Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Notions of separation in graphs of dynamical systems

Abstract:
The concept of d-Separation is a key tool to analyze stochastic models defined by probability distributions of random variables that admit a factorization described by a Directed Acyclic Graph. However, in the area of dynamical systems, and especially control theory, it is common to find network models involving stochastic processes that influence each other according to a directed network where feedback loops may be present as well. These models differ from standard probabilistic models at a fundamental level. Indeed, for a network of dynamical systems it is challenging to introduce an appropriate notion of factorization not only because of the presence of loops, but also because stochastic processes involve an infinite number of random variables. In this article, we show that the concept of d-Separation can still be applied to infer properties of least square estimators defined on subsets of stochastic processes, at least if their mutual influences are described by linear operators. Similar results have been obtained by (Koster, 1999) in the domain of Structural Equation Models for random variables. However, the scenario considered in this article involves stochastic processes and deals with several technical complications, such as noise terms potentially correlated in time and the possibility of causal estimators. The article provides a general framework to overcome all these difficulties that are not present when a graphical model just represents random variables.

Keywords: Graphical models, d-Separation

1. 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. Atalay et al. (2011)), social systems (see e.g. Acemoglu et al. (2011)) biology (see e.g. Eisen et al. (1998); Del Vecchio et al. (2008)), cognitive sciences (see e.g. Brovelli et al. (2004)), and geology (see e.g. Bailly et al. (2006)). 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 Pearl (1988, 2000); Verma and Pearl (1988)) and by many other researchers (see Spirtes et al. (2000), Lauritzen (1996), Koller and Friedman (2009)). However, an approach that is specifically targeted to stochastic processes inteconnected through dynamic relations (in other words considering dependencies occurring at different time instants) is not been fully developed yet. Indeed, for networks of stochastic processes, the presence of dynamic relations 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.

These challenges are leading to new results and techniques which are rapidly emerging (see Nabi-Abdolyousefi and Mesbahi (2010); Sanandaji et al. (2011); Pillonetto et al. (2011); Chowdhary et al. (2011); Materassi and Salapaka (2012); Van den Hof et al. (2012); Quinn et al. (2013)).

The article extends and applies the concept of d-Separation introduced by Judea Pearl (Pearl, 1988) to networks of dynamical systems where loops can be present, as well. The concept of d-Separation is typically defined on stochastic models described by a joint probability distributions of random variables. Such a distribution is assumed to admit a sparse factorization that can be aptly represented by a Directed Acyclic Graph. However, in the case of a network of stochastic processes influencing each other according to a directed network, such a factorization looses its meaning: especially if feedback loops have to be taken into account. We show that the concept of d-Separation can still be applied to infer properties of least square estimators defined on subsets of stochastic processes, at least if their mutual influences are described by linear operators. Similar results have been obtained by (Koster, 1999) in the domain of Structural Equation Models for random variables. However, the scenario considered in this article extends to stochastic processes and deals with several technical complications, such as noise terms potentially correlated in time and the possibility of causal estimators. These complications are not present when a graphical model simply represents random variables. Furthermore, standard graphical model approaches and results are usually derived considering a finite number of random variables. Thus, their application to the realm of stochastic processes (viewed as infinite sequences of random variables) is neither immediate or trivial. The article provides a theoretical framework that overcomes all these difficulties.

NOTATION

- $\{x_i, x_j\}$: unordered pair of two elements x_i, x_j
- (x_i, x_j): ordered pair of two elements x_i, x_j
- $\mathbf{E}[\cdot]$: mean operator
- $R_{XY}(\tau) := \mathbf{E}\left[X(\tau)Y(0)^T\right]$ for two wide sense stationary stochastic vectors X and Y
- $\mathcal{Z}(\cdot)$: $\mathcal{Z}$-transform
- $\Phi_{XY}(z) = \mathcal{Z}(R_{XY}(\tau))$: Power spectral density
- $(\cdot)^*$: transpose conjugate
- $\emptyset$: empty set

2. INTRODUCTORY CONCEPTS AND DEFINITIONS

Aim of this section is to make the reader acquainted with concepts that will be used in the derivation of the main results.

In Section 2.1 we recall basic definitions of graph theory and introduce the notion of *d*-separation (Pearl, 1988). In Section 2.2 we define Linear Dynamic Graphs (LDGs), the main class of dynamical systems that will be considered in this paper. In Section 2.3 we recall some results obtained in (Materassi and Salapaka, 2012, 2013) that allow one to interpret certain variations of Wiener filters as projections in a pre-Hilbert space of stochastic processes.

2.1 d-Separation on directed graphs

We start recalling basic notions of graph theory which are functional to the subsequent developments. First, the standard definition of undirected and directed graphs is provided.

Definition 1. (Directed and Undirected Graphs).
A directed (undirected) graph G is a pair (V, E) where V is a set of vertices or nodes and E is a set of edges or arcs, which are ordered (unordered) subsets of two distinct elements of V.

Given a graph, a sub-graph can be defined with repect to a subset of its nodes.

Definition 2. (Restriction of a Graph). Given a directed graph $G = (V, E)$, its restriction to the node set $V' \subseteq V$ is the graph $G' = (V', E')$ where $E' = \{(x_i, x_j) \,|\, x_i \in V' \text{ and } x_j \in V'\}$

The skeleton of a directed graph is the undirected graph obtained by replacing each directed edge with an undirected one. The formal definition follows.

Definition 3. (Skeleton of a directed graph). Given a directed graph $G = (V, E)$, its skeleton is the undirected graph $\overline{G} = (V, \overline{E})$ where

$$\overline{E} = \{\{x_i, x_j\} \,|\, (x_i, x_j) \in E \text{ or } (x_j, x_i) \in E\}.$$

On a directed graph we also define "chains" and "paths".

Definition 4. (Paths, chains). Consider a directed graph $G = (V, E)$ with vertices $x_1, ..., x_n$ and its skeleton $(V, \overline{E})$. A *chain* starting from x_i and ending in x_j is an ordered set of edges in E ($(x_{\pi_1}, x_{\pi_2}), \ldots, (x_{\pi_{l-1}}, x_{\pi_l})$) where $x_i = x_{\pi_1}$, $x_j = x_{\pi_l}$. A *path* between two vertices, x_i and x_j is an ordered set of edges in $\overline{E}$ ($\{x_{\pi_1}, x_{\pi_2}\}, \ldots, \{x_{\pi_{l-1}}, x_{\pi_l}\}$) where $x_i = x_{\pi_1}$, $x_j = x_{\pi_l}$.

From the concept of chains, we can derive the notions of ancestry and descendance.

Definition 5. (Parents, children, ancestors, descendants). Consider a graph $G = (V, E)$. A vertex x_i is a *parent* of a vertex x_j if there is a directed edge from x_i to x_j. In such a case x_j is a *child* of x_i. Also x_i is an *ancestor* of x_j if there is a chain from x_j to x_i. In such a case x_i is a *descendant* of x_j. Given a set $X \subseteq V$, we define following notation

$$pa\,(X) := \{x_i \in V \,|\, \exists\, x_j \in X : x_i \text{ is a parent of } x_j\}$$
$$ch\,(X) := \{x_j \in V \,|\, \exists\, x_i \in X : x_j \text{ is a child of } x_i\}$$
$$an\,(X) := \{x_i \in V \,|\, \exists\, x_j \in X : x_i \text{ is an ancestor of } x_j\}$$
$$de\,(X) := \{x_j \in V \,|\, \exists\, x_i \in X : x_j \text{ is a descendant of } x_i\}.$$

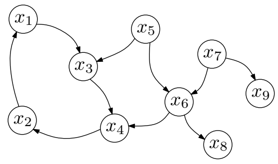


Fig. 1. A directed graph with 9 nodes that is not acyclic.

The Markov blanket of a node is given by the set of parents, children and all the other nodes sharing a child with it.

Definition 6. (Markov blanket). In a directed graph G, the Markov blanket a node is the set of the "parents", "children" and "parents of the children" of the node.

On a given path we define forks and colliders.

Definition 7. (Forks and colliders). A path has a *fork* at x_{π_p} if $x_{\pi_{p-1}}$ and $x_{\pi_{p+1}}$ are both children of x_{π_p} (that is $x_{\pi_{p-1}} \leftarrow x_{\pi_p} \rightarrow x_{\pi_{p+1}}$ appears in the directed graph). A path has an *inverted fork* (or a collider) at x_{π_p} if $x_{\pi_{p-1}}$ and $x_{\pi_{p+1}}$ are both parents of x_{π_p} (that is $x_{\pi_{p-1}} \rightarrow x_{\pi_p} \leftarrow x_{\pi_{p+1}}$ appears in the directed graph).

The following definition introduces a notion of separation on subsets of vertices in a directed graphs (Pearl, 1988).

Definition 8. (d-separation) Consider three mutually disjoint sets of vertices X, Z, Y. The set Z is said to *d*-Separate X and Y if for every $x_i \in X$ and $x_j \in Y$ every path between x_i and x_j meets at least one of the following conditions

(1) the path contains a node $x_k \in Z$ that is not a collider
(2) the path contains a collider at x_k given by $x_{k-1} \rightarrow x_k \leftarrow x_{k+1}$ where neither x_k nor its descendants belong to Z.

If Z *d*-separates X and Y in the graph G, we write $\mathcal{I}_G(X, Z, Y)$ othewise we write $\neg\mathcal{I}_G(X, Z, Y)$.

As an example, consider the graph of Figure 1. In such a graph, we have that $\mathcal{I}_G(x_5, \emptyset, x_7)$; $\mathcal{I}_G(x_1, \{x_5, x_6\}, x_7)$; $\mathcal{I}_G(x_2, \{x_1, x_4\}, x_3)$, and $\neg\mathcal{I}_G(x_5, x_8, x_7)$; $\neg\mathcal{I}_G(x_3, \emptyset, x_9)$.

2.2 Generative class of models: Linear Dynamic Graphs

In this section we describe the class of Linear Dynamic Graphs (LDGs).

First we define the class of processes that we will use in the development of our theoretical framework.

Definition 9. Let $\mathcal{E}$ be a set containing discrete-time scalar, zero-mean, jointly wide-sense stationary random processes such that, for any $e_i, e_j \in \mathcal{E}$, the power spectral density $\Phi_{e_i e_j}(z)$ exists, is real-rational with no poles on the unit circle and given by $\Phi_{e_i e_j}(z) = \frac{A(z)}{B(z)}$, 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 two classes of operators, $\mathcal{F}$ and $\mathcal{F}^+$, transforming rationally related random processes into other rationally related random processes.

Definition 10. 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 11. 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

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

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

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

It is immediate to verify that the following set is closed with respect to the operators defined in $\mathcal{F}$.

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

$$\mathcal{FE} := \left\{ x = \sum_{i=1}^{n} H_i(z) e_i \mid e_i \in \mathcal{E}, H_i(z) \in \mathcal{F}, m \in \mathbf{N} \right\}.$$

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 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 14. (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 processes $\{x_j\}_{j=1}^n$ of the LDG are defined as

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

or in a more compact way

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

Let $V := \{x_1, ..., x_n\}$ and let $E := \{(x_i, x_j) | H_{ji}(z) \neq 0\}$. The pair $G = (V, E)$ 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.

If the operator $(I - H(z))$ 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 in the space $\mathcal{F}e$ will be obtained. For this reason, the following definition is introduced.

Definition 15. A LDG $(H(z), e)$ is well-posed if each entry of $(I - H(z))^{-1}$ belongs to $\mathcal{F}$. Thus, $x = (I - H(z))^{-1} e$. can be written. A LDG $(H(z), e)$ is causally well-posed if all the entries of $(I - H(z))$ and $(I - H(z))^{-1}$ belong to $\mathcal{F}^+$.

2.3 Wiener filtering as a projection

It is possible to introduce an inner product in $\mathcal{FE}$.

Lemma 1. 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 done by inspection checking the properties of vector space and of inner product. □

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

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

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

Definition 17. For a finite number of elements $x_1, ..., x_n \in \mathcal{FE}$, c-tf-span is defined as

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

Lemma 2. The tf-span operator c-tf-span operators define two subspaces of $\mathcal{FE}$.

Proof. The proof is left to the reader. □

The following proposition formulates the problem of non-causal and causal Wiener filtering (Kailath et al., 2000) in terms of projections in the space $\mathcal{FE}$.

Proposition 3. (Wiener Filter). Let $\mathcal{E}$ be a set of rationally related processes Let u and $w_1, ..., w_n$ be processes in the space $\mathcal{FE}$. Define the vector process $W := (w_1, ..., w_n)^T$ and let

$$M := \text{tf-span}\{w_1, ..., w_n\} \quad \text{or}$$
$$M := \text{c-tf-span}\{w_1, ..., w_n\}.$$

Consider the problem

$$\hat{u} := \arg \inf_{q \in M} \|u - q\|^2. \tag{5}$$

If $\Phi_W(e^{i\omega}) > 0$, for $\omega \in [-\pi, \pi]$, then the solution $\hat{u} \in M$ exists, is unique and is the only element in M such that, for any $q \in M$,

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

Proof. See (Materassi and Salapaka, 2012). □

When $M := \text{tf-span}\{w_1, ..., w_n\}$ we have the multivariable non-causal Wiener filter, while when $M := \text{c-tf-span}\{w_1, ..., w_n\}$ we have the multivariable causal Wiener filter. In light of the perpendicularity relation given by (6), the transformation performed by the Wiener filter $\hat{u} := \mathcal{W}(z)W$ can be interpreted as projection of a vector u on a suitable space M.

Definition 18. Let M be a subspace of a pre-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 3 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 19. If $U = [u_1, \ldots, u_n]$ is a vector process and M is a subspace of $\mathcal{H}$ then $\hat{U}_M := [\, \hat{u_1}_M \; \hat{u_2}_M \; \ldots \; \hat{u_n}_M \,]^T$. 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 20. If X is a set of random processes and $Z = [\, Z_1 \; Z_2 \; \ldots \; Z_r \,]'$ is a random vector then $\hat{Z}_X := Proj_M Z$ where M is the span of the random processes X_i in the set X defined for each component of Z. In this case $\tilde{Z}_{\mathbf{X}} = Z - \hat{Z}_{\mathbf{X}}$.

The following definition provides a notion of independence among sets of rationally related stochastic processes.

Definition 21. (Wiener-separation). 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 $I_W(X, Z, Y)$ when the non-causal Wiener filter is considered and as $I_{W^+}(X, Z, Y)$ when the causal Wiener filter is considered.

3. PRELIMINARY RESULTS

The aim of this section is to provide four lemmas that will be helpful in the derivation of the main results. These results follow the derivation presented in (Koster, 1999), but have been suitably modified to fit our framework. The first lemma shows that in every directed graph each path is a subset of the set of ancestors of its extreme points and its colliders.

Lemma 4. Let x_i and x_j be two nodes in a directed graph G. Let $(x_{\pi_0}, ..., x_{\pi_l})$ be a path from x_i to x_j in G and let C be the set of colliders on. Then we have that

$$x_{\pi_p} \in an\,(\{x_i, x_j,\} \cup C)$$

for $p = 0, ..., l$.

Proof. We will prove the statement by induction. If there is no collider on the path the statement is true. Now, assume that the statement is true if there are n_c colliders on a path. Let π be a path from x_i to x_j with $n_c + 1$ colliders. Let c be a collider on *path*. Node c splits the path in two subpaths π_a, from x_j to c and π_b from c to x_j. Both π_a and π_b have less than $n_c + 1$ colliders. Thus, the theorem statement can be applied to both subpaths. This immediately gives the statement also for π. $\square$

The following lemma shows that if three disjoint sets X, Y, Z cover all nodes of a graph then the notion of d-Separation can be characterized by checking paths of at most length 2. This result states that, in order to have d-Separation between two sets X and Y, their nodes can not be in each other's Markov blanket.

Lemma 5. Let $G = (V, E)$ be a directed graph and let $X, Y, Z \subseteq V$ be three disjoints sets satisfying

- $X \cup Y \cup Z = V$.

Then, we have that $\mathcal{I}_G(X, Z, Y)$ if and only if for all nodes $x_i \in X$, $x_j \in Y$ there are no paths of the type

- $x_i \rightarrow x_j$
- $x_i \leftarrow x_j$
- $x_i \rightarrow c \leftarrow x_j$

where $c \in Z$.

Proof. If there is a path of the type $x_i \rightarrow x_j$, of the type $x_i \leftarrow x_j$, or of the type $x_i \rightarrow c \leftarrow x_j$, where $x_i \in X$, $x_j \in Y$ and $c \in Z$, we have that x_i and x_j are not *d*-separated. Thus, the necessity is proven. Now assume that $\neg\mathcal{I}_G(X, Z, Y)$. Then, there is a path π *d*-connecting one node in X with a node in Y. Let x_i be the node with largest index in the path that is in X. Let x_j be the node with smallest index in the path that comes after x_j and is in Y. Consider the subpath π' of π that goes from x_i and x_j. The nodes x_i and x_j are *d*-connected via π'. Also, except for x_i and x_j, all nodes in π' belong to Z. Thus, π' can only be of the type $x_i \rightarrow x_j$, of the type $x_i \leftarrow x_j$, or of the type $x_i \rightarrow c \leftarrow x_j$. $\square$

Lemma 6. Let $G = (V, E)$ be a graph and let Z be a subset of V. Consider $x_i, x_j \in V \setminus Z$. It holds that

$$\mathcal{I}_G(x_i, Z, x_j) \Leftrightarrow \mathcal{I}_{G'}(x_i, Z, x_j)$$

where $G' = (V', E')$ is the restriction of G to the set of nodes $V' = an\,(\{x_i, x_j\} \cup Z)$.

Proof. If there is a path that connects x_i and x_j in G', then the very same path connects x_i and x_j in G. Thus, sufficiency is shown, In order to prove necessity, assume that there is a path in G connecting x_i and x_j. By Lemma 4, we have that all nodes on the path must be ancestors of either x_i, x_j or a collider on the path. Since the path connects x_i, x_j, all its colliders are ancestors of some elements of Z. Thus, all the nodes in the considered path are in V' and such a path connects x_i and x_j in G'. $\square$

Lemma 7. Consider a graph $G = (V, E)$. We have that

$$\mathcal{I}_G(X, Z, Y) \Rightarrow \mathcal{I}_G(X', Z, Y')$$

where

$Y' = \{x \in an\,(X \cup Y \cup Z) \setminus \{X, Z\} | \text{such that } \mathcal{I}(X, Z, x)\}$

$X' = an\,(X \cup Y \cup Z) \setminus \ (Y' \cup Z)$.

Proof. Observe that from Lemma 6 there is no loss of generality by assuming $V = an\,(X \cup Y \cup Z)$. Now, if $x_i \in X$ we have that $\mathcal{I}_G(x_i, Z, Y')$ by the definition of Y'. Instead, if $x_i \in X' \setminus X$, assume by contradiction that there is a path $\pi^{(a)}$ connecting some $x_j \in Y'$ to x_i given Z. Since $x_i \notin Y'$ there must exist a path $\pi^{(b)}$ connecting x_i to some $x_k \in X$. Let

$$\pi^{(a)} = \left(\{x_{\pi_0^{(a)}}, x_{\pi_1^{(a)}}\}, ..., \{x_{\pi^{(a)}_{l^{(a)}-1}}, x_{\pi^{(a)}_{l^{(a)}}}\}\right)$$

$$\pi^{(b)} = \left(\{x_{\pi_0^{(b)}}, x_{\pi_1^{(b)}}\}, ..., \{x_{\pi^{(b)}_{l^{(b)}-1}}, x_{\pi^{(a)}_{l^{(b)}}}\}\right)$$

be these two paths with

$$x_{\pi_0^{(a)}} = x_j; \qquad x_{\pi^{(a)}_{l^{(a)}}} = x_i$$

$$x_{\pi_0^{(b)}} = x_i; \qquad x_{\pi^{(b)}_{l^{(b)}}} = x_k.$$

Let the concatenation of the two paths be

$$\pi = \Big(\{x_{\pi_0^{(a)}}, x_{\pi_1^{(a)}}\}, ..., \{x_{\pi_{p-1}^{(a)}}, x_{\pi_p^{(a)}}\}$$
$$\{x_{\pi_q^{(b)}}, x_{\pi_{q+1}^{(b)}}\}, ..., \{x_{\pi^{(b)}_{l^{(b)}-1}}, x_{\pi^{(a)}_{l^{(b)}}}\}\Big)$$

with $x := x_{\pi_p^{(a)}} = x_{\pi_q^{(b)}}$ for some $0 \leq p \leq l^{(a)}$ and some $0 \leq q \leq l^{(b)}$, defining a path from x_j to x_k. If either $0 = p$ or $q = l^{(b)}$ we have that the path π connects x_j to x_k given Z. This is not possible since $x_j \in Y'$ and $x_k \in X$. Thus it must be that $0 < p$ and $q < l^{(b)}$. All the inner nodes in the concatenated path π are colliders or non-colliders precisely as in the two original paths $\pi^{(a)}$ and $\pi^{(b)}$ with the possible exception of the connecting node $x = x_{\pi_p^{(a)}} = x_{\pi_q^{(b)}}$. Since both $\pi^{(a)}$ and $\pi^{(b)}$ are open paths given Z, the connecting node x is the only node that can block π given Z. If the connecting node x is not a collider on π then it was not a collider on either $\pi^{(a)}$ or $\pi^{(b)}$, thus π is not blocked by Z. If the connecting node x is a collider on π and a collider either on $\pi^{(a)}$ or $\pi^{(b)}$, then π is again not blocked by Z. If

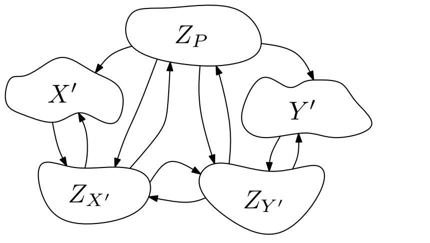


Fig. 2. Schematic representation of the influences between the five subsets of nodes X', Y', $Z_{X'}$, $Z_{Y'}$ and Z_P.

the connecting node x is a collider on π but a non-collider on both $\pi^{(a)}$ and $\pi^{(b)}$, in order to have π blocked by Z we need to have $x \notin an\,(Z)$. Thus $x \in an\,(X)$ or $x \in an\,(Y)$. Assume there is a chain from x to $x_{i'} \in X$. With no loss of generality let $x_{i'} \in X$ be the first node in X in the chain. The chain is not blocked by Z, othewise we would have $x \in an\,(Z)$. Concatenating this chain with $\pi^{(a)}$ gives a path from x_k to $x_{j'} \in X$ that is not blocked by Z. This is a contradiction. By assuming that there is a chain from x to $x_{j'} \in Y$ we obtain a contradiction in a similar way using the path $\pi^{(b)}$. $\square$

4. MAIN RESULTS

The following theorem estabilishes a strong connection between the concept of d-Separation defined on the graph and the notion of Wiener separation that is instead associated with the flow of information through the network.

Theorem 8. In a well-posed LDG with graph G

$$\mathcal{I}_G(X, Z, Y) \Rightarrow \mathcal{I}_W(X, Z, Y).$$

Proof. We have to prove that if X and Y are d-separated by Z according to the graph G, then the two sets of processes X and Y are separated according to the Wiener-separation. From Lemma 6, assume, with no loss of generality, $V = an\,(X \cup Y \cup Z)$. Define X', and Y' as in Lemma 7. By Lemma 7 we have that $\mathcal{I}_{G(X',Z,Y')}$ and also that $X' \subseteq X$, and $Y' \subseteq Y$. Partition the set of processes V in the following sets X', Y', $Z_{X'}$, $Z_{Y'}$, and Z_P where

$$Z_{X'} = ch\,(X')$$
$$Z_{Y'} = ch\,(Y')$$
$$Z_P = Z \setminus (Z_{X'} \cup Z_{Y'}).$$

Thus, reordering the processes, we can write $V = (X^T, Y^T, Z_P^T, Z_{X'}^T, Z_{Y'}^T)^T$. The matrix $H(z)$ that defines the LDG dynamics can now be written as

$$H(z) = \begin{pmatrix} H_{XX} & 0 & H_{XP} & H_{XZ_{X'}} & H_{XZ_{Y'}} \\ 0 & H_{YY} & H_{YP} & H_{YZ_{X'}} & H_{YZ_{Y'}} \\ 0 & 0 & H_{PP} & H_{PZ_{X'}} & H_{PZ_{Y'}} \\ H_{Z_{X'}X} & 0 & H_{Z_{X'}P} & H_{Z_{X'}Z_{X'}} & H_{Z_{X'}Z_{Y'}} \\ 0 & H_{Z_{Y'}Y} & H_{Z_{Y'}P} & H_{Z_{Y'}Z_{X'}} & H_{Z_{Y'}Z_{Y'}} \end{pmatrix}. \quad (7)$$

where the zero-blocks are a consequence of Lemma 5. A graphical representation of the influences between the five subsets X', Y', $Z_{X'}$, $Z_{Y'}$ and Z_P is given in Figure 2.

Exploiting the sparsity pattern of $H(z)$ and the diagonal structure of Φ_e, we find that the inverse of power spectral density has a zero block corresponding to the entries associated with the processes X and Y

$$(\Phi_V)^{-1} = H(z)^* \Phi_e^{-1} H(z) = \begin{pmatrix} * & 0 & * & * & * \\ 0 & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix}$$

where the symbol $*$ denotes a potentially non-zero block. From Lemma 26 in Materassi and Salapaka (2012), we have that the variables X' and Y' are Wiener-separated given Z, namely $\mathcal{I}_W(X', Z, Y')$. Since $X' \subseteq X$ and $Y \subseteq Y'$ we have the assertion. $\square$

The following theorem is the equivalent of Theorem 8 for the causal scenario.

Theorem 9. Consider a causal LDG with graph $G = (V, E)$. Let X, Y and Z three mutually disjoint sets. Define X' and Y' as in Theorem 7. Define $W := V \setminus X' \cup Y'$ and write the LDG dynamics in the form

$$\begin{pmatrix} X' \\ Y' \\ W \end{pmatrix} = \begin{pmatrix} e_{X'} \\ e_{Y'} \\ e_W \end{pmatrix} + \begin{pmatrix} H_{X'X'} & H_{X'Y'} & H_{X'W} \\ H_{Y'X'} & H_{Y'Y'} & H_{Y'W} \\ H_{WX'} & H_{WY'} & H_{WW} \end{pmatrix} \begin{pmatrix} X' \\ Y' \\ W \end{pmatrix}$$

Under the assumption that $(I - H_{X'X'})^{-1}$ and $(I - H_{Y'Y'})^{-1}$ are causal, it holds that $\mathcal{I}_G(X, Z, Y) \Rightarrow \mathcal{I}_{W^+}(X, Z, Y)$.

Proof. We have to prove that if X and Y are d-separated by Z according to the graph G, then the two sets of processes X and Y are separated according to the causal Wiener-separation. From Lemma 6, assume, with no loss of generality, $V = an\,(X \cup Y \cup Z)$. By Lemma 7 we have that $\mathcal{I}_{G(X',Z,Y')}$ and also that $X' \subseteq X$, and $Y' \subseteq Y$. Partition the set of processes V in the following sets X', Y', $Z_{X'}$, $Z_{Y'}$, and Z_P where

$$Z_{X'} = ch\,(X')$$
$$Z_{Y'} = ch\,(Y')$$
$$Z_P = Z \setminus (Z_{X'} \cup Z_{Y'}).$$

Thus, after reordering the processes, we can write $V = (X^T, Y^T, Z_P^T, Z_{X'}^T, Z_{Y'}^T)^T$. Since $(I - H_{X'X'})^{-1}$ and $(I - H_{Y'Y'})^{-1}$ are causal we have no loss of generality by assuming $H_{X'X'} = 0$ and $H_{Y'Y'} = 0$. Indeed, we have that

$$X' = e_{X'} + H_{X'X'}X' + H_{X'\overline{X'}}\overline{X'} \Rightarrow$$
$$X' = (I - H_{X'X'})^{-1} e_X + (I - H_{X'X'})^{-1} H_{X'\overline{X'}}\overline{X'}.$$

A similar argument holds to show that the assumption $H_{Y'Y'} = 0$ does not affect the generality of the proof. The matrix $H(z)$ that defines the LDG dynamics has the same sparsity pattern as in (7) with the addition that $H_{X'X'}(z) = 0$ and $H_{Y'Y'}(z) = 0$. In particular notice that

$$X' = e_{X'} + H_{X'Z_P}Z_P + H_{X'Z_{X'}}Z_{X'} + H_{X'Z_{Y'}}Z_{Y'} \quad (8)$$
$$Y' = e_{Y'} + H_{Y'Z_P}Z_P + H_{Y'Z_{X'}}Z_{X'} + H_{Y'Z_{Y'}}Z_{Y'}. \quad (9)$$

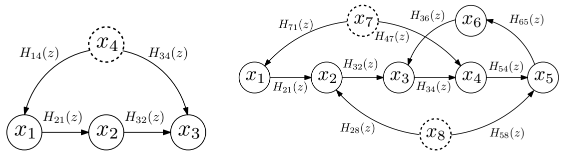
Observe that

$$\arg \min_{u \in \text{c-tf-span}\{Z_P, Z_{X'}, Z_{Y'}\}} \|X' - u\| =$$
$$= H_{X'Z_P}Z_P + H_{X'Z_{X'}}Z_{X'} + H_{X'Z_{Y'}}Z_{Y'} + \arg \min_{u \in \text{c-tf-span}\{Z_P, Z_{X'}, Z_{Y'}\}} \|e_{X'} - u\|.$$

At the same time

$$\arg \min_{u \in \text{c-tf-span}\{Z_P, Z_{X'}, Z_{Y'}, Y'\}} \|X' - u\| =$$
$$= H_{X'Z_P}Z_P + H_{X'Z_{X'}}Z_{X'} + H_{X'Z_{Y'}}Z_{Y'} + \arg \min_{u \in \text{c-tf-span}\{Z_P, Z_{X'}, Z_{Y'}, Y'\}} \|e_{X'} - u\|$$
$$= H_{X'Z_P}Z_P + H_{X'Z_{X'}}Z_{X'} + H_{X'Z_{Y'}}Z_{Y'} + \arg \min_{u \in \text{c-tf-span}\{Z_P, Z_{X'}, Z_{Y'}, e_{Y'}\}} \|e_{X'} - u\|$$

with all these equalities coming as a consequence of (8). Since $e_{X'}$ and $e_{Y'}$ are perpendicular, and $e_{Y'} \in c - tf - span(Z_P, Z_{X'}, Z_{Y'})$ (from Lemma 29 in (Materassi and Salapaka, 2012)), we have that X' and Y' are Wiener separated. $\square$



(a) (b)

Fig. 3. (a) A LDG where the process x_4 plays the role of a confounding process. (b) A LDG with two confounding processes and a loop.

5. APPLICATIONS AND EXAMPLES

5.1 Confounding process

Consider the graph G of Figure 3(a) representing a LDG following the dynamics $x = e + H(z)x$ where the only non-zero entries of $H(z)$ are H_{21}, H_{32}, H_{14} and H_{34}. Assume that the graph is known, but not the transfer functions on the edges. Also assume that only x_1, x_2 and x_3 are observed (but not x_4). The goal is to identify the transfer function $H_{32}(z)$. The task is made difficult by the presence of the confounding process x_4 that is not accessible. However, observe that x_2 and x_4 are d-Separated by $\{x_1\}$, namely $I_G(x_2, \{x_1\}, x_4)$. Let $\hat{x}_{2,1}$ and $\hat{x}_{3,1}$ be respectively the estimate of x_2 and x_3 from x_1 using the non-causal Wiener filter. Also, if x_4 were observable, we could in principle obtain its estimate $\hat{x}_{4,1}$ by Wiener-filtering x_1, as well. However, since x_1 d-Separates x_2 and x_4) we have, from Theorem 8, that $\hat{x}_{2,1}$ and $\hat{x}_{4,1}$ are two non-correlated stochastic processes. Since $\hat{x}_{3,1} = e_3 + H_{32}\hat{x}_{2,1} + H_{42}\hat{x}_{4,1}$ we find that $H_{32}(z) = \frac{\Phi_{\hat{x}_{3,1}}(z)}{\Phi_{\hat{x}_{2,1}\hat{x}_{3,1}}(z)}$.

5.2 Confounding process and a loop

Consider the graph G of Figure 3(b) representing a LDG following the dynamics $x = e + H(z)x$ where the only non-zero entries of $H(z)$ are represented by the edges in the graph. Assume that all processes but x_7 and x_8 are observed and that the goal is to identify $H_{34}(z)$. Since $I_G(x_3, \{x_1, x_2, x_5\}, \{x_7, x_8\})$ we have that the Wiener estimates of x_3 and $\{x_7, x_8\}$ from $\{x_1, x_2, x_5\}$ are not correlated. Thus the effect of the processes $\{x_7, x_8\}$ on x_4 is orthogonal to the effect of the processes x_3, given $\{x_1, x_2, x_5\}$. In a way similar to the previous example we can identify $H_{34}(z)$.

6. CONCLUSIONS

In this paper we have shown how the concept of *d*-Separation (Pearl, 1988) can be usefully extended to networks of stochastic processes interconnected via linear dynamical systems. Thus, many fundamental results from the domain of graphical models can be adapted to the dynamic case, even in presence of feedback loops. The main result is that *d*-Separation is directly related to the sparsity pattern of variations of the Wiener Filter . Under certain hypothesis of well-posedness and regularity, such a relation holds for any topology of the network, even in presence of loops, both in the case of causal and non-causal estimators.

REFERENCES

Acemoglu, D., Dahleh, M., Lobel, I., and Ozdaglar, A. (2011). Bayesian learning in social networks. *The Review of Economic Studies*, 78(4), 1201–1236.

Atalay, E., Hortaçsu, A., Roberts, J., and Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13), 5199.

Bailly, J.S., Monestiez, P., and Lagacherie, P. (2006). Modelling spatial variability along drainage networks with geostatistics. *Mathematical Geology*, 38(5), 515–539.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., and Bressler, S.L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc Natl Acad Sci USA*, 101(26), 9849–9854.

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

Del Vecchio, D., Ninfa, A., and Sontag, E. (2008). Modular cell biology: Retroactivity and insulation. *Nature Molecular Systems Biology*, 4, 161.

Eisen, M., Spellman, P., Brown, P., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, 95(25), 14863–8.

Kailath, T., Sayed, A., and Hassibi, B. (2000). *Linear Estimation*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. The MIT Press.

Koster, J.T. (1999). On the validity of the markov interpretation of path diagrams of gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*, 26(3), 413–431.

Lauritzen, S. (1996). *Graphical models*, volume 17. Oxford University Press, USA.

Materassi, D. and Salapaka, M. (2012). On the problem of reconstructing an unknown topology via locality properties of the wiener filter. *Automatic Control, IEEE Transactions on*, 57(7), 1765–1777.

Materassi, D. and Salapaka, M.V. (2013). Reconstruction of directed acyclic networks of dynamical systems. In *American Control Conference (ACC), 2013*, 4687–4692. IEEE.

Nabi-Abdolyousefi, M. and Mesbahi, M. (2010). Network identification via node knock-out. In *Conference on Decision and Control*, 2239–2244.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Pearl, J. (2000). *Causality: models, reasoning, and inference*, volume 47. Cambridge Univ Press.

Pillonetto, G., Chiuso, A., and De Nicolao, G. (2011). Prediction error identification of linear systems: A non-parametric gaussian regression approach. *Automatica*, 47(2), 291–305.

Quinn, C.J., Kiyavash, N., and Coleman, T.P. (2013). Efficient methods to compute optimal tree approximations of directed information graphs. *IEEE Trans. on Signal Processing*, 61(12), 3173–3182.

Sanandaji, B., Vincent, T., and Wakin, M. (2011). Compressive topology identification of interconnected dynamic systems via clustered orthogonal matching pursuit,. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference.*

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, prediction, and search*, volume 81. The MIT Press.

Van den Hof, P.M., Dankers, A., Heuberger, P.S., and Bombois, X. (2012). Identification of dynamic models in complex networks with prediction error methods-basic methods for consistent module estimates. *Automatica.*

Verma, T. and Pearl, J. (1988). Causal Networks: Semantics and Expressiveness. In *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, 352–359.