Alleviating the influence of weak data asymmetries on Granger-causal analyses

Stefan Haufe[1,2], Vadim Nikulin[3], and Guido Nolte[4]

[1] Berlin Institute of Technology, Machine Learning, stefan.haufe@tu-berlin.de
[2] Bernstein Focus Neurotechnology, Berlin
[3] Charité University Medicine, Berlin, Neurophysics
[4] Fraunhofer Institute FIRST, Berlin, Intelligent Data Analysis

Abstract. We introduce the concepts of weak and strong asymmetries in multivariate time series in the context of causal modeling. Weak asymmetries are by definition differences in univariate properties of the data, which are not necessarily related to causal relationships between time series. Nevertheless, they might still mislead (in particular Granger-) causal analyses. We propose two general strategies to overcome the negative influence of weak asymmetries in causal modeling. One is to assess the confidence of causal predictions using the antisymmetry-symmetry ratio, while the other one is based on comparing the result of a causal analysis to that of an equivalent analysis of time-reversed data. We demonstrate that Granger Causality applied to the SiSEC challenge on causal analysis of simulated EEG data greatly benefits from our suggestions.

Keywords: weak/strong asymmetries, ASR, time inversion, Granger Causality, SiSEC challenge

1 Introduction

Many measures of causal interaction (a. k. a. effective connectivity) are based on the principle that the cause precedes the effect. However, it would be misleading to assume that temporal ordering is necessarily the dominant factor when estimating causal relationship on the basis of the available techniques, such as Granger causality. In fact, methods to estimate causal relations are based on general asymmetries between two (or more) signals out of which the temporal order is just one specific feature. Other asymmetries, like different signal-to-noise ratios, different overall power or spectral details, may in general also affect causal estimates depending on which method is used.

We here propose to distinguish between two different kinds of asymmetries. We call the first type 'strong asymmetries' defined as asymmetries in the relation between two (or more) signals like the temporal ordering. The second type is called 'weak asymmetry' and denotes different univariate properties as given, e. g., by the spectral densities. Weak asymmetries can hence be detected from two signals without estimating any functional relationship between them whereas a strong asymmetry is a property of that functional relationship.

Altough the concepts presented here (but not the test presented below) could be generalized to other cases in a straight forward way we restrict ourselves in the following to the discussion of stationary and Gaussian distributed data. Let $x_j(t)$ be the signal in channel j at time t. Then the statistical properties are completely defined by the cross-covariance matrices

$$C(p) = \left\langle (\mathbf{x}(t) - \widehat{\mu}_{\mathbf{x}}) \left(\mathbf{x}(t-p) - \widehat{\mu}_{\mathbf{x}}\right)^\top \right\rangle \ , \tag{1}$$

where $\langle \cdot \rangle$ denotes expectation. The process is now said to contain a strong asymmetry if for some i, j and some p it is found that $C_{i,j}(p) \neq C_{j,i}(p)$, i. e. $C(p)$ is asymmetric for at least one p. The process is said to contain a weak asymmetry if for some i, j and some p it is found that $C_{i,i}(p) \neq C_{j,j}(p)$, i. e. the diagonals are not all equal. Since the power spectrum of the i-th signal is given by the Fourier transform of $C_{i,i}(p)$ the process contains a weak asymmetry if and only if it contains signals with different power spectra.

Methods to detect causality are typically sensitive to both weak and strong asymmetries. Weak asymmetries can be detected more robustly but can also be considered as weaker evidence for causal relations. This can be illustrated if data are instantaneous mixtures of independent sources. In this case all cross-covariances are weighted sums of auto-covariances of the sources. Since auto-covariances are always symmetric functions of the delay p and since generally $C(-p) = C^\top(p)$ it follows that $C(p) = C^\top(p)$ for mixtures of independent sources [4]. Hence, such mixtures can only contain weak asymmetries but not strong ones.

For methods which are sensitive to both weak and strong asymmetries it is in general difficult to tell on what property of the data an estimate of causal drive is based. However, using empirical estimators of the cross spectra, it is possible to measure the proportions of weak and strong asymmetries in a dataset. In this paper, we demonstrate that a quantity called antisymmetry-symmetry-ratio is a meaningful predictor of the success of the causal estimation for methods that are knowingly affected by weak asymmetries. Moreover, we introduce a procedure based on time inversion, by which it is possible to test whether weak asymmetries are the dominant cause for a given connectivity estimate. We demonstrate that our approaches dramatically reduce the number of wrong predictions of Granger Causality (GC). As a result, GC's performance in the 2011 Signal Separation Evaluation Campaign (SiSEC) challenge on causal analysis of simulated EEG data is significantly improved. Our approaches can be regarded as sanity checks which are applicable in any causal analysis testing temporal delays between driver and receiver.

The paper starts with introducing Granger Causality, the SiSEC challenge dataset and the two novel approaches proposed to improve causal estimations in the Methods section. The Results section confirms that these approaches effectively reduce the number of wrong predictions of Granger Causality on the challenge dataset. In the Discussion section, we elaborate on the applicability of our approaches and draw connections to permutation testing, which is also typically used in conjunction with Granger-causal measures.

2 Methods

2.1 SISEC challenge simulated EEG dataset

To demonstrate our ideas we consider a set of simulated EEG data, which is part of the 2011 Signal Separation Evaluation Campaign. The data consists of 1 000 examples of bivariate data for 6 000 time points. Each example is a superposition of a signal (of interest) and noise. The causally-interacting signals are constructed using a unidirectional bivariate autoregressive (AR) model of order 10 with (otherwise) random AR-parameters and uniformly distributed innovations. The noise is constructed of three independent sources, generated with three univariate AR-models with random parameters and uniformly distributed input, which were instantaneously mixed into the two sensors with a random mixing matrix. The relative strength of noise and signal (i. e. signal-to-noise ratio, SNR) was set randomly. The task of the challenge is to determine the direction of the causal interaction. One point is awarded for every correct prediction, while every wrong prediction causes a penalty of -10 points. If no prediction is given for a dataset, this results in 0 points. The maximum score attainable is 1 000 points, while the minimum score (considering that predictors with less than 50 % accuracy can be improved by sign-flipping) is -4 500 points.

The simulation addresses a conceptual problem of EEG data, namely that the signals of interest are superimposed by mixed noise. However, the actual spectra can be quite different from real EEG data. Volume conduction (i. e., mixing of the signals of interest), which is typically also observed in EEG datasets and poses serious challenges on its own [2], is omitted here in order to facilitate an objective evaluation. We use Matlab code provided by the organizers of the challenge to generate 1 000 new instances of the problem with known directions of causal flow.

2.2 Granger Causality

The *multivariate AR* (MVAR) model is given by

$$\mathbf{x}(t) = \sum_{p=1}^{P} B(p)\mathbf{x}(t-p) + \boldsymbol{\varepsilon}(t) \ , \tag{2}$$

where $B(p)$ are matrices describing the time-delayed influences of $\mathbf{x}(t-\tau)$ on $\mathbf{x}(t)$. Notably, the off-diagonal parts $B_{i,j}(p), i \neq j$ describe time-lagged influences between different time series. Granger Causality [1] involves fitting a multivariate AR model for the full set $\mathbf{x}_{\{1,\ldots,M\}} = \mathbf{x}$, as well as for the reduced set $\mathbf{x}_{\{1,\ldots,M\}\setminus\{i\}}$ of available time series, where $M = 2$ here. Denoting the prediction errors of the full model by $\boldsymbol{\varepsilon}^{\text{full}}$ and those of the reduced model by $\boldsymbol{\varepsilon}^{\setminus i}$, the *Granger score* GC describing the influence of x_i on x_j is defined as the log-ratio of the mean-squared errors (MSE) of the two models with respect to x_j. i. e.,

$$\text{GC}_{i,j} = \log\left(\frac{\sum_{t=P+1}^{T}\left[\varepsilon_j^{\text{full}}(t)\right]^2}{\sum_{t=P+1}^{T}\left[\varepsilon_j^{\setminus i}(t)\right]^2}\right) \ . \tag{3}$$

This definition, which is based on the ratio of prediction errors, is independent of the scale of the time series x_i and x_j. However, as has been demonstrated in [5], [6], it is influenced by asymmetries in the signal-to-noise ratio.

2.3 Exploiting statistical characterics of non-/interacting signals for assessing the reliability causal predictions

Due to additive noise and (in our case) innovation noise introduced by AR modeling, cross-covariances of realistic measurements are never exactly symmetric nor are they exactly antisymmetric. Nevertheless, the amount of symmetric vs. antisymmetric cross-covariance contained in a dataset provides important information about the SNR and hence how difficult the problem of estimating the causal direction is. We propose to use an index called antisymmetry-symmetry ratio (ASR) defined as

$$\text{ASR} = \log\left(\frac{\left\|\left(\widehat{C}(1) - \widehat{C}^\top(1), \ldots, \widehat{C}(P) - \widehat{C}^\top(P)\right)\right\|_{\mathcal{F}}}{\left\|\left(\widehat{C}(1) + \widehat{C}^\top(1), \ldots, \widehat{C}(P) + \widehat{C}^\top(P)\right)\right\|_{\mathcal{F}}}\right) \tag{4}$$

for quantifying the confidence in a given causal estimation, where $(A1, \ldots, A_P)$ is the horizontal concatenation of the matrices $A_1, \ldots, A_P$, $A_{\mathcal{F}}$ denotes the Frobenius norm (sum of squared entries) of a matrix and $\widehat{C}(p)$ are empirical estimates of the cross-covariance matrices. The higher the ASR, the lower the proportion of (potentially misguiding) signal parts with symmetric cross-covariance is. Hence, one strategy to avoid false predictions in Granger- (and other) causal analyses is to evaluate only datasets characterized by high ASR.

2.4 A test for assessing the time-lagged nature of interactions

As a second simple test to distinguish weak from strong asymmetries we here suggest to compare the specific result of a causal analysis with the outcome of the method applied on time-reversed signals. This corresponds to the general intuitive idea that when all the signals are reversed in time, the direction of information flow should also reverse. More specifically, if temporal order is crucial to tell a driver from recipient the result can be expected to be reverted if the temporal order is reverted. The mathematical basis for this is the simple observation that the cross-covariance for the time inverted signals, say $\widetilde{C}(p)$, is given as

$$\widetilde{C}(p) = C(-p) = C^\top(p) \tag{5}$$

implying that time inversion inverts all strong asymmetries but none of the weak asymmetries. If now a specific measure is essentially identical for original and time inverted signals we conclude that the causal estimate in that specific case is based only on weak asymmetry. To avoid estimation biases introduced by weak asymmetries, one may therefore require that a causality measure delivers significant and *opposing* flows on original and time-reversed signals. Alternatively, one may require that the difference of the results obtained on original and time-reversed signals is significant.

2.5 Experiments

As baselines for the numerical evaluation, we apply Granger Causality as well as the Phase-slope Index (PSI) [5] to all 1 000 datasets and compute the respective score according to the rules of the SiSEC challenge. Granger Causality is calculated using the true model order $P = 10$. The Phase-slope Index is calculated using the authors' implementation[5] in a wide-band on segments of length $N = 100$. For both methods, *net flow*, i. e. the difference between the flows in both directions is assessed. Standard deviations of the methods' results are estimated using the jackknife method. Standardized results with absolute values greater than 2 are considered significantly different from zero. Insignificant results are not reported, i. e. lead to zero points in the evaluation. The whole procedure is repeated 100 times for different realizations of the 1 000 datasets to compute average challenge scores and confidence intervals.

The idea introduced in subsection 2.3 is implemented by ordering the datasets according to their ASR (calculated with $P = 30$), and evaluating the competition score attained when only the first K datasets with highest ASR are analyzed. That is, even significant results might be discarded, if the ASR is low. We consider three additional variants of GC, in which results are reported only if additional restrictions are met. The first variant, 'GC inv both' reports a causal net flow only if it is significant, and if the net flow on time-reversed data points to the opposite direction and is also significant. The variant 'GC inv diff' requires that the difference of the net flows estimated from original and time-reversed data is significantly different from zero. Finally, we compare time inversion to general random permutations of the samples (using the same permutation for all channel) according to the 'difference' approach. The resulting procedure is denoted by 'GC perm diff'.

3 Results

Figure 1 illustrates that interacting signals and mixed independent noise are characterized by different proportions of symmetric and antisymmetric parts in their cross-covariances. The upper-left plot depicts the log-norms of symmetric and antisymmetric cross-covariances of normalized signal and noise time series as a scatter plot, while the upper right plot depicts the respective ASR. In both plots, signal and noise are highly separable. In the lower left plot, the ASR of the observation is plotted against the signal-to-noise ratio. Apparently, there exists a quasi-linear functional relationship between the two, which is the basis of our idea to use the ASR as an indicator for the difficulty of causal predictions.

Figure 2 summarizes the results of the numerical evaluation of the various causal prediction strategies according to the rules of the SiSEC challenge. For all methods considered, the challenge score is plotted as a function of the number of datasets analyzed (starting from datasets with highest ASR). The scores depicted on the very right hence correspond to the standard situation that all

[5] http://ml.cs.tu-berlin.de/causality/

1 000 datasets are analyzed. The scores obtained by the six contributors of the SiSec challenge are marked by black horizontal bars.

As in previous analyses [6], PSI outperforms Granger Causality having a total score of 593 ± 3 points compared to -438 ± 11 points after evaluation of all 1 000 datasets. However, as the plot also strikingly shows, the inferior performance of GC is a result of a huge number of false predictions predominantly made on data with low ASR. Hence, by avoiding decisions on low-ASR data, GC's score increases dramatically with the maximum of 384 ± 4 points reached if only the 539 datasets with highest ASR are analyzed. Note that this score is not anymore dramatically worse than the score obtained by PSI for the same amount of data, which is 485 ± 1 points. All three alternative variants of Granger Causality perform better than the conventional GC strategy with scores of 353 ± 2 points, 437 ± 5 points and 79 ± 4 points attained for 'GC inv both', 'GC inv diff' and 'GC perm diff', respectively when all datasets are analyzed. Note that this means that both 'GC inv both' and 'GC inv diff' outperform the winning contribution of the SiSec challenge, which achieved a score of 252 points. At the same time, the difference between the score attained when analyzing all 1 000 datasets and the maximal score attained when analyzing fewer datasets is dramatically reduced. This difference is 2 ± 1 points for 'GC inv both', 36 ± 2 points for 'GC inv diff' and 116 ± 4 points for 'GC perm diff', which is much closer to the value of 11 ± 1 points measured for PSI than to the value of 841 ± 10 points measured for conventional GC. Hence, all three proposed variants can be seen as robustifications of conventional GC, which prevent decisions that are solely based on weak asymmetries. Among the three proposed strategies, 'GC inv diff' performs best with scores that are competitive to those attained by PSI, while 'GC perm diff' performs worst. Note that the curve of 'GC inv diff' is located strictly above the curve of 'GC', which means that the additional restriction imposed by the time inversion causes no loss in performance for high-ASR data.

4 Discussion

Our results confirm that the proposed strategies drastically reduce the number of false predictions for methods that are prone to be dominated by weak asymmetries in the data such as Granger Causality. While for conventional Granger Causality the inclusion of the ASR as an additional criterion guiding the prediction is highly benefical, this is less helpful for modified variants that take the results obtained on time-reversed (or permuted) data into account. These modifications make GC behave more similarly to PSI, which is itself robust to many weak asymmetries by construction and in particular rather unaffected by dominant symmetric cross-covariances as indicated by low ASR. The choice of the ASR threshold remains an open problem, which is outside the scope of this paper. Empirical strategies to adjust the threshold are, however, conceivable.

Notably, the idea of performing pairwise testing of results obtained on original and time-reversed signals is a special case of permutation testing, as proposed, for example, by [3] in the context of Granger-causal analysis of EEG data

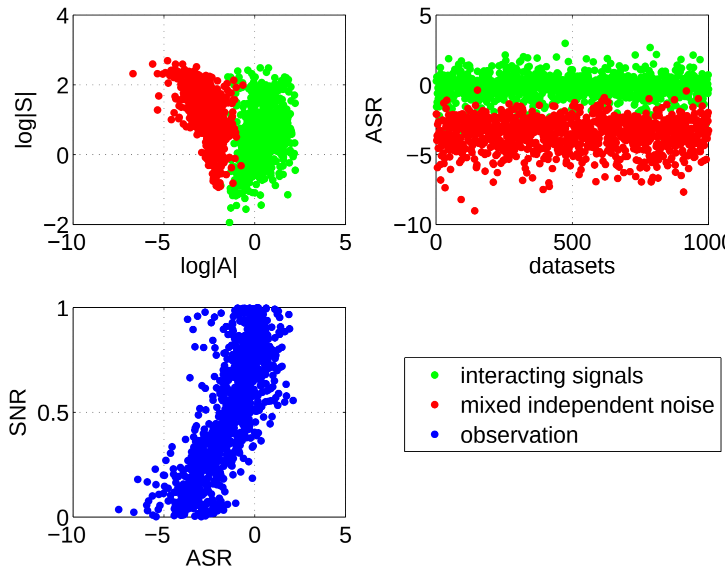

Fig. 1. Upper left: characterization of interacting signals and mixed independent noise by means of the log-norms of the symmetric and antisymmmetric parts of the cross-covariance matrices. Upper right: separation of signal and noise by means of the antisymmetry-symmetry ratio (ASR). Lower left: approximately linear relationship between the ASR of the observations and the signal-to-noise ratio (SNR).

using the directed transfer function (DTF). Both approaches have in common that the reordered data shares certain weak asymmetries with the original data, which are likely to cancel out in pairwise comparisons. However, time-reversed data additionally contains strong asymmetries in the opposite direction, which increases the statistical power of the comparison of original and time-reversed data. Consequently, our empirical results indicate that time inversion outperforms permutation testing by far and should be a viable alternative also when using DTF. Interestingly, PSI exactly flips its sign (direction) upon time inversion, for which reason pairwise testing against time-reversed data cannot be used to improve PSI.

5 Conclusion

We proposed two strategies for robustifying Granger-causal analyses, which boost its performance in the SiSEC 2011 challenge.

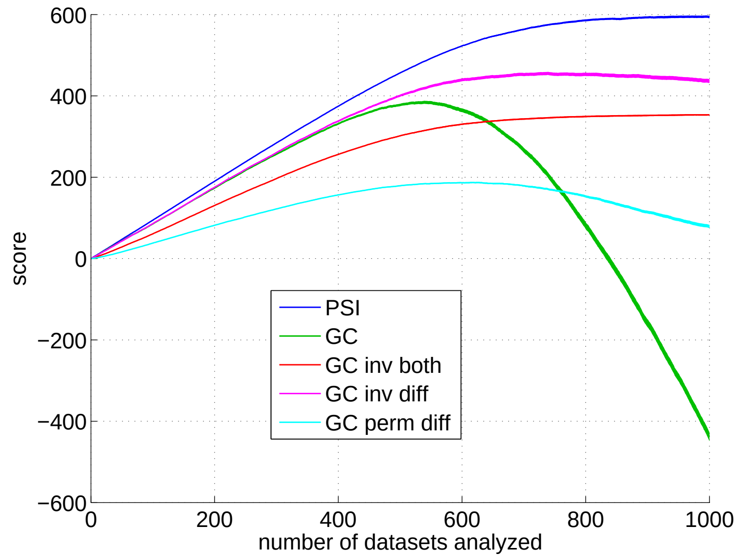

Fig. 2. Score according to the rules of the Signal Separation Evaluation Campaign (SiSEC) 2011 challenge on causal analysis of simulated EEG data as a function of the number of datasets analyzed for the Phase-slope Index (PSI) and different variants of Granger Causality (GC). Confidence intervals are indicated by linewidths. GC: original approach, requiring significant net flow. GC inv both: improved approach, requiring significant net flow and significant opposing net flow on time-reversed data. GC inv diff: improved approach, requiring significantly different net flows on original and time-reversed data. GC perm diff: improved approach, requiring significantly different net flows on original and temporally permuted data. Datasets are ordered by their antisymmetry-symmetry ratio (ASR) to illustrate that the analysis of datasets with low ASR with conventional Granger Causality is error-prone.

References

1. C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.
2. S. Haufe. *Towards EEG* source *connectivity analysis.* PhD thesis, TU Berlin, 2011.
3. M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol Cybern*, 85:145–157, 2001.
4. G. Nolte, F. C. Meinecke, A. Ziehe, and K.-R. Müller. Identifying interactions in mixed and noisy complex systems. *Phys Rev E*, 73:051913, 2006.
5. G. Nolte, A. Ziehe, V. V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and K. R. Müller. Robustly estimating the flow direction of information in complex physical systems. *Phys Rev Lett*, 100:234101, 2008.
6. G. Nolte, A. Ziehe, N. Krämer, F. Popescu, and K.-R. Müller. Comparison of Granger causality and phase slope index. *JMLR W&CP*, 6:267–276, 2010.