Estimating complex causal structure: There and back again

Jaroslav Hlinka[1,2]

[1]Institute of Computer Science, Academy of Sciences, Prague, Czech Republic
[2]National Institute of Mental Health, Klecany, Czech Republic

Acknowledgement: JH is supported by the Czech Health Research Council projects NV15-29835A and NV15-33250A.

Detecting causal interactions among system parts

In the study of complex systems, one of the key quests is that for being able to disentangle the structure of interactions between the observed subsystems. Based on the intuition that cause precedes the effect, we can use the formalization by Granger (1969): X causal with respect to Y, iff the X_t improves the prediction of $Y_{t+\delta t}$.
In particular for a linear stochastic process: Let X_t and Y_t be two jointly stationary stochastic processes satisfying:

$$X_t = \sum_{j=1}^{\infty} a_{1j} X_{t-j} + \epsilon_{1t}, \qquad \mathrm{var}(\epsilon_{1t}) = \Sigma_1, \tag{1}$$

$$Y_t = \sum_{j=1}^{\infty} d_{1j} Y_{t-j} + \eta_{1t}, \qquad \mathrm{var}(\eta_{1t}) = \Gamma_1, \tag{2}$$

with the joint autoregressive representation:

$$X_t = \sum_{j=1}^{\infty} a_{2j} X_{t-j} + \sum_{j=1}^{\infty} b_{2j} Y_{t-j} + \epsilon_{2t}, \tag{3}$$

$$Y_t = \sum_{j=1}^{\infty} c_{2j} X_{t-j} + \sum_{j=1}^{\infty} d_{2j} Y_{t-j} + \eta_{2t}, \tag{4}$$

where the covariance matrix of the noise terms is:

$$\Sigma = \mathrm{Cov}\begin{pmatrix}\epsilon_{2t}\\ \eta_{2t}\end{pmatrix} = \begin{pmatrix}\Sigma_2 & \Upsilon_2\\ \Upsilon_2 & \Gamma_2\end{pmatrix}. \tag{5}$$

The causal influence from Y to X is then quantified based on the decrease in the residual model variance when we include the past of Y in the model of X:

$$F_{Y\to X} = \ln \frac{\Sigma_1}{\Sigma_2}. \tag{6}$$

Similarly, the causal influence from X to Y is defined as:

$$F_{X\to Y} = \ln \frac{\Gamma_1}{\Gamma_2}. \tag{7}$$

Causality – nonlinear generalization

Generalization of linear Granger causality: *transfer entropy* (TE, Schreiber, 2001): X causes Y iff the knowledge of past of X decreases the uncertainty about Y.
For two discrete random variables X, Y with sets of values Ξ and Υ and probability distribution functions (PDFs) $p(x), p(y)$ and joint PDF $p(x,y)$, the Shannon entropy $H(X)$ is defined as

$$H(X) = -\sum_{x\in\Xi} p(x)\log p(x), \tag{8}$$

The conditional entropy $H(X|Y)$ of X given Y is

$$H(X|Y) = -\sum_{x\in\Xi}\sum_{y\in\Upsilon} p(x,y)\log p(x|y). \tag{9}$$

The amount of common information contained in the variables X and Y is quantified by the mutual information $I(X;Y)$ defined as

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{10}$$

The conditional mutual information $I(X;Y|Z)$ of the variables X, Y given the variable Z is given as

$$I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z). \tag{11}$$

Transfer entropy from process X_t to process Y_t then corresponds to the conditional mutual information between X_t and Y_{t+1} conditional on Y_t:

$$T_{X\to Y} = I(X_t, Y_{t+1}|Y_t). \tag{12}$$

Interestingly, it can be shown that for linear Gaussian processes, transfer entropy is equivalent to linear Granger causality, up to a multiplicative factor (Barnett et al., 2009):

$$\mathcal{T}_{X\to Y} = \frac{1}{2}\mathcal{F}_{X\to Y}. \tag{13}$$

Technical Problem of Transfer Entropy: not robust estimation

Granger causality outperforms Transfer Entropy (Hlinka et al., 2013).

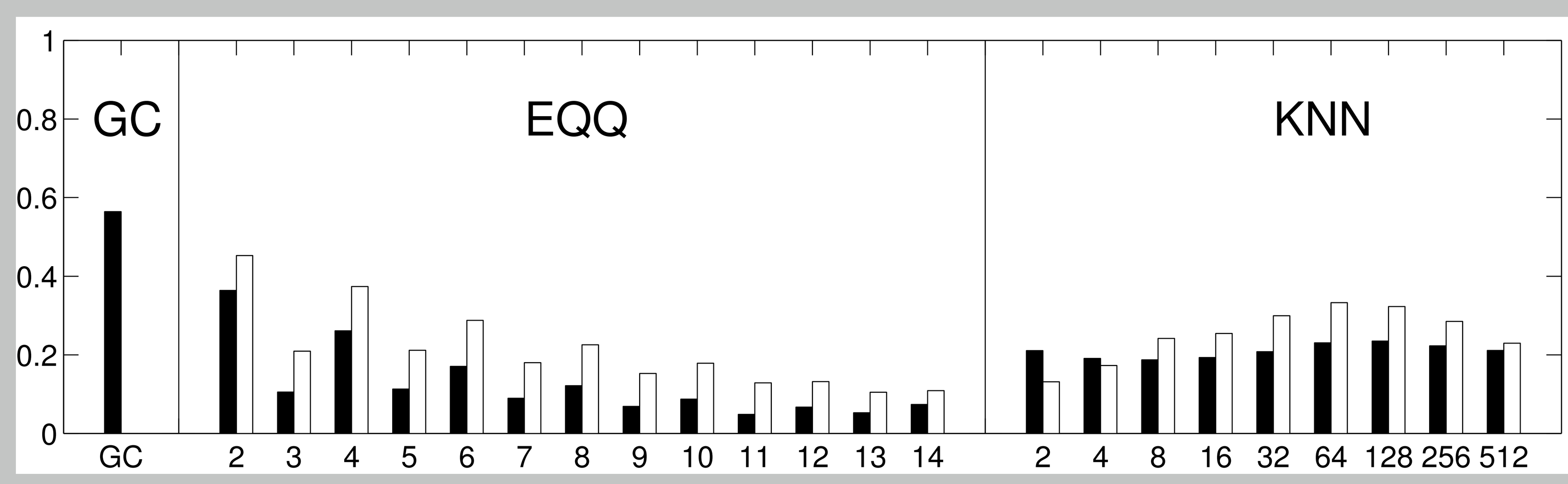

Figure: Reliability of causality network detection using different causality estimators, and the similarity to linear causality network estimates. Black: average Spearman's correlation across all 15 pairs of decades. White: average Spearman's correlation of nonlinear causality network and linear causality network across 6 decades.

Moreover, many complex system timeseries close to 'linearity': for example brain activity (Hlinka et al, 2011), climate data (Hlinka et al, 2015).

Principal problem of TE: obscurity wrt. system equations

Solution: Nonlinear Granger causality by linearization

We recently suggested approach (Wahl et al., 2016) to generalize Granger causality concept to diffusion processes given by the Langevin equation.

$$\dot{Y}_i = h_i(\mathbf{Y}) + [g(\mathbf{Y})]_{ij}\Gamma_j(t),$$
$$\langle\Gamma_i(t)\rangle \equiv 0,\ \langle\Gamma_i(t)\Gamma_j(t')\rangle = 2\delta_{ij}\delta(t-t'),\ i = 1,2. \tag{14}$$

With corresponding Fokker-Planck equation

$$\frac{\partial P(\mathbf{y},t)}{\partial t} = -\sum_{i=1}^{N}\frac{\partial}{\partial y_i}\left[D_i^{(1)}(\mathbf{y})P(\mathbf{y},t)\right] + \sum_{i,j=1}^{N}\frac{\partial^2}{\partial y_i\partial y_j}\left[D_{ij}^{(2)}(\mathbf{y})P(\mathbf{y},t)\right]. \tag{15}$$

It relates to the Langevin equation for Itô's interpretation of the stochastic integral by $\mathbf{h} = \mathbf{D}^{(1)}$ and $gg^T = D^{(2)}$. A local approximation in state space around $\mathbf{y}_0$ yields an Ornstein-Uhlenbeck process

$$\dot{\tilde{Y}}_i = \sum_{j=1}^{N}\gamma_{ij}\tilde{Y}_j + \Theta_i(t),\ i = 1,2,$$
$$\langle\Theta_i(t)\rangle \equiv 0,\ \langle\Theta_i(t)\Theta_j(t')\rangle = \theta_{ij}\delta(t-t'). \tag{16}$$

Temporal discretization gives rise to an autoregressive process with well defined Granger causality. Granger-causality maps $\mathbf{y}\to\mathcal{F}_{\tilde{Y}_2\to\tilde{Y}_1}(\mathbf{y})$ can be drawn in the subset M_{stable} of state space with locally stable OU process. Global Granger-causality defined as the average of local causality over M_{stable} with respect to the stationary probability density $P^*(\mathbf{y})$:

$$\mathcal{I}_{Y_2\to Y_1} := \frac{\int_{M_{\mathrm{stable}}}\mathcal{F}_{\tilde{Y}_2\to\tilde{Y}_1}(\mathbf{y})P^*(\mathbf{y})\,\mathrm{d}y_1\mathrm{d}y_2}{\int_{M_{\mathrm{stable}}}P^*(\mathbf{y})\,\mathrm{d}y_1\mathrm{d}y_2}. \tag{17}$$

Summary: Convenient properties of the generalized concept

- allows local mapping of causality throughout state-space
- consistent with linear Granger causality on linear stochastic processes
- invariant under the same transformations

References

Barnett, L.; Barrett, A. B. & Seth, A. K.: Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables, Physical Review Letters, 2009, 103, 238701

Granger, C. W. Investigating causal relations by econometric model and cross-spectral methods, Econometrica, 1969, 37, 424-438

Hlinka, J.; Palus, M.; Vejmelka, M.; Mantini, D. and Corbetta, M. Functional connectivity in resting-state fMRI: Is linear correlation sufficient? NeuroImage, 2011, 54, 2218-2225

Hlinka, J.; Hartman, D.; Vejmelka, M.; Runge, J.; Marwan, N.; Kurths, J. and Palus, M. Reliability of Inference of Directed Climate Networks Using Conditional Mutual Information Entropy, 2013, 15, 2023-2045

Hlinka, J.; Hartman, D.; Vejmelka, M.; Novotna, D. and Palus, M. Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity. Climate Dynamics, 2014, 42, 1873-1886

Schreiber, T. Measuring information transfer, Physical Review Letters, 2000, 85, 461-464

Wahl, B.; Feudel, U.; Hlinka, J.; Wächter, M.; Peinke, J.; Freund, J. A. Granger-causality maps of diffusion processes, Physical Review E, 2016, 93, 022213