2013 American Control Conference (ACC)
Washington, DC, USA, June 17-19, 2013

Reconstruction of directed acyclic networks of dynamical systems

Donatello Materassi Murti V. Salapaka

Abstract— **Determining the relation structure of various interconnected entities from multiple time series data is of significant interest to many areas. Knowledge of such a structure can aid in identifying cause and effect relationships, clustering of similar entities, detecting representative elements for an aggregate and determining reduced order models. Current methods tend to treat observations in a static manner by modeling the measured time series as repeated realizations of as many random variables that are independent over time. This amounts to assume static relationships among the measurements, making these techniques ill-suited for detecting propagative and dynamic phenomena that can be fundamental for the understanding of the system. In this paper we extend techniques for the identification of networks of random variables connected through static relations to the case of random processes with dynamic relations. This is achieved by showing that the Wiener filter defines a relationship among jointly stationary stochastic processes that has the properties of a semi-graphoid.**

I. Introduction

The adoption of networks as a modeling tool has become ubiquitous in science. Interconnections of simple systems are commonly used to explain and describe complicated phenomena. We find examples in many fields, such as Economics (see e.g. [1], [2]), Sociology (see e.g. [3]) Biology (see e.g. [4], [5], [6]), Cognitive Sciences (see e.g. [7]), and Geology (see e.g. [8], [9]). Indeed, when a complex system is observed, multiple measurements can usually be obtained, where each measurement represents the behavior of an individual unit of the whole system. Since the various components of a complex systems are often related to each other through dynamic relations, the resulting observations might show asynchronous correlation and dependencies. The literature on graphical models is extensive, but it is principally focused on random variables inteconnected through static relations. Fundamental work in this area has been pioneered by Judea Pearl and his group (see [10], [11], [12], [13]) and by many other researchers (see [14], [15], [16]). However, an approach that is specifically targeted to stochastic processes inteconnected through dynamic relations (in other words considering dependencies occurring at different time instants) has not been fully developed yet. Indeed, when dealing with the problem of reconstructing networks of stochastic processes, the dynamic scenario poses several challenges. Compared to random variables, the amount of data required to obtain information about joint probability distributions for stochastic processes is prohibitive even for small networks because of the additional "time dimension". Dependencies at different time instants have to be identified, limiting the applicability of non-parametric bayesian methods. Also, compared to a scenario where the random variables are connected through static functions, the presence of a "time dimension" makes it meaningful to consider structures with cycles: the well-posedness of a system is guaranteed if, for example, there is positive delay in each loop. Thus, not only more data is needed in order to accurately estimate joint probabilities, but also the class of structures to identify is significantly larger since it comprises models with feedback loops. In addition, the potential presence of cycles in the structure leads to more complicated probabilistic dependencies that need to be taken into account.

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

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

These challenges are leading to new results and techniques which are rapidly emerging (see [17], [18], [19], [20], [21], [22]). The motivations behind the approach followed in this paper stem primarily from the results developed in [22]. In particular, the results in [22] show striking similarities with results already developed in the area of machine learning for Bayesian Networks (BNs). One of the main results obtained in the BNs literature (see [10]) is that the probability distribution of a random variable conditioned on the rest of the random variables of an acyclic network is equal to the probability distribution of the random variable conditioned only on the random variables within the so-called "Markov Blanket" of the random variable. Given an oriented graph the Markov Blanket of a node is the set of the "parents", "children" and "parents of the children" of the node. The work in [22] shows that in order to estimate or predict a node signal in a network only the signals within the Markov Blanket of the node are required, even when cycles are present. This property allows one to reconstruct the Markov Blanket of each node providing an estimated structure for the system topology derived using only the second order statistics.

Given these similarities, the main goal of this paper is to formulate the problem of reconstructing a network of stochastic processes on the ground of general theoretical foundations.

To this aim, we will rely on the mathematical framework provided by the theory of "semi-graphoids" developed by Judea Pearl [23]. We will translate familiar concepts of Control Theory, such as Wiener filtering, in the language of semi-graphoids. This endevour will bring several positive outcomes: the results, mainly on acyclic networks, already obtained for semi-graphoids in the Artificial Intelligence community will seamlessly extend to the case of networks of stochastic processes; the Control Theory community will be exposed to new concepts that can be readily applied to stochastic systems; finally, it will motivate research towards new significant directions, such as the development of network identification/reconstruction techniques where feedback loops are present.

The paper is organized as follows. TODO: In Section II we describe a class of models that will suite our purposes of describing an interconnected structure of dynamical systems. In Section III we will recall some basic concepts of Wiener filtering. In Section IV we will show that the projections obtained using Wiener filters induce a semigraphoid structure on a space of stochastic processes. In Section V ans Section VI we will use the semigraphoid properties to obtain reconstruction results on sets of stochastic processes.

Notation:

The symbol := denotes a definition
$< \cdot, \cdot >$: inner product $\|x\|$: 2-norm of a vector x
A^T: the transpose of a matrix or vector 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;
$\mathcal{Z}(\cdot)$: Z-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;

II. A CLASS OF MODELS FOR DYNAMIC NETWORKS

In this section we describe the class of Linear Dynamic Graphs (LDGs), that will be functional for the formulation of the problem of reconstructing a network of dynamical systems.
First we define the class of processes that we will use in the development of our theretical framework.

Definition 1: 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 $B(z) \neq 0$ for any $z \in \mathbf{C}$, with $|z| = 1$. Then, $\mathcal{E}$ is a set of rationally related random processes.
We define a class operators under which a set of rationally related random processes is closed.

Definition 2: 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 3: Let $\mathcal{E}$ be a set of rationally related random processes. The set $\mathcal{FE}$ is defined as

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

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 other "parent nodes" through dynamic links. The noise acting on each node is assumed to be unrelated to other noise components. If a certain agent "influences" another a directed edge is drawn between the nodes representing the agents and a directed graph is obtained.

Definition 4 (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 "node processes" $\{x_j\}_{j=1}^n$ of the LDG are the processes defined as

$$x_j = e_j + \sum_{i=1}^{n} H_{ji}(z)x_i,$$

or in a more compact way

$$x(t) = e(t) + H(z)x(t). \tag{1}$$

Let $V := \{x_1, ..., x_n\}$ and let $E := \{(x_i, x_j) | H_{ji}(z) \neq 0\}$. 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. Also, we say that a LDG is *topologically identifiable* if $\Phi_e(e^{i\omega})$ is positive definite for every ω.

Intuitively, a LDG is an interconnection of stochastic processes via linear transfer functions $H_{ji}(z)$ according to a graph G and forced by stationary additive mutually uncorrelated noise.

A. Linear Dynamic Directed Acyclic Graphs

The definition of an LDG allows for the presence of cycles. The purpose of this paper is twofold. On one hand, we want to translate familiar concepts of Control Theory into the language of semi-graphoids. On the other hand we show how the semi-graphoid property results in significant results on a network of stochastic processes interconnected via dynamical systems. The PC algorithm (see [11], [14]) is a well-known algorithm for the reconstruction of Directed Acyclic Graphs of random variables. In this paper we will illustrate how it can be suitably modified in order to reconstruct acyclic networks of dynamic systems. For this reason we introduce the following definition.

Definition 5 (Linear Dynamic DAG): A LDG $(H(z), e)$ is a Linear Dynamic DAG (LDDAG) if its associated graph is a Directed Acyclic Graph, that is a directed graph with no loops.
A graphical representation of a LDDAG is given in Figure 1.

III. WIENER FILTERING

In this section we recall the concept of Wiener filter, showing that the projection it operates is closed with respect to the semi-Hilbert space $\mathcal{FE}$, where $\mathcal{E}$ is a set of rationally related processes.

Definition 6 (Semi-Hilbert space of random processes): Let $\mathcal{H}$ denote the space of rationally related random variables with the inner product defined by $< x, y >= R_{XY}(0)$.

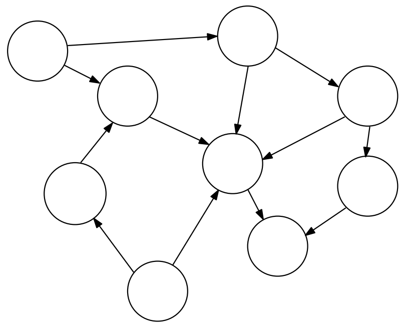


Fig. 1. A Directed Acyclic Graph.

Proposition 7: Under the assumption that two random processes are identical if they are equal to each other almost everywhere, the space $\mathcal{H}$ with its inner product is a semi-Hilbert space.

Proof: The proof is left to the reader. ∎

First, we introduce a way of creating a vector space from a finite set of generators.

Definition 8: For a finite number of elements $x_1, ..., x_m \in \mathcal{F}\mathcal{E}$, tf-span is defined as

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

We make the following observation.

Lemma 9: The tf-span operator defines a subspace of $\mathcal{F}\mathcal{E}$.

Proof: The proof is left to the reader. ∎

The following proposition formulates the Wiener filter in the spaces that we have defined.

Proposition 10 (Wiener Filter): Let $\mathcal{E}$ be a set of rationally related processes Let u and $w_1, ..., w_n$ be processes in the space $\mathcal{F}\mathcal{E}$. Define the vector process $W := (w_1, ..., w_n)^T$ and $M :=$ tf-span$\{w_1, ..., w_n\}$. Consider the problem

$$\inf_{q \in M} \|u - q\|^2. \tag{2}$$

If $\Phi_W(e^{i\omega}) > 0$, for $\omega \in [-\pi, \pi]$, then the solution $\hat{u} \in M$ exists, is unique and is given by $\hat{u} = \mathcal{W}(z)x$ where

$$\mathcal{W}(z) = \Phi_{uW}(z)\Phi_W(z)^{-1}.$$

Moreover, $\hat{u}$ is the only element in M such that, for any $q \in M$,

$$< u - \hat{u}, q >= 0. \tag{3}$$

Proof: See [22]. ∎

The operation performed by the Wiener filter $\mathcal{W}(z)$ can be interpreted as projection of a vector u on the space M.

Definition 11: Let M be a subspace of the semi-Hilbert space $\mathcal{H}$ and let u be a vector in $\mathcal{H}$. Then $\hat{u}_M := Proj_M u$ denotes the projection of u on M if such a projection exists. We will use interchangeably the notation $\hat{u}_M := Proj_M u$ to denote the projection performed by the Wiener filter on a subspace M. Also, notice that Theorem 10 guarantees that such a projection always exists in the spaces of rationally related processes that we have defined. We extend the notation to include the projections of single components of a vector process U.

Definition 12: If $U = [u_1, \ldots, u_n]$ is a vector process and M is a subspace of $\mathcal{H}$ then $\hat{U}_M := \begin{bmatrix} \hat{u}_{1M} & \hat{u}_{2M} & \ldots & \hat{u}_{nM} \end{bmatrix}^T$. We also define $\tilde{U}_M := U - \hat{U}_M$.

We also overload the notation allowing the subspace M to be replaced by a set X of random processes. In such a case, we mean that the projection happens on tf-span(X).

Definition 13: If X is a set of random processes and $Z = \begin{bmatrix} Z_1 & Z_2 & \ldots & Z_r \end{bmatrix}'$ is a random process then $\hat{Z}_X := Proj_M Z$ where M is the span of the random processes X_i in the set X. In this case $\tilde{Z}_{\mathbf{X}} = Z - \hat{Z}_{\mathbf{X}}$.

IV. Semi-graphoids induced by Wiener projections

The notion of semi-graphoid was introduced in [23] for a more general and abstract description of graphical models. Intuitively, a semi-graphoid is given by a set χ of mathematical entities and a relation $I(\cdot, \cdot, \cdot)$ defined on 3-tuple of disjoint subsets of χ. The relation $I(X, Z, Y)$ indicates whether the subset Z separates the subsets X and Y according to certain properties.

Definition 14: Consider the set $\chi = \{x_1, ..., x_n\}$, with a relation of "separation" $I(X, Y, Z)$ defined for 3-tuples of disjoint subsets of χ. The pair (χ, I) is a semi-graphoid if the following properties are met

- Simmetry: $I(X, Z, Y) \Leftrightarrow I(Y, Z, X)$
- Decomposition: $I(X, Z, Y \cup W) \Rightarrow I(X, Z, Y)$
- Weak union: $I(X, Z, Y \cup W) \Rightarrow I(X, Z \cup W, Y)$
- Contraction: $I(X, Z \cup Y, W)$ and $I(X, Z, Y) \Rightarrow I(X, Z, Y \cup W)$

where $X, Y, Z \subseteq \chi$ are three disjoint subsets. The relation $I(X, Z, Y)$ is read as "the set X is separated from the set Y given the set Z in the model I".

The relation of separation is the key component of a semi-graphoid and lends itself to a representation in the form of a directed graph. This is the reason why, by defining a separation I on a set χ and obtaining a semi-graphoid, it is possible to represent χ in the form of a directed graph.

Here we will establish that the projection operation defined by the Wiener estimate on the processes generated by a LDG induces a separation on them that satisfies the properties of a semi-graphoid. Also we will prove that if the LDG is a LDDAG, it will be possible to recover information about its underlying structure.

Definition 15: Let X, Y and Z be disjoint subsets in a set of rationally related random processes. We say that X is Wiener-separated from Y given Z if

$$\hat{x}_{Y,Z} = \hat{x}_Z$$

for all $x \in X$ and we denote this relation as $I(X, Z, Y)$.

Theorem 16: The Wiener-separation satisfies the property of Symmetry.

$$I(X, Z, Y) \Leftrightarrow I(Y, Z, X) \tag{4}$$

Proof: From $I(X, Z, Y)$ we have that $\hat{x}_{Y,Z} = \hat{x}_Z$ for all $x \in span(X)$. This implies that

$$< x - \hat{x}_Z, y >=< \tilde{x}_Z, y >= 0 \text{ for all } y \in span(\mathbf{Y}). \tag{5}$$

Now consider some $y \in span(Y)$ and $\hat{y}_Z$ exist (this will be true if $span(Z)$ is closed). Then consider any $x \in X$.

$$\begin{aligned}
< y - \hat{y}_Z, x > &= < \tilde{y}_Z, \hat{x}_Z + \tilde{x}_Z > \\
&= < \tilde{y}_Z, \tilde{x}_Z > \text{ as } \hat{x}_Z \in span(Z) \\
&= < y - \hat{y}_{\mathbf{Z}}, \tilde{x}_Z > \\
&= < y, \tilde{x}_Z > \text{ as } \hat{y}_Z \in span(Z) \\
&= 0 \text{ as } I(X, Z, Y).
\end{aligned}$$

Thus, since all the projections are guaranteed to exist, the theorem holds. ∎

Theorem 17: The Wiener-separation satisfies the property of Decomposition.

$$I(Y, Z, X \cup W) \Rightarrow I(Y, Z, X) \text{ and } I(Y, Z, W).$$

Proof: This follows trivially from the definition. ∎

Theorem 18: The Wiener-separation satisfies the property of Weak Union.

$$I(X, Z, Y \cup W) \Rightarrow I(X, Z \cup W, Y)$$

Proof: Suppose $I(X, Z, Y \cup W)$. Then from the theorem on symmetry $I(Y \cup W, Z, X)$. Thus $\hat{x}_{Y,W,Z} = \hat{x}_Z$. In particular $\hat{x}_Z \in span(Z)$ with $\tilde{x}_Z$ is such that $< \tilde{x}_Z, z >=< \tilde{x}_Z, y >=< \tilde{x}_Z, w >= 0$ for any $y \in Z$, $w \in W$ and $z \in Z$. Now as $\hat{x}_Z \in span(Z) \subset span(Z, W)$ and $< \tilde{x}_Z, w >= 0$ for all $w \in span(W)$ it follows that $\hat{x}_Z = \hat{x}_{Z,W}$. Thus $\hat{x}_{Y,Z,W} = \hat{x}_Z = \hat{x}_{Z,W}$ and therefore $\hat{x}_{Y,W,Z} = \hat{x}_{Z,W}$. That is $I(Y, Z \cup W, X)$ and from the symmetry theorem $I(X, Z \cup W, Y)$. This proves the theorem. ∎

Theorem 19: The Wiener-separation satisfies the property of Contraction.

$$I(X, Z, Y) \text{ and } I(X, Z \cup Y, W) \Rightarrow I(X, Z, Y \cup W)$$

Proof: Suppose $I(X, Z, Y)$ and $I(X, Z \cup Y, W)$. Then it follows that for any $y \in span(Y)$ and $w \in span(W)$, $\hat{y}_{Z,X} = \hat{y}_Z$ and $\hat{w}_{X,Z,Y} = \hat{w}_{Z,Y}$. Let $\hat{w}_{Z,Y} = z_w + y_w$ where $z_w \in span(Z)$ and $y_w \in span(Y)$. Further let $y_w = y_{wz} + \tilde{y}_{wz}$ where $y_{wz} \in span(Z)$ and $\tilde{y}_{wz} \in span(Z)^{\perp}$. Now $(w - \hat{w}_{Z,Y}) \perp span(Z)$ which implies $(w - z_w - y_w) \perp span(Z)$. Thus $w - z_w - y_{wz} \perp span(Z)$. Now as $z_w \in span(Z)$ and $y_{wz} \in span(Z)$ it follows that the projection of w on $span(Z)$, $\hat{w}_Z = z_w + y_{wz}$. Now as $\hat{w}_{Z,Y} = \hat{w}_{X,Z,Y}$ it follows that $w - \hat{w}_{Z,Y} \perp span(X)$.

$$\begin{aligned}
& w - \hat{w}_{Z,Y} \perp span(X) \\
\Rightarrow\ & w - z_w - y_w \perp span(X) \\
\Rightarrow\ & w - z_w - y_{wz} - \tilde{y}_{wz} \perp span(X) \\
\Rightarrow\ & w - z_w - y_{wz} \perp span(X) \text{ as } Proj_Z y_w = Proj_{Z,X} y_w \\
\Rightarrow\ & w - \hat{w}_Z \perp span(X) \\
\Rightarrow\ & \tilde{w}_Z \perp span(X) \\
\Rightarrow\ & \hat{w}_Z = \hat{w}_{X,Z} \\
\Rightarrow\ & I(X, Z, W).
\end{aligned}$$

Thus $I(X, Z, W)$ and $I(X, Z, Y)$. This proves the theorem ∎

The previous results are summarized in the following theorem.

Theorem 20: Let χ be a set of rationally related stochastic processes and let $I(\cdot, \cdot, \cdot)$ be the separation relation induced by the Wiener-separation. (χ, I) is a semi-graphoid.

Proof: As the above properties hold, the projection based relationship I defined on a set of stochastic processes is a semi-graphoid. ∎

V. I-GRAPHS AND d-SEPARATION

In many areas of interest, multiple time-series data are monitored and a relationship between these time-series data is sought. A natural way of organizing the data is in terms of graphs where the vertices represent the processes generating the time series and the links between vertices denote the "closeness" of one process to another. Here the focus will be to determine a *directed graph* which consists of a set of vertices V with directed edges E. Each edge consists of a head vertex and a tail vertex with the arrow directed from the tail to the head.

Definition 21 (Skeleton of a directed graph): The *skeleton* of the directed graph is obtained by removing the directions of the edges thereby yielding an undirected graph.

Definition 22: Given a graph (V, E) with vertices $x_1, ..., x_n$, a chain starting from x_j and ending in x_i is a sequence $(x_{c_1}, x_{c_2}, \ldots, x_{c_l})$ where $x_j = x_{c_1}$, $x_j = x_{c_l}$ and directed edges from $x_{c_{i-1}}$ to x_{c_i}, $i = 2, \ldots, l$ exist in the directed graph.

Definition 23 (Parents, children, ancestors, descendants): Vertex x_j is a *parent* of vertex x_i if there is a directed edge from x_j to x_i. In such a case x_i is a *child* of x_j. x_j is an *ancestor* of x_i if there is a chain from x_j to x_i. In such a case x_i is a *descendant* of x_j.

Definition 24 (Paths, forks and inverted forks): A *path* between two vertices, x_1 and x_n is an ordered set of vertices $(x_{c_1}, x_{c_2}, \ldots, x_{c_l})$ where $x_j = x_{c_1}$, $x_j = x_{c_l}$ and there is an edge between $x_{c_{i-1}}$ and x_{c_i} for all $i = 2, \ldots, n$ in the skeleton of the directed graph. A path has a *fork* at x_{c_m} if $x_{c_{m-1}}$ and $x_{c_{m+1}}$ are both children of m (that is $x_{c_{m-1}} \leftarrow x_{c_m} \rightarrow x_{c_{m+1}}$ appears in the directed graph). A path has an *inverted fork* at x_{c_m} if $x_{c_{m-1}}$ and $x_{c_{m+1}}$ are both parents of m (that is $x_{c_{m-1}} \rightarrow x_{c_m} \leftarrow x_{c_{m+1}}$ appears in the directed graph).

The following definition introduces a notion of separation on subsets of vertices in a directed graphs.

Definition 25: (d-separation) Consider three set of vertices X, Z, Y. The set Z is said to d separate X and Y if every path between a pair of vertices, one from X and another from Y, meets at least one of the following conditions

1) the path contains a chain $x_i \rightarrow x_m \rightarrow x_j$ where $x_m \in Z$.
2) the path contains an inverted fork at x_m given by $x_{m-1} \rightarrow x_m \leftarrow x_{m+1}$ where neither x_m nor its descendants belong to Z.

If Z d-separates X and Y we write $dSep(X, Z, Y)$.

Definition 26: (I-graph) A directed graph $G = (\chi, E)$ provides an *independence graph* on a semigraphoid (χ, I) if every separation of vertices in the graph implies an separation in the model I. Thus

$$dSep(X, Z, Y) \implies I(X, Z, Y).$$

The smallest I-graph is called the minimal I-graph.

We remark that the minimal I-graph need not capture all separations in the model I. To describe a situation where a graph G represent exactly the the separation in a semi-graphoid we introduce the notion of faithfulness

Definition 27: If for the graph $G = (\chi, A)$ we have that $I(X, Z, Y) \Leftrightarrow dSep(X, Z, Y)$, we say that G is faithful to I (or that G is a perfect map for I).

The following theorem is a fundamental consequence of the algebraic structure provided by a semi-graphoid.

Theorem 28: Consider a finite set of stochastic processes χ. Consider an ordering on the χ given by $x_1, \dots, x_n$. $A_i = \{x_1, \dots, x_{i-1}\}$ form the predecessors of x_i in this order. Suppose $B_{(i)} \subset A_i$ separates x_i from all other predecessors in the model I defined by the Wiener-separation (that is $I(x_i, B_{(i)}, A_i \backslash B_{(i)})$ holds) and $B_{(i)}$ is the minimal such set. Form a DAG by assigning as parents of x_i all vertices in $B_{(i)}$. Then the DAG so created will be a minimal I-graph of the model I.

Proof: The proof follows from [12] as we have shown that the projection model of I is a semi-graphoid. ■

The above theorem provides a means to create the minimal I-graph of a set of time-series by following the steps

1) Consider any ordering $x_1, \dots, x_n$ of the time series set. Determine for each i the minimal subset $B_{(i)} \subset A_i$ where $A_i = \{v_1, \dots, v_{i-1}\}$ such that
$$Proj_{A_i} v_i = Proj_{B_{(i)}} x_i$$
2) Create a graph such that for each i, the only parents of x_i are the elements of $B_{(i)}$.

We note that the minimal I-graph for a model I is not unique, but in general depends on the ordering of the nodes.

The following theorem holds

Theorem 29: Consider a vertex v_i of a DAG with vertices in S. Consider the set of vertices K_i consisting of (a) parents of v_i (b)children of v_i (c) spouses of v_i. Then K_i d-separates v_i from rest of the vertices. Also, K_i is the minimal set that d-separates v_i from the rest of the graph. That is, every set T that satisfies
$$dSep(v_i, T, S \backslash \{T \cup v_i\})$$
is such that $T \supset K_i$.

The above theorem provides a means of identifying the Markov blanket of a time series v_i in the model I. Indeed if G is an I-graph of model I then every separation in the graph implies a separation in the model I and thus the set of spouses, children and parents will separate the node i from the rest of time-series in the model I. Notice how this theorem is the equivalent (but limited to DAGs) of main result in [22] in the non-causal scenario. We have been able to prove it with virtually no effort just by relying on the algebraic structure of semi-graphoids.

VI. Illustration: reconstruction of the pattern of a LDDAG

First observe that the graph underlying a LDDAG satisfies the following property.

Theorem 30: The graph G of a LDDAG $\mathcal{G} = (H(z), e)$ is an I-graph for the semi-graphoid generated by the processes of $\mathcal{G}$ under Wiener-separation.

Proof: Since G is a DAG, it defines a partial order on the processes where a node is a predecessor of another it is one of its ancestors. Define a global order $\chi = \{x_1, ..., x_n\}$ that is compatible with the partial order determined by G. For every x_i define $B_{(i)}$ as the parents of x_i. Observe that $I(x_i, B_{(i)}, \{x_1, x_{i-1}\} \setminus B_i)$ from the diagonal structure of $\Phi_e(z)$. The application of Theorem 28 gives the assertion. ■

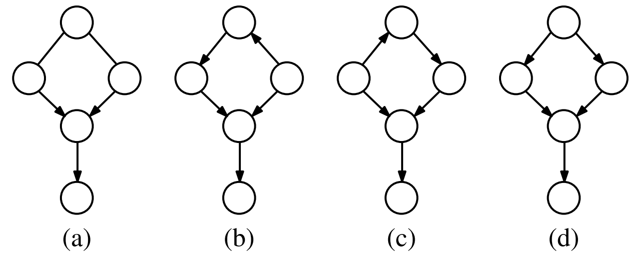


Fig. 2. A pattern represented as semi-directed graph (a) and the three DAGs sharing that pattern (b), (c), (d).

Thus, assuming that the underlying graph G of a LDDAG $\mathcal{G}$, is a perfect map of $\mathcal{G}$ opens the possibility of recovering G only by using the second order statistics of the processes necessary to compute the Wiener projections. This would be possible, according to the proof of Theorem 30, if a global order on the processes that is compatible with the partial order defined by the DAG G were known. Unfortunately such information is in general not available. Since the DAG obtained from Theorem 28 is order dependent, this procedure is not viable.

There is also a deeper reason why, only from Wiener-separation, an exact reconstruction of the underlying graph of a DAG is not possible: different DAGs can d-separate the same subsets. So they can not be disringuished from each other.

Definition 31: Two DAGs with the same set of vertices have the same *pattern* if they have the same skeleton and the same inverted forks.

Theorem 32: Let D_1 and D_2 be two DAGs with the same pattern. Let X, Y amd Z be three disjoint subsets of the set of vertices of D_1 and D_2. If Z d-separates X and Y in D_1, then Z d-separates X and Y in D_2.

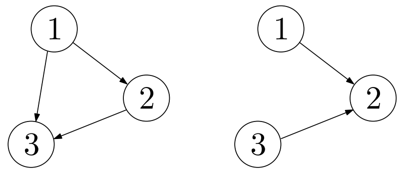
Proof: This theorem is proven in [13]. ■

Given this result, the reconstruction of the topology of a LDDAG can not go beyond the pattern of its underlying graph. Indeed, since DAGs with identical patterns d-separate the same sets, from the knowledge of the semi-graphoid associated with the LDG it is not possible to determine the original DAG. A pattern can be represented as semidirected graph where the only directed edges are the ones creating an inverted fork (see Figure 2(a)). A DAG belongs to the class of equivalence defined by a pattern if, by orienting the unoriented edges, no new inverted forks are originated (see Figure 2(b,d,d)).

In [13], an algorithm is provided that allows to obtain from a semi-graphoid the pattern of a DAG, if this DAG is a perfect map of the semigraphoid. The reconstruction of the pattern of a LDDAG can be performed by the Inductive Causation (IC) algorithm (see [13]).

Inductive Causation Algorithm

1) Initialize the set of edges $E = \emptyset$.



(a) (b)

Fig. 3. Two DAGs considered in the example of a non-faithful graph.

2) For each pair (x_i, x_j) of variables in $\chi = \{x_1, x_n\}$, determine a set $S_{i,j}$ such that $I(x_i, S_{i,j}, x_j)$. If no such a set can be found, then add (x_i, x_i) to E
3) For each pair (x_i, x_j) of variables that are not connected in E and have a common neighbour x_k, check if $x_k \in S_{i,j}$.
   - If $x_k \in S_{i,j}$, the inverted fork $x_i \rightarrow x_k \leftarrow x_j$ is in the pattern.
4) In the partially directed graph fix the orientation of as many undirected edges as possible, as long as
   - there are no directed cycles
   - there are no new inverted forks.

The IC algorithm takes as an input the semi-graphoid (χ, I) generated by a LDDAG under Wiener-separation and gives as an output the pattern of a DAG. The following theorem guarantees that the computed pattern matches the pattern of the original LDDAG, if the original DAG is faithful to the generated semi-graphoid.

Theorem 33: If the DAG D of the LDDAG is faithful to the separation I induced by the Wiener projection, the IC algorithm provides the pattern of D.

Proof: The proof follows from the fact that the separation I is a semi-graphoid and from the results in [13]. ∎

As observed in [11], the faithfulness condition is quite mild and is verified in most practical scenarios. Indeed, the graph associated to a LDDAG is not faithful to the generated graphoid only in pathological cases. An example is given by considering the following 3-node LDG

$$x_1 = e_1; \qquad x_2 = x_1 + e_2 x_3 = x_2 - x_1 + e_3. \quad (6)$$

The underlying DAG is represented in Figure 3(a), but, because of a cancellation of the effect of x_1 on x_2 we have $x_3 = e_2 + e_3$ that implies $I(x_1, \emptyset, x_3)$. Thus the DAG is not faithful to the Wiener-separation. Indeed, the application of the IC algorithm produces the DAG in Figure 3(b).

VII. Conclusions

In this paper the problem of identifying the unknown structure of a network of stochastic processes has been considered. The main result consists in showing that a set of jointly stationary stochastic processes is a semi-graphoid, if we introduce in the set a notion of separation derived from Wiener filtering. This property allows one to to seamlessly apply results from the area of Bayesian network to the more complicated scenario of network of stochastic processes inteconnected through dynamic trasfer functions. We have illustrated the effectiveness of the developed theoretical framework by extending the algorithm of Inductive Causation to reconstruct linear dynamic networks with no loops.

References

[1] E. Atalay, A. Hortaçsu, J. Roberts, and C. Syverson, "Network structure of production," *Proceedings of the National Academy of Sciences*, vol. 108, no. 13, p. 5199, 2011.

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

[3] D. Acemoglu, M. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[4] 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.

[5] E. Ravasz, A. Somera, D. Mongru, Z. Oltvai, and A. Barabasi, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, p. 1551, 2002.

[6] D. Del Vecchio, A. Ninfa, and E. Sontag, "Modular cell biology: Retroactivity and insulation," *Nature Molecular Systems Biology*, vol. 4, p. 161, 2008.

[7] 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.

[8] 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.

[9] P. Monestiez, J.-S. Bailly, P. Lagacheriec, and M. Voltz, "Geostatistical modelling of spatial processes on directed trees: Application to fluvisol extent," *Geoderma*, vol. 128, pp. 179–191, 2005.

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

[11] ——, *Causality: models, reasoning, and inference*. Cambridge Univ Press, 2000, vol. 47.

[12] T. Verma and J. Pearl, "Causal Networks: Semantics and Expressiveness," in *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, pp. 352–359.

[13] J. Pearl and T. Verma, "A theory of inferred causation," *Studies in Logic and the Foundations of Mathematics*, vol. 134, pp. 789–811, 1995.

[14] P. Spirtes, C. Glymour, and R. Scheines, *Causation, prediction, and search*. The MIT Press, 2000, vol. 81.

[15] S. Lauritzen, *Graphical models*. Oxford University Press, USA, 1996, vol. 17.

[16] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. The MIT Press, 2009.

[17] M. Nabi-Abdolyousefi and M. Mesbahi, "Network identification via node knock-out," in *Conference on Decision and Control*, 2010, pp. 2239–2244.

[18] B. Sanandaji, T. Vincent, and M. Wakin, "Compressive topology identification of interconnected dynamic systems via clustered orthogonal matching pursuit,âĂİ," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, 2011.

[19] G. Pillonetto, A. Chiuso, and G. De Nicolao, "Prediction error identification of linear systems: A nonparametric gaussian regression approach," *Automatica*, vol. 47, no. 2, pp. 291–305, 2011.

[20] 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*, vol. 30, no. 1, pp. 17–44, 2011.

[21] G. Chowdhary, M. Egerstedt, and E. Johnson, "Network discovery: An estimation based approach," in *American Control Conference (ACC), 2011*. IEEE, 2011, pp. 1076–1081.

[22] D. Materassi and M. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the wiener filter," *Automatic Control, IEEE Transactions on*, vol. 57, no. 7, pp. 1765–1777, 2012.

[23] J. Pearl and A. Paz, "Graphoids: a graph-based logic for reasoning about relevance relations. ucla computer science department technical report 850038," *Advances in Artificial Intelligence-II, North-Holland Publishing Co*, 1987.