Low-dimensional stationary subspace representation of multivariate time series

Anass El Yaagoubi

King Abdullah University of Science and Technology
STATISTICS Program
Biostatistics Group
http://biostats.kaust.edu.sa

Collaborators

- R. Sundararajan
- H. Ombao
- R. Frostig

Overview

Problem motivation

SSA: Existing methods

- SSA and DSSA
- Shortcomings

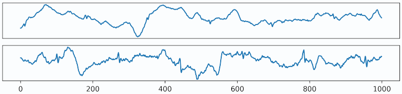
How to overcome the challenge of high dimensions

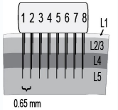
- Community structure based SSA - CSSA
- Efficient gradient descent algorithms
- Application to LFP recordings

Summary and ongoing work

Problem motivation

- Time series data is not always stationary
- Challenges to analyzing nonstationary time series:
  - Modeling covariance
  - Predictions
- Local Field Potential of rats: Stroke Experiment at UC Irvine





- A reasonable approach is to decompose nonstationary signals into: stationary + nonstationary latent sources
- Nonstationarity can be regarded as background activity (Kaplan et al.). In some applications removing nonstationarity can improve classification accuracy (P. von Bunau et al.)

Model setup

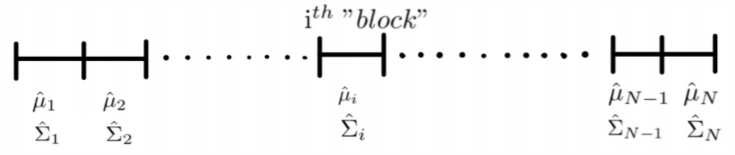
Let $X_t \in \mathbb{R}^p$ be the observed signal:

$$\begin{aligned} X_t &= A_1 Y_t^s + A_2 Y_t^n, & Y_t^s \in \mathbb{R}^d,\ Y_t^n \in \mathbb{R}^{p-d} \\ &= [A_1 | A_2] \begin{bmatrix} Y_t^s \\ Y_t^n \end{bmatrix} & \\ &= A Y_t & A^T A = I_p \end{aligned}$$

Goal: recover $\mathcal{C}(A_1)$ and $\mathcal{C}(A_2)$

Remark: A_1 and A_2 can not be recovered uniquely

SSA: P. von Bunau et.al, Physical Review Letters (2009)



- Under weak stationarity the first two moments should remain constant
- Use the Kullback-Leibler divergence (D_{KL}) to compare the first two moments (measure on nonstationarity)
- $\hat{\mu}_i^s = B_1 \hat{\mu}_i$, such that $B = [B_1^T | B_2^T]^T$
- $\hat{\Sigma}_i^s = B_1 \hat{\Sigma}_i B_1^T$
- $J_{SSA}(B) = \sum_{i=1}^N D_{KL}\left[\mathcal{N}(\hat{\mu}_i^s, \hat{\Sigma}_i^s) || \mathcal{N}(\hat{\mu}^s, \hat{\Sigma}^s)\right]$

DSSA: Sundararajan and Pourahmadi, JTSA (2018)

Define DFT and sample autocovariance in frequency domain as

- $d_X(\omega_k) = [d_{X_1}(\omega_k), \ldots, d_{X_p}(\omega_k)]^T = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t e^{-it\omega_k}$
- $\hat{\Gamma}_r^X = \frac{1}{T} \sum_{k=1}^{T} d_X(\omega_k) d_X^*(\omega_{k+r})$

Key idea

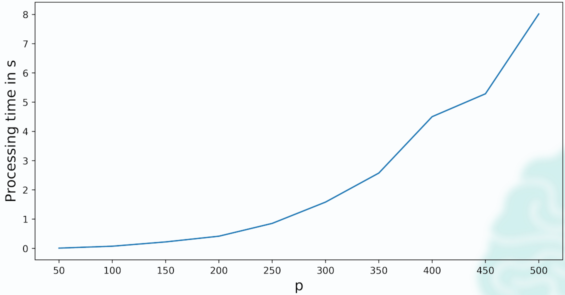
- Under weak statioarity, the Fourier coefficients at different frequencies are uncorrelated, i.e., $\mathrm{Cov}\left[d_X(\omega_k), d_X(\omega_{k'})\right] = \mathcal{O}(1/T)$
- $J_{DSSA}(B) = \sum_{r=1}^{m} \left|\left|B_1 \hat{\Gamma}_r^X B_1^T\right|\right|_F^2$, such that $B = [B_1^T | B_2^T]^T$

Optimization problem

- $\hat{B} = \underset{B}{\mathrm{argmin}} \;\; J_{DSSA}(B)$, subject to $BB^T = I_p$

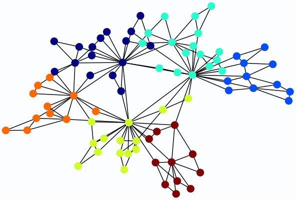
Theoretical and Empirical time complexity

- $\mathcal{T}(\hat{\Gamma}^X) \in \mathcal{O}\left(mp^2T + pT\log(T)\right)$
- $\mathcal{T}(\nabla J_{DSSA}) \in \mathcal{O}\left(m\left(d^2p + dp^2\right)\right)$
- Processing time becomes an issue as dimension p grows



Key ideas

- Divide and conquer strategy
  - Community detection in nonstationary time series
  - Community structure based SSA - CSSA



- Develop efficient optimization algorithms
- Clustering epochs based on canonical angles between subspaces

How do we find the communities?

Pseudo optimal solution for DSSA: Consider the special case d=1 (number of stationary components)

$$
\begin{aligned}
\sum_{r=1}^{m} \left| b\hat{\Gamma}_r^X b^T \right| &\leq \sum_{r=1}^{m} \left| b\Re(\hat{\Gamma}_r^X) b^T \right| + \left| b\,\Im(\hat{\Gamma}_r^X) b^T \right| \\
&= \sum_{r=1}^{m} \left| bP_r^R (D_{r+}^R - D_{r-}^R) {P_r^R}^T b^T \right| + \left| bP_r^I (D_{r+}^I - D_{r-}^I) {P_r^I}^T b^T \right| \\
&\leq b \left[\sum_{r=1}^{m} P_r^R (D_{r+}^R + D_{r-}^R) {P_r^R}^T + P_r^I (D_{r+}^I + D_{r-}^I) {P_r^I}^T \right] b^T \\
&= bUb^T
\end{aligned}
$$

Clustering based on eigenvectors of U

Key idea: A and B share information

$$\begin{pmatrix} \vdots \\ X_{i,t} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \underline{a_i}' \\ \vdots \end{pmatrix} \times Y_t \qquad \Longleftrightarrow \qquad Y_t = \begin{pmatrix} \dots & \underline{b_i} & \dots \end{pmatrix} \times \begin{pmatrix} \vdots \\ X_{i,t} \\ \vdots \end{pmatrix}$$

Key steps for doing community structure based SSA

1. Compute eigen decomposition $U = PDP^T$
2. Rows of P carry similar information to the columns of B
3. Cluster the rows of P
4. Run SSA on each cluster
5. Combine the results

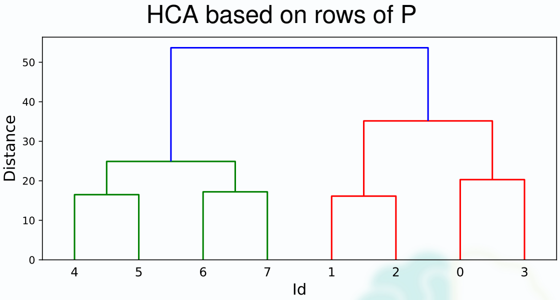
Simulation

- T=1024, p=8, d=4, m=4
- Stationary components: Different AR(2) processes corresponding to theta, alpha, beta, gamma bands (Gao et al., Statistica Sinica 2018)
- Nonstationary components: time varying AR(1)

$$A = \begin{pmatrix} 10 & 10 & 10 & 10 & 1 & 1 & 1 & 1 \\ 10 & 10 & -10 & -10 & 1 & 1 & -1 & -1 \\ 10 & -10 & -10 & 10 & 1 & -1 & -1 & 1 \\ 10 & -10 & 10 & -10 & 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 & 10 & 10 & 10 & 10 \\ -1 & -1 & 1 & 1 & 10 & 10 & -10 & -10 \\ -1 & 1 & 1 & -1 & 10 & -10 & -10 & 10 \\ -1 & 1 & -1 & 1 & 10 & -10 & 10 & -10 \end{pmatrix} / \sqrt{404}$$

Community detection and running time comparison

P	Community SSA	SSA
20	0.5s	0.2s
40	0.9s	0.9s
60	1.5s	3.0s
80	2.1s	6.6s
100	3.1s	13.0s
120	4.1s	20.9s



Gradient Descent with momentum[1]

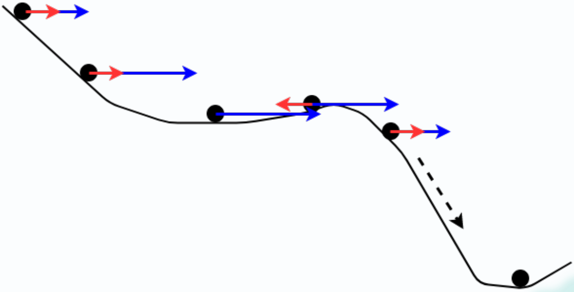
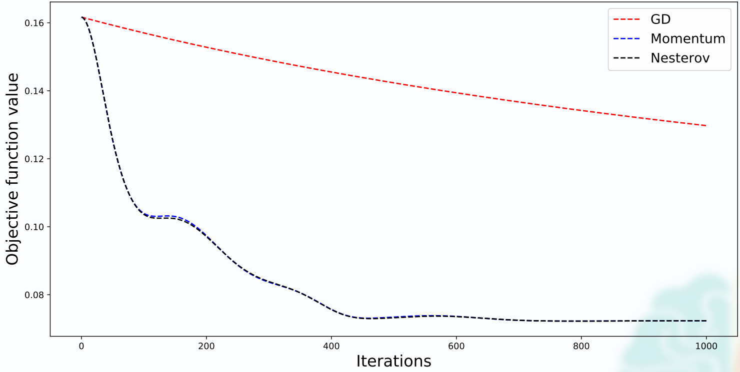


Figure: Gradient in red and momentum part in blue

[1] On the Momentum Term in Gradient Descent Learning Algorithms: Ning Qian, Neural Networks (1999)

Acceleration in practice



Gradient descent with multiplicative updates

Accelerated multiplicative updates

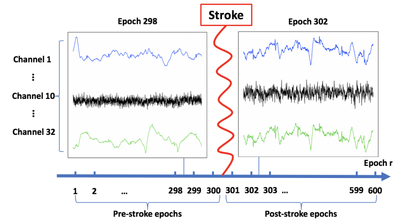
- Let $\{B_k\}$ be the sequence of iterates
- Let G_k be the gradient at B_k
- Let α be the step size and β be the momentum coefficient

$$\begin{aligned} H_k &= G_k B_k^T - B_k G_k^T \\ B_{k+1} &= R B_k \\ &= e^{-\alpha H_k} B_k, \alpha \in \mathbb{R}_+ \end{aligned}$$

$$\begin{aligned} H_k &= \tilde{G}_k B_k^T - B_k \tilde{G}_k^T \\ \tilde{H}_k &= \beta \tilde{H}_{k-1} + \alpha H_k \\ B_{k+1} &= e^{-\tilde{H}_k} B_k \end{aligned}$$

Local Field Potentials

- LFP signal: $X_t \in \mathbb{R}^{32}$ recorded over 10 minutes
- 1 epoch = 1 second
- Pre-stroke: 5 minutes (300 epochs)
- Post-stroke: 5 minutes (300 epochs)
- Sampling Rate: 1000 Hz (T=1000 per epoch)
- Data Source: Ron Frostig, UC Irvine Neurobiology.



LFP channel clusters for epochs 299, 300 and 301

Biostatistics at KAUST

Epoch 299 (pre-stroke)

0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31

Epoch 300 (post-stroke)

0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31

Epoch 301 (post-stroke)

0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31

Discriminating between pre and post stroke epochs

Key questions

- How to discriminate between pre and post stroke epochs ?
- How to use the stationary subspace as a ”feature” for discrimination ?
- Are there any changes in the stationary subspace that are induced by the stroke ?

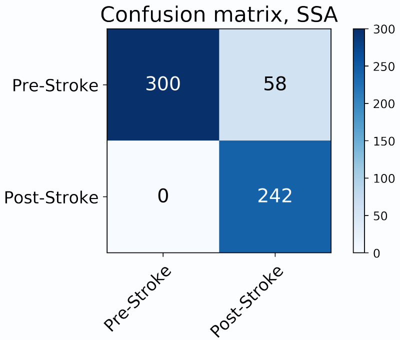
Solution

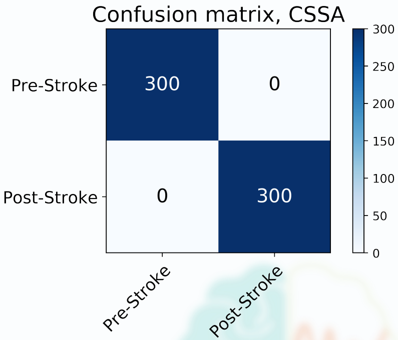
Step 1: Apply DSSA and get the demixing matrix B_1 for each epoch

Step 2: Compute the canonical angles between all pairs of recovered stationary subspaces

Step 3: Cluster the recovered subspaces using the canonical angles between the subspaces

Discriminating between pre and post stroke epochs





Summary and ongoing work

Summary

1. Computational challenge of high dimensions
2. Community detection on nonstationary time series
3. Community structure based SSA (CSSA)
4. Efficient optimization algorithms
5. Stationary subspaces can be used to discriminate between epochs in Local Field Potential recordings

Future work

- How to consistently estimate the dimension and the subspace itself
- Investigate stochastic gradient descent

Thank you for your attention!

Anass El Yaagoubi Bourakna

anass.bourakna@kaust.edu.sa

Github: github.com/A-EL-YAAGOUBI

Group Website: biostats.kaust.edu.sa

Post-doctoral position inquiries:

Dr. Hernando Ombao

hernando.ombao@kaust.edu.sa

References

P. von Bunau et.al, (2009)
Finding Stationary Subspaces in Multivariate Time Series
Physical Review Letters.

Sundararajan and Pourahmadi, (2018)
Stationary Subspace Analysis of Nonstationary Processes
Journal of Time Series Analysis.

Mark D. Plumbley, (2005)
Geometrical Methods for Non-negative ICA: Manifolds, Lie Groups and Toral Subalgebras.Neurocomputing
Neurocomputing.

Ning Qian, (1999)
On the Momentum Term in Gradient Descent Learning Algorithms
Neural Networks.