Sparse estimation in large multiple time series

An adaptation of the Union of Intersections method to vector autoregression

Trevor Ruiz, Oregon State University; MS degree capstone project advised by Dr. Sharmodeep Bhattacharyya

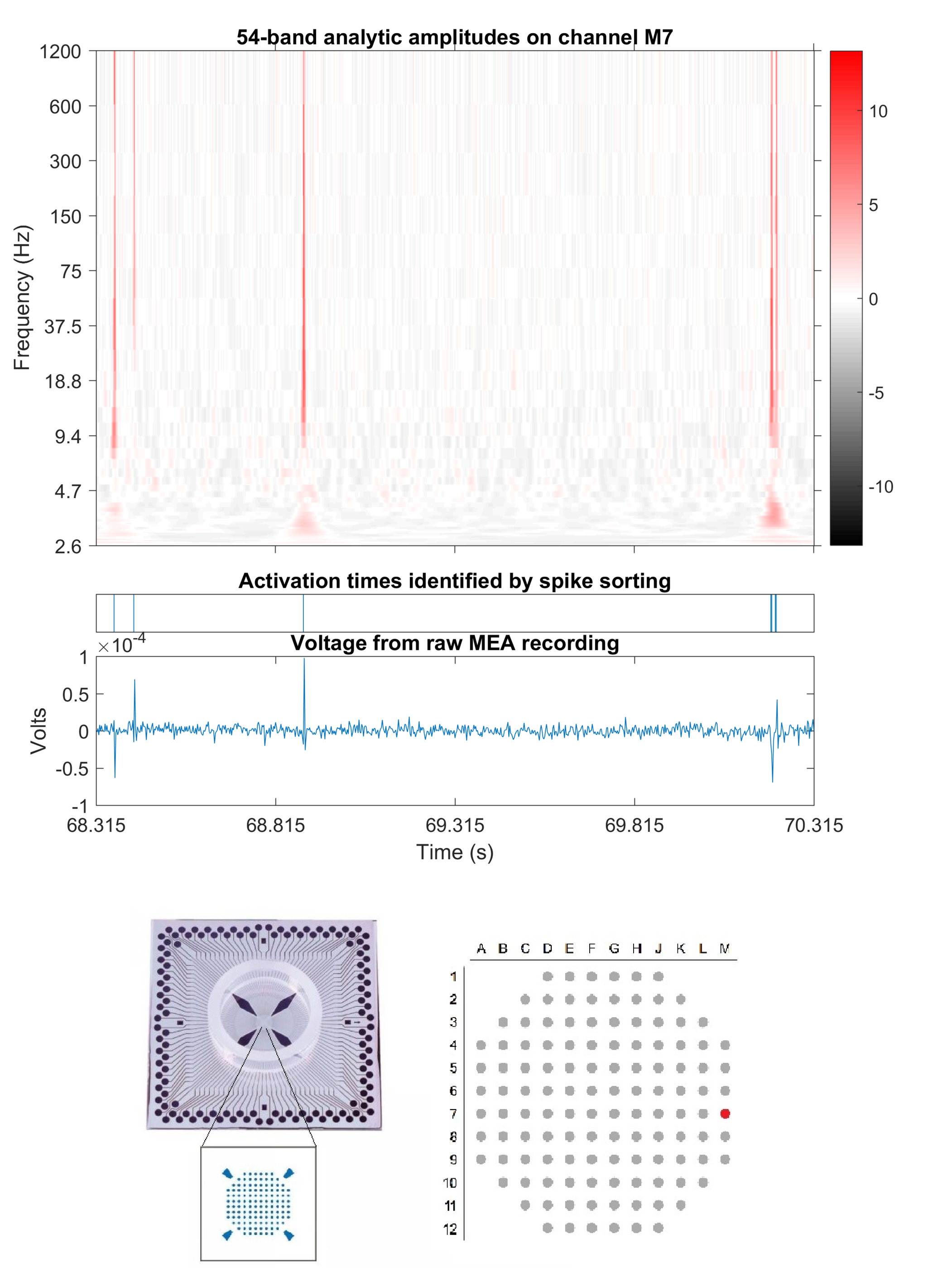
Overview

Results from a simulation study of a method of estimating transition matrices for high-dimensional autoregressive time series models are presented. The simulation study is part of a broader project in which such autoregressive models are candidate methods of graph construction for modeling neural functional connections from multi-electrode array electrocorticography recordings.

Background

Multielectrode array (MEA) devices record electrical voltage activity in a brain region or neural culture over time, and can comprise anywhere from 100 to over 1000 electrodes. One problem arising from MEA data is how to form graphs on the electrode array that estimate functional connections – areas or neurons that activate in sequence.

Figure 1: MEA data example – raw voltage, spike train, and time-frequency data for two seconds on one channel of 120-channel device, shown below with electrode configuration.



Estimation problem

A broad aim is to use a combination of the high-frequency components of time-frequency data from MEA recordings to estimate temporal correlations between electrodes.

$X_t \in \mathbb{R}^p$: vector of measurements on p channels at time t. A vector autoregressive model of order one is:

$$X_t = \boldsymbol{A}X_{t-1} + U_t\,, \qquad U_t \overset{iid}{\sim} \mathbb{N}_p(0,\Sigma)\,, \qquad \mathbf{det}(\boldsymbol{I} - \boldsymbol{A}z) \neq 0 \quad \forall\, |z| \leq 1$$

In graph construction the transition matrix coefficients are interpreted:

$$a_{ij} \neq 0 \iff \text{electrode } i \text{ at time } t \leftarrow \text{electrode } j \text{ at time } t-1$$

With data comprising observations at $N+1$ evenly-spaced times, the $VAR(1)$ model is

$$\begin{pmatrix} X_t' \\ X_{t-1}' \\ \vdots \\ X_1' \end{pmatrix}_{N\times p} = \begin{pmatrix} X_{t-1}' \\ X_{t-2}' \\ \vdots \\ X_0' \end{pmatrix}_{N\times p} \boldsymbol{A}'_{p\times p} + \begin{pmatrix} U_t' \\ U_{t-1}' \\ \vdots \\ U_1' \end{pmatrix}_{N\times p}$$

writing this compactly as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}' + \boldsymbol{E}$ and vectorizing yields a regression with correlated errors:

$$\text{vec}\boldsymbol{Y} = (\boldsymbol{I}_p \otimes \boldsymbol{X})\text{vec}\boldsymbol{A}' + \text{vec}\boldsymbol{E} \qquad i.e., \qquad \mathcal{Y} = \mathcal{X}\beta + \mathcal{E}$$

To obtain a sparse graph, a natural approach is to use an L_1-penalized least-squares-type estimator

$$\hat{\beta} = \arg\min_{\beta}\left\{\frac{1}{N}\|\mathcal{Y} - \mathcal{X}\beta\|_2^2 + \lambda\|\beta\|_1\right\} \qquad \text{(LS)}$$

which is consistent (under technical conditions on $\boldsymbol{X}, \boldsymbol{Y}$, and λ) per [1].

Union of Intersections for Vector Autoregression (UoI_{VAR})

UoI_{VAR} is a two-step procedure for estimating the transition matrix $\boldsymbol{A}$: first (**selection**) compute support sets on a regularization path $\boldsymbol{\lambda}$ using bolasso; then (**estimation**) compute bagged OLS estimates on supports chosen by cross-validation. This is a modification of the UoI method [2] adapted to vector autoregression.

Selection

Result: Support sets $S_1, \dots, S_k$ on regularization path
Data: $(\boldsymbol{X}, \boldsymbol{Y})$;
path $(\lambda_1, \dots, \lambda_k)$;
boostrap samples B_1
for $b = 1$ *to* B_1 **do**
  bootstrap sample $(\boldsymbol{X}^*, \boldsymbol{Y}^*)$;
  compute $(\mathcal{Y}^*, \mathcal{X}^*) = (\text{vec}\boldsymbol{Y}, \boldsymbol{I} \otimes \boldsymbol{X}^*)$;
  for $i = 1$ *to* k **do**
    compute lasso estimate $\hat{\beta}^{(\lambda_i)}$;
    compute lasso support $S^b_{\lambda_i} = \{j \mid \hat{\beta}^{(\lambda_i)}_j \neq 0\}$;
  end
end
compute bolasso support $S_i = \cap_{b=1}^{B_1} S^b_i$;

Estimation

Result: UoI estimate $\hat{\beta}$
Data: $(\boldsymbol{X}, \boldsymbol{Y})$; $(\lambda_1, \dots, \lambda_k)$; $S_1, \dots, S_k$;
boostrap samples B_2
for $b = 1$ *to* B_2 **do**
  training bootstrap sample $(\boldsymbol{X}^*, \boldsymbol{Y}^*)_1$;
  test bootstrap sample $(\boldsymbol{X}^*, \boldsymbol{Y}^*)_2$;
  compute $(\mathcal{Y}^*, \mathcal{X}^*)_i = (\text{vec}\boldsymbol{Y}^*, \boldsymbol{I} \otimes \boldsymbol{X}^*)_i$;
  for $i = 1$ *to* k **do**
    compute OLS estimate $\hat{\beta}^{(\lambda_i)}$ on $(\mathcal{Y}^*, \mathcal{X}^*_{S_i})_1$;
    compute prediction error e_i on $(\mathcal{Y}^*, \mathcal{X}^*)_2$;
  end
  compute $J = \arg\min_j e_j$;
  set $\hat{\beta}_b = \hat{\beta}^{\lambda_J}$;
end
compute $\hat{\beta} = B_2^{-1}\sum_{b=1}^{B_2}\hat{\beta}_b$

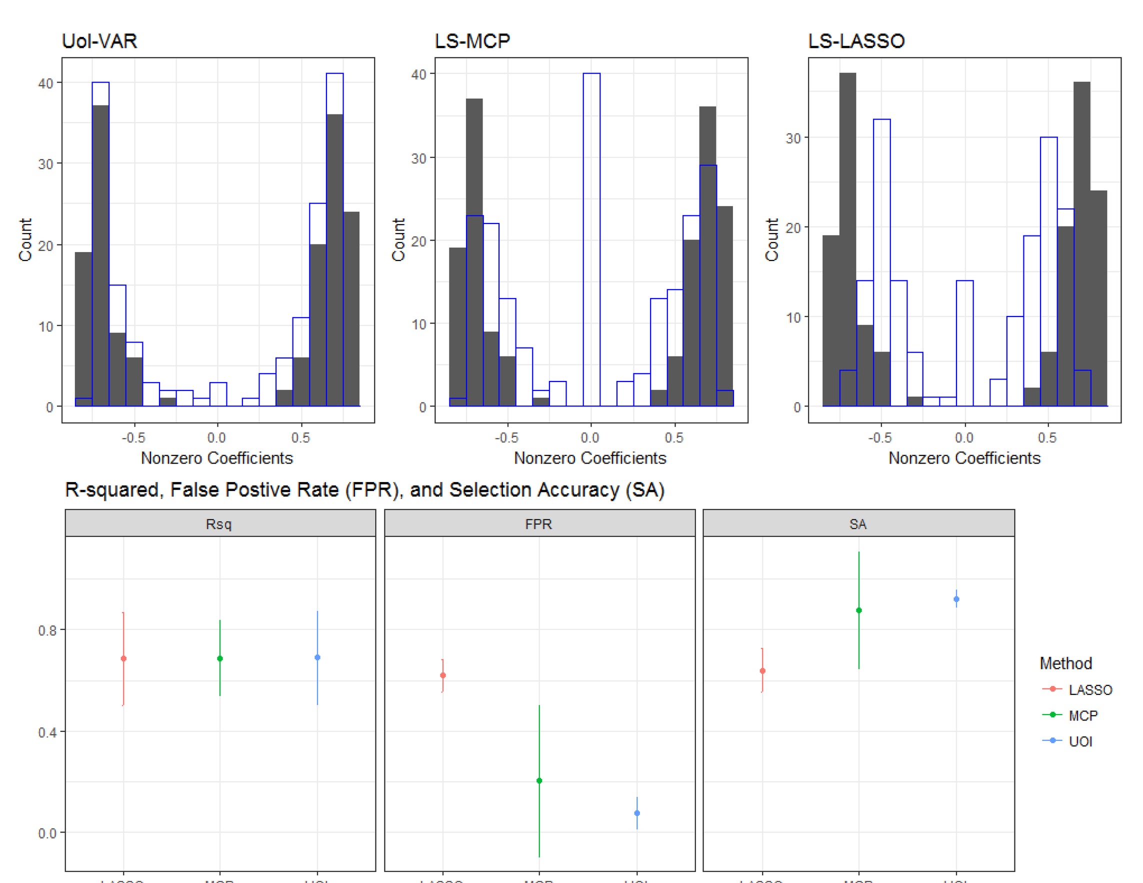
Numerical experiment

UoI_{VAR} estimates were computed on 100 synthetic datasets and compared with (LS) and a modified (LS) estimator using the minimax concave penalty.

Simulation

- Time series dimension $p = 160$
- Block-diagonal covariance $\Sigma = \boldsymbol{B} \otimes \boldsymbol{I}$
- Fixed transition matrix $\boldsymbol{A}$ comprising p randomly positioned nonzero coefficients (edges)
- Methods are compared with respect to R^2, false positive rate $\frac{|S_{\hat{\beta}} \setminus S_\beta|}{|S_{\hat{\beta}}|}$, selection accuracy $1 - \frac{|(S_{\hat{\beta}} \setminus S_\beta) \cup (S_\beta \setminus S_{\hat{\beta}})|}{|S_{\hat{\beta}}| + |S_\beta|}$, and the average estimate across all datasets (Fig. 2 histograms, top)

Figure 2: UoI_{VAR} outperforms other methods on synthetic data with respect to selection accuracy and bias.



References

[1] S. Basu and G. Michailidis.
Regularized estimation in sparse high-dimensional time series models.
The Annals of Statistics, 43(4):1535–1567, 2015.

[2] K. E. Bouchard et al.
Union of intersections (uoi) for interpretable data driven discovery and prediction.
arXiv:1705.07585v1 [stat.ML], 2017.
Submitted 22 May 2017.