IDENTIFYING DIRECTIONAL CONNECTIONS IN BRAIN NETWORKS VIA MULTI-KERNEL GRANGER MODELS

G. V. Karanikolas and G. B. Giannakis

Dept. of ECE and DTC, Univ. of Minnesota, USA; Emails: {karan029,georgios}@umn.edu

ABSTRACT

Granger causality based approaches are popular in unveiling directed interactions among brain regions. The present work advocates a multi-kernel based nonlinear model for obtaining the effective connectivity between brain regions, by wedding the merits of partial correlation in undirected topology identification with the ability of partial Granger causality (PGC) to estimate edge directionality. The premise is that existing linear PGC approaches may be inadequate for capturing certain dependencies, whereas available nonlinear connectivity models lack data adaptability that multi-kernel learning methods can offer. The proposed approach is tested on both synthetic and real resting-state fMRI data, with the former illustrating the gains in directed edge presence detection performance, as compared to existing PGC methods, and with the latter highlighting differences in the estimated test statistics.

Index Terms— fMRI, partial Granger causality, partial correlation, kernel-based regression, multiple kernel learning.

1. INTRODUCTION

Functional (f)MRI is an imaging modality used in estimating brain activity that has greatly improved our understanding of brain functionality [15]. Many contemporary fMRI studies deal with functional connectivity networks, that is networks formed by pairs of regions whose activities exhibit some form of statistical dependence [30]. These dependencies are assessed by means of functional connectivity measures, which are also the main focus of this work.

Popular approaches to estimating undirected functional connectivity graphs include Pearson's correlation [29], and partial correlation (PC) coefficients [21]. In cases where edge directionality is additionally required, (linear) Granger causality and its variants, such as partial Granger causality, are typical alternatives [12, 24].

Although the aforementioned approaches are linear ones, motivation for developing nonlinear methods is provided by observations indicating that the relationship between the blood-oxygen-level dependent (BOLD) response and the underlying neural activity may be nonlinear [19].

This work was supported by NSF 1500713, 1514056, and NIH 1R01GM104975-01

In this work, kernel-based *nonlinear* regression will be employed first for obtaining the PGC test statistics, with the aim of unveiling nonlinear interactions that linear models are unable to capture. Instead of all node pairs, these test statistics will only be estimated for pairs of nodes whose kernel-based PC (KPC) coefficient [17] is deemed statistically significant. This intermediate step reduces the number of hypothesis tests required to be jointly performed for directed edge inference and thus alleviates the loss in statistical power that correction for multiple comparisons results in. Finally, multi-kernel learning is employed in order to choose the kernel, or more precisely learn an optimal combination of kernels from a preselected set, based on the data. This choice critically affects the performance of any kernel-based method.

A kernel-based variant of Granger causality has been described in [18, 20], and a nonlinear version of PGC can be found in [12]. In contrast with the present work, [18, 20] do not handle the (critical) problem of kernel selection, whereas [12] does not a employ a reproducing kernel Hilbert space formulation. Moreover, none of the aforementioned approaches attempts to reduce the number of hypothesis tests that are jointly performed. Regarding multi-kernel learning techniques, they have been employed in fMRI based tasks, with examples including classification and feature selection [6, 16] as well as functional connectivity estimation [17].

2. NONLINEAR DIRECTED LINK MODELS

Let $\mathcal{V}$ denote the set of nodes in the sought after brain network, and $|\mathcal{V}|$ the cardinality of this set. Each node corresponds to a region, defined by means of a data-driven or anatomical parcellation of the brain. Moreover, let $\mathbf{x}_\nu := [x_\nu[1] \dots x_\nu[T]]^\top$ stand for the representative vector of the region $\nu \in \mathcal{V}$, obtained from the timecourses of the voxels belonging thereto.

Here we will focus on partial Granger causality (PGC), that is known to be more robust to exogenous inputs and latent variables relative to the "ordinary" (multivariate) Granger causality [12]. Let $\mathcal{S} := \mathcal{V}\backslash\{i,j\}$ denote the set of all nodes except for i and j, $\{n_{1\backslash ij}, \dots, n_{|V|-2\backslash ij}\}$ be an indexing set for the nodes in $\mathcal{S}$, and let $\boldsymbol{\chi}_{\backslash ij}[t] := [x_{n_{1\backslash ij}}[t], \dots, x_{n_{|V|-2\backslash ij}}[t]]^\top$ collect observations across these nodes

per slot t. PGC relies on the residuals $\{\epsilon_{i|V\backslash j}[t]\}_{t=d+1}^{T}$, $\{\epsilon_{i|V}[t]\}_{t=d+1}^{T}$ of the following d-th order linear regression models [1]

$$x_i[t] = \bar{\boldsymbol{\chi}}_{\backslash ij}^{\top}[t]\boldsymbol{\gamma}_i + \epsilon_{i|V\backslash j}[t] \tag{1a}$$

$$x_i[t] = \bar{\boldsymbol{\chi}}_{\backslash ij}'^{\top}[t]\boldsymbol{\delta}_i + \epsilon_{i|V}[t] \tag{1b}$$

where $\bar{\boldsymbol{\chi}}_{\backslash ij}[t] := [\boldsymbol{\chi}_{\backslash ij}^{\top}[t], \ldots, \boldsymbol{\chi}_{\backslash ij}^{\top}[t-d], x_i[t-1], \ldots, x_i[t-d]]^{\top}$ and $\bar{\boldsymbol{\chi}}_{\backslash ij}'[t] := [\bar{\boldsymbol{\chi}}_{\backslash ij}^{\top}[t], x_j[t-1], \ldots, x_j[t-d]]^{\top}$; note that $\bar{\boldsymbol{\chi}}_{\backslash ij}'[t]$ augments $\bar{\boldsymbol{\chi}}_{\backslash ij}[t]$ with the d past values of $\{x_j[t]\}$.

If model (1b) is deemed statistically more valid than (1a), then a (directed) edge from j to i is declared to be present in the estimated graph. Intuitively, if augmenting the regressors for predicting $x_i[t]$ with past values of $\{x_j[\tau]\}_{\tau<t}$ lowers the residual error variance, then (1b) is in effect, and we assert that j Granger causes i [11]; that is, improved prediction accuracy is captured by lower residual variance in the direction of the edge under consideration. In particular, the following hypothesis test, that uses $F_{ij} := \text{var}(\epsilon_{i|V\backslash j})/\text{var}(\epsilon_{i|V})$ as a test statistic, is performed

$$\text{H}_0 : F_{ij} \leq 1; \quad \text{H}_1 : F_{ij} > 1\,. \tag{2}$$

According to H_1, $\{x_j[t]\}$ "Granger causes" $\{x_i[t]\}$, whereas according to H_0 it does not.

2.1. Kernel-based PGC

Although (1a) and (1b) postulate that the relationships between $x_i[t]$ and $\bar{\boldsymbol{\chi}}_{\backslash ij}[t]$, $\bar{\boldsymbol{\chi}}_{\backslash ij}'[t]$ are linear, we surmise that allowing for a much-broader class of nonlinear functions will better capture the presence (or absence) of dependencies between nodal time series.

In order to model the aforementioned nonlinear functions, we will rely on a reproducing kernel Hilbert space (RKHS) formulation [27]. In particular, we will consider nonlinear functions from the space $\mathcal{H}$ described as

$$\mathcal{H} := \{f : f(\boldsymbol{\chi}[t]) = \sum_{\tau=1}^{\infty} \beta_\tau \kappa(\boldsymbol{\chi}[t], \boldsymbol{\chi}[\tau])\} \tag{3}$$

where κ denotes a (preselected) similarity function, known as a kernel. Typical examples of kernel functions, which uniquely define their associated RKHSs $\mathcal{H}$, include the linear kernel $\kappa_L(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2) := \boldsymbol{\chi}_1^{\top}\boldsymbol{\chi}_2$, and the Gaussian kernel [27]

$$\kappa_{\text{G}}(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2) := e^{\frac{-\|\boldsymbol{\chi}_1 - \boldsymbol{\chi}_2\|_2^2}{2\sigma^2}}. \tag{4}$$

Kernels such as κ_{G} are reproducing, that is $\forall f \in \mathcal{H}$ and with $\langle \cdot, \cdot \rangle$ denoting the inner product in $\mathcal{H}$, it holds that $f(\boldsymbol{\chi}[t]) = \langle R_{\boldsymbol{\chi}[t]}, f \rangle$, where $R_{\boldsymbol{\chi}[t]}(\boldsymbol{\chi}[\tau]) := \kappa(\boldsymbol{\chi}[t], \boldsymbol{\chi}[\tau])$ [27].

With $\mathcal{H}$ defined, we can now formulate our kernel-based estimators for PGC. In particular, consider replacing the (linear) regression models (1a) and (1b) respectively with

$$x_i[t] = f_{i|V\backslash j}(\bar{\boldsymbol{\chi}}_{\backslash ij}[t]) + \epsilon_{i|V\backslash j}[t] \tag{5a}$$

$$x_i[t] = f_{i|V}(\bar{\boldsymbol{\chi}}_{\backslash ij}'[t]) + \epsilon_{i|V}[t] \tag{5b}$$

where the nonlinear functions $f_{i|V\backslash j}, f_{i|V}$ belong to $\mathcal{H}$.

Given $\{x_i[t], \bar{\boldsymbol{\chi}}_{\backslash ij}[t])\}_{t=d+1}^{T}$, kernel ridge regression seeks a function $f \in \mathcal{H}$ that optimally fits the data while controlling for the smoothness of f. Specifically, the function in e.g., (5a) will be estimated as

$$\hat{f}_{i|V\backslash j} = \underset{f\in\mathcal{H}}{\arg\min} \sum_{\tau=d+1}^{T} (x_i[\tau] - f(\bar{\boldsymbol{\chi}}_{\backslash ij}[\tau]))^2 + \lambda\|f\|_{\mathcal{H}}^2 \tag{6}$$

where λ is a regularization parameter and $\|\cdot\|_{\mathcal{H}}$ denotes the norm of $\mathcal{H}$. From the representer theorem [28], it follows that the optimal solution to (6) will be of the form

$$\hat{f}_{i|V\backslash j}(\bar{\boldsymbol{\chi}}_{\backslash ij}[t]) = \sum_{\tau=d+1}^{T} \beta_{i\tau}\, \kappa(\bar{\boldsymbol{\chi}}_{\backslash ij}[t], \bar{\boldsymbol{\chi}}_{\backslash ij}[\tau]). \tag{7}$$

Plugging now (7) into (6), the functional minimization problem boils down to estimating the vector $\boldsymbol{\beta}_i := [\beta_{i(d+1)}, \ldots, \beta_{iT}]^{\top}$, as

$$\hat{\boldsymbol{\beta}}_i = \underset{\boldsymbol{\beta}_i\in\mathbb{R}^{(T-d)}}{\arg\min} \|\mathbf{x}_i^{(d)} - \mathbf{K}_{i|V\backslash j}\boldsymbol{\beta}_i\|^2 + \lambda\boldsymbol{\beta}_i{}^{\top}\mathbf{K}_{i|V\backslash j}\boldsymbol{\beta}_i \tag{8}$$

where $\mathbf{x}_i^{(d)} := [x_i[d+1], \ldots, x_i[T]]^{\top}$, $[\mathbf{K}_{i|V\backslash j}]_{t\tau} := \kappa(\bar{\boldsymbol{\chi}}_{\backslash ij}[t], \bar{\boldsymbol{\chi}}_{\backslash ij}[\tau])$ stands for the kernel matrix, and we also used the fact that $\|f\|_{\mathcal{H}}^2 = \boldsymbol{\beta}_i{}^{\top}\mathbf{K}_{i|V\backslash j}\boldsymbol{\beta}_i$. A closed-form solution is available for (8), and it is given by

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{K}_{i|V\backslash j} + \lambda\mathbf{I})^{-1}\mathbf{x}_i^{(d)}. \tag{9}$$

From (5a), (7), and (9), it follows that the $\boldsymbol{\epsilon}_{i|V\backslash j} := [\epsilon_{i|V\backslash j}[d+1], \ldots, \epsilon_{i|V\backslash j}[T]]^{\top}$ kernel-based prediction residual can be expressed as $\boldsymbol{\epsilon}_{i|V\backslash j} = [\hat{f}_{i|V\backslash j}(\bar{\boldsymbol{\chi}}_{\backslash ij}[d+1]) \ldots \hat{f}_{i|V\backslash j}(\bar{\boldsymbol{\chi}}_{\backslash ij}[T])]^{\top} - \mathbf{x}_i^{(d)} = \mathbf{K}_{i|V\backslash j}\hat{\boldsymbol{\beta}}_i - \mathbf{x}_i^{(d)}$, which after using (9) yields

$$\boldsymbol{\epsilon}_{i|V\backslash j} = \mathbf{K}_{i|V\backslash j}(\mathbf{K}_{i|V\backslash j} + \lambda\mathbf{I})^{-1}\mathbf{x}_i^{(d)} - \mathbf{x}_i^{(d)}. \tag{10}$$

Likewise, upon replacing $\mathbf{K}_{i|V\backslash j}$ by $\mathbf{K}_{i|V}$ in (10), with $[\mathbf{K}_{i|V}]_{t\tau} := \kappa(\bar{\boldsymbol{\chi}}_{\backslash ij}'[t], \bar{\boldsymbol{\chi}}_{\backslash ij}'[\tau])$, yields $\boldsymbol{\epsilon}_{i|V}$. The test statistic F_{ij} in (2) can then be obtained as the ratio of the sample variances of $\epsilon_{i|V\backslash j}$ and $\epsilon_{i|V}$.

So far, we have seen how to assess the directional presence of an edge between *any pair* of nodes, while allowing for nonlinear models capturing directional nodal dependencies that reflect corresponding links. But can one afford checking all possible pairs and directions in a brain network? This is the subject addressed in the ensuing section.

2.2. Efficient inference of directional network links

In a $|\mathcal{V}|$-node network there are $m = |\mathcal{V}|(|\mathcal{V}| - 1)$ potential edges, and thus an equal number of hypothesis tests (cf. (2)) to be performed. When multiple tests are jointly performed, there is a loss in statistical power (joint true positive

rate) which becomes more severe as the number of tests increases [2]. In essence, in order to guard against a growing number of false alarms, threshold adjustment procedures require an increased deviation of F_{ij} from 1 in order to decide H_1, as m increases. This implies that for some tests that H_1 is true albeit with F_{ij} close to 1, H_0 will (incorrectly) be chosen instead, which in turn will lead to a decrease of the joint true positive rate.

On the other hand, brain networks are typically sparse [26, 25], and thus for the majority of tests H_0 will be in effect. For future use, let $\mathcal{E}_0$ denote the pairs of nodes for which edges are absent in both directions. Had we known $\mathcal{E}_0$ a priori, it would only be required to perform tests for the rest of the potential edges. Our idea, here, is to estimate $\mathcal{E}_0$ using kernel-based partial correlations [17], which are particularly well suited to identifying edge absence. Letting $\tilde{\mathbf{x}}_i := \mathbf{x}_i - \hat{\mathbf{x}}_i$, the sample KPC coefficient of $\mathbf{x}_i, \mathbf{x}_j$ with respect to $\{\mathbf{x}_k\}_{k\in\mathcal{S}}$ is given by

$$\hat{\rho}_{ij|\mathcal{S}} := \frac{(\tilde{\mathbf{x}}_i - \bar{\tilde{\mathbf{x}}}_i)^\top (\tilde{\mathbf{x}}_j - \bar{\tilde{\mathbf{x}}}_j)}{\|\tilde{\mathbf{x}}_i - \bar{\tilde{\mathbf{x}}}_i\|_2 \, \|\tilde{\mathbf{x}}_j - \bar{\tilde{\mathbf{x}}}_j\|_2} \tag{11}$$

where $\hat{\mathbf{x}}_i$ is a kernel-based estimate of $\mathbf{x}_i$ based on $\{\boldsymbol{\chi}_{\backslash ij}^\top[t]\}_{t=1}^T$ and $\bar{\tilde{\mathbf{x}}}_i := T^{-1}\sum_{t=1}^T \tilde{x}_i[t]\mathbf{1}$, with $\mathbf{1}$ denoting the all-ones vector. In short, (kernel-based) partial correlation aims at being robust against mediated dependencies between nodes. As an example, consider a network of three nodes (i,k,j) [29], where nodes i and j are mediated through node k. Avoiding the false alarm of declaring an edge (i,j) in this case, is the goal of KPC. In particular, KPC regresses $\mathbf{x}_k$ out of $\mathbf{x}_i$ and $\mathbf{x}_j$, and as a result the correlation (and therefore the spurious edge) between i and j vanishes. The proposed approach can now be summarized in the following steps:

S1. Use (11) to find $[\mathbf{P}]_{ij} = \hat{\rho}_{ij|\mathcal{S}}$.

S2. Hard threshold $[\mathbf{P}]_{ij}$ to form the binary adjacency matrix $[\mathbf{B}]_{ij}$

S3. For each (i,j) s.t. $[\mathbf{B}]_{ij} = 1$, estimate F_{ij}, as per Sec. 2.1.

S4. For each F_{ij}, estimate the p-value as detailed below.

S5. Using $\{p_{ij}\}$ jointly perform $N_B :=$ number of nonzero entries of $\mathbf{B}$ tests (cf. (2)), correcting for multiple comparisons as in [2].

S1 and S2 are performed using the methods described in [17]. With Φ_0 denoting the cumulative distribution function of F_{ij} under H_0, the p-value of F_{ij} is $p_{ij} := 1 - \Phi_0(F_{ij})$. To estimate the latter, Φ_0 is estimated using the stationary bootstrap scheme in [22] with the mean block length being selected using the data-driven method of [23].

Rather than testing $\mathcal{O}(|\mathcal{V}|^2)$ hypotheses (cf. (2)), we have now to test only for $\mathcal{O}(N_B)$ directed edges. For the latter, the criterion we adopt in S5 to jointly account for the multiple comparisons is the false discovery rate $\mathrm{FDR} := \mathbb{E}[N_{FA}/(N_{FA}+N_{TP})]$, where N_{FA}, N_{TP} stand for the number of false alarms and true positives, respectively. In particular, we will follow the Benjamini-Hochberg method in [2], which is skipped due to space limitations.

2.3. Multiple kernel learning

From (3) and (7) it is evident that the predictors $\{f_{i|V\backslash j}, f_{i|V}\}$ are highly dependent on the choice of the kernel κ. Multi-kernel learning aims at optimally choosing a combination of kernels, from a preselected dictionary $\{\kappa_p\}_{p=1}^P$, based on the data [10]. Here we will consider nonnegative combinations of the basis kernels, that is $\kappa \to \sum_{p=1}^P \theta_p \kappa_p$ with $\theta_p \geq 0 \;\; \forall p$. Moreover, an ℓ_2 regularizer will be applied on the weight vector $\boldsymbol{\theta} := [\theta_1, \ldots, \theta_P]^\top$ [7]. A detailed description of the optimization task involved in estimating the optimal $\boldsymbol{\theta}$, as well as a solution algorithm, can be found in [17].

3. NUMERICAL TESTS

3.1. Synthetic data

To evaluate the performance of the proposed approach, synthetic fMRI datasets based on the (forward) dynamic causal model (DCM) [9] were generated in a setup similar to that of [29]. Letting $\{\psi_i(t)\}, \{u_i(t)\}$ stand for the neural and input time series of node i respectively, and $\boldsymbol{\psi}(t) := [\psi_1(t), \ldots, \psi_{|\mathcal{V}|}(t)]^\top$, $\mathbf{u}(t) := [u_1(t), \ldots, u_{|\mathcal{V}|}(t)]^\top$ denote the corresponding vector time series, the DCM neural network model can be described by

$$\dot{\boldsymbol{\psi}}(t) = \delta \mathbf{A}\boldsymbol{\psi}(t) + \mathbf{u}(t) \tag{12}$$

with $\mathbf{A}$ denoting the (ground truth) connectivity matrix, and δ being a scalar that adjusts neural lags; here $\delta = 20$.

In order to simulate resting-state fMRI data, similar to [29], $u_i(t)$ is obtained as a binary pulse train generated by a Markov chain (20% average duty cycle), corrupted by zero-mean additive white Gaussian noise of variance 10^{-2}. Each $\{\psi_i(t)\}$ obtained as a solution to (12) is provided as the input to the nonlinear balloon model for vascular dynamics [3], the output of which is downsampled with period $TR = 0.5s$, yielding the ith node data vector $\mathbf{x}_i$, comprising $T = 190$ samples.

The matrix $\mathbf{A}$ was chosen to be of dimensions 10×10, with fixed diagonal entries $\mathbf{A}_{ii} = -1$ and randomly placed non-zero entries, each drawn uniformly at random from the interval $[0.25, 0.6]$. Finally, the choice of the DCM parameters is described in detail in [17].

The set of basis kernels used consisted of a linear kernel and 19 Gaussian kernels with variances $\{\sigma_p^2\}_{p=1}^{19}$ taken from the interval $[10^{-6}, 1]$. The model order d was chosen using cross-validation for the proposed approach, whereas the Akaike Information Criterion was applied in the linear and nonlinear PGC cases, as per [12]. Finally, for each potential edge $i \to j$, five-fold cross-validation was employed in order to choose the regularization parameter λ, from the set $\{0.1, 1, 10\}$.

Our tests were performed on 4 randomly generated networks, that on average comprised 6.5 edges. Moreover, the

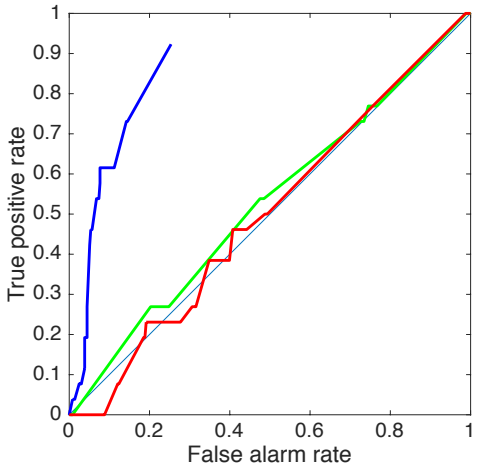

Fig. 1. ROC curves for directed edge inference obtained on DCM-based synthetics. The blue curve corresponds to the proposed approach, whereas the green and red ones stand for linear and nonlinear PGC respectively.

distribution of the test statistics $\{F_{ij}\}$ was obtained using 100 bootstrap realizations. The performance of the proposed approach was compared to that of linear and nonlinear PGC using receiver operating characteristics (ROC) curves, obtained by gradually increasing the maximum FDR level. Fig. 1 illustrates the gain in (directed) edge detection performance offered by our novel method. It is worth mentioning that both existing PGC approaches yielded zero-edge graphs, for any reasonable (up to 0.5) maximum FDR level, in 3 out of 4 simulated networks.

3.2. Real data

The behavior of our novel approach on real data was assessed using resting-state (RS) fMRI data from a single subject. A detailed description of the data is available in [14]. For the purpose of our test, we focused on the (33, anatomically defined) regions comprising the left hemisphere[1] of the brain, and used the first 6 mins. of the scan.

In particular, KPC was used first to estimate $\mathbf{B}$, and thus the pairs of nodes for which both potential edges were absent. Let $p_{ij}^{(l)}$ stand for the p-value of the test statistic estimated by linear PGC for the potential edge $j \to i$, and $\mathcal{E}_B$ denote the set of all (i,j) s.t. $[\mathbf{B}]_{ij} = 1$. The differences between the p-values of the test statistics estimated by the proposed approach and linear PGC are highlighted in Fig. 2 , which plots $p_{ij}^{(l)} - p_{ij}$ for all $(i,j) \in \mathcal{E}_B$. The lack of ground truth, however, precludes us from assessing relative merits. Note also that after correcting for multiple comparisons the proposed approach identified 35 edges as opposed to zero for linear PGC, in the aforementioned set; the maximum FDR level was set to 0.4.

[1] In order to reduce the number of latent variables the observations from the right hemisphere were also included in the regressors.

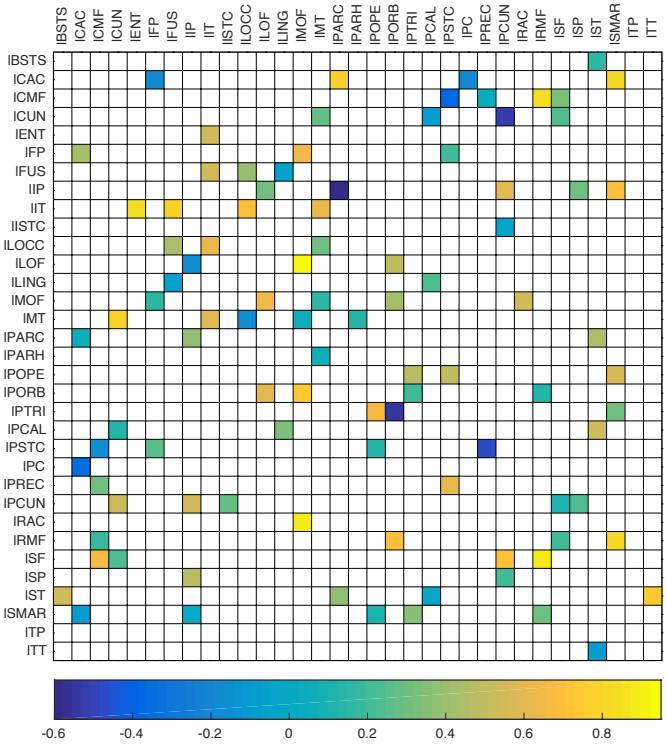

Fig. 2. Comparing the proposed approach to linear PGC on real RS-fMRI data. The (i,j)-th entry color-codes $p_{ij}^{(l)} - p_{ij}$. Full names of the abbreviated regions can be found in [4, SI].

4. CONCLUSIONS

Motivated by the presumably nonlinear nature of the BOLD signal, as well as the challenges arising in multiple testing as the number of brain regions under consideration grows, the present work proposes a kernel-based nonlinear approach for identifying directed graphs by combining KPC with kernel-based estimates of the PGC test statistics. The task of choosing the kernel is tackled using multi-kernel learning. The proposed approach outperformed both linear and nonlinear PGC on DCM-based synthetics. Finally, tests on real resting-state fMRI data demonstrate significant differences between the proposed approach and linear PGC.

5. ACKNOWLEDGEMENTS

The authors wish to thank Prof. O. Sporns for kindly providing the real data used in this work. Toolboxes used in this paper include [13, 5, 8].

6. REFERENCES

[1] A. B. Barrett, L. Barnett, and A. K. Seth, "Multivariate Granger causality and generalized variance," *Phys. Rev. E*, vol. 81, p. 041907, 2010.

[2] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[3] R. B. Buxton, E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: The balloon model," *Magnetic Resonance in Medicine*, vol. 39, no. 6, pp. 855–864, 1998.

[4] J. Cabral, E. Hugues, O. Sporns, and G. Deco, "Role of local network oscillations in resting-state functional connectivity," *NeuroImage*, vol. 57, no. 1, pp. 130–139, 2011.

[5] A. Canty and B. D. Ripley, *boot: Bootstrap R (S-Plus) Functions*, 2016, r package version 1.3-18.

[6] E. Castro, V. Gómez-Verdejo, M. Martínez-Ramón, K. A. Kiehl, and V. D. Calhoun, "A multiple kernel learning approach to perform classification of groups from complex-valued fMRI data analysis: Application to schizophrenia," *NeuroImage*, vol. 87, pp. 1–17, 2014.

[7] C. Cortes, M. Mohri, and A. Rostamizadeh, "L2 regularization for learning kernels," in *Proc. Conf. on Uncertainty in Artificial Intelligence*, Arlington, VA, USA, 2009, pp. 109–116.

[8] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and their Applications*. Cambridge University Press, 1997.

[9] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, 2003.

[10] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

[11] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[12] S. Guo, A. K. Seth, K. M. Kendrick, C. Zhou, and J. Feng, "Partial Granger causality—Eliminating exogenous inputs and latent variables," *Journal of Neuroscience Methods*, vol. 172, no. 1, pp. 79–93, 2008.

[13] T. Hayfield and J. S. Racine, "Nonparametric econometrics: The np package," *Journal of Statistical Software*, vol. 27, no. 5, 2008.

[14] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proc. of the Ntl. Academy of Sciences*, vol. 106, no. 6, pp. 2035–2040, 2009.

[15] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional Magnetic Resonance Imaging*. Sinauer Associates, 2004.

[16] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 576–589, 2014.

[17] G. V. Karanikolas, G. B. Giannakis, K. Slavakis, and R. M. Leahy, "Multi-kernel based nonlinear models for connectivity identification of brain networks," in *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Shanghai, China, Mar. 2016, pp. 6315–6319.

[18] W. Liao, D. Marinazzo, Z. Pan, Q. Gong, and H. Chen, "Kernel Granger causality mapping effective connectivity on fMRI data," *IEEE Trans. Medical Imaging*, vol. 28, no. 11, pp. 1825–1835, Nov. 2009.

[19] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, vol. 412, no. 6843, Jul. 2001.

[20] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Physical Review Letters*, vol. 100, no. 14, pp. 144–103, 2008.

[21] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, and H. Benali, "Partial correlation for functional brain interactivity investigation in functional MRI," *NeuroImage*, vol. 32, no. 1, pp. 228–237, 2006.

[22] D. N. Politis and J. P. Romano, "The stationary bootstrap," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1303–1313, 1992.

[23] D. N. Politis and H. White, "Automatic block-length selection for the dependent bootstrap," *Econometric Reviews*, vol. 23, no. 1, pp. 53–70, 2004.

[24] A. Roebroeck, E. Formisano, and R. Goebel, "Mapping directed influence over the brain using Granger causality and fMRI," *NeuroImage*, vol. 25, no. 1, pp. 230–242, 2005.

[25] S. Ryali, T. Chen, K. Supekar, and V. Menon, "Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty," *NeuroImage*, vol. 59, no. 4, pp. 3852 – 3861, 2012.

[26] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, 2005.

[27] B. Schölkopf and J. A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press, 2002.

[28] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Computational Learning Theory: 14th Annual Conf. on Computational Learning Theory, COLT and 5th European Conf. on Computational Learning Theory, EuroCOLT*. Amsterdam, The Netherlands: Springer Berlin Heidelberg, Jul. 2001, pp. 416–426.

[29] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "Network modelling methods for fMRI," *NeuroImage*, vol. 54, no. 2, pp. 875–891, 2011.

[30] O. Sporns, *Networks of the Brain*. MIT press, 2011.