Detecting Nonlinear Granger Causality via the kernelization of Partial Directed Coherence

Lucas Massaroppe*
University of São Paulo, São Paulo, Brazil — lucasmassaroppe@usp.br

Luiz A. Baccalá
University of São Paulo, São Paulo, Brazil — baccala@lcs.poli.usp.br

Abstract

Though by now classic for inferring causal relations between time series, linear vector autoregressive models fail to capture certain types of nonlinear coupling. Here we explore a new concept *kernel-nonlinear-Partial Directed Coherence* and its ability to overcome those limitations, via a simulated example of its asymptotic behaviour.

Keywords: Nonlinear-Granger causality; Partial Directed Coherence; Inference; Detection; Kernel Processing.

1 Introduction

Thanks to the well understood properties of linear vector autoregressive (VAR) models, convergence and asymptotics among them (Lütkepohl, 2005), they have long been the methods of choice for infering Granger causality (Granger, 1969) even when nonlinear time series are under investigation (Schelter et al., 2006) despite their proven inability to portray polynomial couplings of even power (Massaroppe et al., 2011). Alternatives exist, both parametric (Faes et al., 2008; He et al., 2014; Pereda et al., 2005) and nonparametric (Hlaváčková-Schindler et al., 2007; Marinazzo et al., 2008; Paluš et al., 2001), they share the weakness of requiring many data points for proper convergence or, sometimes, require explicit prior knowledge of the coupling structure. Additional shortcomings stem from having to deal with many parameters and/or the minimization of nonconvex functionals.

Here we examine a new parametric approach, termed *kernel-nolinear-Partial Directed Coherence* (Kannan and Tangirala, 2014) based on computing the partial directed coherence (PDC) (Baccalá and Sameshima, 2001) of properly 'kernelized' data. The present approach differs from Kannan and Tangirala (2014) in two important ways: (a) the use of polynomial kernels instead of the gaussian kernel adopted by the latter and (b) in the use of implicit kernel mapping generalizing Kumar and Jawahar (2007) to the multivariate time series case.

After a brief recap (Section 2), simulation of a simple model borrowed from Massaroppe et al. (2011) illustrates the potential viability (Section 3) of applying our recently developed asymptotic causality detection criteria (Baccalá et al., 2013; Takahashi et al., 2010) to this new form of PDC. We end with a brief discussion (Section 4).

2 The Method

To extend partial directed coherence (PDC) (Baccalá and Sameshima, 2001) to capture more general couplings, we propose to represent the observed data $x_i(n)$ $(1 \leq i \leq N)$ through a *Kernel Vector Autoregressive* (*kVAR*) model

$$\boldsymbol{\phi}[\mathbf{x}(n)] = \sum_{r=1}^{p} \mathbf{A}^{\phi}(r)\boldsymbol{\phi}[\mathbf{x}(n-r)] + \boldsymbol{\vartheta}^{\phi}(n), \{\boldsymbol{\vartheta}^{\phi}(n)\}_{n\in\mathbb{Z}} \sim \text{i.i.d.WN}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}^{\phi}}\right), \quad (1)$$

where $\phi(\cdot)$ stands for a mapping (Parzen, 1959) taking the data from its input space $\boldsymbol{\mathcal{X}}$ to a suitably defined feature space $\mathbf{F}$, by chosing $\phi(\cdot)$ so that $\mathbb{E}\{\phi[x_i(n)]\phi[x_i(n-k)]\} = \mathbb{E}\{\kappa[x_i(n), x_i(n-k)]\}$ where $\kappa(\cdot)$ is a Mercer kernel.
This leads naturally to the definition of *kernel-nonlinear-PDC* as

$$_{\kappa\eta}\pi_{ij}(f) = \frac{\overline{A}^{\boldsymbol{\phi}}_{ij}(f) \Big/ \sqrt{\sigma^{\phi}_{ii}}}{\sqrt{\bar{\mathbf{a}}^{\boldsymbol{\phi}\mathsf{H}}_j \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\vartheta}^{\phi}} \bar{\mathbf{a}}^{\boldsymbol{\phi}}_j}}, \tag{2}$$

where

$$\overline{A}^{\boldsymbol{\phi}}_{ij}(f) = \delta_{ij} - \sum_{l=1}^{p} a^{\phi}_{ij}(l)\mathrm{e}^{-\mathbf{i}2\pi f l}, \ (\mathbf{i}^2 = -1), \tag{3}$$

with $a^{\phi}_{ij}(l)$ representing adequately fit kVAR model coefficients, while $\bar{\mathbf{a}}^{\boldsymbol{\phi}}_j(f)$ represent the columns of the $[\overline{A}^{\boldsymbol{\phi}}_{ij}(f)]$ matrix.
To fit (1) we extend the univariate model fitting approach from Kumar and Jawahar (2007) to the multivariate case by rewriting it as:

$$\boldsymbol{\Phi}_i = \mathbf{A}^{\boldsymbol{\phi}} \left(\mathbf{X}\mathbf{J}_i\right)^{\top} + \boldsymbol{\Theta}^{\boldsymbol{\phi}}_i \tag{4}$$

where

$$\begin{gathered}\boldsymbol{\Phi}_i = (\boldsymbol{\phi}[\mathbf{x}(1)], \cdots, \boldsymbol{\phi}[\mathbf{x}(N)]) \quad (K \times N), \quad \mathbf{A}^{\boldsymbol{\phi}} = (\mathbf{A}^{\boldsymbol{\phi}}(1), \cdots, \mathbf{A}^{\boldsymbol{\phi}}(p)) \quad (K \times Kp), \\ \mathbf{X}(n) = \begin{bmatrix} \phi[x(n)] \\ \vdots \\ \phi[x(n-p+1)] \end{bmatrix} \qquad (Kp \times 1), \quad \mathbf{J}_i = \begin{bmatrix} \mathbf{0}_{(i-Kp-1)\times Kp} \\ \mathbf{I}_{Kp} \\ \mathbf{0}_{(N-i+1)\times Kp} \end{bmatrix},\end{gathered} \tag{5}$$

which upon the application of the **vec** operator

$$\boldsymbol{\varphi}_i = \mathrm{vec}(\boldsymbol{\Phi}_i) \quad (KN \times 1), \quad \boldsymbol{\alpha}^{\boldsymbol{\phi}} = \mathrm{vec}(\mathbf{A}^{\boldsymbol{\phi}}) \quad (K^2 p \times 1), \quad \boldsymbol{\theta}^{\boldsymbol{\phi}}_i = \mathrm{vec}(\boldsymbol{\Theta}^{\boldsymbol{\phi}}_i) \quad (KN \times 1), \tag{6}$$

leads to

$$\boldsymbol{\varphi}_i = (\mathbf{X}\mathbf{J}_i \otimes \mathbf{I}_K)\boldsymbol{\alpha}^{\boldsymbol{\phi}} + \boldsymbol{\theta}^{\boldsymbol{\phi}}_i, \tag{7}$$

whose least squares solution

$$\widehat{\boldsymbol{\alpha}}^{\boldsymbol{\phi}} = \left(\sum_{i=p+1}^{N} \mathbf{J}_i^{\top}\mathbf{X}^{\top}\mathbf{X}\mathbf{J}_i\right)^{-1} \left(\sum_{i=p+1}^{N} \mathbf{J}_i^{\top}\mathbf{X}^{\top}\boldsymbol{\varphi}_i\right) = \left(\sum_{i=p+1}^{N} \mathbf{J}_i^{\top}\mathbf{K}\mathbf{J}_i\right)^{-1} \left(\sum_{i=p+1}^{N} \mathbf{J}_i^{\top}\mathbf{K}^i\right), \tag{8}$$

where $\mathbf{K} = \mathbf{X}^{\top}\mathbf{X}$ is the block-Kernel matrix and the matrix $\mathbf{X}^{\top}\boldsymbol{\varphi}_i$ equals $\mathbf{K}$'s i-th column — $\mathbf{K}^i$. In computing $\mathbf{K}$, the usual kernel centering procedure (Kumar and Jawahar, 2007; Shawe-Taylor and Cristianini, 2004) was adopted together with proper ridge regression stabilization procedures, with $\lambda = 10^{-6}$. Model orders were estimated using a suitably adapted Hannan-Quinn criterion.
Note that (4,7) generalize the univariate time series formulations of Kallas et al. (2013); Kumar and Jawahar (2007), while (2) employs the information theoretical formulation of PDC (Takahashi et al., 2010), which more closely portrays the notion of information flow between series (Baccalá and Sameshima, 2014a,b).

3 Numerical Illustration

To gauge approach performance we simulated the following system:

$$\begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} 2R\cos(2\pi f_0) & 0 \\ 0 & \xi \end{bmatrix} \begin{bmatrix} x_1(n-1) \\ x_2(n-1) \end{bmatrix} + \begin{bmatrix} -R^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(n-2) \\ x_2(n-2) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ c & 0 \end{bmatrix} \begin{bmatrix} x_1^2(n-1) \\ x_2^2(n-1) \end{bmatrix} + \begin{bmatrix} w_1(n) \\ w_2(n) \end{bmatrix}, \tag{9}$$

with $R = 0.99$ chosen to produce a sharp resonance at $f_0 = 0.1$ and for $\xi = -0.9$, where $w_i(n)$ are independent zero mean and unit variance innovations. The quadratic coupling constant c was in turn allowed

to take different values ($\{0.10, 0.50, 1.00\}$) and realization lengths N in $\{100, 250, 500, 750, 1000\}$ taken after sufficiently long burn-in times as in Massaroppe et al. (2011), which also shows a sample run of this example and its allied result of linear PDC's inability to capture coupling without prior kernelization.
Figure 1 shows the result of applying the detection criteria described in Baccalá et al. (2013) to the computed $knPDC$ of a typical realization under a bi-quadratic kernel, showing that it is possible detect the effect $\phi[x_1(n)] \rightarrow \phi[x_2(n)]$ at $f = f_0 = 0.1$. Threshold computations were made based on the covariance structure from the data before kernelization.

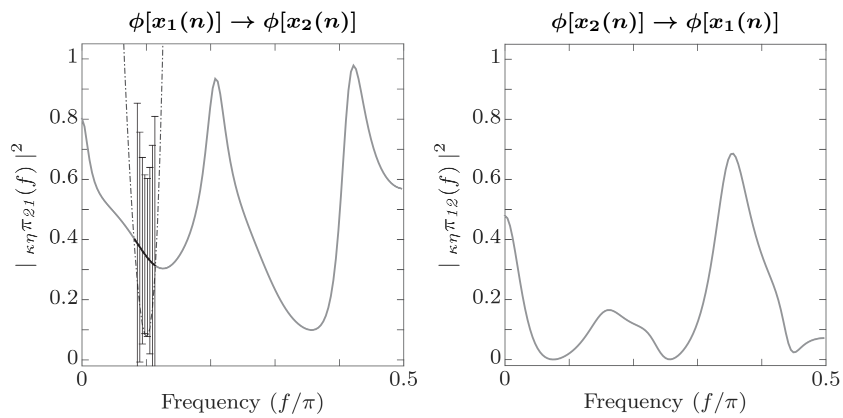

Figure 1: knPDC for a sample run of (9) under $\kappa(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^\top \mathbf{y})^4$ capable of capturing the right direction ($\phi[x_1(n)] \rightarrow \phi[x_2(n)]$) of connectivity in the feature space ($c = 1.00$ and $N = 750$) at confidence level of $\alpha = 0.01$.

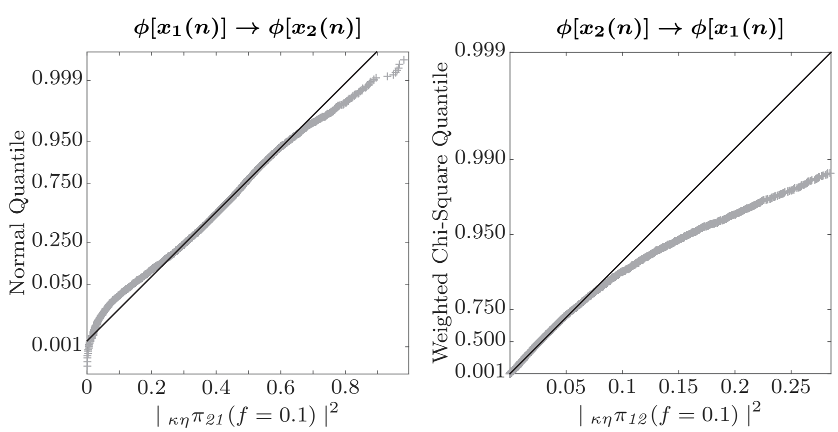

Figure 2: Comparison between the actually observed knPDC with that inferred based on the sample run knPDC of Figure 1.

To investigate Baccalá et al. (2013)'s detection capabilities we performed a Monte Carlo simulation comprising $10,000$ repetitions whose distribution behaviour is exemplified in Figure 3 for $c = 1.00$ and $N = 750$. More details for other parameters are shown on Table 1 for both quadratic and bi-quadratic kernels.
Whereas Figure 2 contrasts knPDC ensemble behaviour with that predicted , based on the sample run result of Figure 1 at $f = f_0 = 0.1$, Figure 3 portrays the actual ensemble data distribution against the best fits to a normal distribution ($\phi[x_1(n)] \rightarrow \phi[x_2(n)]$) and an $a\chi^2_\nu$ distribution ($a = 0.05, \nu = 1.08$) for $\phi[x_2(n)] \rightarrow \phi[x_1(n)]$. Both Figures 2 and 3 point to a slightly high predicted **FP** rate at $N = 750$ which is confirmed by Table 1 results where one immediatley sees the **FP** convergence to 1% as N increases.

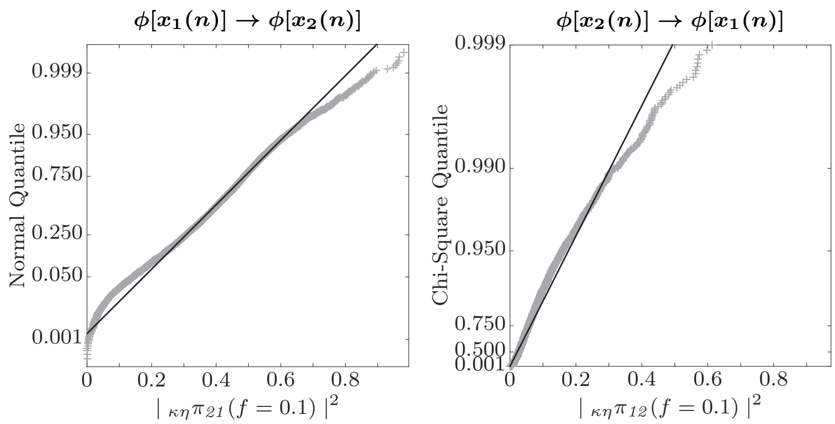

Figure 3: Ensemble asymptotic behaviour of applying the criteria from Baccalá et al. (2013), at $f = f_0 = 0.1$ using $\kappa(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^\top \mathbf{y})^4$ ($N = 750$ and $c = 1.00$), showing the approximate normal behaviour for the existing connection $\phi[x_1(n)] \to \phi[x_2(n)]$, against the distribution of the non-existing $\phi[x_2(n)] \to \phi[x_1(n)]$, compared to the expected $a\chi^2_\nu$ distribution that best fits the data ($a = 0.05, \nu = 1.08$).

		TP — 100 × $(1 - \widehat{\beta})$					**FP — 100 × $(\widehat{\alpha})$**				
Kernel	***c*** \ ***N***	**100**	**250**	**500**	**750**	**1000**	**100**	**250**	**500**	**750**	**1000**
$\kappa(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle^2$	0·10	98·18	99·97	100·00	100·00	100·00	46·50	34·92	28·57	24·98	22·75
	0·50	99·98	100·00	100·00	100·00	100·00	57·01	43·59	34·56	29·22	26·32
	1·00	99·98	100·00	100·00	100·00	100·00	57·24	43·57	34·55	29·23	26·47
$\kappa(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle^4$	0·10	72·38	90·60	96·05	97·71	99·03	34·60	15·07	5·24	1·94	0·97
	0·50	82·93	94·92	97·95	98·81	98·71	43·05	19·16	6·70	2·56	1·29
	1·00	83·21	95·02	97·92	98·81	98·68	43·26	19·16	6·75	2·52	1·32

Table 1: Observed knPDC true positive (**TP**) and false positive (**FP**) detection rates under quadratic and bi-quadratic homogenous polynomial kernels of orders $d = 2$ and 4, respectively, for the connections $\phi[x_1(n)] \to \phi[x_2(n)]$ and $\phi[x_2(n)] \to \phi[x_1(n)]$ ($f = f_0 = 0.1$).

4 Discussion

After introducing a generalized form of PDC and how to compute the required time series coupling model (8), we used an example to investigate the possibility that connectivity detection criteria developed for treating linear VAR models (Baccalá et al., 2013) could still be nonetheless be useful, even if only as an approximate guideline. Our simulations in the example point to reasonably good performance, approaching the expected **FP** rate as $N \to \infty$.

The choice of (9) was in part dictated by the possibility of making ready comparisons with results in Massaroppe et al. (2011) whose approach was based on preprocessing the data into time varying estimates of the time series approximate entropy (Pincus, 2008) and whose relationships were then investigated using linear models. Whereas that methodology managed to capture quadratic connections, in all counts, the present one outperforms the latter by dispensing with the need for approximate entropy reconstruction parameters whose values proved responsible for the large degrees of sensitivity inherent to the latter method's results.

Our current efforts are geared towards expanding this investigation to systems whose dynamics goes beyond that of the narrow bandwidth represented by (9) and towards rigorously establishing connectivity criteria based for connectivity detection including time domain criteria as in Baccalá et al. (1998), which applied Wald type likelihood tests.

An important open issue regards kernel choice. Our preliminary evidence suggests that polyspectral analysis can be useful in this regard (Nikias and Petropulu, 1993).

Finally it is interesting to mention that bivariate causality representations involving nonlinear systems have appeared before (Marinazzo et al., 2014, 2008) though they are more akin to what we called *influentiability* in Baccalá and Sameshima (2014a,b) rather than to causality, because of the data partialization capability of vector autoregressive modeling.

For the present case, the optimal kernel is the (bi-)quadratic one, due to the conditions of the coupling, as (perhaps) expected. Therefore, this points to studying the present technique specially in cases of models in larger dimensions. Exploratory data analysis of further examples is under way, as is the use of kernels other than homogenous polynomials.

5 Acknowledgments

L.M. gratefully acknowledges support from an institutional Ph.D. CAPES Grant. L.A.B. is also gratefully acknowledged to CNPq Grant $307163/2013-0$.

References

Baccalá, L. A., de Brito, C. S. N., Takahashi, D. Y., and Sameshima, K. (2013). Unified asymptotic theory for all partial directed coherence forms. *Philos. T. Roy. Soc. A*, 371(**1997**):20120158.

Baccalá, L. A. and Sameshima, K. (2001). Partial directed coherence: A new concept in neural structure determination. *Biol. Cybern.*, 84(**6**):463–474.

Baccalá, L. A. and Sameshima, K. (2014a). Causality and influentiability: The need for distinct neural connectivity concepts. In *Brain Informatics and Health*, volume 8609, pages 424–435.

Baccalá, L. A. and Sameshima, K. (2014b). *Methods in Brain Connectivity Inference through Multivariate Time Series Analysis*, chapter Multivariate time series brain connectivity: A sum up, pages 245–251. CRC Press, Boca Raton.

Baccalá, L. A., Sameshima, K., Ballester, G., do Valle, A. C., and Timo-Iara, C. (1998). Studying the interaction between brain structures via directed coherence and Granger causality. *Applied Sig. Process.*, (**5**):40–48.

Faes, L., Nollo, G., and Chon, K. H. (2008). Linear and nonlinear parametric model identification to assess Granger causality in short-term cardiovascular interactions. *Comput. Cardiol.*, 35:793–796.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(**3**):424–438.

He, F., Billings, S. A., Wei, H.-L., and Sarrigiannis, P. G. (2014). A nonlinear causality measure in the frequency domain: Nonlinear partial directed coherence with applications to EEG. *J. Neurosci. Meth.*, (**225**):71–80.

Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., and Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.*, 441(**1**):1–46.

Kallas, M., Honeine, P., Francis, C., and Amoud, H. (2013). Kernel autoregressive models using Yule-Walker equations. *Signal Process.*, 93(**11**):3053–3061.

Kannan, R. and Tangirala, A. K. (2014). Correntropy-based partial directed coherence for testing multivariate Granger causality in nonlinear processes. *Physical Review E*, 89(**6**):062144.

Kumar, R. and Jawahar, C. V. (2007). Kernel approach to autoregressive modeling. In *Thirteenth National Conference on Communications (NCC 2007)*, Kanpur, India. Available in: <`http://cvit.iiit.ac.in/papers/ranjeeth07Kernel.pdf`>. Accessed in: October/06/2014.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Marinazzo, D., Liao, W., Pellicoro, M., and Stramaglia, S. (2014). *Methods in Brain Connectivity Inference through Multivariate Time Series Analysis*, chapter Nonlinear parametric granger causality in dynamical networks, pages 135–159. CRC Press, Boca Raton.

Marinazzo, D., Pellicoro, M., and Stramaglia, S. (2008). Kernel method for nonlinear Granger causality. *Phys. Rev. Lett.*, 100(**14**):144103.

Massaroppe, L., Baccalá, L. A., and Sameshima, K. (2011). Semiparametric detection of nonlinear causal coupling using partial directed coherence. In *Eng. Med. Biol. Soc. Ann.*, pages 5927–5930.

Nikias, C. and Petropulu, A. P. (1993). *Higher order spectra analysis: A non-linear signal processing framework*. Prentice Hall.

Paluš, M., Komárek, V., Hrnčíř, Z., and Štěrbová, K. (2001). Synchronization as adjustment of information rates: Detection from bivariate time series. *Phys. Rev. E*, 63(**4**):046211.

Parzen, E. (1959). Statistical inference on time series by Hilbert space method, I. Technical Report 23, Applied Mathmatics and Statistics Laboratory, Stanford University.

Pereda, E., Quiroga, R. Q., and Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.*, 77(**1–2**):1–37.

Pincus, S. M. (2008). Approximate entropy as an irregularity measure for financial data. *Econometric Reviews*, 27(**4–6**):329–362.

Schelter, B., Winterhalder, M., Eichler, M., Peifer, M., Hellwig, B., Guschlbauer, B., Lücking, C. H., Dahlhaus, R., and Timmer, J. (2006). Testing for directed influences among neural signals using partial directed coherence. *J. Neurosci. Meth.*, 152(**1–2**):210–219.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge University Press.

Takahashi, D. Y., Baccalá, L. A., and Sameshima, K. (2010). Information theoretic interpretation of frequency domain connectivity measures. *Biol. Cybern.*, 103(**6**):463–469.