Robust extraction of dependence structures in high-dimensional nonlinear time series

Supervisor: Alex Shestopaloff

Research Group: Statistics and Data Science

Funding: For September 2021 entry: Funding may be available through QMUL Principal's Postgraduate Research Studentships, School of Mathematical Sciences Studentships, and EPSRC DTP, in competition with all other PhD applications.

Studentships will cover tuition fees, and a stipend at standard rates for 3-3.5 years.

We welcome applications for self-funded applicants year-round, for a January, April or September start.

Project description:

When observed over time, natural physical systems often produce data recorded as high-dimensional, nonlinear time series. Examples are numerous and range from economics and finance to medicine, such as EEG data. Key insights about the physical system can be obtained by discovering latent structures present in the data [4]. An example of a latent structure is correlation relationships which refer to the dependence between two time series (potentially shifted in time), and can be used for the purposes of control or forecasting. The importance of this problem is broadly recognized; however, limited progress on robust detection of latent structures in high-dimensional time series data has been made to date. As an underexplored area with potentially high impact, it is well-suited to conduct novel research in.

The main goal of this project is to explore strategies to detect complex relationships in high-dimensional time series data. A basic strategy is to look at pairwise variables in isolation, e.g. pair-wise correlations. In many applications, this may only detect weak dependencies. However, when considering correlations between groups of variables (many-to-many, or many-to-one), the resulting correlations become significantly stronger [3]. This project explores several routes to discovering relationships in groups of time series, by leveraging a number of techniques broadly construed as *inverse problems on graphs*. The application domain is targeted at financial time series data, but there is scope to explore applications to other areas of interest as well, where high-dimensional time series data are ubiquitous.

This project will draw on three related areas. First, multi-reference alignment [1, 2], with a focus on the heterogeneous setting, and their application to time series alignment. Second, vector diffusion maps [6] as a dimensionality reduction tool in large panels of time series data. Third, the technique of angular synchronization [5] applied to time series, based on standard spectral methods; potential approaches based on graph neural networks can also be further explored.

References

[1] A. S. BANDEIRA, M. CHARIKAR, A. SINGER, AND A. ZHU, Multireference alignment using semidefinite program-ming, in Proceedings of the 5th conference on Innovations in theoretical computer science, 2014, pp. 459–470.

[2] N. BOUMAL, T. BENDORY, R. R. LEDERMAN, AND A. SINGER, Heterogeneous multireference alignment: a single pass approach, 2018 IEEE 52nd Annual Conference on Information Sciences and Systems, Princeton, NJ, (2018).

[3] H. V. NGUYEN, E. M ¨ULLER, J. VREEKEN, P. EFROS, AND K. B ¨OHM, Multivariate maximal correlation analysis, in International Conference on Machine Learning, 2014, pp. 775–783.

[4] J. RUNGE, P. NOWACK, M. KRETSCHMER, S. FLAXMAN, AND D. SEJDINOVIC, Detecting and quantifying causal associations in large nonlinear time series datasets, Science Advances, 5 (2019), pp. 1–15.

[5] A. SINGER, Angular synchronization by eigenvectors and semidefinite programming, Appl. Comput. Harmon. Anal., 30 (2011), pp. 20–36.

[6] A. SINGER AND H. T. WU, Vector diffusion maps and the connection Laplacian, Comm. Pure Appl. Math., (2012).

Further information:

How to apply

Entry requirements

Fees and funding