# METHODOLOGIES

## CAUSALITY MEASURES

**GrangerCausalityIndex (GCI)**

The standard Granger causality index (GCI) is estimated based on VARs (Granger, 1969). In the bivariate case, the variable $X_2$ Granger causes $X_1$, if the knowledge of past values of $X_2$ significantly improves the prediction of $X_1$. A bivariate autoregressive model of order $P$ (unrestricted / full model) is fitted to the time series $\{x_{1t}\}$:

$$x_{1t} = \sum_{j=1}^{P} \alpha_{1j} x_{2t-j} + \sum_{j=1}^{P} \alpha_{2j} x_{1t-j} + \varepsilon_{1t},$$

where $\alpha_{1j}, \alpha_{2j}$ are the coefficients of the VAR model and $\varepsilon_{1t}$ the residuals from fitting the model.The restricted model is similarly defined (but without variable $X_2$). If the variance $s_U^2$ of the residuals of the unrestricted model is significantly lower than the corresponding variance $s_R^2$ of the restricted model, then $X_2$ Granger causes $X_1$. The magnitude of the effect of $X_2$ on $X_1$ is given by the Granger Causality Index (GCI)

$$GCI_{X_2 \rightarrow X_1} = \ln \frac{var\,(s_R^2)}{var\,(s_U^2)}.$$

**Reference**

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

**ConditionalGrangercausality Index (CGCI)**

The Conditional Granger Causality Index (CGCI) extends GCI to themultivariate case(Geweke, 1982). Considering $K$ variables in total, with $X_1$ the response variable and $X_2$ the driving one,additional $K-2$conditioning variables are included in the models, denoted as$Z = \{Z_1, Z_2, .., Z_{K-2}\}$. The CGCI is given as:

$$CGCI_{X_2 \to X_1|Z} = \ln \frac{var(s_R^2)}{var(s_U^2)}.$$

**Reference**

J. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77(378):304–313, 1982.

**PartiallyConditionedGrangercausality (PCGC)**

The PartiallyConditionedGrangercausality (PCGC) is an extension of the ConditionalGrangerCausality Index(CGCI) (Marinazzo*et al.*, 2012). It quantifies the direct causality and is based on variable selection to reduce the dimensionality. A limited subset of the most informative variables for the driving one is set as the conditioning variables' ensemble using information gain as a criterion.

**Reference**

D. Marinazzo, M. Pellicoro, and S. Stramaglia. Causal information approach to partial conditioning in multivariate data sets. *Computational and Mathematical Methods in Medicine*, No. 2012, 2012.

**Partial Granger Causality (PGC)**

The Partial Granger Causality (PGC) is a multivariate direct causality measure defined again on VARs but inspired by the definition of partial correlation(Guo *et al.*, 2008). It is an extension of the standard Granger causality test appropriate for cases with latent and exogenous variables. The idea of PGC stems from the fact that the influence of exogenous and/or latent variables on a system will be reflected by correlations among the residuals of a VAR model of the measured variables.

We consider again the unrestricted and restricted multivariate VARs and denote the corresponding residual covariance matrixes of the two models as $\Sigma$ and $\rho$. In case of three variables $X_1, X_2, X_3$, the PGC is given as

$$PGC_{X_2 \to X_1|X_3} = ln\frac{\rho_{11} - \rho_{12}\rho_{22}^{-1}\rho_{21}}{\Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31}}.$$

Note that $\Sigma_{11} = var(s_{1U}^2)$and $\rho_{11} = var(s_{1R}^2)$.PGC is an improved estimation of the CGCI when the residuals of the VAR models are correlated. Otherwise, it is identical to the CGCI.

**Reference**

S. Guo, A.K. Seth, K.M. Kendrick, C. Zhou, & J. Feng. Partial Granger causality-eliminating exogenous inputs and latent variables. *Journal of neuroscience methods*, 172(1):79-93, 2008.

**Restricted ConditionalGrangerCausality Index (CGCI)**

The Restricted Conditional Granger Causality Index (RCGCI) is an extension of CGCI using dimension reduction to face the curse of dimensionality (Siggiridou&Kugiumtzis, 2016). For its computation, the VAR is restricted by a modified backward-in-time selection (mBTS) method. Autoregressive models are again fitted to the data, such as for CGCI, however the selected lagged variables are different for each variable, instead of having all lagged variables for a common maximum order for all variables, as in the definition of CGCI. To determine the suitable subset of lagged variables for each variable, an inherent property of time series is exploited, i.e. the fact that the dependence structure is closely related to the temporal order of the variables. The RCGCI is computed similarly to CGCI however makes use of the mBTS algorithm and therefore the determined autoregressive models have much fewer lagged terms. RCGCI is particularly effective in case of many observed variables and relatively short time series.

Specifically, we consider the vector $\boldsymbol{w}_1$ that includes the optimal lagged terms for predicting $X_1$ extracted from mBTS method selected from the original ensemble of lagged terms of all the observed variables $\{X_{1,t-1}, X_{1,t-2}, \dots, X_{1,t-p_{max}}, \dots, X_{K,t-1}, \dots, X_{K,t-p_{max}}\}$, where $p_{max}$ is the largest lag considered for each variable. Then, we check whether any lagged variable of $X_2$ enters in vector $\boldsymbol{w}_1$. If not, then $\text{RCGCI}_{X_2 \to X_1} = 0$. If there are lagged terms of $X_2$within$\boldsymbol{w}_1$, then we consider the unrestricted VAR model for

$X_1$ based on the lagged terms of $\boldsymbol{w}_1$. The corresponding restricted model for $X_1$is formed by excluding all lagged terms of variable $X_2$. In this case, RCGCI from $X_2$ to $X_1$is given similarly to CGCI, i.e., is the logarithm of the ratio of the variances of the residuals of the two models.

**Transferentropy (TE)**

Transferentropy is a bivariate causality measure from information theory that defines Granger causality based on entropy instead of VARs(Schreiber, 2000). It is model free and indicates both linear and nonlinear causal effects.

Computation of TE involves the formulation of uniformly spaced embedding vectors from each variable, e.g. for $X_1$: $\boldsymbol{x}_{1t} = [x_{1t}, x_{1t-\tau}, \dots, x_{1t-(m-1)\tau}]'$, where $m$ is the embedding dimension and $\tau$is the time lag. TE appreciates the effect of variable $X_2$ (driving variable) on$X_1$(response variable) by quantifying the improvement of the prediction of the future of $X_1$in time $t+1$, $x_{1t+1}$, when using additional information from the past of $X_2$ instead of utilizing only the past of $X_1$

$$\mathrm{TE}_{\mathrm{X}_2 \to \mathrm{X}_1} = H(x_{1t+1}|\boldsymbol{x}_{1t}) - H(x_{1t+1}|\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}),$$

where$H(x)$istheShannon entropy of a discrete variable$X$:$H(x) = -\Sigma p(x) \ln p(x)$.

The computation of TE involves joint and marginal probability functions. It has been proposed as a convenient method the k-nearest neighbors' technique (Kraskov*et al.*, 2004)which is proved to be stable and efficient, especially when interested to capture possible nonlinear causal effects.

**References**

A. Kraskov, H. Stogbauer, and P. Grassberger. Estimating mutual information. *Physical review E, 69*(6):066138, 2004.

T. Schreiber. Measuring information transfer. *Physical Review Letters*, *85*(2):461, 2000.

**Partialtransferentropy (PTE)**

Partialtransferentropyis the multivariate extension of TE that indicates only the direct causal effects in case of $K$ variables

$$\mathrm{PTE}_{\mathrm{X}_2\rightarrow \mathrm{X}_1|\mathrm{X}_3,\ldots,\mathrm{X}_K} = H(x_{1t+1}|\boldsymbol{x}_{1t}, \boldsymbol{x}_{3t}, \ldots, \boldsymbol{x}_{Kt}) - H(x_{1t+1}|\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}, \boldsymbol{x}_{3t}, \ldots, \boldsymbol{x}_{Kt}).$$

The k-nearest neighbors' estimator is used (Papana *et al.*, 2012).

**Reference**

A. Papana, D. Kugiumtzis, and P.G. Larsson. Detection of direct causal effects and application in the analysis of electroencephalograms from patients with epilepsy. *International Journal of Bifurcation and Chaos*, 22(9):1250222, 2012.

**Partial Transfer Entropy variants (PTE variants)**

The calculation of the partial entropy transfer (PTE) is based on the estimation of marginal and joint probability distributions. The larger the number of variables of a multivariate system, the greater the computational complexity of estimating the above probability functions while the accuracy of the estimation decreases. Taking into consideration this remark and regardless of the above main simulation study, some variants of the PTE areintroducedbased on variable selectionin order to address this issue.

To overcome the large-dimensional problem that occurs when calculating the PTE, one can reduce the number of conditioning variables based on a suitable criterion.The PTE variants consider a subset of the original conditioning variables of PTE with respect to connectivity pattern of each system (Papana *et al.*, 2020). Finding the optimal subset of conditioning variables is based on the connectivity of the variablesand in particular,the following two correlation measures are used:

i. linear correlation coefficient$r = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

ii. mutual information $\mathrm{I} = \mathrm{I}(\mathrm{X},\mathrm{Y}) = H(X) - H(X|Y)$

We examine how informative are the conditioning variables in relation to the driving variable and / or the response variable. Based on the above consideration, we define the following PTE variants in terms of the formulation of the subset of conditioning variables:

**1A**: Choose the most correlated variables to the driving one based on the linear correlation

**1B**: Choose the most correlated variables to the driving one based on mutual information

**2A**: Choose the least correlated variables to the driving one based on linear correlation

**2B**: Choose the least correlated variables to the driving one based on mutual information

**3A**: Choose the most correlated variables to the response one based on the linear correlation

**3B**: Choose the most correlated variables to the responseone based on mutual information

**4A**: Choose the least correlated variables to the response one based on linear correlation

**4B**: Choose the least correlated variables to the response one based on mutual information

**5A**: Choose the most correlated variables to the driving one based on random forest

**5B**: Choose the most correlated variables to the response one based on random forest

**5C**: Choose the most correlated variables to the driving and response one based on random forest

Inamultivariatesystemof$K$variables, letusdenote$\mathbf{Z}$asubsetof$\{X_3, \ldots, X_K\}$thatincludes$nc$conditioningvariables and has emerged from one of the above considered cases. Then the PTE variant is defined asPTE$_{X_2 \to X_1 | \mathbf{Z}}$.Depending on the estimated correlations, a smaller number of binding variables can be chosen than given $nc$.If there are no correlated variables, then the binary variable TE is calculated.

**Reference**

Papana, A., Papana-Dagiasis, A., &Siggiridou, E. Shortcomings of transfer entropy and partial transfer entropy: Extending them to escape the curse of dimensionality. *International Journal of Bifurcation and Chaos*, 30(16), 2050250, 2020.

## Mutual information on mixed embedding (MIME)

Mutual information on mixed embedding (MIME) is derived directly from a mixed embedding scheme based on the conditional mutual information criterion (Vlachos & Kugiumtzis, 2010). Therefore, it is a bivariate measure using dimension reduction.

An optimized mixed embedding vector of past terms fromtheobserved variables is formed, which best explains the future of the response variable.The terms chosen by each variable are not necessarily uniformly spaced. The mixed embedding vector $\boldsymbol{w}_t = [\boldsymbol{w}_t^{X_1}, \boldsymbol{w}_t^{X_2}]$with varying delays from the variables $X_1$ and $X_2$ is progressively formed regarding a conditional mutual information (CMI) criterion. The maximum lag for searching in $X_1$, $X_2$ is denoted as $L_{max}$ and is the only free parameter of this measure. However, it has been shown that MIME is not affected by $L_{max}$, if it is sufficiently large. Setting $L_{max}$ equal to a very large number may only affect the computational time of the measure and not its performance. Starting by an empty vector $\boldsymbol{w}_t^0$, a new vector $\boldsymbol{w}_t^j$ is formed at each step $j$, by adding a component $w_t^j$ (from $X_1$ or $X_2$), so that the future of the response variable $X_1$, $x_{1t+1}$, is best explained:

$$w_t^j = argmax_{w_t^j}\{I(x_{1t+1}; \boldsymbol{w}_t^j | \boldsymbol{w}_t^{j-1}\},$$

where $I(X;Y|Z)$ denotes the conditional mutual information of $X$ and $Y$ conditioning on $Z$.

The stopping criterion for determining the mixed embedding vector is based on an adjusted threshold. Specifically, a significance test is performed for the CMI of the new component selected as best. For the significance test, an ensemble of $nsur$surrogate time series is formed, and the test decision is made at a given significance level (we set $a = 0.05$). The surrogates are obtained using random permutations of the time indices of thecandidate lag variable and independent random permutations of the lagged variables already selected.To decide on a significant CMI, the p-value is obtained by the rank ordering the original and surrogate CMI values,applying the suggested correction of Yu& Huang (2001)

$$p - value = (1 - r_0 - 0.326)/(nsur + 1 + 0.348)$$

(one-sided test), where $r_0$ is the rank of the original CMI value.

The MIME is defined as

$$\mathrm{MIME}_{X_2\to X_1} = \frac{I(x_{1t+1};\, \boldsymbol{w}_t^{X_1}|\boldsymbol{w}_t^{X_2})}{I(x_{1t+1};\boldsymbol{w}_t)}.$$

Probability densitiesare computed using the k-nearestneighbors'estimator (KNN –Kraskov*et al.*, 2004). If there is no causal effect from $X_2$ to $X_1$, then $\mathrm{MIME}_{X_2\to X_1}$ is zero, otherwise is positive.

**Reference**

A. Kraskov, H. Stogbauer, and P. Grassberger. Estimating mutual information. *Physical review E*, *69*(6):066138, 2004.

I. Vlachos, and D. Kugiumtzis. Nonuniform state-space reconstruction and coupling detection. *Physical Review E*, 82(1): 016207, 2010.

G.H. Yu, and C.C. Huang. A distribution free plotting position. *Stochastic Environmental Research and Risk Assessment*, 15(6):462-476, 2001.

## Partial mutual information on mixed embedding (PMIME)

Partial mutual information on mixed embedding **(**PMIME) extends MIME to the multivariate case considering $K$ variables in total (Kugiumtzis, 2013). The NUE scheme is utilized to reduce dimensionality. A mixed embedding vector is formed by the inclusion of properly selected lagged terms from all the observed variables based on the CMI criterion, i.e., we form$\boldsymbol{w}_t = [\boldsymbol{w}_t^{X_1}, \boldsymbol{w}_t^{X_2}, \boldsymbol{w}_t^{Z}]$. For a predefinedmaximum lag $L_{max}$ , the ensemble of lagged terms for forming $\boldsymbol{w}_t$ is

$$B = \{x_{1t}, x_{1t-1}, \dots, x_{1t-L_{max}}, x_{2t}, x_{2t-1}, \dots, x_{2t-L_{max}}, \dots..\}.$$

The PMIME is equal to:

$$\text{PMIME}_{X_2 \rightarrow X_{1|Z}} = \frac{I(x_{1t+1};\ \boldsymbol{w}_t^{X_1} | \boldsymbol{w}_t^{X_2}, \boldsymbol{w}_t^{Z})}{I(x_{1t+1}; \boldsymbol{w}_t)}.$$

As for MIME, positive values of PMIME indicate the existence of causality, otherwise the measure is equal to zero. Surrogates are utilized in the stopping criterion to identify $\boldsymbol{w}_t$.

**Reference**

D. Kugiumtzis. Direct-coupling information measure from nonuniform embedding. *Physical Review E*, 87(6):062918, 2013.

## Partial transfer entropy with non-uniform embedding (PTENUE)

Partial transfer entropy with non-uniform embedding (PTEknnNUE) is again defined using the NUE scheme, similarly to MIME / PMIME (Montalto *et al.*, 2014).Its’ estimation procedure is like PMIME’s. However, the probability densities arecomputedonadifferent k-nearestneighbors’estimator (KNN), again introduced inKraskov*et al.* (2004). The stopping criterion relies on randomization of the driving variable. The decision for a significant CMI is made by comparing the CMI of the original data with the $1 - a$ percentile of the surrogate CMI values.

PTEknnNUE measures the direct effect of $X_2$ on $X_1$ in the presence of the "appropriate" past terms of the remaining variables:

$$\text{PTEknnNUE}_{X_2 \rightarrow X_{1|Z}} = I(x_{1t+1};\ \boldsymbol{w}_t^{X_2} | \boldsymbol{w}_t).$$

The measure is zero in case of no causality, otherwise is positive.

**Reference**

A. Kraskov, H. Stogbauer, and P. Grassberger. Estimating mutual information. *Physical review E*, *69*(6):066138, 2004.

A. Montalto, L. Faes, & D. Marinazzo. MuTE: a MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy. *PloS one*, 9(10):e109462, 2014.

## Nonlinear RCGCI (NRCGCI)

The linear measure Restricted Conditional Granger Causality Index (RCGCI) is extended to be able to capture also nonlinear causal effects within the framework of PoCoTe project. The nonlinear RCGCI (NRCGCI) in introduced to overcome the limitation of RCGCI to capture only linear couplings. Its' estimation procedure isimplementedin the basis of the RCGCI. The optimal mixed vector $\boldsymbol{w}_1$ for predicting variable $X_1$ is defined based on mBTS algorithm extracted from the original set of lagged terms that includes also nonlinear terms of second order, i.e.,products of lagged terms and lagged terms in the second power. Thus, the lagged terms of the unrestricted model are chosen from the following set

$$\boldsymbol{w} = \{\sum_{p=1}^{p_{\max}} \sum_{i=1}^{K} \mathrm{X}_{i,t-p}, \sum_{p_m=1}^{p_{max}} \sum_{p_n=1}^{p_{max}} \sum_{m=1}^{K} \sum_{n=1}^{K} \mathrm{X}_{m,t-p_m} X_{n,t-p_n}\}.$$

In the restricted model, we exclude all possible lagged terms including $X_2$, i.e. linear terms, squares of lagged terms of $X_2$ and products of lagged terms of $X_2$ with any other variable.

For example, let us consider the case of $K$=3 variables and $p_{\max} = 2$. Then, we determine $\boldsymbol{w}_1$ by choosing lagged terms from the set

$$\boldsymbol{w} = \{X_{1,t-1}, X_{2,t-1}, X_{3,t-1},$$

$$X_{1,t-2}, X_{2,t-2}, X_{3,t-2}, X_{1,t-1}^2, X_{2,t-1}^2, X_{3,t-1}^2, X_{1,t-2}^2, X_{2,t-2}^2, X_{3,t-2}^2, X_{1,t-1} * X_{2,t-1}, X_{1,t-1} * X_{3,t-1}, X_{2,t-1} * X_{3,t-1}, X_{1,t-2} * X_{2,t-1}, X_{1,t-2} * X_{3,t-1}, X_{2,t-2} * X_{3,t-1}, X_{1,t-1} * X_{2,t-2}, X_{1,t-1} * X_{3,t-2}, X_{2,t-1} * X_{3,t-2}\}.$$

To identify $\boldsymbol{w}_1$,i.e.the set of lagged variables for the unrestricted model, we apply the mBTS method in $\boldsymbol{w}$**.** To form the restricted model, we exclude from $\boldsymbol{w}$ any presence of the driving variable $X_2$ ($\boldsymbol{w}_{\boldsymbol{r}}$):

$$\boldsymbol{w}_r = \{X_{1,t-1}, \cancel{X_{2,t-1}}, X_{3,t-1}, X_{1,t-2}, \cancel{X_{2,t-2}}, X_{3,t-2}, X_{1,t-1}^2, \cancel{X_{2,t-1}^2}, X_{3,t-1}^2, X_{1,t-2}^2, \cancel{X_{2,t-2}^2},$$

$$X_{3,t-2}^2, \cancel{X_{1,t-1} * X_{2,t-1}}, X_{1,t-1} * X_{3,t-1}, \cancel{X_{2,t-1} * X_{3,t-1}}, \cancel{X_{1,t-2} * X_{2,t-1}}, X_{1,t-2} * X_{3,t-1}, \cancel{X_{2,t-2} * X_{3,t-1}},$$

$$\cancel{X_{1,t-1} * X_{2,t-2}}, X_{1,t-1} * X_{3,t-2}, \cancel{X_{2,t-1} * X_{3,t-2}}\}.$$

and we apply mBTS algorithm in $\boldsymbol{w}_r$. So, the mBTS method is used to define the optimal lagged vector for both the restricted and unrestricted model. In each step of mBTS method, the lagged variable with the least value, based on Bayesian Information Criteria (BIC – Schwarz, 1978), is randomized 100 times and is selected if the initial value of BIC lies inin the tails of distribution of the 100 BIC values (estimated from the 100 surrogate time series).

**References**

G. Schwarz. Estimating the dimension of a model. *The Annals of Atatistics*, 6(2): 461–464, 1978.

## Nonlinear Fast-Approximate Causal Discovery Algorithm (NLFACDA)

Hlinka &Kořenek (2020) introduced the Fast-Approximate Causal Discovery Algorithm for causality detection, which is based on conditional testing. It is a hybrid method that takes into consideration the estimation procedure of PMIME and of Runge's algorithm (Runge, 2012;Runge *et al.*, 2019).

Let us consider a random process $\{\boldsymbol{X}_t | t \epsilon \mathbb{Z}\}$, where$\boldsymbol{X}_t$is a multivariate random variable $\boldsymbol{X}_t = (X_t^1, X_t^2, \dots, X_t^K)^{\mathrm{T}}$, with the random variable $X_t^i$ indicating the state of element $i$ at the time $t$. The previous states of the system are expressed by$\boldsymbol{X}_t^- = (\boldsymbol{X}_{t-1}, \dots, \boldsymbol{X}_{t-\tau_{max}})$ while for each variable is defined as $X_t^{i-} = (X_{t-1}^i, \dots, X_{t-\tau_{max}}^i)$. To quantify the causal effect of the variable $X_{t-\tau}^j$ on the variable $X_t^i$ conditioned on all other elements of the system $\boldsymbol{X}_t$, the conditional mutual information $I(X_t^i; X_{t-\tau}^j | \boldsymbol{X}_t^- \setminus X_{t-\tau}^j)$ is used.

To face the problem of estimating high-dimensional information functionals, as the above CMI, dimension reduction methods are required. FACDA algorithm determines the set of causal parents $N_{X_t^i}$ which contains the elements $X_{t-\tau}^j$ that have a causal effect on element $X_t^i$:

$$N_{X_t^i} = \{X_{t-\tau}^j | I\left(X_t^i; X_{t-\tau}^j \middle| \boldsymbol{X}_t^- \backslash X_{t-\tau}^j\right) > 0$$

in three phases.In each step, iteratively includes in the parent set, the link with the strongest conditional mutual information, and limits the candidate set to only those links that have significant conditional mutual information (conditioned on the already established parents), limiting this way the computational cost of each step.

More detailed, the algorithm starts with an initial phase that generates a set ofcandidate causal parents for the target / responsevariable. This is doneiteratively by evaluating the conditional mutual information on the already identified candidate parents. In the second phase, the potential candidates are removed by iterative testing of their added value (conditional mutual information) with respect its subsets of increasing size. Finally, in the last step of the algorithm, additional partial correlation tests are performed to decide whether an element should be removed from the set of causal parents. The significanceof CMI is assessed based on permutation tests, which does not assume normal distribution and independence of samples.

**Reference**

Kořenek, J., & Hlinka, J. Causal network discovery by iterative conditioning: comparison of algorithms. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *30*(1), 013117, 2020.

J. Runge, J. Heitzig, V. Petoukhov, et al. Escaping the curse of dimensionality in estimating multivariate transfer entropy. Physical Review Letters, 108(25): 258701, 2012.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., &Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. Science Advances, 5(11), eaau4996, 2019.

## Linear Fast-Approximate Causal Discovery Algorithm (LFACDA)

The linear version of FACDA, namely denoted as LFACDA, is defined in terms of the partial correlation coefficient (Hlinka &Kořenek, 2020). Authors exploit the fact that under the Gaussian hypothesis, an estimate of the conditional mutual information $I(X;Y|Z)$can be obtained based on partial correlation$\rho(X,Y|Z)$:

$$I(X;Y|Z) = -\frac{1}{2}\log(1-\rho(X,Y|Z)^2),$$

Thus,LFACDA is computed on partial correlation instead of conditional mutual information. This way, the computational cost is minimal. To further speed up the simulations, significance test for partial correlation is assessed by Student's t-test, instead of considering random permutations.

**Reference**

Kořenek, J., & Hlinka, J. Causal network discovery by iterative conditioning: comparison of algorithms. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *30*(1), 013117, 2020.

## Low-dimensional approximation searching algorithm for Transfer Entropy (LATE)

Zhang (2018) introducedan estimator of transfer entropy (TE) based on the non-uniform embedding (NUE) scheme that exploits thelow-dimensional approximation algorithms for the estimation of the multivariate conditional mutual information (CMI).

The estimation procedure of LATE is like PMIME's, however the selection of thelagged terms that form the mixed embedding vectors considers sums of low-dimensional CMIs instead of a high-dimensional one. Thus, we substitute the relationship regarding $w_t^j$ (as given in MIME) by

$$w_t^j = argmax_{w_t^j}\left\{I\left(x_{1t+1};\, w_t^j\right) - \frac{2}{|B|}\sum_{w_t^i \in B} I\left(w_t^j;\, w_t^i\right) - \frac{2}{|B|}\sum_{w_t^i \in B} I\left(w_t^j;\, w_t^i | x_{1t+1}\right)\right\}.$$

After defining the mixed embedding vector $\boldsymbol{w}_t$, LATE is expressed as

$$\text{LATE}_{X_2 \to X_{1|Z}} = I(x_{1t+1};\, \boldsymbol{w}_t^{X_2} | \boldsymbol{w}_t).$$

In the termination criteria, the same low-dimensional approximation scheme is considered for the computation of the surrogate LATE values.

**Reference**

J. Zhang. Low-dimensional approximation searching strategy for transfer entropy from non-uniform embedding. *PloS one*, 13(3): e0194382, 2018.

## Partial Transfer Entropy on Rank Vectors (PTERV)

The Partial Transfer Entropy on Rank Vectors (PTERV)utilizes rank points instead of the original time delayed vectorsof the time series(Kugiumtzis, 2013).From the embedding vectors of each observed time series, we form the corresponding rank-points. For example, for the time series $\{x_{1t}\}, t = 1, .., n$, the embedding vectors are of the form $\boldsymbol{x_{1t}} = (x_{1t}, x_{1t-\tau}, \dots, x_{1t-(m-1)\tau})'$, where $m$ is theembedding dimension and $\tau$ is the time delay. The rank-point$\boldsymbol{\hat{x}_{1t}} = (r_1, \dots, r_m)'$ ($r_j \epsilon \{1, \dots, m\}, j = 1, \dots, m$) is formed by the ranks of the corresponding amplitude values of the embedding vector $\boldsymbol{x_{1t}}$when arranged in ascending order. The future$x_{1t+1}$of the response variable$X_1$of one step ahead at time t, is given by the rank$\boldsymbol{\hat{x}_{1t+1}}$of $x_{1t+1}$, when sorting the observations of the joint vector $[x_{1t+1}, \boldsymbol{x_{1t}}]$. ThePTERV is defined as

$$\begin{aligned}\text{PTERV}_{X_2 \to X_{1|Z}} &= I(\boldsymbol{\hat{x}_{1t+1}};\, \boldsymbol{\hat{x}_{2t}} | \boldsymbol{\hat{x}_{1t}}, \boldsymbol{\hat{z}_t}) \\ &= \sum_{t=(m-1)\tau+1}^{n-1} p(\boldsymbol{\hat{x}_{1t+1}}, \boldsymbol{\hat{x}_{2t}}, \boldsymbol{\hat{x}_{1t}}, \boldsymbol{\hat{z}_t})\, log\, \frac{p(\boldsymbol{\hat{x}_{1t+1}} | \boldsymbol{\hat{x}_{2t}}, \boldsymbol{\hat{x}_{1t}}, \boldsymbol{\hat{z}_t})}{p(\boldsymbol{\hat{x}_{1t+1}} | \boldsymbol{\hat{x}_{1t}}, \boldsymbol{\hat{z}_t})}\end{aligned}$$

where $I(\cdot)$ denotes the mutual information and $p(\cdot)$ the probability mass function.

**Reference**

D. Kugiumtzis. Partial transfer entropy on rank vectors. *The European Physical Journal Special Topics*, 222(2): 401-420, 2013.