High Dimensional Data Analysis With Dependency and Under Limited Memory

Yiming Sun

Cornell University

Based on joint work with
Madeleine Udell (Cornell), Sumanta Basu(Cornell)
Yang Guo (UW Madison), Charlene Luo (Columbia)
Joel Tropp (Caltech), Amy Kuceyeski(Cornell University)
Yige Li(Harvard University)

September 9, 2019

Outline

Multivariate Spectral Density Estimation Under Weak Sparsity

Low Rank Tucker Approximation of a Tensor from Streaming Data

A Connection of These Two

*

Motivation

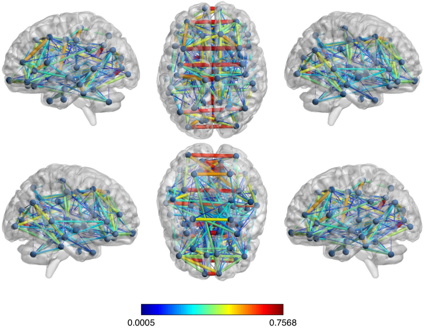

Figure: Interactions Between Regions in Brain

Weakly & Strongly Stationary Time

Definition (Weak Stationarity)

p-variate time series X is weakly stationary, if $\mathbb{E}X_t = \mathbb{E}X_s$ for any t, s and $\Gamma(\ell) := \mathbb{E}X_t X_{t-\ell}^\top$ only depends on the lag ℓ.

Definition (Strong Stationarity)

p-variate time series X is strongly stationary, if for any sequence $t_1, \cdots, t_n$, $X_{t_1} \cdots X_{t_n}$ has the same distribution of $X_{t_1+\tau} \cdots X_{t_n+\tau}$ for any integer τ,

Gaussian Process

Definition (Gaussian Process)

p-variate time series X is Gaussian process if for any sequence $t_1, \cdots, t_n$, $X_{t_1} \cdots X_{t_n}$ are jointly Gaussian distributed.

For Gaussian process, weak stationarity is equivalent to strong stationarity.

Spectral Density

Given a weakly stationary p-variate time series X, the spectral density at frequency $\omega \in [-\pi, \pi)$ is defined

$$f(\omega) = \sum_{\ell=-\infty}^{\infty} \Gamma(\ell) e^{-i\omega\ell}$$

where $\Gamma(\ell) = \mathbb{E} X_0 X_{-\ell}^{\top}$. X_t is independent with X_s, $t \neq s$ iff $f_{rs}(\omega) = 0$ for any ω.

Thresholding Estimator Under Weak Sparsity- A Example

Suppose that we have n observation of p-variate Gaussian distribution as follows.

$$y_i \overset{i.i.d}{\sim} \mathcal{N}\left(\mu, \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \sigma_p^2 \end{bmatrix}\right),$$

$i = 1, \cdots, n$.

A Example

The maximum likelihood estimator for μ_j is $\bar{y}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$.
Does not Work Well Under Weak Sparsity

$$\mu \in \left\{ \mu \in \mathbb{R}^p, \sum_{j=1}^{p} |\mu_j|^q \leq c_0(p) \right\}.$$

for some $0 \leq q < 1$ and $c_0(p)$ measures the weak sparsity.

Solution : Thresholding

Suppose $\sigma_i \leq B$, define element-wise thresholding operator

$$T_\lambda(x) = \begin{cases} x & |x| \geq \lambda \\ 0 & \text{else} \end{cases}$$

. Hard thresholding estimator $T_\lambda(\bar{y}_j)$ can be shown asymptotically consistent under weak sparsity where we set

$$\lambda \propto B c_0(p) \sqrt{\frac{\log p}{n}}$$

and assume $\lambda \to 0$.

Two Key Ingredients for Thresholding

Two key ingredients under above example assuming weak sparsity. Cai & Liu 2011

- An element-wise concentration inequality :

$$\mathbb{P}(|\bar{y}_j - \mu_j| \geq \eta) \leq 2\exp(-n\eta^2/2\sigma_j^2).$$

- σ_j are uniformly bounded.

Shortcomings for Hard Thresholding

- σ_j may variate much
- B will appear in the thresholding value making convergence rate slow

Two Key Ingredients for Thresholding

Two key ingredients under above example assuming weak sparsity.

- An element-wise concentration inequality :

$$\mathbb{P}(|\bar{y}_j - \mu_j| \geq \eta) \leq 2\exp(-n\eta^2/2\sigma_j^2).$$

- σ_j are uniformly bounded.

Solution: Adaptive Thresholding

Simply estimate σ_j, say with sample standard deviation:

$$\hat{\sigma}_j = \sqrt{1/(n-1)\sum_{i=1}^{n}(y_{ij} - \bar{y}_j)^2}$$

and replace B: $\lambda_j \propto \hat{\sigma}_j\sqrt{\frac{\log p}{n}}$. Now we can relax constraint in upper bound for σ_j and upper bound will not appear in rate of convergence.

An Similar Example: Covariance Matrix

$$y_i \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_{p\times p})$$

Goal: Estimate Σ assuming weak sparsity, $\|\Sigma\|_1 \leq c_0(p)$. Bickel & Levina 2008

An Similar Example: Covariance Matrix

Estimate the expectation of a vector of length p^2:

$$[(y_1 y_1^\top)_{rs}, 1 \leq r, s \leq p].$$

MLE: $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} y_i y_i^\top$. But we need to perform thresholding. Remember two ingredients:

- $\mathbb{P}(|\hat{\Sigma}_{rs} - \Sigma_{rs}| \geq \eta) \leq c_1 \exp(-c_2 n \eta^2)$
- $\mathbf{var}((y_1 y_1^\top)_{rs}) = \Sigma_{rr}\Sigma_{ss} + \Sigma_{rs}^2 \leq 2\max_{r=1}^{p} \Sigma_{rr}^2$

Thus Bickel & Levina 2008 presents an assumption $\max_{r=1}^{p} \Sigma_{rr}$ is bounded.

An Similar Example: Covariance Matrix

hard thresholding: $\lambda_{rs} \propto (\max_{r=1}^{p} \Sigma_{rr})\sqrt{\frac{\log p}{n}}$

adaptive thresholding: $\lambda_{rs} \propto \sqrt{\frac{\log p}{n}}$

Outline

Multivariate Spectral Density Estimation Under Weak Sparsity

Low Rank Tucker Approximation of a Tensor from Streaming Data

A Connection of These Two

*

Outline

Multivariate Spectral Density Estimation Under Weak Sparsity

Low Rank Tucker Approximation of a Tensor from Streaming Data

A Connection of These Two

*

Outline

Multivariate Spectral Density Estimation Under Weak Sparsity

Low Rank Tucker Approximation of a Tensor from Streaming Data

A Connection of These Two

*

References

Bickel, P. J. & Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 2577–2604.

Cai, T. & Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, *106*(494), 672–684.