CORRENTROPY BASED GRANGER CAUSALITY

*Il Park, José C. Príncipe**

Computational NeuroEngineering Laboratory
University of Florida
Gainesville, FL 32611, USA

ABSTRACT

We propose a novel nonlinear extension to Granger causality. It is derived from a nonlinear mapping of a stochastic process using the recently introduced generalized correlation measure called correntropy. The method is demonstrated by detecting the direction of coupling in a chaotic system where the original Granger causality failed.

Index Terms— Nonlinear systems, Causality

1. INTRODUCTION

Resolving the dependencies among time series is challenging because the system dynamics mixes information through time. One of the most successful tools is the Granger causality which utilizes the temporal order of dependence of a bivariate stochastic process (for review of causality measures see [1]). Given two time series $X(t)$ and $Y(t)$, if the one step prediction error of $X(t)$ using univariate autoregressive model is significantly larger than the bivariate autoregressive model of both time series, $Y(t)$ is said to be Granger causal to $X(t)$. Although the original formulation can be robustly applied to various signals, the method as explained is limited to linear causality because of the linear prediction scheme.

We propose a nonlinear extension of Granger causality by mapping the original stochastic process into a novel stochastic process induced by *centered correntropy* [2]. Correntropy is a recently introduced generalized correlation measure that can capture the higher-order statistics of random variables. It was Parzen that showed the possibility; for any positive semi-definite kernel on random variables, there exists a mapping from a stochastic process to a Gaussian process [3]. We can perform Granger causality in the nonlinearly related Gaussian process that is effectively a nonlinear causality measure of the original stochastic process. In the algorithm we present, the mapping is implicitly used–only the covariance functions of the transformed bivariate process is required, and moreover it is immune to the approximations needed to get back to the input space [4].

*This work is supported by NSF grants ECS-0422718 and ECS-0601271.

2. GRANGER CAUSALITY AND ITS EXTENSIONS

Let $X(n)$ and $Y(n)$ be real valued discrete time stochastic process for $n \in \mathbb{N}$. We will model the stochastic process as autoregressive (AR) processes. Univariate and bivariate AR models of order L can be written as

$$\begin{aligned}
X(n) &= \sum_{k=1}^{L} w_X(k)X(n-k) + \epsilon_X \\
Y(n) &= \sum_{k=1}^{L} w_Y(k)Y(n-k) + \epsilon_Y \\
X(n) &= \sum_{k=1}^{L} w_{XX}(k)X(n-k) \\
&\quad + \sum_{k=1}^{L} w_{XY}(k)Y(n-k) + \epsilon_{XY} \\
Y(n) &= \sum_{k=1}^{L} w_{YY}(k)Y(n-k) \\
&\quad + \sum_{k=1}^{L} w_{YX}(k)X(n-k) + \epsilon_{YX} \qquad (1)
\end{aligned}$$

where w_X, w_Y denote the coefficients for the univariate case, while $w_{XX}, w_{XY}, w_{YY}, w_{YX}$ corresponds to the bivariate case. The random variables for the one step prediction errors are denoted as ϵ. If $\mathrm{var}(\epsilon_{XY}) < \mathrm{var}(\epsilon_X)$ in a significant way, then $Y(n)$ is Granger causal to $X(n)$ and similarly for $\mathrm{var}(\epsilon_{YX}) < \mathrm{var}(\epsilon_Y)$. In the next section, we will use the same scheme for a different stochastic process.

Above described Granger causality assumes that the signals are jointly wide-sense-stationary and fit an AR model. In reality these assumptions can fail either due to non-stationarity, or due to nonlinear dynamics or coupling that cannot be well approximated by an AR model. There have been several extensions to Granger causality to overcome these problems. In order to deal with non-stationarity, averaging over locally linear models is a possible solution [5]. For nonlinear dynamics, one can introduce a nonlinear predictor with invariant properties [6]. And finally there are nonparametric methods that directly deals with conditional joint distributions or information flow [1]. The proposed method which will be introduced in the next section still requires the assumption of non-stationarity, but is able to deal with mild nonlinearity.

3. CORRENTROPY GRANGER CAUSALITY

Given a family of random variables (or vectors) $\{\mathbf{X}(t)\}$, and a symmetric positive semi-definite kernel $V(\mathbf{X}(t), \mathbf{X}(s)) : (\Omega \to \mathbb{R}) \times (\Omega \to \mathbb{R}) \to \mathbb{R}$, a unique reproducing kernel Hilbert space (RKHS) $\mathcal{H}_V$ can be induced following Parzen [3]. Parzen also proves that there exists a Gaussian process $\{\mathbf{Z}(t)\}$ where the covariance function is given by $\mathrm{E}_{\mathbf{Z}_t,\mathbf{Z}_s}[\mathbf{Z}_t\mathbf{Z}_s] = V(\mathbf{X}(t), \mathbf{X}(s))$ [3]. Let $V(\cdot,\cdot)$ be the correntropy as defined [7],

$$V(\mathbf{X}(t), \mathbf{X}(s)) = \mathrm{E}_{\mathbf{X}_t\mathbf{X}_s}[K(\mathbf{X}(t), \mathbf{X}(s))] \tag{2}$$

where $K(\cdot,\cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a symmetric positive semi-definite kernel. By choosing K to be a nonlinear kernel such as the Gaussian kernel, we obtain a Gaussian process that is nonlinearly related to the original stochastic process. Although closely related, the induced nonlinear process should not be confused with the stochastic process in the RKHS mapped only via the deterministic kernel K as widely used in kernel methods.

We can quantify the time series $\mathbf{Z}(t)$ via autoregressive modeling. Then, knowing the mean and covariance of the nonlinearly related random process, and assuming wide-sense stationarity and ergodicity, we can perform linear regression via least squares to derive the variance of the error [8]. In prediction, due to the fact that the desired signal is in the RKHS, the variance of the error can be estimated without calculating the system output in the input space which cannot be done without approximation in the correntropy filter proposed by Pokharel and coworkers [4]. Therefore, for Granger causality the correntropy RKHS is particularly useful.

If we choose the Gaussian kernel which is widely used in kernel methods, the Taylor expansion can be used to show the relation to the moments. Assuming a correntropy sense stationary process [2],

$$\begin{aligned} V_{\mathbf{XY}}(\tau) &= \mathrm{E}\left[G(\mathbf{X}(t), \mathbf{Y}(t+\tau))\right] \quad (3)\\ &= \frac{1}{\sqrt{2\pi}\sigma}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2\sigma^2)^k k!}\mathrm{E}_{\mathbf{XY}}\left[(\mathbf{X}(t)-\mathbf{Y}(t+\tau))^2\right]\\ &= \frac{1}{\sqrt{2\pi}\sigma}\left[1 - \frac{\mathrm{E}[\mathbf{X}(t)^2]+\mathrm{E}[\mathbf{Y}(t)^2]}{2\sigma^2}\right.\\ &\qquad \left. + \frac{\mathrm{E}[\mathbf{X}(t)\mathbf{Y}(t+\tau)]}{\sigma^2} + \cdots\right], \quad (4)\end{aligned}$$

where $G(x, y) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-y)^2}{2\sigma^2})$ is the Gaussian kernel with kernel size σ. From Eq. (4), we can infer that if the kernel size σ is large compared with the scale of the signal, the value of correntropy is dominated by the cross-correlation (the second order term of the expansion).

Since correntropy does not yield a zero mean process, we remove the mean by centering in the RKHS, obtaining effectively the centered correntropy. Approximating the expected value by time averages, centered correntropy is estimated by

$$U_{\mathbf{XY}}(\tau) = \frac{1}{N}\sum_{i=1}^{N} G(X_i, Y_{i-\tau}) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G(X_i, Y_j), \tag{5}$$

where X_i and Y_i denotes the i-th sample of the time series. The centered correntropy function is a positive semi-definite function [2], therefore the above derivation of nonlinear Granger causality can be directly applied, that is, we use the centered correntropy function as the covariance function and estimate the AR coefficients via solving the Yule-Walker equation. Furthermore, the AR coefficients and the covariance function is enough for the estimation of error variance of the model given by Eq. (1).

4. RESULT

To demonstrate the utility of the proposed method, we chose the coupled nonlinear dynamical system proposed by Chen and coworkers, for which the authors showed that the linear Granger causality fails [5].

$$\begin{aligned} x(n) &= 3.4x(n-1)\left(1 - x^2(n-1)\right)e^{-x^2(n-1)} + 0.8x(n-2)\\ y(n) &= 3.4y(n-1)\left(1 - y^2(n-1)\right)e^{-y^2(n-1)}\\ &\qquad + 0.5y(n-2) + cx^2(n-2), \quad (6)\end{aligned}$$

where c is the coupling strength. From the equations it is obvious that $x(n)$ is driving $y(n)$ but not vice versa. Also note that the coupling term is nonlinear. A sample signal is plotted in Fig. 1. The signal $x(n)$ has two distinct regions and switches between them which resembles the behavior of the well known Lorenz attractor. The reconstructed attractor is shown in Fig. 2.

The centered correntropy is a real valued function that can be estimated given samples by Eq. (6), therefore we can plot the autocovariogram to intuitively understand the nonlinear transformation as demonstrated in Fig. 3. As a guideline, the kernel size given by the Silverman's rule of thumb is 0.22 since this attractor has a variance of 0.8 when the coupling is 0.5. When the kernel size is 10, the normalized centered correntropy is almost identical to the normal autocorrelation and cross-correlation function as expected from Eq. (4). However, as the kernel size decreases, the relative amplitude decay in autocorrentropy is faster than autocorrelation. Moreover, strikingly we observed that the centered cross correntropy function is often inverted for small kernel sizes. This is because the Gaussian kernel weights the neighborhood of the

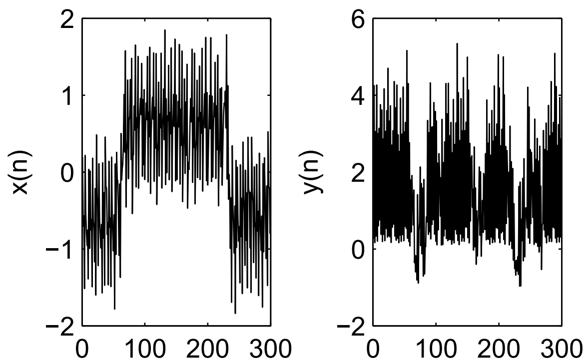


Fig. 1. A sample of time series Eq. (6). The coupling constant was 0.8. Note that $x(n)$ has two visually distinct value range. Transient behavior is removed and the index is arbitrary.

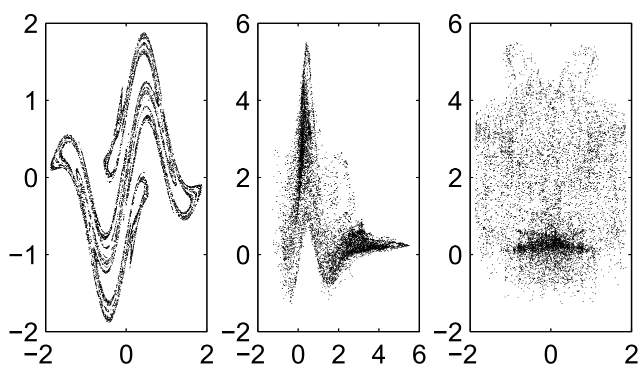


Fig. 2. The time embedded attractor of the time series Eq. (6). *(Left)* The driving time series. $x(n)$ versus $x(n-1)$ is plotted. *(Middle)* The driven time series with coupling $c = 0.8$. *(Right)* Instantaneous correlation $x(n)$ versus $y(n)$.

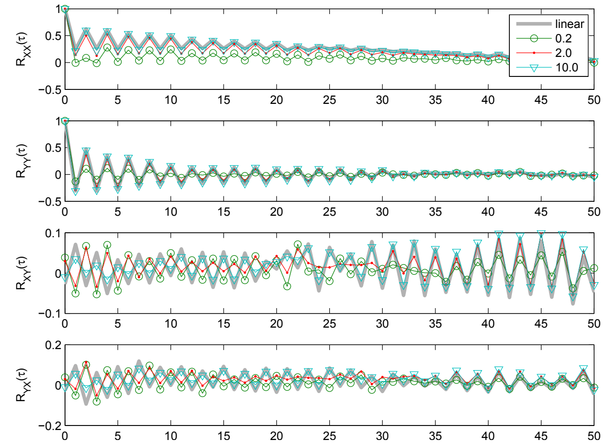


Fig. 3. Autocorrelation function and centered correntropy for different kernel sizes $\sigma = 0.2, 2, 10$. From top to bottom, the graphs correspond to normalized autocorrelation/autocorrentropy of $x(n)$, $y(n)$, cross-correlation/cross-correntropy of $x(n)y(n+\tau)$, $x(n)y(n-\tau)$.

signal and ignores if the distance is large. As seen in Fig. 1, the coupled chaotic system is highly uncorrelated for most of the time. The small kernel size effectively weights primarily regions where the differences in signal samples are close to zero (unlike correlation that weights small and large differences equally), therefore it provides a very different quantification of similarity between signals.

We compare the proposed method against (linear) Granger causality method. Figure 5 shows the performance of the methods. The correntropy based Granger causality can detect the direction of coupling correctly if it is stronger than 0.3 while the classic Granger causality does not show significant deviation from the baseline for all coupling values we tested.

The surrogate baseline was generated by shuffling the pairs of $x(n)$ and $y(n)$ for different trials so that the temporal structure of individual time series is maintained but the coupling is destroyed. Note that in the ideal situation, the baseline in principle should have the value of 1. However, in the proposed method, the coupling of the system changes the scale and structure of the attractors, which means the kernel size should be appropriately modified.

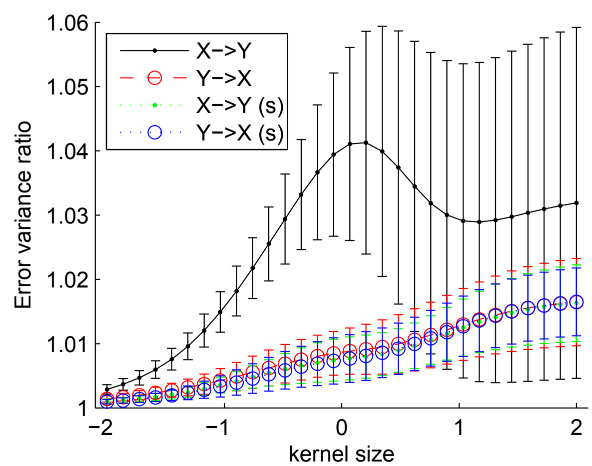


Fig. 4. Effect of kernel size σ. X-axis is on 10 base log scale. Note that the significance is higher for smaller kernel sizes.

The high variability of the linear Granger causality observed in Fig. 5(a) is also observable in the proposed nonlinear extension when the kernel size σ gets larger as shown in Fig. 4. The kernel size essentially determines the scale of analysis. In the case of the dynamical system under investigation, a smaller kernel size is preferred. However, it should be noted that smaller kernel sizes would generally require more data for estimation.

5. DISCUSSION

We proposed a nonlinear extension to Granger causality by simply substituting the covariance function with centered correntropy function. It implicitly utilizes the nonlinearly related stochastic process induced by correntropy. We demonstrated that the proposed method can detect causality of a nonlinear dynamical system where the linear Granger causality failed.

The kernel size parameter σ is the only additional free parameter, which is related to the scale of the dynamical system. The proposed method approaches the linear Granger causality when the kernel size is large. When the kernel size is smaller than the dynamic range of the signal, the correntropy weights the samples that are close to each other. In this sense, the proposed method is similar to local linear averaging methods for Granger causality [5], but without the assumptions and complexity of estimating the neighborhoods over the attractor. However, the selection of optimal kernel size and model order are left as open problems. [1]

6. REFERENCES

[1] Katerina Hlavàčkovà-Schindler, Milan Palus, Martin Vejmelka, and Joydeep Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, pp. 1–46, Mar. 2007.

[2] Jian-Wu Xu, *Nonlinear signal processing based on reproducing kernel Hilbert space*, Ph.D. thesis, University of Florida, 2007.

[3] Emanuel Parzen, *Time Series Analysis Papers*, Holden-Day, 1967.

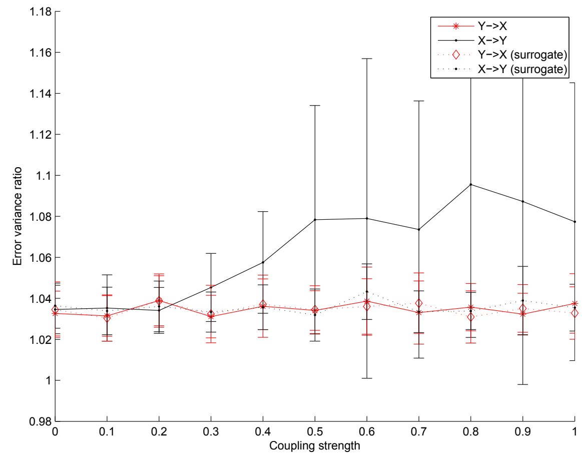
[4] Puskal P. Pokharel, Jian-Wu Xu, Deniz Erdogmus, and José C. Príncipe, "A closed form solution for a nonlinear Wiener filter," in *International Conference on Acoustics, Speech and Signal Processing*, 2006.

[5] Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, and Mingzhou Ding, "Analyzing multiple nonlinear time series with extended granger causality," *Physics Letters A*, vol. 324, pp. 26–35, 2004.

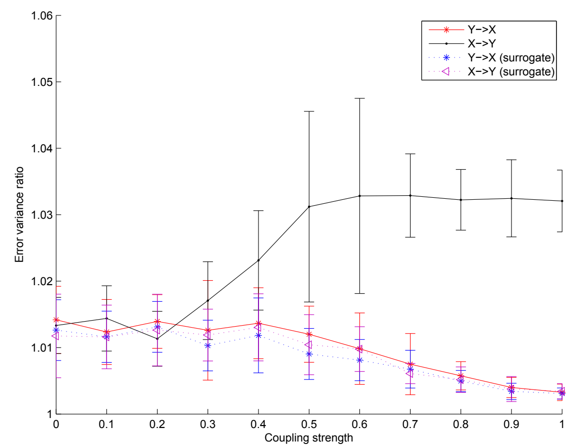
[6] Nicola Ancona and Sebastiano Stramaglia, "An invariance property of predictors in kernel-induced hypothesis spaces," *Neural Computation*, vol. 18, pp. 749–759, 2006.

[7] I. Santamaria, P. Pokharel, and J. C. Príncipe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006.

[8] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, New Jersey, 4th edition, 2002.



(a) Linear Granger causality



(b) Correntropy based Granger causality

Fig. 5. Performance comparison between linear and the proposed nonlinear Granger causality. Higher error variance ratio means stronger causality. The actual causality direction is from X to Y (see Eq.(6)). The coupling was 0.5 and order 16 AR models were used for estimation.

[1]Memming thanks JianWu Xu and Mingzhou Ding.