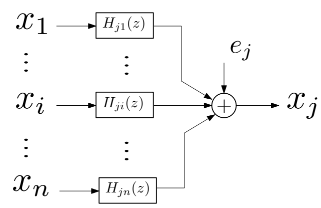


Fig. 1. A block diagram representation of Equation 2.

definition for Linear Dynamic Graph (LDG).

Definition 6: Consider a vector of n zero-mean, time-discrete, wide sense jointly stationary processes $e = (e_1, ..., e_n)^T$ with diagonal power spectral density $\Phi_e(z)$ that is strictly positive on the unit circle $|z| = 1$. Let $H(z) \in \mathcal{F}^{n \times n}$ be a rational transfer matrix, with $H_{jj}(z) = 0$, for $j = 1, ..., n$. The pair $\mathcal{G} = (H(z), e)$ is a *Linear Dynamic Graph* (LDG). The vector process $x = (x_1, ..., x_n)^T$ satisfying the dynamic relation

$$x = e + H(z)x \tag{3}$$

is the LDG *observation* or *output*. The LDG is *well-posed* if the operator $(\mathcal{I} - H(z))^{-1}$ is invertible, making the definition of the LDG observation x meaningful. A LDG is *causal* if each entry of $H(z)$ and each entry of $(\mathcal{I} - H(z))^{-1}$ is causal.

Observe that Equation (3) is an equivalent vector notation for Equation (2) describing the dynamics of all the processes x_j, $j = 1, ..., n$, in a compact manner.

Given a LDG it is possible to associate a graph G to it in the following way. Let $G = (V, E)$ be a graph where V is the set of nodes and E is the set edges. Define $V := \{x_1, ..., x_n\}$ and $E := \{(x_i, x_j) | H_{ji}(z) \neq 0\}$. Thus, each observation process x_j is a node in this graphical model, while a link connection (x_i, x_j) is determined by a non-identically zero transfer function $H_{ji}(z)$. Observe that the matrix $H(z)$ plays the role of an adjacency matrix, where entries are not limited to be just either 0 or 1, but can be rational transfer functions. The following definition establishes "kinship relations" among the nodes of a LDG.

Definition 7: Given a LDG $(H(z), e)$ with observation $x_1, ..., x_n$, we say that

- x_i is a parent of x_j if $H_{ji}(z) \neq 0$
- x_i is a child of x_j if $H_{ij}(z) \neq 0$
- x_i and x_j are co-parents if they have a common child

Furthermore, in each of these cases, we say that x_i and x_j are kins.

Estimation for jointly stationary processes

In this section we recall basic notion of linear estimation for wide-sense jointly stationary processes with specific attention to Wiener filtering.

Definition 8: Let $x = (x_1, ..., x_n)^T$ be a vector of wide-sense jointly stationary processes with $\Phi_x(z)$ real-rational and positive definite on $|z| = 1$. Consider the problem

$$\hat{x}_j = \arg \min_{q \in X_j} \|x_j - q\|^2. \tag{4}$$

where X_j is a vector space of stochastic processes that are wide-sense jointly stationary processes with x_j. According to different choices for X_j we have the following definitions

- $X_j = \text{tfspan}\{x_i\}_{i \neq j}$
  Let $\hat{x}_j := \sum_{i \neq j} W_{ji}(z) x_i$. The transfer functions $W_{ji}(z)$ are the entries of the **non-causal Wiener filter** estimating x_j from $\{x_i\}_{i \neq j}$
- $X_j = \text{ctfspan}\{x_i\}_{i \neq j}$
  Let $\hat{x}_j := \sum_{i \neq j} W_{ji}^{(c)}(z) x_i$. The transfer functions $W_{ji}^{(c)}(z)$ are the entries of the **causal Wiener filter** estimating x_j from $\{x_i\}_{i \neq j}$
- $X_j = \text{ctfspan}\{z^{-1} x_i\}$
  Let $\hat{x}_j := \sum_{i \neq j} W_{ji}^{(g)}(z) x_i$. The transfer functions $W_{ji}^{(g)}(z)$ are the entries of the **one-step predictor** (or Granger filter) estimating x_j from $\{x_i\}_{i=1,...,n}$

In each case of the previous definition, it can be proven that if $\{x_1, ..., x_n\}$ are the output of a LDG, the solution to (4) exists and is unique.

III. Estimators in Linear Dynamic Graphs

Given a LDG $(H(z), e)$ the structure of which is unknown, we intend to investigate what properties of its topology can be inferred just by observing its output signal.

In this section we describe certain properties that hold for mean least square estimators on the observation of an LDG. This properties are directly connected to the structure of the graph associated with the LDG, thus they can be used in order to infer the underlying connectivity of the network. Indeed, all the estimators that we have introduced can be computed just by the knowledge of second order statistics of the LDG output $x_1, ..., x_n$.

Wiener filtering

First we prove that, for a LDG, the Wiener filter estimating x_j using the other signals $\{x_i\}_{i \neq j}$ only requires the signals that are kins of x_j.

The definition of conditional non-causal Wiener-uncorrelation is given.

Definition 9: Let $x_1, ..., x_n$ be wide-sense jointly stationary processes. Define $X_j :=$ tf-span$\{x_k\}_{k \neq j}$. The process x_j is conditionally non-causally Wiener-uncorrelated with x_i given the processes $\{x_k\}_{k \neq i,j}$ if the estimate

$$\hat{x}_j = \arg \min_{q \in X_j} \|x_j - q\| = \sum_{k \neq j} W_{jk}(z) x_k \tag{5}$$

does not depend on x_i, that is $W_{ji}(z) = 0$.

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 $a_1, ..., a_n$ is zero if and only if a_i and a_j are conditionally independent given the other variables.

Lemma 10: Let $x = (x_1, ..., x_n)^T$ be a vector of wide-sense jointly stationary processes. Assume that $\Phi_x(z)$ has full normal rank. The process x_j is non-causally Wiener-uncorrelated with x_i 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 $i \neq j$, $b_j^T \Phi_x^{-1} b_i = b_i^T \Phi_x^{-1} b_j = 0$.

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

$$x_n = \varepsilon_n + W_{n\overline{n}}(z) x_{\overline{n}} \tag{6}$$

where, from the Hilbert projection theorem [14], 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

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

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 11: Consider a well-posed LDG $(H(z), e)$. 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

$$\hat{x}_j = \arg \min_{q \in X_j} \|x_j - q\|^2 . \tag{7}$$

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: The LDG dynamics is given by $x = (\mathcal{I} - H(z))^{-1} e$ implying that $\Phi_x^{-1} = (\mathcal{I} - H)^* \Phi_e^{-1} (\mathcal{I} - H)$. Consider the j-th row of Φ_x^{-1}. We have $b_j^T \Phi_x^{-1} = (b_j^T - H_{*j}^*) \Phi_e^{-1} (\mathcal{I} - H)$. 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 co-parents), it follows that the entry (j, i) of $\Phi_{xx}^{-1}(z)$ is zero. Using Lemma (10) the assertion is proven. ■

Causal Wiener filtering (Wiener-Hopf filtering)

The same property of sparsity of the Wiener filter in a LDG provided by Theorem 11 holds for the causal Wiener filter as well. The proof is not as straightforward. We first need to introduce a lemma.

Lemma 12: 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,

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

Proof: Define

$$\begin{aligned}
&\varepsilon_j := 0 \\
&\varepsilon_k := e_k + H_{kj}(z)e_j && \text{if } k \in C \\
&\varepsilon_k := e_k && \text{if } k \notin \{C\} \cup \{j\} \\
&\xi_k := \sum H_{ki}(z)x_i && \text{if } k = j \\
&\xi_k := x_k && \text{if } k \neq j
\end{aligned} \tag{8}$$

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 $x_i = \xi_i \in$ tf-span$\{\varepsilon_k\}_{k=1,...,n} =$ tf-span$\{\varepsilon_k\}_{k\neq j}$. where the first equality follows from (8) and the last one follows from the fact that $\varepsilon_j = 0$. The causality of $\mathcal{G}$ also implies that $x_i =$ ctfspan$\{\varepsilon_k\}_{k\neq j}$. This proves the assertion. ■

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 13: Consider a well-posed and causal LDG with observation $x = (x_1, ..., x_n)^T$ Define $X_j = \text{ctfspan}\{x_i\}_{i\neq j}$. 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 $\hat{x}_j$ is $\hat{x}_j = \sum_{i\neq j} W_{ji}^{(c)}(z)x_i$, where $W_{ji}^{(c)}(z) \neq 0$ implies $(x_i, x_j) \in kin(G)$.

Proof: For any $i \neq j$, define ε_i as in (8) and observe that ε_i can be represented as

$$\varepsilon_i = e_i + H_{ij}(z)e_j. \tag{9}$$

Also note that

$$e_j := x_j - \sum_i H_{ji}(z)x_i. \tag{10}$$

Consider $\hat{e}_j$ defined as

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

where the transfer functions $C_{ji}^{(c)}(z)$ are given by the causal Wiener filter estimating e_j from $\{\varepsilon_i\}_{i\neq j}$. Notice that, by (9), $C_{ji}^{(c)}(z)$ is equal to zero if x_i is not a child of x_j. Now, let us consider the optimization problem

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

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

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

where the first equality derives from (10) and the last one is obtained by using Lemma 12. 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 expression of ε_i, $i \neq j$, as a function of x_i, $i \neq j$, the assertion is proven. ■

One-step predictor (or Granger-causality)

The following theorem proves the sparsity of the one-step predictor (or Granger operator) [15]. If the stronger hypothesis of strictly causal transfer functions $H_{ji}(z)$ is met, the one-step predictor provides an exact reconstruction of parent-child links in a LDG.

Theorem 14: Consider a well-posed, causal detectable LDG $(H(z), e)$ with output $(x_1, ..., x_n)^T$. Assume that each entry of $H(z)$ is strictly causal. Define $X_j =$ ctfspan$\{x_1, ..., x_n\}$. Consider the problem of approximating the signal zx_j with an element $\hat{x}_j \in X_j$, as defined below

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

Then the optimal solution $\hat{x}_j$ is

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

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

Proof: For any $i \neq j$, define ε_i as in (8) and observe that ε_i can be represented as in (9). Also define $\varepsilon_j := e_j$. Note that

$$e_j := x_j - \sum_i H_{ji}(z)x_i. \tag{11}$$

Consider the minimization problem

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

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

$$\begin{aligned}
&\arg \min_{q \in \text{ctfspan}\{\varepsilon_i\}_{i=1}^n} \|ze_j - q\| = \\
&= \arg \min_{q \in \text{ctfspan}\{e_i\}_{i=1}^n} \|ze_j - q\| = \\
&= \arg \min_{q \in \text{ctfspan}\{e_j\}} \|ze_j - q\|.
\end{aligned}$$

Conversely, from the expression of the causal Wiener filter [16], we find $C_{jj}^{(g)}(z) = \{zS_j(z)\}_C z^{-1}S_j(z)$ where $S_j(z)$ is the spectral factor of e_j. Now, let us consider the problem

$$\arg \min_{q \in \text{ctfspan}\{x_i\}_{i\neq j}} \|zx_j - q\|.$$

Its solution $\hat{x}_j$ is

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

This proves the assertion. ■

IV. A RECONSTRUCTION ALGORITHM

The results of the previous section can be briefly summarized in three main points

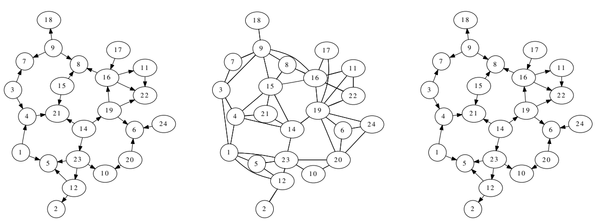
- for a LDG, the standard Wiener filter estimating a signal x_j from the other signals $\{x_i\}_{i\neq j}$ only requires the signals that are kins of x_j
- for a causal LDG the analogous property holds also for the causal implementation of the Wiener filter (making it possible, for example, to track slow varying networks in real-time)
- if the LDG is defined by strictly causal links, the Granger filter that estimates x_j only requires its parents.

Since all these three mean least estimators can be computed just by using second order statistics in the form of (cross)-power spectral densities (see [16], [17]), an algorithm to infer the structure of an unknown network is now suggested.

`Reconstruction algorithm`

Initialize the set of edges $A = \{\}$
Determine the Power Spectral Density $\Phi_x(z)$ from the realization of $\{x_1, ..., x_n\}$
For any signal x_j
    Determine the optimal filter entries $W_{ji}^{(opt)}(z)$ (non-causal Wiener, causal Wiener or one-step predictor) from the Power Spectral Densities
    For any $W_{ji}^{(opt)}(z) \not\cong 0$
        add $\{N_i, N_j\}$ to A
    end
end
return A

If $W_{ji}(z)^{(opt)}$ is given by the Wiener filter of the Wiener-Hopf filter, we can not guarantee an exact reconstruction of the network, but only the reconstruction of the so-called Markov Blanket of the network (that is the network defined by the links between kins of the original network) [13]. Furthermore, the link orientation can not be detected. In the case of the one-step predictor (and if the links are strictly causal), the network is reconstructed exactly and the link orientation is determined, as well. The situation is represented by an example in Figure 2



(a) (b) (c)

Fig. 2. Original generative network of 24 nodes (a); the reconstructed topology using non-causal and causal Wiener filters (b); and the reconstructed topology using the one-step predictor (c). In (b) every single link in the original network is detected, but the topology also contains the additional links between the "kins". In (c) the reconstruction is exact (with detection of the link orientation) but this technique can be successfully applied only when the dynamic links of the network are strictly causal.

V. Connections with Compressive Sensing

In the previous sections we have highlighted certain sparsity properties of linear least square estimators in LDGs. These sparsity properties hold for the solution of least square optimization problems where a sparse solution would not be expected. It is possible to use them in order to infer structural characteristics of the LDG introducing a simple algorithm that checks non-zero entries of a suitable transfer vector. This observation constitutes the main motivation of this section: *since in a LDG the connectivity structure is associated with the sparsity of a certain transfer vector, given an intricate network, the most important connections can be inferred by forcing a certain degree of sparsity in the computation of a least square estimator*.

In the recent few years, sparsity problems have attracted the attention of researchers in the area of Signal Processing. Applications are numerous, ranging from data-compression to high-resolution interpolation, and noise filtering [18], [19]. There are many formalizations of the problem, but one of the most common is to cast it as

$$\min_w \|x_0 - \Psi w\|_2 \quad \text{subject to} \quad \|w\|_0 \le m \ , \tag{12}$$

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

$$\|w\|_0 := |\{i \in \mathbb{N} | w_i \neq 0\}| \tag{13}$$

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 $j = 1, ..., n$ define the following sets

$$\mathcal{W}^{(j)} = \{W(z) \in \mathcal{F}^{1\times n} | W_j(z) = 0\} \ , \tag{14}$$

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

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

Then, for example a Wiener filtering problem with sparsity enforced by a node-dependent parameter m_j can be formally cast as

$$\min_{W\in\mathcal{W}_j} \|x_j - Wx\|^2 \quad \text{subject to} \quad \|W\|_0 \le m_j \tag{15}$$

which is, from a formal point of view, equivalent to the standard l_0 problem as defined in (12).

This formal equivalence shows how the problem of determining a suitable simplified 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 (12), the problem of a topology reconstruction cast as in (15) still has its own characteristics. The significant difference between (12) and (15) is that in (15) we are looking for the sparsity of a transfer vector instead of real vector. Since the projection procedure in (15) is given by the estimation of a transfer vector, it is computationally more expensive than a standard projection in the space of vectors of real numbers. For this reason greedy algorithms offer a good approach to tackle such a problem where speed could become a relevant factor. Moreover, since the complexity of the network model is here one of final goal, 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. As an alternative approach to greedy algorithms we also describe a strategy based on iterated reweighed optimizations as described in [20].

A modified Orthogonal Least Squares (Cycling OLS)

Orthogonal Least Squares (OLS) is a greedy algorithm proposed for the first time in [21] and in many ways it resembles the algorithm of Matching Pursuit developed in [22]. It basically consists of iterated orthogonal projections on elements of a (possibly redundant) base to approximate a given vector. For the details of this algorithm we remand the reader to [21]. For the sake of clarity, we simply reformulate in the terms our problem. Given the node signal x_j, we intend to approximate it with at most m_j signals chosen from $\{x_i\}_{i\neq j}$. At the $l-$th iteration $\Gamma^{(l)}$ is the set of the chosen signals. The initialization occurs defining $\Gamma^{(0)} = \emptyset$. At the $l-$th iteration step, a new element from $\{x_i\}_{i\neq j}$ is added to $\Gamma^{(l)}$ with respect to $\Gamma^{(l-1)}$ in order to achieve the largest reduction of the cost function

$$\Gamma^{(l)} = \Gamma^{(l-1)} \cup \left\{ \arg\min_{x_i \neq x_j} \left[\min_{q \in \text{tfspan}\{\Gamma^{(l-1)} \cup x_i\}} \|x_j - q\| \right] \right\}.$$

The standard OLS goes on at every step introducing a new vector until a stopping condition is met (usually if a degree of approximation for x_j is achieved or on the number of iterations).

We propose an algorithm which derives directly from OLS but it doesn't increase the number of vectors approximating x_j above

m_j. Our variation of OLS, named Cycling OLS (COLS), 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 pseudo-code.

`Cycling Orthogonal Least Squares:`

0. define $x_0 := 0$ (null time series) and $c = 0$.
1. initialize the m_j-tuple $S = (x_0, x_0..., x_0)$ and $k = 1$
2. while $c \le m_j$
   - 2a. for $i = 1, ..., n$, $i \neq j$
     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 best approximation of x_j using the signals in S_i
   - 2b. $\alpha = \arg\min_i \|x_j - \hat{x}_j^{(i)}\|$
   - 2c. if $x_\alpha = S[k]$ then $c = c + 1$
   - 2d. else $S[k] = x_\alpha$, $c = 1$,
   - 2d. $k = k \bmod m_j$, $k = k + 1$
3. return S

The reason of our modification is simple. 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.

Reweighted least squares

Another possible approach to "encourage" sparse solutions is provided by reweighed minimization algorithms as proposed in [20] and [18]. A comparison between reweighed norm-1 and norm-2 methods is performed in [19]. We consider only reweighed least squares, because such an algorithm is easier to implement, but the intuition behind the two techniques is basically the same.
Consider the optimization problem

$$\min_{W \in \mathcal{W}_j} \|x_j - Wx\|^2 \quad \text{subject to} \quad \|W\|_\mu^2 \le 1 \qquad (16)$$

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

$$\|W\|_\mu^2 := \sum_1^n \int_{-\pi}^{\pi} \mu_k W_{j,k}^*(\omega) W_{j,k}(\omega) d\omega \le m_j.$$

Let us pretend that the non-zero entries $\alpha_{j,k}$'s of the optimal $W_{\alpha_{j,k}}$ solving (15) are known. We could set

$$\mu_l := \left(\int_{-\pi}^{\pi} W_l^*(\omega) W_l(\omega) d\omega \right)^{-1}, \qquad (17)$$

if $l = \alpha_{j,k}$ for some $k = 1, ..., m_j$ and $\mu_l = +\infty$ otherwise. With such a choice of weights, the two problems (15) and (16) would be equivalent since they would provide the same solutions. However, Problem (16) 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 (17). An iterative approach to estimate the weights (17) has been proposed in [18].

`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
     $$\min_{W \in \mathcal{W}_j} \|x_j - Wx\|^2 \quad \text{subject to} \quad \|W_j\|_\mu^2 \le 1$$
   - 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 on the number of iterations), the final solution can be obtained by selecting the m_j largest entries of each W_j.

VI. Conclusions

In this paper we have introduced a class of models, Linear Dynamic Graphs, that is apt to describe a network of linear interconnected systems. Least mean estimators such as Wiener, Wiener-Hopf, and Granger filters have locality and sparsity properties that are directly connected to the topology of Linear Dynamic Graphs. Such properties are used to develop algorithms capable of inferring the underlying connectivity structure of a Linear Dynamic Graphs. The main advantage is that only second order statistics only are required for the estimate. The sparsity properties of the transfer vectors defining the least mean estimators considered in the article allow to draw analogies and comparisons between Compressive Sensing and the problem of reconstructing a network.

References

[1] J. Liu, V. Yadav, H. Sehgal, J. M. Olson, H. Liu, and N. Elia, "Phase transitions on fixed connected graphs and random graphs in the presence of noise," *IEEE Transactions on Automatic Control*, vol. 53, p. 1817, 2008.

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

[3] M. Eisen, P. Spellman, P. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proc. Natl. Acad. Sci. USA*, vol. 95, no. 25, pp. 14 863–8, 1998.

[4] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality." *Proc Natl Acad Sci USA*, vol. 101, no. 26, pp. 9849–9854, June 2004.

[5] D. Urban and T. Keitt, "Landscape connectivity: A graph-theoretic perspective," *Ecology*, vol. 82, no. 5, pp. 1205–1218, 2001.

[6] J.-S. Bailly, P. Monestiez, and P. Lagacherie, "Modelling spatial variability along drainage networks with geostatistics," *Mathematical Geology*, vol. 38, no. 5, pp. 515–539, 2006.

[7] 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.

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

[9] 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.

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

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

[12] 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.

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

[14] D. G. Luenberger, *Optimization by vector space methods*. Hoboken, New Jersey: John Wiley & Sons Inc., 1969.

[15] C. Quinn, T. Coleman, N. Kiyavash, and N. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *Journal of Computational Neuroscience*, pp. 1–28, 2010.

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

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

[18] I. Daubechies, R. DeVore, M. Fornasier, C. S&idot *et al.*, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2009.

[19] D. P. Wipf and S. Nagarajan, "Iterative reweighted l1 and l2 methods for finding sparse solutions," in *Signal Processing with Adaptive Sparse Structured Representations*, Rennes, France, April 2009.

[20] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[21] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Intl. J. Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[22] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, 1993.