Norbert Wiener’s legacy in the study and inference of causation

Donatello Materassi

Abstract—Norbert Wiener has pioneered the study of causality and his ideas are still shaping this active area of research. The present article explores the concept of “inferred causation” and demonstrates the continued impact of Wiener’s work. It shows how the statistical notion of causation introduced by Clive Granger in the 70’s and the counterfactual notion of causation introduced by Judea Pearl in the 90’s can be interpreted and merged under a single unifying framework that is based on Wiener filtering.

Determining whether a given process “causes” another has been a central subject of philosophical debate since even before the time of David Hume and Immanuel Kant. Aside from purely philosophical or gnoseological considerations, practicing scientists are content to be able to infer causation in operative ways. In practical terms, if it is possible to isolate the processes of interest within a controlled laboratory environment, causality can be inferred by performing a sufficient number of experiments. However, a deeper understanding of the meaning of causation is required if the processes can not be perfectly isolated, not all variables can be measured or only passive observations are available. Traditionally, Clive Granger is credited with the introduction of the first quantitative statistical tests to infer causation from passive observations. In Granger’s formulation, one variable is causal to another, if the ability to predict the second variable is significantly improved by incorporating information about the first. As Granger himself acknowledges in his seminal paper, both the mathematical formalization and the fundamental ideas of his notion of causality have been deeply inspired by the contributions of Wiener in the area of estimation and prediction. More recently, a different approach to the study of causality has been developed by the computer scientist and philosopher Judea Pearl. Central to Pearl’s work is the description of probability distributions and conditional independence among random variables by using directed acyclic graphs. Over this class of graphs, Pearl has introduced a semantic language and computational rules to formally describe the concept of causality in a network of partially observed variables. These computational rules are usually known as “do-calculus.” Typically, Pearl’s approach is static, in the sense that there is no time variable defined in the probabilistic model, and is not well-suited to consider feedback loops. Granger’s approach, instead, is based on one-step ahead predictors exploiting the fact that the processes are dynamical systems, but is not capable of dealing with unobserved quantities.

In this article, we provide a combination of both approaches. We first show how, by using certain variations of the Wiener filter, specific sparsity properties can be represented over directed graphs where loops can be present as well. These results are in line with the results obtained by Pearl for his graphical models. Under relatively mild hypotheses, we, as well, provide an interpretation for a generalized (multivariate) notion of Granger causality over these graphs. Finally, we prove that these Wiener filters can be used within the framework provided by do-calculus and seamlessly take into account scenarios where feedback loops are also present.

I. HISTORICAL BACKGROUND AND INTRODUCTION

Determining whether a given process “causes” another has always been a central subject of philosophical debate. Aristotle (384-322 BCE) distinguished categorically between what was to be considered scientific knowledge (scientia) and what was simply belief (opinio). Scientific knowledge was the knowledge of deterministic relations of cause and effect between natural phenomena. Given this knowledge, only logic was necessary to determine the future outcomes of a certain situation or condition. Aristotle’s view of causation as knowledge in purely deterministic terms persisted for several centuries until the time of Descartes (1596-1650) and his follower Malebranche (1638-1715) [1]. However, this notion of causation fails to provide practicing scientists with an operative way to determine from data if a certain process affects another, since it presupposes a perfect knowledge about the state of the world. Indeed, only a perfect knowledge of the state of world, along with a perfect knowledge of its physical laws, would allow us to verify it. Furthermore, most typical causal statements tend to involve general classes of events and not a single one. Statements, such as “smoking cigarettes causes lung cancer” or “higher inflation causes lower unemployment”, do not take into account single events, but collection of events. Indeed, from a purely logical perspective, “smoking cigarettes” and “lung cancer” are neither sufficient nor necessary conditions to each other. Thus, it is easy to see how it would be difficult, if not impossible, to frame statements of this kind under such a strict and deterministic approach. A more modern view of causation was given by David Hume (1711-1776) in his “Treatise on Human Nature” and later in his “An Enquiry Concerning Human Understanding” when he was discussing what Immanuel Kant referred to as the “Problem of Induction”. In Hume’s view, we are left simply observing that one event follows another and they seem connected, but we do not know how or why. We can only rely upon repeated observation (induction) to determine the causal laws implicitly assuming that the future is akin to the past.

“All events seem entirely loose and separate. One event follows another; but we never can observe any tie between them. They seem conjoined but never connected. And as we can have no idea of anything, which never appears to our outward sense or inward sentiment, the necessary conclusion seems to be, that we have no idea of connexion or power at all, and that these words are absolutely without any meaning, when employed either in philosophical reasoning, or in private life”
As an empiricist, he was, though, mostly interested in the problem of how it is possible to determine from observations the existence of “causal connections” between classes of events occurring one after the other. Even with Hume’s empiricism, though, we are still far from discussing about causality in scientific, or better quantitative, terms. Traditionally, the first scientist credited with the introduction of a quantitative notion of causality is Clive Granger (1934-2009). In Granger’s formulation, one variable is causal to another, if the ability to predict the second variable is improved, in a statistically significant way, by incorporating information about the first. However, as Granger has always acknowledged, his formalization was suggested by a paper by Norbert Wiener (1894-1964) [2]. This is Granger’s personal account of how he conceived his notion of causality [3].

“In the early 1960’s I was considering a pair of related stochastic processes which were clearly inter-related and I wanted to know if this relationship could be broken down into a pair of one way relationships. It was suggested to me to look at a definition of causality proposed by a very famous mathematician, Norbert Weiner, so I adapted this definition into a practical form and discussed it.”

[Clive Granger]

Indeed, we find in nuce the very same idea in one of Norbert Wiener’s papers [2].

“[W]e note that [the just derived mathematical expression] represents that part of the normalized innovation of [the first stochastic process] if we have as information the past of [the second stochastic process] as well as the past of the [the first stochastic process].

[...]

This may be considered as a measure [...] of the causality effect of [the second stochastic process on the first one].”

[Norbert Wiener]

More recently, a different approach to the study of causality has been developed by the computer scientist and philosopher Judea Pearl (1936-). Central to Pearl’s work is the description of probability distributions and conditional independence among random variables by using directed acyclic graphs. Over this class of graphs, Pearl has introduced a semantic language and computational rules to formally describe the concept of causality in a network of partially observed variables. These computational rules are usually known as “do-calculus.” Typically, Pearl’s approach is static, in the sense that there is no time variable defined in the probabilistic model, and is not well-suited to consider feedback loops. Granger’s approach, instead, is based on one-step ahead predictors exploiting the fact that the processes are dynamical systems. However, Granger’s formalization is not capable of dealing with unobserved quantities.

In this article, we provide a combination of both approaches. We first show how, by using certain variations of the Wiener filter, specific sparsity properties can be represented over directed graphs where loops can be present as well. These results are in line with the results obtained by Pearl for his graphical models. Under relatively mild hypotheses, we, as well, provide an interpretation for a generalized (multivariate) notion of Granger causality over these graphs. Finally, we prove that these Wiener filters can be used within the framework provided by do-calculus and seamlessly take into account scenarios where feedback loops are also present.

II. LINEAR DYNAMIC GRAPHS AS THE GENERATIVE CLASS OF MODELS

The aim of this section is to introduce a relatively broad class of network models that will provide the formal framework for the derivation of quantitative results about the inference of causal relations. We will refer to this class of mathematical models as Linear Dynamic Graphs (LDGs) [4], [5]. Informally speaking, a LDG is a collection of linear dynamical systems connected according to a directed graph and forced by stationary additive mutually uncorrelated noise processes. We stress that the assumption of exclusively linear relations and additive noise processes are not required to develop most of the theoretical results discussed in the article. However, they provide the advantage of a simpler notation making the exposition clearer.

Linear Dynamic Graphs: motivations and definition

Let us consider two discrete-time stochastic processes $x_i$ and $x_j$. As a working assumption, let $x_i$ and $x_j$ be jointly wide-sense stationary with zero mean. Also, let $H_{ji}(z)$ be a transfer function such that the Laurent expansion

$$H_{ji}(z) = \sum_{t=-\infty}^{\infty} h_{ji}(t)z^{-t}$$

is well defined on an open domain containing the unit circle \( \{ z \in \mathbb{C} : |z| = 1 \} \) for some bi-infinite sequence \( \{ h_{ji}(t) \}_{t \in \mathbb{Z}} \). We also say that $H_{ji}(z)$ is the (bilateral) $Z$-transform of $h_{ji}(t)$ and denote this as $H_{ji}(\cdot) = \mathcal{Z}\{h_{ji}(\cdot)\}$. Let us assume, for the moment, that the process $x_j$ is deterministically obtained by filtering $x_i$ via $H_{ji}(z)$. Namely, this means that

$$x_j(t) = \sum_{\tau=-\infty}^{\infty} h_{ji}(t-\tau)x_i(\tau) \quad \forall \ t \in \mathbb{Z}. \quad (2)$$

We denote Relation (2) as $x_j = H_{ji}(z)x_i$ and represent it in a block diagram as in Figure 1(a). In the following, we will make use of the concept of time-causality.

Definition 1 (Time-causality): A transfer function $H(\cdot) = \mathcal{Z}\{h(\cdot)\}$ is time-causal if $t \leq 0$ implies $h(t) = 0$. If, in
The output processes density mutual influence, we postulate that the power spectral
G Linear Dynamic Graph the following definition.

Behaviors of the processes $e$ represent it in a block diagram as in Figure 1(b).

Fig. 1. The signal $x_j$ as a deterministic function of $x_i$ (a) and the signal $x_j$ as a superposition of its independent behavior $e_j$ and the “influence” from $x_i$ (b).

Addition, $h(0) = 0$ the transfer function is strictly time-causal.

Note that the concept of time-causality is not the same of causality (or influence) that we are going to introduce for an LDG.

From Equation 2 it follows that each realization of $x_i$ determines univocally a realization of $x_j$. Thus, in this case, if $x_i$ is taken as an autonomous and “free” process, we can definitely say that $x_j$ is caused by $x_i$. Now, it is possible to extend this relation in order describe a non-deterministic form of “influence” (or “causality”) of $x_i$ on $x_j$ in the following manner

$$x_j(t) = e_j(t) + \sum_{\tau=-\infty}^{\infty} h_{ji}(t-\tau)x_i(\tau) \quad \forall t \in \mathbb{Z} \quad (3)$$

where $e_j$ is another zero-mean wide sense stationary process that represents the behavior of $x_j$ with no “influence” from $x_i$. We denote this relation as $x_j = e_j + H_{ji}(z)x_i$ and represent it in a block diagram as in Figure 1(b).

Now, let as assume that there are a finite number $n$ of processes and that they can influence each other. The extension to this multivariate scenario can be now described by the relations

$$x_j = e_j + \sum_{i \neq j} H_{ji}(z)x_i \quad \forall j = 1, ..., n.$$

Since the processes $e_1, ..., e_n$ represent the autonomous behaviors of the processes $x_1, ..., x_n$ in the absence of any mutual influence, we postulate that the power spectral density $\Phi_e(z)$ of the stochastic vector $e = (e_1, ..., e_n)^T$ is diagonal and positive definite for all $|z| = 1$. This motivates the following definition.

**Definition 2 (Linear Dynamic Graph):**
A Linear Dynamic Graph $G$ 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 defined in an open set containing the unit circle $\{z \in \mathbb{C} : |z| = 1\}$ and such that $H_{ji}(z) = 0$, for $j = 1, ..., n$.

The output processes $\{x_j\}_{j=1}^n$ of the LDG are defined by the implicit relation

$$x = e + H(z)x \quad (4)$$

where $x = (x_1, ..., x_n)^T$, under the hypothesis that the operator $I - H(z)$ is causally invertible. 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.

Thus, given a LDG $(H(z), e)$ with output processes $x_1, ..., x_n$, we can interpret the matrix $H(z)$ as an adjacency matrix of a graph where a directed edge from the node $x_i$ to the node $x_j$ is present if and only if the entry $H_{ji}(z)$ is not identically zero.

In the following we will make extensive use of these relations between nodes in a graph

**Definition 3:** Given a graph $G = (V, E)$ where $V = \{x_1, ..., x_n\}$ we say that
- $x_i$ is a parent of $x_j$ or $x_j$ is a child of $x_i$ if $(x_i, x_j) \in E$.
- $x_i$ and $x_j$ are spouses if $x_i$ and $x_j$ share a child.

We denote as $Pa(x), Ch(x)$ and $Sp(x)$ the set or parents, children and spouses of a node $x \in V$.

**Problem statement**

Take the class of LDGs as the generative class of models representing, via its associated graph, the relations of direct causality among the stochastic processes. Assume that the observable processes $x_1, ..., x_n$ are the output processes of a LDG. Determine for any pair $(x_i, x_j)$ if there is a relation of direct influence among them.

**III. SPARSITY PROPERTIES OF WIENER FILTERS IN LINEAR DYNAMIC GRAPHS**

In this section we report sparsity properties that hold for certain variations of the Wiener filter [6] when applied to the estimation of a process of a LDG using the other processes of the LDG. Indeed, it happens that, in a LDG, minimum least square estimates of a process $x_i$ do not necessarily depend on all the other processes $\{x_j\}_{j \neq i}$. Instead only a subset of the processes $\{x_j\}_{j \neq i}$ is required to estimate $x_j$. Also, such a subset is strictly related to the graph of the LDG. Thus, we show how the computation from observed data (via spectral or statistical methods) of different flavors of the Wiener filter can provide ways to reconstruct the causal structure of the LDG.

**A. One step-ahead predictor in Linear Dynamic Graphs**

We consider a first result that provides a connection between the structure of the LDG graph and the one step-ahead predictor, the specific variant of the Wiener filter that Clive Granger utilized in the definition of his test for causality.

**Theorem 4:** Consider a well-posed LDG $G = (H(z), e)$ where $\Phi_e(e^{i\omega})$ is positive definite for $\omega \in [-\pi, \pi]$, and each entry of $H(z)$ is strictly time-causal. Let $x_1, ..., x_n$ be the output signals associated with the $n$ nodes of its graph. Define

$$\chi = \left\{ \sum_{i=1}^n W_i(z)x_i : W_i(z) \text{ is strictly time-causal } \forall i \right\}. \quad (5)$$
Consider the problem of approximating the signal $x_j$ with an element $\hat{x}_j \in \chi$, as defined below

$$\min_{\hat{x}_j \in \chi} ||x_j - \hat{x}_j||^2$$  \hfill (6)

where the norm is to be interpreted in terms of the expected square error. Then the optimal solution $\hat{x}_j$ exists, is unique and

$$\hat{x}_j = \sum_{i=1}^{n} \hat{W}_{ji}(z)x_i$$  \hfill (7)

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

\textit{Proof:} See [5].

The interpretation of this result is straightforward. Assume that data have been generated by a LDG where the effect of one node on another always occurs with at least one step of delay ($H_{ji}(z)$ is strictly time-causal for all $i$ and $j$). Let $Pa(x_j)$ be the set of processes directly influencing $x_j$. Then the one-step-ahead predictor to estimate $x_j(t + 1)$ from the knowledge the processes $x_1, ..., x_n$ up to time $t$ makes use only of the processes in $Pa(x_j)$. Since the causality test proposed by Granger [7] is based precisely on one-step-ahead predictors of this form, Theorem 4 shows that such a test consistently determines the causality structure of the LDG. Observe that in Theorem 4 no assumption is made on the structure of the graph. Thus, as long as there are delays in the transfer functions modeling the influence among nodes, the Granger-causality test consistently works, even in the presence of cycles and feedback loops.

B. Wiener-Hopf filter in Linear Dynamic Graphs

A second result involves the so-called Wiener-Hopf filter when used to time-causally estimate a process $x_j$ in a LDG.

\textbf{Theorem 5:} Consider a well-posed LDG $G = (H(z), e)$ where $\Phi_v(e^{\omega})$ is positive definite for $\omega \in [-\pi, \pi]$ and each entry of $H(z)$ is time-causal. Let $x_1, ..., x_n$ be the output signals associated with the LDG. Define

$$\chi_j^* = \left\{ \sum_{i \neq j} W_{ji}(z)x_i : W_{ji}(z) \text{ is time-causal } \forall \ i \right\}$$.  \hfill (8)

Consider the problem of approximating the signal $x_j$ with an element $\hat{x}_j \in \chi_j^*$ (non-causal Wiener filtering), as defined below

$$\min_{\hat{x}_j \in \chi_j^*} ||x_j - \hat{x}_j||^2$$  \hfill (9)

where the norm is to be interpreted in terms of the expected square error. Then, the optimal solution $\hat{x}_j$ exists, is unique and

$$\hat{x}_j = \sum_{i \neq j} \hat{W}_{ji}(z)x_i$$  \hfill (10)

where $\hat{W}_{ji}(z) \neq 0$ implies at least one of the following conditions

- $x_i$ is a parent of $x_j$
- $x_i$ is a child of $x_j$

\textit{Proof:} See [5].

The interpretation of this result is slightly more intricate than the interpretation of Theorem 4. Assume that data have been generated by a LDG where the effect of one node on another always occurs at the same time or with at least one step of delay ($H_{ji}(z)$ is time-causal for all $i$ and $j$). The Wiener-Hopf filter used to estimate $x_j$ from the other nodes makes use only of the nodes influencing $x_j (Pa(x_j))$, the nodes influenced by $x_j (Ch(x_j))$, and the nodes influencing the nodes influenced by $x_j (Sp(x_j))$. The Wiener-Hopf filter can be determined from data, but we find that it does not directly provide the oriented structure of the original network as in the case of Theorem 4 for the Granger causality test. Instead, the sparsity pattern of the Wiener-Hopf filter detects the parents and the children of the node $x_j$ along with the spouses of $x_j$, that are not necessarily connected with it.

IV. SEPARATION IN LINEAR DYNAMIC GRAPHS

A different approach to determine causation has been developed by Judea Pearl [8] for the study of interconnected random variables. Pearl’s framework is static in the sense that, having been developed in the domain of random variables, there is no time variable to be taken into account and as a consequence no concept of time-causality. Central to Pearl’s work is a notion of separation among random variables.

\textbf{Definition 6:} Let $V = \{v_1, ..., v_n\}$ be a set of $n$ random variables. Two random variables $v_i$ and $v_j$ are separated by a set of random variables $V_s \subset V \setminus \{v_i, v_j\}$ if $v_i$ and $v_j$ are independent given the set $V_s$.

Pearl also introduces a graphical representation for the joint probability distribution of a finite set of random variables.

\textbf{Definition 7:} Let $V = \{v_1, ..., v_n\}$ be a set of $n$ random variables and assume that the joint probability distribution can be factorized as

$$p(v_1, ..., v_n) = \prod_{j=1}^{n} p(v_j | Pa(v_j))$$  \hfill (11)

where $p(v_j | Pa(v_j))$ is the conditional probability distribution of $v_j$ given the set of random variables $Pa(v_j) \subseteq \{v_1, ..., v_{j-1}\}$. The directed acyclic graph obtained linking each variable in $Pa(v_j)$ to $v_j$ is the graph associated with the factorization of the distribution.

It would be tempting to say that, in the probabilistic model (11), each variable in $Pa(v_j)$ is a cause for the variable $v_j$. However, as Pearl shows, the factorization depends on the order of the variables. Thus, rearranging the variables in a different order would provide a different graph and, as a consequence, a different causal structure. This, of course, would not acceptable since the very concept of causal structure needs to be independent of the order of the variables.

It is possible to show the following result.

\textbf{Theorem 8:} Let the set random variables $V = \{v_1, ..., v_n\}$ be defined according to a structural equation model

$$v_j = f_j(g_j, Pa_G(v_j)) \quad \forall \ j = 1, ..., n$$  \hfill (12)
where \( g_1, ..., g_n \) are independent random variables and 
\( P_{AG}(v_j) \) is the set of nodes that are parents of \( v_j \) according to a directed acyclic graph \( G \). If \( v_i \notin P_{AG}(v_j) \) and \( v_j \notin P_{AG}(v_i) \), then there exists a set \( V_{ji} \subseteq V \setminus \{v_i, v_j\} \) separating \( v_i \) and \( v_j \).

**Proof:** It is a special instance of Verma-Pearl Theorem. See [9].

Theorem 8 is a fundamental property since it provides a connection between the structure of the generative model (12) and the notion of separation given by Definition 6.

For a LDG we can define an analogous notion of separation based on Wiener filtering.

**Definition 9 (Wiener-separation):** Let \( x_j, x_i \) be two processes and let \( X \) be a set of random processes that does not contain neither \( x_i \) nor \( x_j \). We say that \( x_j \) is Wiener-separated from \( x_i \) given \( X \) if the Wiener filter estimating \( x_j \) from the processes in \( X \) gives the same estimate of the Wiener filter estimating \( x_j \) from the processes in \( X \cup \{x_i\} \).

Given this notion of separation, an analogous of Theorem 8 can be derived for LDGs.

**Theorem 10:** Consider a LDG with output processes \( x_1, ..., x_n \) and associated graph \( G \) that has no directed cycles. If \( x_i \notin P_{AG}(x_j) \) and \( x_j \notin P_{AG}(x_i) \), then there exists a set \( X_{ji} \) separating \( x_i \) and \( x_j \).

**Proof:** See [10].

Theorem 10 provides a quantitative method to determine if the two processes \( x_i \) and \( x_j \) are directly connected in the graph \( G \) associated with the LDG: search for a set of processes \( X_{ji} \) that separates them according to the notion given by the Wiener-Hopf filter. If such a process can not be determined, then either \( x_i \) causes \( x_j \) or \( x_j \) causes \( x_i \).

V. RELATION BETWEEN GRANGER-CAUSALITY AND PEARL-CAUSALITY

In the case of LDGs we can compare the notion of causality provided by Pearl and the one provided by Granger. In Pearl’s framework (reinterpreted via LDGs) there is a relation of direct causality between the process \( x_i \) and the process \( x_j \) if it is not possible to find other processes such that \( x_i \) and \( x_j \) become separated according to the notion given by the Wiener-Hopf filter. Instead, according the definition, a process \( x_i \) Granger-causes a process \( x_j \) if the information in \( x_i \) improves the one-step-ahead prediction of \( x_j \). However, Granger-causality can be also interpreted in terms of a notion of separation that is based on the one-step-prediction instead of the Wiener-Hopf filter. The two notions of causality provide different results in the following sense. Granger-causality is capable of reconstructing the directed structure of a LDG, as long as there is at least one step of delay in the dynamics of the transfer functions. The notion of causality provided by Pearl, in general, is not capable of determining the directed structure of a LDG, but only its associated graph without link orientation. Also, the reconstruction is successful only if the associated graph of the LDG has no cycles.

As a fundamental fact, in the case of LDGs, we have found that different \textit{a priori} assumptions on a LDG model require different notions of separation in order to reconstruct its causal structure. If it is known that the LDG dynamics involves only strictly time-causal transfer functions, then the notion of separation given by the one-step-ahead predictor needs to be considered. If it is known that the LDG dynamics is only time-causal but has no loops, then the notion of separation given by the Wiener-Hopf filter has to be considered instead.

At least in the domain of LDGs, we can interpret Granger causality and Pearl causality as a two notions of separation defined using two different variants of the Wiener filter (because of a-priori assumptions on the LDG).

In light of this interpretation, we now show how a novel notion of separation that can be applied to reconstruct the causal structure of LDGs with not necessarily strictly time-causal dynamics and in presence of loops. In Control Theory particular relevance is given to the concept of well-posedness in the presence of feedback loops [11]. If a feedback loop is not well-posed, it means that it is not possible to define an injective map from the input signals to the output signals of the system. Thus, the behavior of the system is not uniquely specified. It is well-known that a sufficient condition to have a well-posed system is the presence of a delay in every loop [12].

**Theorem 11:** Consider a LDG with output processes \( x_1, ..., x_n \) and associated graph \( G \). Assume that for every directed cycles of the LDG there is at least one strictly causal transfer function. If \( x_i \notin P_{AG}(x_j) \) and \( x_j \notin P_{AG}(x_i) \), then there exist a subset \( X_{ji} \) such that the set of processes \( X_{ji} \cup \{x_1, ..., x_n, x_j, x_i\} \) separates \( x_i \) and \( x_j \) in the Wiener-Hopf sense.

**Proof:** The proof is omitted because of space issues, but techniques similar to the ones used to prove Theorem 5 and Theorem 4 are used in the demonstration.

As we can see, this last results creates a bridge between the notion of causality by Granger and by Pearl defining a notion of separation based on a set \( X_{ji} \) of processes to be filtered in a causal way to infer the presence of direct causality between \( x_i \) and \( x_j \).

VI. CONCLUSIONS

In this article, we have provided a synthesis between two approaches for the study of causality that are considered quite distinct. The key point has been the interpretation of Judea Pearl’s notion of causality in terms of Wiener filtering and Clive Granger’s notion of causality in terms of separation on a projective space. Pearl’s notion of causality is shown to work consistently for networks of dynamical systems, as long as loops are not present. Granger’s notion of causality consistently works for network of dynamical systems, as long as delays are present on every link of the network. This interpretation has allowed to provide a novel criterion to consistently infer causality in well-posed networks of dynamical systems in presence of feedback loops. This criterion comes as a natural combination of the two approaches.
ACKNOWLEDGEMENTS

The author wants to thank Prof. Anil Seth for useful references and Prof. Maxim Raginsky for stimulating discussions.

REFERENCES