Causality testing: A Data Compression Framework

Aditi Kathpalia and Nithin Nagaraj
National Institute of Advanced Studies,
Indian Institute of Science Campus,
Bengaluru 560012, India

(Dated: October 13, 2017)

Causality testing, the act of determining cause and effect from measurements, is widely used in physics, climatology, neuroscience, econometrics and other disciplines. As a result, a large number of causality testing methods based on various principles have been developed. Causal relationships in complex systems are typically accompanied by entropic exchanges which are encoded in patterns of dynamical measurements. A data compression algorithm which can extract these encoded patterns could be used for inferring these relations. This motivates us to propose, for the first time, a generic causality testing framework based on data compression. The framework unifies existing causality testing methods and enables us to innovate a novel Compression-Complexity Causality measure. This measure is rigorously tested on simulated and real-world time series and is found to overcome the limitations of Granger Causality and Transfer Entropy, especially for noisy and non-synchronous measurements. Additionally, it gives insight on the ‘kind’ of causal influence between input time series by the notions of positive and negative causality.

Keywords: Causality testing, data compression, compression-complexity, negative causality.

Causality testing is used widely in various disciplines including neuroscience [1], physics [2], climatology [3], econometrics [4], epidemiology [5]. A number of causality testing methods exist which can be applied to estimate the magnitude and direction of causality between two time series, some of which are listed in Fig. 1. These methods make very different model assumptions about input time series, making an appropriate choice often difficult for a given context. Thus, there is a need for a unifying framework to explain the working of these measures and suitably guide their application.

According to Wiener [6], if incorporating the past of a time series $Y$ helps to improve the prediction of a time series $X$, then $Y$ causes $X$. Existing methods are based on notions of improved predictability (Granger Causality and its variations), reduction of uncertainty (Transfer Entropy, Information Flow), dynamical modeling (Dynamical Causal Modeling) and estimation based on proximity in attractor manifold (Convergent Cross Mapping) (refer Fig. 1). All these notions are closely related to information transfer of one form or the other. By establishing the relation of information to the reality of our world, the basis of causality on information transfer is increasingly being shown to have a rigorous physical foundation [2, 8]. In [7], information transfer in complex systems (based on Shannon Entropy) is fundamentally linked to energy and entropy (Boltzmann) flows. In case of living and other complex systems, information is created by self-organization, resulting in a re-ordering of entropy with a reduction in their interior and a corresponding increase in the external environment. Causal relationships based on these entropic exchanges are encoded in patterns of dynamical measurements. A ‘data compression’ framework that captures these patterns seems to be the most natural approach to extract these causal relationships.

Inspired by decades of work in data compression [9], we propose for the first time, a generic framework for causality testing. Data compression is concerned with encoding information either by way of modeling statistical redundancy (eg., Huffman coding [10]) or learning patterns algorithmically (eg., Lempel-Ziv coding [11]) with the aim of reducing resources to store or transmit data. This framework is well founded mathematically owing to a close link between data compression, information entropy (Shannon’s source coding theorem [12]) and algorithmic complexity (Kolmogorov complexity [13]). Fig. 1 gives the block diagram for the framework with a list of possible choices (not exhaustive) for each block below. The table describes the work-flow of various causality testing methods indicating their choice for each block. All these methods clearly fit within the proposed framework.

The framework makes no assumptions about the nature of data, is flexible and easily configurable for a given application by an appropriate choice for each block. Furthermore, the framework lends itself to the invention of novel methods of causality testing. In this work, we propose a new measure – Compression-Complexity Causality (CCC) which is described below.

The minimum description length principle [21] formalizes the Occam’s razor and states that the best hypothesis (model and its parameters) for a given set of data is the one that leads to its best compression. Extending this principle for causality estimation, if the compressibility of time series $X$ remains unchanged even upon incorporating information from time series $Y$, then we
Input Data

\[ X \rightarrow Y \]

Filtering, De-trending, Binning, Fourier, Wavelet, Kernel feature space, Linear/Non-Linear Transform

Pre-processing/Transform

Model

Quantizer

Entropy Coding

Testing Criteria

Estimated Causality

| Causality Methods                  | Brief Description                                                      |
|------------------------------------|------------------------------------------------------------------------|
| Granger Causality (GC) [14]        | Autoregressive (AR) \rightarrow F-statistic, Hypothesis testing       |
| GC using Fourier and Wavelet Transform [15] | Fourier/ Wavelet Transform \rightarrow AR \rightarrow F-statistic, Hypothesis-testing |
| Kernel (Non-linear) GC [16]        | Kernel feature space \rightarrow AR \rightarrow F-statistic, Hypothesis-testing |
| Transfer Entropy (TE) [17]         | Binning \rightarrow Markov \rightarrow Entropy rate                   |
| Dynamical Causal Modelling (DCM) [18] | Non-linear state-space model \rightarrow Bayesian criteria            |
| Convergent Cross Mapping (CCM) [19] | Time-delay embedding \rightarrow Non-linear attractor reconstruction   |
| Information Flow [20]              | Differentiable vector fields \rightarrow Rate of information Flow      |
| Compression-Complexity Causality (CCC) | Binning \rightarrow Pair substitution \rightarrow Compression-complexity rate |
| Zip Causality (ZC)*                | Binning \rightarrow Dictionary-based \rightarrow Lempel-Ziv coding \rightarrow Zipped file sizes |

FIG. 1: A unifying data compression framework for causality testing. Top: Block diagram of the proposed framework indicating various stages of causality testing. Bottom: Table depicting the flow for various existing and proposed (in bold) causality testing methods. *ZC is under development.

Conclude that there is no causal influence from \( Y \) to \( X \) (implying that time series \( Y \) has no role in modeling \( X \)). However, if there is a \textit{change in compressibility} of \( Y \) when \( Y \) is included in its model, then we infer a causality from \( Y \) to \( X \).

CCC uses lossless data compression algorithms, e.g. Lempel-Ziv (LZ) [11, 22] and Effort-to-Compress (ETC) [23], to estimate compressibility using the notion of \textit{compression-complexity} [24]. While either LZ/ETC or any other compression-complexity measure could be used to compute CCC, in this work we use ETC as it has been found to perform better than LZ for short and noisy time series [25]. The given series are first binned — converted to a sequence of symbols using ‘\( B \)’ uniformly sized bins for the application of these complexity measures. For binned time series \( X \) and \( Y \) of length \( N \), to determine whether \( Y \) causes \( X \) or not, we consider a moving window \( \Delta X \) of length \( w \) and define compression-complexity rates as follows:

\[
CC(\Delta X|X_{\text{past}}) = ETC(X_{\text{past}} + \Delta X) - ETC(X_{\text{past}}), \tag{1}
\]

\[
CC(\Delta X|X_{\text{past}} Y_{\text{past}}) = ETC(X_{\text{past}} + \Delta X, Y_{\text{past}} + \Delta X) - ETC(X_{\text{past}}, Y_{\text{past}}), \tag{2}
\]

where the compressibility of \( \Delta X \) is estimated based on windows of immediate past \( L \) values, \( X_{\text{past}} \) and \( Y_{\text{past}} \), taken from \( X \) and \( Y \) respectively (‘+’ refers to appending). Eq. \( 1 \) gives the compression-complexity rate defined as the effort-to-compress \( \Delta X \), knowing the recent past of \( X \) alone. Eq. \( 2 \) is the compression-complexity rate for \( \Delta X \) knowing the recent pasts of both \( X \) and \( Y \). We now define Compression-Complexity Causality \( CCC_{Y \rightarrow X} \) as

\[
CCC_{Y \rightarrow X} = CC(\Delta X|X_{\text{past}}) - CC(\Delta X|X_{\text{past}}, Y_{\text{past}}), \tag{3}
\]

which is the difference between the two time averaged compression-complexity rates over the entire length of the time series with the window \( \Delta X \) being slid by a step-size of \( \delta \). If \( CC(\Delta X|X_{\text{past}}, Y_{\text{past}}) \approx CC(\Delta X|X_{\text{past}}) \), then \( CCC_{Y \rightarrow X} \) is statistically zero, implying no causal influence from \( Y \) to \( X \). If \( CCC_{Y \rightarrow X} \) is statistically significant different from zero, then we infer that \( Y \) causes \( X \).

Our formulation has a striking resemblance to Transfer Entropy [17]. In fact, the terms \( CC(\Delta X|X_{\text{past}}, Y_{\text{past}}) \) and \( CC(\Delta X|X_{\text{past}}) \) asymptotically \((N \rightarrow \infty)\) approach entropy rates used in TE formulation (see Eq. \( 3 \) and \( 4 \) of [17]) for stationary ergodic processes when CC is computed using an optimal lossless data compression algorithm [12, 13].

However, there are important differences between TE and CCC. In TE, \( Y \) is said to cause \( X \) if there is a reduction in entropy rate of \( X \) when \( Y \) is included in the model of \( X \). However, TE is limited by the fact...
that (i) the model strictly assumes Markovian property which may not be valid for the given data, (ii) entropy rate of $X$ when $Y$ is included can never increase (conditioning always reduces entropy). In contrast, for CCC, (i) the model is non-linear and generic since it is based on optimal lossless data compression algorithms, (ii) compression-complexity rate of $X$ when $Y$ is included can either decrease, indicating positive causality, or increase, indicating negative causality from $Y$ to $X$.

The notion of positive and negative causality, which we propose, parallels that of positive and negative correlation. It is a richer characterization of causality than discussed before in literature. Consider the following two cases. Case (A): Minimally coupled autoregressive (AR) processes: $X(t) = aX(t-1) + cY(t-1) + \epsilon_{X,t}$, $Y(t) = bY(t-1) + \epsilon_{Y,t}$, where $a = 0.9$, $b = 0.8$, $c = 0.8$, $t = 1$ to 1000. Noise terms, $\epsilon_{Y}, \epsilon_{X}$ are Gaussian with mean zero and $\eta$ follows standard normal distribution. Here, $CCC_{Y \rightarrow X} = 0.0782$, $CCC_{X \rightarrow Y} = 0.0004$ (settings: $L = 150$, $w = 10$, $\delta = 80$, $B = 2$). Case (B): Non-linearly coupled deterministic processes: $X(t) = \sin^3(t/10)$, $Y(t) = \sin^3(t/50)$, $t = 1$ to 1000. Here, $CCC_{Y \rightarrow X} = -0.0732$, $CCC_{X \rightarrow Y} = -0.0441$ (settings: $L = 100$, $w = 50$, $\delta = 50$, $B = 8$). In both cases, $X$ and $Y$ depend linearly on their past. In (A), $X$ depends linearly on the past of $Y$ while in (B), this dependence is non-linear. For (A), the kind of information transferred from $Y$ to $X$ is such that $CC(\Delta X|X_{\text{past}}, Y_{\text{past}}) < CC(\Delta X|X_{\text{past}})$ whereas in (B), $CC(\Delta X|X_{\text{past}}, Y_{\text{past}}) > CC(\Delta X|X_{\text{past}})$. From this, we infer that compressibility of $\Delta X$ is reduced (as compression-complexity is increased), due to non-linear influence of $Y$ in case (B). Here, $CCC_{Y \rightarrow X} < 0$, and we say that $Y$ negatively causes $X$. On the other hand, for case (A), $CCC_{Y \rightarrow X} > 0$, and we say that $Y$ positively causes $X$. Now, we can not only speak of $Y$ causing $X$, but also infer the kind of information that is transferred from $Y$ to $X$ based on the sign of $CCC_{Y \rightarrow X}$.

Testing on simulations: The performance of CCC was compared with that of TE and GC for the case of minimally coupled AR processes simulated as defined before. Spurious causalities using GC and TE in case of noise and low temporal resolution have been discussed in literature \[20,28\]. Here, we move a step ahead and present results for non-uniformly sampled/ non-synchronous measurements common in real-world physiological data acquisition due to jitters/ motion-artifacts as well as in economics \[29\]. Mean causality estimated for 50 trials using the three measures with increasing noise intensity ($\nu$), are shown in Fig. 2 (left column), and with increasing $\alpha$, which is the percentage of non-uniform sampling in $X$, while $\nu = 0.07$, are shown in Fig. 2 (right column). CCC settings used: $L = 150$, $w = 10$, $\delta = 80$, $B = 2$, TE settings used: $B = 6$. For TE estimation, Markovian models of order 1 are assumed throughout this paper. CCC estimates positive causality from $Y$ to $X$ and is statistically zero in the opposite direction. The values are stable for all cases but show a mildly increasing trend for $Y$ to $X$ causation when $\alpha$ is increased. In contrast, both TE and GC show confounding values of estimated causality in the two directions for increasing $\alpha$.

The second test involved simulation of linearly and non-linearly coupled chaotic tent maps where the independent process, $Y(t) = 2Y(t-1)$ if $0 \leq Y(t-1) < 1/2$ and $Y(t) = 2 - 2Y(t-1)$ if $1/2 \leq Y(t-1) \leq 1$. For linear coupling, the dependent process, $X(t) = \epsilon Y(t-1) + (1 - \epsilon)h(t)$ where $h(t) = 2X(t-1)$ if $0 \leq X(t-1) < 1/2$ and $h(t) = 2 - 2X(t-1)$ if $1/2 \leq X(t-1) \leq 1$. For non-linear coupling, $X(t) = 2f(t)$ if $0 \leq f(t) < 1/2$ and $X(t) = 2 - 2f(t)$ if $1/2 \leq f(t) \leq 1$, where $f(t) = \epsilon Y(t-1) + (1 - \epsilon)X(t-1)$. The strength of coupling, $\epsilon$ is varied from 0 to 0.9 for both simulations. Fig. 3 shows the mean values of causality for 50 trials estimated using CCC and TE for linear and non-linear coupling (left and right columns respectively). CCC settings used: $L = 50$, $w = 20$, $\delta = 20$, $B = 8$, TE settings used: $B = 8$. The assumption of a linear model for estimation of GC was proved to be erroneous for most trials and hence GC values are not displayed. As $\epsilon$ is increased for both linear and non-linear coupling, $TE_{Y \rightarrow X}$ increases in the positive direction and then falls to zero when the two series become completely synchronized at $\epsilon = 0.5$. The trend of the magnitude of CCC values is similar to TE, how-
ever, $CCC_{Y \rightarrow X}$ increment is in negative direction. A non-linear influence of $Y$ on $X$ results in negative causality. We also note from Fig. 4 that standard deviation values for CCC are less than that for TE in the direction of causation and comparable in the opposite direction.

**Testing on real-world data:** CCC was applied to estimate causality on measurements from two real-world systems and compared with TE. System (a) comprised of short time series for dynamics of a complex ecosystem, with 71 point recording of predator (Didinium) and prey (Paramecium) populations, reported in [30] and originally acquired for [31], with first 9 points from each series removed to eliminate transients (Fig. 5(a)). CCC settings used: $L = 40$, $w = 15$, $\delta = 4$, $B = 8$. CCC is seen to aptly capture the higher (and direct) causal influence from predator to prey population and lower influence in the opposite direction (see Fig. 5). The latter is expected, owing to the indirect effect of the change in prey population on predator. CCC results are in line with that obtained using CCM [19]. TE, on the other hand, fails to capture the correct causality direction.

System (b) comprised of raw single-unit neuronal membrane potential recordings ($V$, in 10V) of squid giant axon in response to stimulus current ($I$, in V, 1V=5 $\mu$A/cm$^2$), recorded in [32] and made available by [33]. We test for the causation from $I$ to $V$ for three axons (1 trial each) labeled ‘a3t01’, ‘a5t01’ and ‘a7t01’, extracting 5000 points from each recording. CCC settings used: $L = 50$, $w = 10$, $\delta = 20$, $B = 4$, TE settings used: $B = 4$. We find that $CCC_{I \rightarrow V} < CCC_{V \rightarrow I} < 0$ for the three axons (Fig. 5), indicating negative causality in both directions. This implies bidirectional non-linear dependence between $I$ and $V$. However, since $|CCC_{I \rightarrow V}| > |CCC_{V \rightarrow I}|$, the expected behavior that stimulus current drives membrane potential is well supported by CCC. The behaviour of TE across the three axons is inconsistent and hence inconclusive.
ity (CCC) measure that outperforms GC and TE for noisy and non-uniformly sampled simulated stochastic data and real-world time series. CCC can be negative, as in the case of linearly and non-linearly coupled chaotic maps, giving rise to the notion of negative causality. Such an insightful characterization is absent in existing measures including TE.

Existing methods do not use all the blocks in the framework (Fig. 1). It is hoped that our framework would inspire novel measures other than CCC to be invented in the future. We indicate here a few possibilities. As an example, we have started developing Zip Causality (ZC) testing, based on the popular LZ77 compression algorithm [11] (used by zip and gzip). ZC measure has all the blocks in the framework except for the quantizer. Our initial testing of ZC measure gave promising preliminary results, but needs further experimentation. Another example of a novel measure would be a compressed sensing [34] based measure for sparse signals which can efficiently model a wide class of compressible signals. A more futuristic example would be a Deep Learning (DL) based causality measure that is also conceivable within our framework. These futuristic measures suggested here are merely illustrative and limited only by our own imagination.

We provide free open access to the CCC MATLAB toolbox developed as a part of this work. It can be downloaded from the following URL: https://sites.google.com/site/nithinnagaraj2/journal/ccc

[1] A. K. Seth, A. B. Barrett, and L. Barnett, Journal of Neuroscience 35, 3293 (2015).
[2] J. P. Crutchfield and C. R. Shalizi, Physical review E 59, 275 (1999).
[3] A. Stips, D. Macias, C. Coughlan, E. Garcia-Gorriz, and X. San Liang, Scientific reports 6 (2016).
[4] C. R. Chen, P. P. Lung, and N. S. Tay, Review of Financial Economics 14, 1 (2005).
[5] K. Sachs, O. Perez, D. Pe’er, D. A. Lauffenburger, and G. P. Nolan, Science 308, 523 (2005).
[6] N. Wiener, Modern mathematics for engineers 1, 125 (1956).
[7] W. Ebeling, The European Physical Journal Special Topics 226, 161 (2017).
[8] X. San Liang, Physical Review E 94, 052201 (2016).
[9] D. Salomon and G. Motta, Handbook of data compression (Springer Science & Business Media, 2010).
[10] D. A. Huffman, Proceedings of the IRE 40, 1098 (1952).
[11] J. Ziv and A. Lempel, IEEE Transactions on information theory 23, 337 (1977).
[12] C. E. Shannon, Bell Syst. Tech. J. 27, 623 (1948).
[13] T. M. Cover and J. A. Thomas, Elements of information theory (John Wiley & Sons, 2012).
[14] C. Granger, Econometrica 37, 424 (1969).
[15] S. P. Ellner, Proceedings of the Royal Society of London B: Biological Sciences 267, 1611 (2000).
[16] D. Paydarfar, D. B. Forger, and J. R. Clay, Journal of Neurophysiology 96, 3338 (2006).
[17] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, Circulation 101, e215 (2000).
[18] E. J. Cand` es and M. B. Wakin, IEEE signal processing magazine 25, 21 (2008).