Supplementary Material for "Graphical Model-Based Lasso for Weakly Dependent Time Series of Tensors"

Dorcas Ofori-Boateng[1,⋆], Jaidev Goel[2,⋆], Ivor Cribben[3], and Yulia R. Gel[2,4]

[1] Portland State University, Portland OR 97201, USA
dorcas.oforiboateng@pdx.edu
[2] University of Texas at Dallas, Richardson TX 75080, USA
[3] University of Alberta, Edmonton, Alberta, T6G 2R3, Canada
[4] National Science Foundation, Alexandria, VA 22314, USA

Appendix A

In this section we provide proofs for Lemma 1 and Theorem 1.

Lemma 1

Proof. Following the path of Lemma 10.1.c in Lin and Bai [2] and the condition **C**.2, we have

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{T} \boldsymbol{\mathcal{Z}}_{(k)_{ij}} \boldsymbol{t}_i\right)^2 &= \sum_{i=1}^{T} \boldsymbol{\mathcal{Z}}_{(k)_{ij}} \mathbb{E}(\boldsymbol{t}_i^2) + 2 \sum_{1 \le i \le w \le T} \boldsymbol{\mathcal{Z}}_{(k)_{ij}} \boldsymbol{\mathcal{Z}}_{(k)_{wj}} \mathbb{E}(\boldsymbol{t}_i \boldsymbol{t}_w) \\
&\le R^2 \left(\sum_{i=1}^{T} \mathbb{E}(\boldsymbol{t}_i^2) + 2 \sum_{1 \le i \le w \le T} \rho(w-i) \mathbb{E}(\boldsymbol{t}_i^2)^{1/2} \mathbb{E}(\boldsymbol{t}_w^2)^{1/2} \right) \\
&\le R^2 \left(\sum_{i=1}^{T} \mathbb{E}(\boldsymbol{t}_i^2) + \sum_{i=1}^{T} \sum_{s=1}^{T} \rho(s) \left[\mathbb{E}(\boldsymbol{t}_i^2) + \mathbb{E}(\boldsymbol{t}_s^2) \right] \right) \\
&\le TR^2 \left(1 + 2 \sum_{s=1}^{T} \rho(s) \right)
\end{aligned}$$

and by inducing the condition **C**.1, we have

⋆ Dorcas Ofori-Boateng and Jaidev Goel are co-first authors with equal contribution and importance.

$$\mathbb{E}\left(\sum_{i=1}^{T}\boldsymbol{\mathcal{Z}}_{(k)_{ij}}\boldsymbol{t}_i\right)^2 \leq TR^2\left(1+2\sum_{s=1}^{T}\rho(s)\right)$$
$$= TR^2\left(1+\frac{2a_1}{1-e^{-a_2}}\right) = TR^2C_1^2/8,$$

where C_1 term contains the mixing coefficient ρ.

Utilizing results of [3], let $P_k = \frac{\boldsymbol{\mathcal{Z}}_{(k)}{}^T\boldsymbol{t}}{\sqrt{\mathbb{E}\left(\boldsymbol{\mathcal{Z}}_{(k)}{}^T\boldsymbol{t}\right)^2}}$, which follows the standard gaussian distribution.Therefore,

$$\boldsymbol{P}(\mathscr{A}^c) = \boldsymbol{P}\left(\max_{1\leq k\leq p}\left\{\left|\boldsymbol{\mathcal{Z}}_{(k)}{}^T\boldsymbol{t}\right| \geq \frac{T\lambda_k}{2}\right\}\right)$$
$$\leq \sum_{k=1}^{p}\boldsymbol{P}\left(\left|\frac{\boldsymbol{\mathcal{Z}}_{(k)}{}^T\boldsymbol{t}}{\sqrt{\mathbb{E}\left(\sum_{i=1}^{T}\boldsymbol{\mathcal{Z}}_{(k)_{ij}}\boldsymbol{t}_i\right)^2}}\right| \geq \frac{T\lambda_k}{2\sqrt{TR^2C_1^2/8}}\right)$$
$$= 2p\cdot\boldsymbol{P}\left(P_k \geq \frac{\sqrt{2T}\lambda_k}{RC_1}\right)$$

Therefore,

$$\boldsymbol{P}(\mathscr{A}^c) \leq p\cdot\ \exp\left(-\frac{T\lambda_k^2}{R^2C_1^2}\right),$$

and with $\lambda_k = A_0C_1R\sqrt{\frac{v_k log\, p}{T}}$ for all k, we will have that $\boldsymbol{P}(\mathscr{A}) \leq p^{1-A_0^2}$.

Theorem 1

For Theorem (1), we note the following:

$$
\begin{aligned}
&||L_T(\hat{\boldsymbol{\theta}}, \boldsymbol{\mathcal{Z}})||_2^2 + \sum_{k=1}^{K} P_{\lambda_k}(\hat{\boldsymbol{\theta}}_k) \leq ||L_T(\boldsymbol{\theta}, \boldsymbol{\mathcal{Z}})||_2^2 + \sum_{k=1}^{K} P_{\lambda_k}(\boldsymbol{\theta_k}) \\
&||L_T(\hat{\boldsymbol{\theta}}, \boldsymbol{\mathcal{Z}})||_2^2 - ||L_T(\boldsymbol{\theta}, \boldsymbol{\mathcal{Z}})||_2^2 \leq \sum_{k=1}^{K} P_{\lambda_k}(\boldsymbol{\theta_k}) - \sum_{k=1}^{K} P_{\lambda_k}(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}) \\
&||L_T(\hat{\boldsymbol{\theta}}, \boldsymbol{\mathcal{Z}}) - L_T(\boldsymbol{\theta}, \boldsymbol{\mathcal{Z}})||_2^2 \leq \sum_{k=1}^{K} P_{\lambda_k}(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \boldsymbol{\theta_k}) \\
&||L_T(\hat{\boldsymbol{\theta}}, \boldsymbol{\mathcal{Z}}) - L_T(\boldsymbol{\theta}, \boldsymbol{\mathcal{Z}})||_2^2 \leq \sum_{k=1}^{K} P_{\lambda_k}(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \boldsymbol{\theta_k}) \\
&= \sum_{k=1}^{K} \lambda_k ||\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \boldsymbol{\theta_k}||_{1,off}.
\end{aligned}
$$

Furthermore, by applying Karush-Kuhn-Tucker (KKT) conditions and Bernstein inequality for ρ mixing [1], for any constant $C^*(\boldsymbol{\theta}, \kappa) > 0$,

$$\sum_{k=1}^{K} \lambda_k ||\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \boldsymbol{\theta_k}||_{1,off} \leq C^*(\boldsymbol{\theta}, \kappa) K \max_k q_k \lambda_k^2, \tag{1}$$

which means that

$$||L_T(\hat{\boldsymbol{\theta}}, \boldsymbol{\mathcal{Z}}) - L_T(\boldsymbol{\theta}, \boldsymbol{\mathcal{Z}})||_2 \leq C^*(\boldsymbol{\theta}, \kappa)\sqrt{K} \max_k \sqrt{q_k} \lambda_k \tag{2}$$

$$= C^*(\boldsymbol{\theta}, \kappa)\sqrt{K} \max_k \sqrt{q_k} \left(A_0 C_1 R \sqrt{\frac{v_k log\, p}{T}} \right) \tag{3}$$

and with Lemma (1), the above inequality will hold with probability no less than $1 - p^{1-A_0^2}$.

Again, from Lemma (1), we see that

$$
\begin{aligned}
||\hat{\boldsymbol{\theta}}_{\mathscr{A}_k} - \boldsymbol{\theta}_{\mathscr{A}_k}||_1 &\leq C^*(\boldsymbol{\theta}, \kappa) ||\hat{\boldsymbol{\theta}}_{\mathscr{A}_k} - \boldsymbol{\theta}_{\mathscr{A}_k}||_1 \\
&\leq C^*(\boldsymbol{\theta}, \kappa)\sqrt{q_k} ||\hat{\boldsymbol{\theta}}_{\mathscr{A}_k} - \boldsymbol{\theta}_{\mathscr{A}_k}||_2.
\end{aligned} \tag{4}
$$

Armed with the above inequality and that of the inequality of (3), any solution $\hat{\boldsymbol{\theta}}_{\mathcal{A}}$ is within

$$\left\{ ||\hat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{\mathcal{A}}||_2 \leq C^*(\boldsymbol{\theta}, \kappa)\sqrt{K} \max_k \sqrt{q_k} \left(A_0 C_1 R \sqrt{\frac{v_k log\, p}{T}} \right) \right\}$$

with probability no less than $1 - p^{1-A_0^2}$.

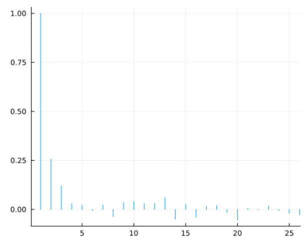
B Simulations

Simulated structures

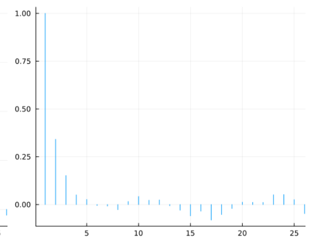
1. **AR(1) with γ coefficient** : Covariance matrix of the form $|\gamma_{i,j}^{i-j}|$
2. **Star Block (SB)**: A block-structured covariance matrix with equal dimension blocks whose inverses correspond to star-structured graphs.
3. **Uniformly weighted (UW)**: Weighted counterpart of the Erdos-Renyi random graph where the weights are generated as per the Uniform(0,1).

Dependence Check

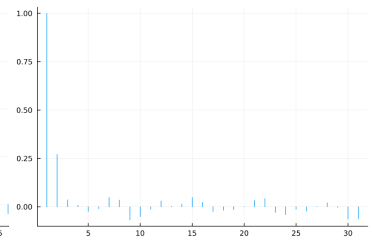
We use Autocorrelation function plots in Fig 1 to do a quality check of the simulated data. The plots suggest that the synthetic data has a fast decaying correlation as the lag between covariates increases which shows our data closely follows a weakly dependent mixing process.



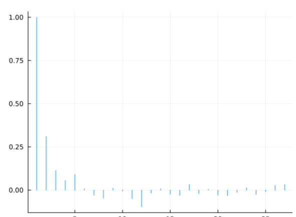
(a) KS,AR(1)



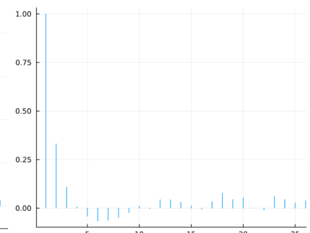
(b) KS , SB



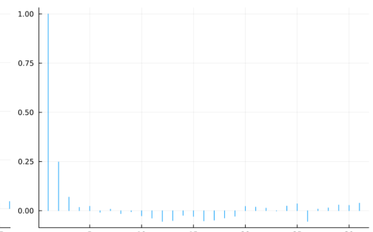
(c) KS, Uniform



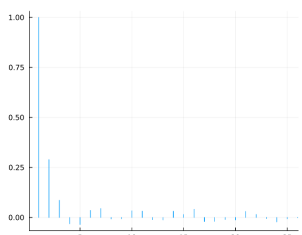
(d) KP, AR(1)



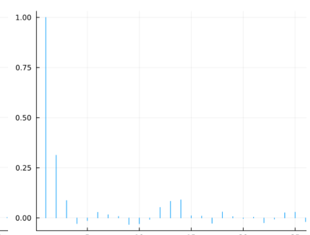
(e) KP, SB



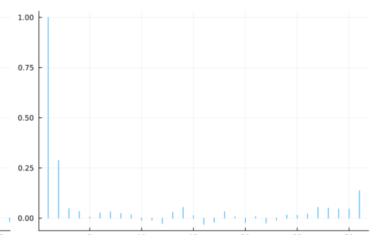
(f) KP, Uniform



(g) Sylvester, AR(1)



(h) Sylvester, SB



(i) Sylvester, Uniform

Fig. 1: Autocorrelation function plots to check the dependence structure of the simulated **AR(1)**, **SB** and **UW** graph data

C Code and Reproducibility

The code for the data is available at `https://drive.google.com/drive/folders/1C_qYTZXbZ2QgZ5eC8GuqJv0wvXgrhvK8?usp=sharing`. All the simulations were done on a system with Linux OS, 32GB RAM, 13th Gen Intel(R) Core(TM) i9-13900HX, and RTX 4080 GPU.

Bibliography

[1] Doukhan, P., Neumann, M.H.: Probability and moment inequalities for sums of weakly dependent random variables, with applications. Stochastic Processes and their Applications **117**(7), 878–903 (2007)

[2] Lin, Z.: Probability inequalities. Springer (2010)

[3] Xie, F., Xiao, Z.: Square-root lasso for high-dimensional sparse linear systems with weakly dependent errors. Journal of Time Series Analysis **39**(2), 212–238 (2018)