

Abstract

Obtaining a sparse representation of high dimensional data is often the first step towards its further analysis. Conventional Vector Autoregressive (VAR) modelling methods applied to such data results in noisy, non-sparse solutions with a too many spurious Coefficients. Computing auxiliary quantities such as the Power Spectrum, Coherence and Granger causality (GC) from such non-sparse models is slow and gives wrong results. Thresholding the distorted values of these quantities as per some criterion, statistical or otherwise, does not alleviate the problem.

We propose two sparse Vector Autoregressive (VAR) modelling methods that work well for high dimensional time series data, even when the number of time points is relatively low, by incorporating only statistically significant Coefficients. In numerical experiments using simulated data, our methods show consistently higher accuracy compared to other contemporary methods in recovering the true sparse model. The relative absence of spurious Coefficients in our models permits more accurate, stable and efficient evaluation of auxiliary quantities. Our VAR modelling methods are capable of computing Conditional Granger causality (CGC) in datasets consisting of tens of thousands of variables with a speed and accuracy that far exceeds the capabilities of existing methods.

Using the Conditional Granger causality computed from our models as a proxy for the weight of the edges in a network, we use community detection algorithms to simultaneously obtain both local and global functional connectivity patterns and community structures in large networks.

We also use our VAR modelling methods to predict time delays in many-variable systems. Using simulated data from non-linear delay differential equations, we compare our methods with commonly used delay prediction techniques and show that our methods yield more accurate results.

We apply the above methods to the following real experimental data:

1. Analysis of data from the Human Connectome Project (HCP): fMRI data from the HCP database is used to compute sparse brain functional connectivity networks. The network and community structures obtained are consistent over independent recording sessions and show good spatial correspondence with known functional and anatomical regions of the brain.
2. Analysis of ADHD-200 data: fMRI data from children with ADHD (Attention Deficit Hyperactive Disorder) is used to compute sparse brain functional connectivity networks. Analysis of the network measures obtained provide new ways of differentiating between ADHD and typically developing children using global and node-level network measures. They also enable refinement of published results relating to the rFIC-ACC interaction in fMRI resting state data.
3. Time-delay prediction from LFP recordings: When applied to Local Field Potential (LFP) recordings from the rat and monkey, our methods predict consistent

delays across a range of sampling frequencies.

4. Application to the Hela gene interaction dataset: The network obtained by applying our methods to this dataset yields results that are at least as good as those from a specialized method for analyzing gene interaction. This demonstrates that our methods can be applied to any time series data for which VAR modelling is valid.

In addition to the above methods, we apply non-parametric Granger causality analysis (originally developed by A. Nedungadi, G. Rangarajan et al) to mixed point process and local field potential data. Extending the computations to Conditional GC and by increasing the efficiency of the original algorithm, we can compute the Conditional GC spectrum in systems consisting of hundreds of variables in a relatively short period. Further, combining this with VAR modelling provides an alternate faster route to compute the significance level of each element of the GC and CGC matrices. We use these techniques to analyze mixed spike train and LFP data from monkey electrocorticography (ECoG) recordings during a behavioral task. A preliminary interpretation of the results of the analysis is provided.