Correcting Multivariate Auto-Regressive Models for the Influence of Unobserved Common Input

Vicenç GÓMEZ [a], Mohammad GHESHLAGHI AZAR [b] Hilbert J. KAPPEN [c]

[a] *Department of Information and Communication Technologies, Universitat Pompeu Fabra. Barcelona, Spain (vicen.gomez@upf.edu)*
[b] *Rehabilitation Institute of Chicago, Northwestern University Chicago, USA*
[c] *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, the Netherlands*

Abstract. We consider the problem of inferring connectivity from time-series data under the presence of time-dependent common input originating from non-measured variables. We analyze a simple method to filter out the influence of such confounding variables in multivariate autoregressive models (MVAR). The method learns the parameters of an extended MVAR model with latent variables. Using synthetic MVAR models we characterize where connectivity reconstruction is possible and useful and show that regularization is convenient when the common input has strong influence. We also illustrate how the method can be used to correct partial directed coherence, a causality measure used often in the neuroscience community.

Keywords. MVAR, common input, expectation maximization, connectivity

1. Introduction

Estimating connectivity from time series data is of fundamental interest, for instance, in economics or neuroscience [1]. Many techniques rely on parametric linear models, e.g. multivariate autoregressive (MVAR) models, which are learned from time series data as a first step. From the parameters of the estimated model, a variety of different coupling measures such as coherence or Granger causality, can be computed [2,3].

There are several factors that can affect the accuracy of a MVAR model, e.g. nonstationarities or nonlinearities in the data or a wrong model order selection. A more general problem is the influence of unobserved latent variables (e.g. common input). Strong influence of such confounding variables can lead to bias in the connectivity measures, resulting, for instance, in the identification of spurious dependencies. In this paper, we focus on how common input can influence model estimation in a way that is not originally captured in a MVAR model, e.g. as temporally and spatially uncorrelated Gaussian noise.

The standard approach to deal with unobserved common input is using latent factors. For example, in neuroscience the common input problem is related to the fact that each sensor measures a noisy mixture of cortical sources (volume conduction problem) [4]. Latent variables in this case model the brain activity sources. However, due to model degeneracy, the interpretability of the estimated connectivity in the latent space has limitations. Standard ways to address this identifiability problem rely on physical models of electro- and magnetoencephalography (EEG and MEG) that are inverted under physiologically motivated constraints (see [5,6,7] for a few examples). Such models require very specialized prior knowledge that, if incorrect, can affect dramatically the results. Although the problem is less critical for intra-cortical recordings (also known as ECoG) due to better signal quality, it is often the case that this type of recordings only covers partially the brain [8], thus relevant activity from other brain regions is overlooked.

We propose a simple method that provides directly the connectivity estimates between the observed variables while correcting for the influence of the unobserved common input. Specifically, we consider the problem of learning a MVAR for which part of the variables are latent. We thus assume a linear model for the common input process. This choice permits to capture more complex confounding influences than just uncorrelated Gaussian noise.

Throughout this paper, we use the term connectivity to denote the matrix of linear coefficients of the MVAR model. In the next section, we describe the problem formulation and our parameter learning method. In section 3, we characterize under which conditions one can recover the true underlying connectivity pattern and illustrate the applicability of the proposed method to correct partial directed coherence, a causality measure used often in the neuroscience community.

2. MVAR Models with Common Input

A $p-$th order multivariate autoregressive model, MVAR(p), describes the dynamic interactions among a vector of n variables $\mathbf{y}_t$ where each variable is regressed on p of its own lags as well as on p lags of each of the other variables. We consider a MVAR(p) process in which observations $\mathbf{y}_t$ are affected by a $m \times 1$ vector of (unobserved) common-input $\mathbf{z}_t$. The influence of $\mathbf{z}_t$ on $\mathbf{y}_t$ is instantaneous, but non-homogeneous:

$$\mathbf{y}_t = \sum_{k=1}^{p} A_y^{(k)} \mathbf{y}_{t-k} + B\mathbf{z}_t + \boldsymbol{\xi}_t. \tag{1}$$

Here $\boldsymbol{\xi}_t \sim \mathcal{N}(0, Q_y)$. The common-input process is assumed to evolve as a first-order Markov process:

$$\mathbf{z}_t = A_z \mathbf{z}_{t-1} + \boldsymbol{\eta}_t \tag{2}$$

with $\boldsymbol{\eta}_t \sim \mathcal{N}(0, Q_z)$ and initial state $\mathbf{z}_0 \sim \mathcal{N}(0, Q_z^0)$. For multidimensional $\mathbf{z}_t$, such common-input model can represent rich and complex dynamics, including oscillations with correlations at multiple time-scales.

This model can be seen as an extension of a standard MVAR(p) model with additional latent variables $\mathbf{z}_t$ (for $B = 0$), but it can also be seen as an extension of an ordinary linear dynamical system (LDS) with coupled observations. In an ordinary LDS, the latent factors $\mathbf{z}_t$ form a Markov chain as in (2) but the observed variables are independent of the previous observations ($A_y^{(k)} = 0, k = 1, \ldots, p$) and only depend in a linear-Gaussian manner on the latent values.

The common input MVAR model can be also expressed as an *extended* LDS

$$\begin{aligned} \mathbf{x}_t &= A\mathbf{x}_{t-1} + \mathbf{w}_t \\ \mathbf{y}_t &= C\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \tag{3}$$

with a new latent space that includes both the common input and the observed process

$$\underset{((np+m)\times 1)}{\mathbf{x}_t} = \begin{bmatrix} \mathbf{y}_t & \mathbf{y}_{t-1} & \cdots & \mathbf{y}_{t-p} & \mathbf{z}_t \end{bmatrix}^\top, \quad \underset{(n\times(np+m))}{C} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \end{bmatrix}$$

$$\underset{((np+m)\times(np+m))}{A} = \begin{bmatrix} A_y^{(1)} & \ldots & A_y^{(p-1)} & A_y^{(p)} & B \\ \mathbf{I}_n & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \ldots & \mathbf{I}_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & A_z \end{bmatrix}, \qquad \underset{(np\times 1)}{\mathbf{w}_t} = \begin{bmatrix} \boldsymbol{\xi}_t & \mathbf{0}\ldots\mathbf{0} & \boldsymbol{\eta}_t \end{bmatrix}^\top$$

$$\underset{((np+m)\times(np+m))}{Q} = \begin{bmatrix} Q_y & \mathbf{0} \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \ldots & \mathbf{0} \\ \vdots & \vdots \ddots & \vdots \\ \mathbf{0} & \mathbf{0} \ldots & Q_z \end{bmatrix}, \qquad \underset{(n\times n)}{R} = \mathbf{0}$$

with $\mathbf{w}_t \sim \mathcal{N}(0, Q)$ and $\mathbf{v}_t \sim \mathcal{N}(0, R)$ a dummy variable. We call this model the extended LDS (eLDS).

Denote the parameter vector $\theta = \theta_o \cup \theta_h$ where $\theta_o = \{A_y^{(1)} \ldots A_y^{(p)}, Q_y\}$ and $\theta_h = \{A_z, B, Q_z\}$. The existence of the latent structure introduces model degeneracy: θ_h can be estimated up to a similarity transformation. Typically, one sets the noise covariance Q_z to the identity matrix without loss of generality [9]. Note that the degeneracy only affects θ_h. Because $\mathbf{y}_t$ is observed, the parameters θ_o, which are the ones required for computing the connectivity between observables, are identifiable.

2.1. Parameter Estimation

The parameters θ of the eLDS of Eq. (3) or, equivalently, the extended MVAR(p) model of Eqs. (1) and (2) can be learned using maximum likelihood. Our interest is to maximize the complete log-likelihood function for observed $\mathcal{Y}$ and common input $\mathcal{Z}$ processes. However, since $\mathcal{Z}$ is not observed, this optimization can not be performed in closed form. We use expectation-maximization (EM) [9,10] or,

alternatively, expected gradient based methods which are more convenient than EM for low observation noise [11].

A specific problem that appears in this scenario occurs when the influence of the common input is strong (*large* B). In those cases, the optimization tends to overestimate the entries of A_y. To prevent this, we introduce a ℓ-2 norm regularization term in the objective function that penalizes large values of A_y. The optimization problem becomes (for a given value of the regularizer λ)

$$\theta^*_{ML} = \text{argmax}_\theta \mathcal{L}(\theta; \mathcal{Y}, \mathcal{Z}) = \text{argmax}_\theta \log \prod_t p(\mathbf{y}_t, \mathbf{z}_t|\theta) + \frac{\lambda}{2} \sum_{ij} |A_y|^2_{ij}. \quad (4)$$

Optimization in the presence of latent variables strongly depends on the initialization. We define our initialization strategy considering two independent optimizations on the simpler models, namely, the simple LDS with $A_y^{(k)} = 0, k = 1, \ldots, p$ (independent outputs given the common input) and the MVAR(p) model with $B = 0$ (no common input) to obtain initial values for all parameters.

First, in the case of the simple LDS, we use subspace identification (SSID) methods [12]. SSID methods first construct an estimate of the latent state sequence by decomposing (using singular value decomposition) an observation matrix (Hankel matrix) and then solve a least squares problem. The dimension of the latent space can be selected based on the spectrum of the decomposition, see $N4SID$ method [13]. With this initialization, we run EM or expected gradient method [9] to obtain the initialization of the simple LDS.

Second, we initialize coefficients $A_y^{(k)}, k = 1, \ldots, p$ as the solution of the standard MVAR(p) problem without common input. This is a convex problem and the maximum likelihood estimator for $\mathcal{A} = \left[A_y^{(1)}, \ldots, A_y^{(p)}\right]$ is given by

$$\widehat{\mathcal{A}} = \left(\sum_{t=1}^{T} \mathbf{y}_t \boldsymbol{\nu}_{t-1}^\top\right) \left(\sum_{t=1}^{T} \boldsymbol{\nu}_t \boldsymbol{\nu}_t^\top\right)^{-1} \quad (5)$$

where $\underset{(np\times 1)}{\boldsymbol{\nu}_t} = [\mathbf{y}_t \quad \mathbf{y}_{t-1} \quad \ldots \quad \mathbf{y}_{t-p}]^\top$. We use $\widehat{\mathcal{A}}$ as an initialization for our procedure and also as a baseline for comparison in the next section.

Summarizing, we initialize parameters A_z, B and Q_y using SSID+{EM or gradient method} and parameters $\mathcal{A}$ using Eq. (5). Once we have initial values for θ, we can finally optimize for the full eLDS of Eq. (3). This is a standard parameter estimation procedure with minor changes. We reproduce here the equations for completeness. For fixed θ, the E-Step computes the following expectations under the posterior $p(\mathbf{x}|\mathbf{y}, \theta)$ using Kalman filtering and smoothing:

$$M_{ij} := \frac{1}{T-1} \sum_{t=1}^{T} \langle \mathbf{x}_{t+i} \mathbf{x}_{t+j}^\top \rangle, \quad i, j = \{0, 1\}. \quad (6)$$

Given these expectations (6) we can either solve for θ in closed form (M-Step) and alternate both E-Step and M-Steps until convergence or use a gradient based procedure on the expected log-likelihood. The gradients and update equations are

- Common input dynamics matrix A_z:

$$\frac{\partial \mathcal{L}}{\partial A_z} = -Q_z \sum_{t=1}^{T} M_{10}^{zz} + Q_z \sum_{t=1}^{T} A_z M_{00}^{zz}, \quad A_z = \left(\sum_{t=1}^{T} M_{10}^{zz}\right)\left(\sum_{t=1}^{T} M_{00}^{zz}\right)^{-1}$$

- Observed dynamics matrix:

$$\frac{\partial \mathcal{L}}{\partial A_y} = -Q_y \sum_{t=1}^{T} M_{10}^{yy} + Q_y \sum_{t=1}^{T} A_y M_{00}^{yy} + \lambda |A_y|,$$

$$A_y = \left(\sum_{t=1}^{T} M_{10}^{yy} - \lambda \boldsymbol{I}_y\right)\left(\sum_{t=1}^{T} M_{00}^{yy}\right)^{-1}$$

- Common input mapping:

$$\frac{\partial \mathcal{L}}{\partial B} = -Q_y \sum_{t=1}^{T} M_{10}^{yz} + Q_y \sum_{t=1}^{T} B M_{00}^{zy}, \quad B = \left(\sum_{t=1}^{T} M_{10}^{yz}\right)\left(\sum_{t=1}^{T} M_{00}^{zy}\right)^{-1}$$

- Output covariance matrix Q_y:

$$\frac{\partial \mathcal{L}}{\partial Q_y} = -\frac{T-1}{2} Q_y + \frac{1}{2} \sum_{t=1}^{T} \left(M_{00}^{yy} - B M_{10}^{\cdot y} - M_{01}^{y\cdot} B^\top + B M_{00} B^\top\right)$$

$$Q_y = \frac{1}{T-1}\left(\sum_{t=1}^{T} M_{00}^{yy} - B \sum_{t=1}^{T} M_{10}^{\cdot y}\right)$$

where superindices on the matrices M_{ij} denote indexing on the corresponding latent z or observed y components (index $\cdot$ corresponds to entire row/column(s)).

To optimize the regularization parameter λ we use an annealing approach: we start with a strong regularizer (high λ) and decrease it progressively. At each decreasing step, we use the solution of the previous step as initialization (warm start). This helps the convergence of the algorithm and works best in practice. The results we report are using this procedure.

3. Numerical Results

We show results using simulated data. First, we analyze a minimal example that can be characterized in detail. We then compare the performance of the algorithm on larger systems as we vary the relation between different model parameters. We also show the convenience of using regularization and illustrate the method using a higher-order system. For comparison, we will refer to the solution obtained from the proposed method as *corrected* solution as opposite to the *naive* (or not corrected) solution obtained from the maximum likelihood (ML) estimator that ignores the common input, Eq. (5). We measure the error as the area under the ROC curve. We found this measure preferable to other measures, i.e. the norm of the differences, since the ROC is invariant under rescalings.

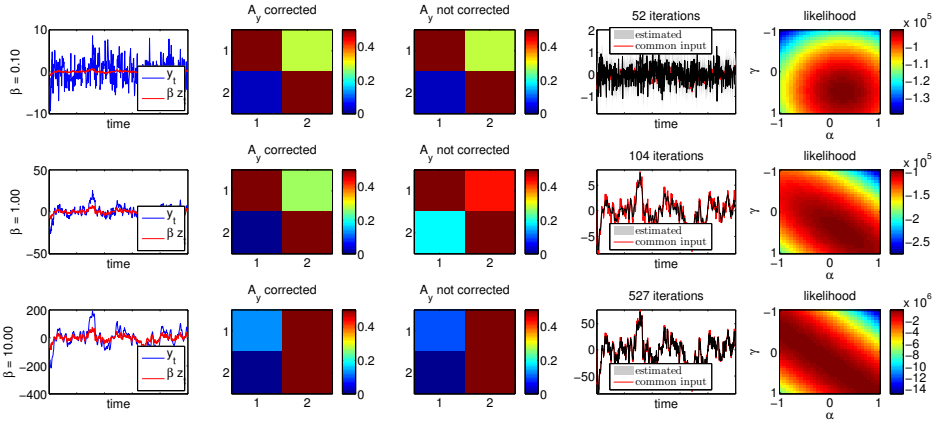

Figure 1. Three scenarios as a function of β for the minimal model for $\gamma = 0.5$ and $\alpha = 0.25$. **Top row** ($\beta = 0.1$): The common input has weak influence on the observed variables. The true connectivity can be recovered without correction. **Middle row** ($\beta = 1.0$): The influence of the common input is of the same order as the coefficients in A_y. No correction results in overestimated couplings and correction allows to fully recover the true connectivity pattern. **Bottom row** ($\beta = 10$): The common input has too much influence. **From left to right**: simulated time series; A_y estimated using the corrected method; A_y estimated using the naive method; real and estimated common input projected onto the first dimension of the output; expected likelihood as a function of γ and α. **Other parameters**: $T = 10^3, \sigma_y^2 = \sigma_z^2 = 1$.

3.1. Minimal Example

We consider the following two dimensional example with a one dimensional common input $A_y = [\gamma \quad \alpha; \quad 0 \quad \gamma]\,, B = \beta\,[1.05 \quad 0.95]^\top$ and $A_z = 0.95$. The influence of the common input is thus determined by the parameter β. We recognize three different scenarios, which are illustrated in the different rows of Fig. 1. The first column shows examples of time series generated from this toy model.

For $\beta = 0.1$ (top row) the common input (red) has little influence in the observed process (blue). In this case, the naive solution, Eq. (5), correctly estimates A_y. Correcting for the common input is harmless, as shown in the second and third columns respectively. The fourth column shows the estimated latent process projected onto the first observed variable (as obtained from the Kalman filtering/smoothing step) compared to the projected real common input (in red). The common input is not well approximated, but this has no effect on the predicted connectivity. The rightmost column shows the expected likelihood as a function of γ and α. We observe a clearly defined global optimum.

For $\beta = 1$ (middle row), the influence of the common input is significant. It is in this scenario where we expect best performance of the proposed approach. The naive method (third column) overestimates the off-diagonal elements of A_y whereas the corrected estimate (second column) recovers the true values.

For $\beta = 10$ (bottom row), the common input has too strong influence on the observed process. The method requires a large number of steps to initialize the parameters and little corrections are being made afterwards. Both naive and corrected estimates are biased. The common input, however, is correctly identified

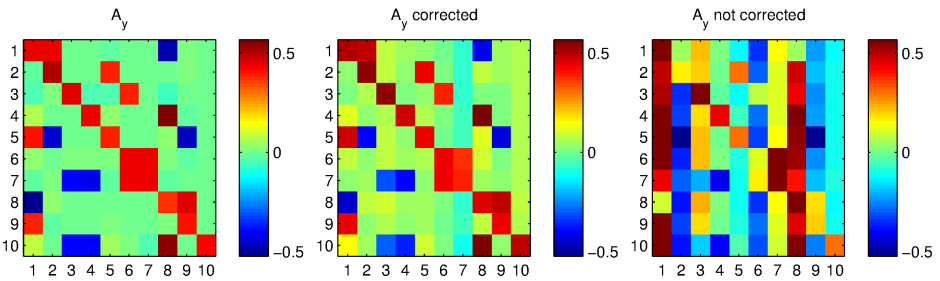

Figure 2. Illustration of the method: A_y: Randomly generated connectivity matrix; A_y corrected: solution recovered by the algorithm ; A_y not corrected: solution obtained from the standard approach, Eq. (5). Parameter values: $n = 10$, $m = 4$, $\beta = 1$, $\lambda_{y\min} = 0.1, \lambda_{y\max} = 0.8$, $\lambda_{z\min} = 0.9, \lambda_{z\max} = 0.95, \sigma_y^2 = 0.25^2$ and $T = 10^5$ time-steps.

(fourth column), since it has a strong signal. The expected likelihood (rightmost column) shows a ridge, where the maximum occurs for $\alpha = \gamma/2$. The algorithm in this situation requires a large number of iterations to initialize A_z, B and Q_y. After initialization, B contains large entries and no further significant corrections are made to the initial A_y. In these cases, regularization can help to find the correct solution.

3.2. Large, First Order Systems

We generate random MVAR(1) models for given dimensions m, n using the following procedure: A_y is obtained from a random $n \times n$ matrix with 60% of nonzero, Gaussian distributed entries with random sign. To create systems with prescribed dynamical time-scales, we compute the eigendecomposition of this initial matrix and rescale the eigenvalues within a prescribed range $[\lambda_{y\min}, \lambda_{y\max}]$. The final matrix A_y is obtained by multiplying again the rescaled eigenvalues with the eigenvectors. For a fully observable system, the dynamical time-scale is determined by $\tau = -1/\log \lambda_{y\max}$. We generate a random A_z in a similar way. The mixing matrix B is generated orthogonal (via Gram-Schmidt procedure) without loss of generality (non orthogonal B can be made orthogonal by redefining $\mathbf{z}_t$). The final B is rescaled by β. Finally, we fix the noise $Q_z = \mathbf{I}_m$ and parameterize the observation noise by σ_y^2, $Q_y = \sigma_y^2 \mathbf{I}_n$. The models are evolved until they reach stationary dynamics and then time series are obtained for estimation.

The full matrix A for such randomly generated eLDS has a block-triangular form $A = [A_y \quad B; \quad \mathbf{0} \quad A_z]$ and its eigenvalues are simply those of A_y and A_z. Hence, we can analyze the performance of the algorithm as a function of the interplay between the independent time-scales of the common input and the observed process by considering different values of $\lambda_{z\max}$ and $\lambda_{y\max}$.

Fig. 2 illustrates the method for $n = 10$ observed variables and $m = 4$ latent dimensions for a case where the exact pattern of connectivity can be reconstructed. In this example, $\lambda_{z\min} = 0.9$ and $\lambda_{y\max} = 0.8$, which corresponds to the situation in which the common input process has longer timescales than the observed process.

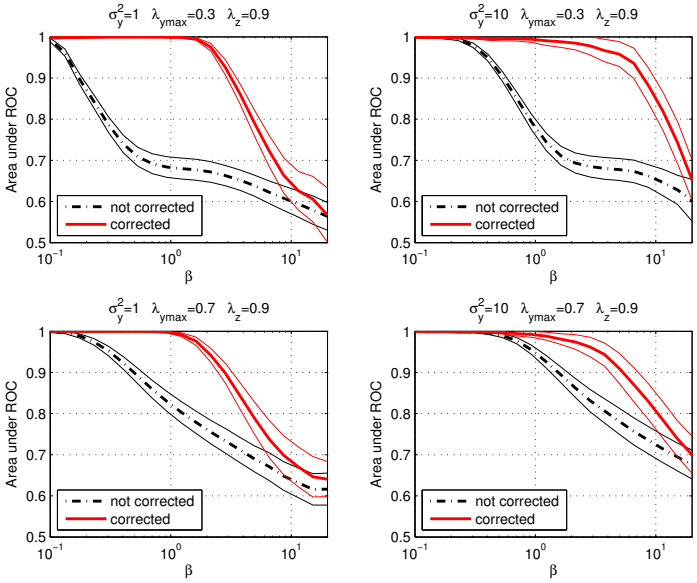

Figure 3. Comparison between corrected and non-corrected solutions in terms of area under the ROC curve. Plots show 95% confidence intervals of 50 random experiments, as a function of β for different time scales (λ_{ymax} vs λ_z) and noise values σ_y^2.

We now analyze the performance of the method for different conditions. Fig. 3 shows 95% confidence intervals of 50 random experiments for each condition as a function of β. In all figures, the performance of the naive method starts to decrease earlier than the corrected method.

The top rows show results for $\lambda_{y\max} = 0.3$ and $\lambda_{z\max} = \lambda_{z\min} = 0.9$. In these cases, the common input process evolves at a much longer time-scale than the observed process. Not correcting for the common input results in large bias. The improvement becomes less significant as $\lambda_{y\max}$ approaches $\lambda_{z\max}$. The bottom plots show results for $\lambda_{z\max} = 0.7$, where the improvement is still notable. Eventually, both curves collapse for $\lambda_{y\max} \approx \lambda_{z\max}$. Importantly, for $\lambda_{z\max} < \lambda_{y\max}$, correcting for the common input does not worsen the naive method (results not shown). In those cases, the time-scale of the common input is short and the process can be captured as noise by the naive method. In general, we can conclude that the improvement of the corrected over the naive method is proportional to the difference in $\lambda_{z\max} - \lambda_{y\max}$.

The behavior for different noise levels is illustrated by comparing the left ($\sigma_y^2 = 1$) and the right ($\sigma_y^2 = 10$) plots. In general, higher noise in the observations results in easier problems, as can be seen from the shift to the right of both curves. Also, higher noise levels yields more variability in the estimates. Thus, the improvement of the method becomes more pronounced for low observational noise. Interestingly, for this situation the EM algorithm tends to *freeze* and becomes impractical. In these cases, the expected gradient method converges faster [11]. The results shown are computed using the best approach in both conditions.

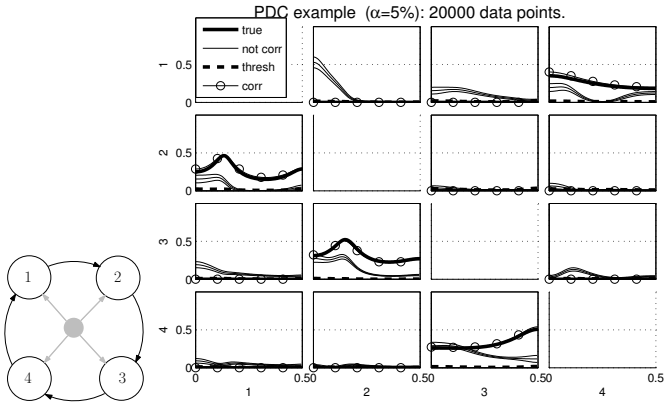

Figure 4. Higher-order system example. **Left**: underlying connectivity: a loop with a common input source. **Right**: comparison between partial directed coherence calculated from a MVAR(p) which is learned using the standard procedure (continuous line) and the corrected one using our approach (circles) against the ground truth (thick line). The corrected measure reproduces better the true PDC in terms of the frequency dependent values and also removes the spurious interactions detected by the naive method ($m = 2, n = 4, T = 20^3, p = 3, \beta = 10, \sigma_y = 1$).

3.3. Higher-Order Systems

We now look at higher-order systems ($p > 1$) and consider the partial directed coherence (PDC), a frequency-domain measure for Granger-causality [14] which is often used within the neuroscience community [15,1]. Given a MVAR(p) model estimated from time-series data, the PDC is defined as

$$\pi_{ij}(f) = \frac{|\Lambda_{ij}(f)|}{\sqrt{\sum_{\ell=1}^{n} |\Lambda_{\ell j}(f)|^2}}$$

where $\Lambda(f) = \sum_{k=0}^{p} A_y^{(k)} e^{-\mathrm{i}kf2\pi\Delta t}$ is the z-transform that converts the time-dependent coefficients $A_y^{(k)}, k = 1, \ldots, p$ to the frequency domain. Δt is the sampling interval and i is the imaginary number. Our example is a four dimensional observed process affected by two dimensional common input. The observed variables interact at different delays according to the following mixing matrices

$$A_y^{(1)} = \begin{bmatrix} 1/3 & 0 & 0 & 0 \\ 1/4 & 1/3 & 0 & 0 \\ 0 & 1/3 & -1/3 & 0 \\ 0 & 0 & 0 & 1/3 \end{bmatrix}, A_y^{(2)} = \begin{bmatrix} 1/10 & 0 & 0 & 0 \\ 0 & -1/10 & 0 & 0 \\ 0 & 0 & 1/10 & 0 \\ 0 & 0 & 1/3 & 0 \end{bmatrix}, A_y^{(3)} = \begin{bmatrix} -2/5 & 0 & 0 & 1/4 \\ 0 & -1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The underlying structure is a loop shown in Fig. 4(left). The common input process is governed by $A_{z_{ij}} = 0.99$ for $i = j$ and zero otherwise and B is composed of two orthonormal vectors and rescaled by β to force a strong influence of the confounding variables.

Fig. 4 shows comparison results between the PDC for the full system (denoted as `true`), the observed one (denoted as `not corr`, with 95% confidence intervals) and the corrected one [1]. We observe that the corrected measure reproduces better the true PDC in terms of the frequency dependent values and also removes the spurious interactions detected by the standard method.

4. Conclusions

In this paper, we have addressed the problem of estimating the dynamical structure of a MVAR model when some of the variables are not observed. We have assumed a linear-Gaussian model for the unobserved variables. The resulting model can be solved using a constrained form of EM or expected gradient methods. Using synthetic data, we have characterized the conditions in which the proposed method can successfully be applied. Future work includes applying the method to real-world data and incorporating sparsity-inducing regularization, as in [16].

References

[1] R. E. Greenblatt, M. E. Pflieger, and A. E. Ossadtchi. Connectivity measures applied to human brain electrophysiological data. *J Neurosci Meth*, **207** (2012), 1–16.

[2] L. Faes, S. Erla, and G. Nollo. Measuring connectivity in linear multivariate processes: definitions, interpretation and practical analysis. *Comput Math Method M*, 2012.

[3] A. Schlögl and G. Supp. Analyzing event-related EEG data with multivariate autoregressive parameters. In *Event-Related Dynamics of Brain Oscillations*, **159** (2006), 135–147.

[4] J.-M. Schoffelen and J. Gross. Source connectivity analysis with MEG and EEG. *Hum Brain Mapp*, **30** (2009), 1857–1865.

[5] B. L. P. Cheung, B. A. Riedner, G G. Tononi, and B. Van Veen. Estimation of cortical connectivity from EEG using state-space models. *Biomedical Engineering, IEEE Transactions on*, **57** (2010), 2122–2134.

[6] S. Haufe, R. Tomioka, G. Nolte, K. R. Müller, and M. Kawanabe. Modeling sparse connectivity between underlying brain sources for EEG/MEG. *IEEE Bio-Med Eng*, **57** (2010), 1954–1963.

[7] G. Gómez-Herrero, M. Atienza, K. Egiazarian, and J. L. Cantero. Measuring directional coupling between EEG sources. *Neuroimage*, **43** (2008), 497–508.

[8] A. T. Valderrama, R. Oostenveld, M. J. Vansteensel, G. M. Huiskamp, and N. F. Ramsey. Gain of the human dura in vivo and its effects on invasive brain signal feature detection. *J Neurosci Meth*, **187** (2010), 270–279.

[9] S. Roweis and Z. Ghahramani. A unifying review of linear Gaussian models. *Neural Comput*, **11** (1999), 305–345.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via de EM algorithm. *J. Roy. Statist. Soc. Ser. B*, **39** (1977), 1–38.

[11] R. K. Olsson, K. B. Petersen, and T. Lehn-Schiøler. State-Space Models: From the EM Algorithm to a Gradient Approach. *Neural Comput*, **19** (2007), 1097–1111.

[12] T. Katayama. *Subspace methods for system identification.* Springer Verlang, 2005.

[13] P. van Overschee and B. de Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, **30** (1994), 75–93.

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

[15] D. Y. Takahashi, L. A. Baccalá, and K. Sameshima. Connectivity inference between neural structures via partial directed coherence. *J Appl Stat*, **34** (2007), 1259–1273.

[16] H. J. Kappen and V. Gómez. The variational garrote. *Mach. Learn.*, **96** (2014), 269–294.

[1] We use the software from `http://www.lcs.poli.usp.br/~baccala/` to compute the PDC.