We model the strength of a causal relationship with directed information. We propose a new method of discovering causal relationships based on the notion of causal compression. We follow the approach proposed in [4] which motivates sparsity-inducing methods in modelling causality. We present two applications of the proposed method: causal time series segmentation which selects time points capturing the incoming and outgoing causal flow between time points belonging to different signals, and causal bipartite graph recovery. We show that modelling of causality in the adopted set-up only requires estimating the copula density of the data distribution and thus does not depend on its marginals. We evaluate the method on time resolved gene expression data.

Causal graphs. Causal relationships in graphical models are frequently represented with directed acyclic graphs (DAGs), where the arrows can be imbued with causal interpretation in different ways. We follow the approach proposed in [4]. It requires that one be able to perform, or think of performing, an intervention on any node or collection of nodes in the graph. An intervention means that the variable intervened upon has its value set externally, while the influence of any other variables in the DAG (most importantly its parents) upon it is suppressed. This process corresponds to the subset exclusion of directed information between a subset \(S\) of \(X\) and \(Y\) denoted with \(I(X \rightarrow Y; S)\). The directed information is obtained by passively observing the values of \(X\) and \(Y\).\footnote{For the same value of directed information between a subset \(S\) of \(X\) and \(Y\), adding more variables to the subset \(S\) means adding variables which do not exhibit causal (in the sense of Pearlian graphs) relationships with \(X\) and \(Y\).}\footnote{This distribution is contrasted with the observational distribution \(P_X\), which represents the conditional independence relations of the underlying probability distribution of \(X\), its conditional distribution given its parents does not depend on interventions on any other nodes in the DAG. Pearlian DAGs are an intuitive extension of conditional independence representation in graphical models to causality: the absence of a direct causal relationship between two nodes implies the absence of a direct causal relationship between them.}

We model the strength of a causal relationship with directed information\footnote{Directed information is a measure of the influence of a chosen set of variables on the rest of the system.} (a concept also introduced as causal conditioning\footnote{Causal conditioning is a measure of the influence of a chosen set of variables on the rest of the system.}), directed stochastic kernels\footnote{Directed stochastic kernels are measures of the influence of a chosen set of variables on the rest of the system.}, and by other authors\footnote{Directed stochastic kernels are measures of the influence of a chosen set of variables on the rest of the system.}. We show that choosing the most sparse time series representation is equivalent to main results. We show that choosing the most sparse time series representation is equivalent to excluding the nodes that do not contribute to the direct causal relationships in the Pearl graph, i.e. for \(A, B \subseteq X\), \(A \cap B = \emptyset\) [8]:

\[
I(A, B \rightarrow Y) = I(A \rightarrow Y) \iff I(B \rightarrow Y | A) = 0
\]

For the same value of directed information between a subset \(S\) of \(X\) and \(Y\), adding more variables which do not exhibit causal (in the sense of Pearlian graphs) relationships with \(X\) and \(Y\).
relations with $Y$ other than via the original $S$. Therefore, the optimal causal compression at a given level of directed information is ensured by the sparsity of the compressed representation of $X$, i.e. by selecting as few nodes as possible. Note that this can be interpreted in the spirit of Granger causality: the variables in $X$ that are not selected by the sparsity requirement do not Granger-cause the effect $Y$.

We subsequently demonstrate how to use the principle of causal compression to circumvent the estimation of a full causal network and compute two partial sub-structures of it, presented schematically in Figure 1. The first one is the causal segmentation of time points of one time series into those that exhibit outgoing or incoming causal flow (orange and green nodes in Figure 1 respectively) to the other time series and those involved in instantaneous information exchange (blue nodes in Figure 1). Another sub-structure is causal bipartite graph estimation, e.g. computing a mixed bipartite graph between the two time series, where the arrows mean causal dependence and edges mean instantaneous information exchange. We achieve this by defining and solving a LASSO-type constrained optimisation problem that finds a sparse representation of a set of nodes such that directed information is optimised.

We also show that for continuous $(X, Y)$, any causal relationship described with directed information only depends on the entropy of copula density of $(X, Y)$ [8]. This means that for inference we only have to estimate the copula part of the distribution. In particular, for Gaussian distributed data only the correlation matrices have to be identified. More importantly, modelling of causality in the framework of Pearlian graphs only requires knowing the copula structure of the modelled data and is independent of their marginals.

**Experiments.** We evaluate our method on a human hepatitis C virus (HCV) dataset containing time-resolved gene expression profiles from patients with chronic HCV genotype 1 infection [7]. Gene expression was profiled at six time points after initiation of treatment with pegylated alpha interferon and ribavirin. For our analysis, we focused on two genes that are known to have a crucial interacting role in interferon signalling, namely the transcription factor STAT1 and the interferon-induced antiviral gene IFIT3. Based on the observed decrease in HCV RNA levels on the last day, patients were labelled to have a “marked” (27 patients) or “poor” response to treatment (25 patients). The analysis was carried out separately for the “marked” and the “poor” responders, see Figure 2. There are pronounced differences between the two groups: both groups show causal pre-treatment/post-treatment interactions, but for the marked responders, the influence of initial IFIT3 on late STAT1 values is much more prominent. This might be particularly interesting, since pre-treatment expression levels of interferon-induced genes are known to be strong predictors of treatment response [2], but the underlying mechanism of this effect is largely unknown.

---

**Figure 1:** Direct computation of causal segmentation and causal bipartite graph estimation.

**Figure 2:** Time-resolved gene expression data from HCV patients: reconstructed causal graphs for the groups of poor and marked responders.
References


