Thought about SupCP

Jiaxin Hu

May 28, 2021

1 SupCP covariance

Consider the observation $\mathcal{Y} \in \mathbb{R}^{d\times d\times d}$, the covariance $\boldsymbol{X} \in \mathbb{R}^{d\times R}$. Recall the SupCP model,

$$\mathcal{Y} = [\![\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3]\!] + \mathcal{E}, \quad \boldsymbol{A}_1 = \boldsymbol{X}\boldsymbol{B} + \mathcal{E}',$$

where $\boldsymbol{A}_k \in \mathbb{R}^{d\times R}$, $\boldsymbol{B} \in \mathbb{R}^{p\times R}$ is the coefficient matrix, $\mathcal{E} \in \mathbb{R}^{d\times d\times d}$ has $i.i.d.$ entries from $N(0, \sigma_e^2)$, and $\mathcal{E}' \in \mathbb{R}^{d\times R}$ has $i.i.d.$ rows from $\mathcal{N}(0, \Sigma)$.

Note that

$$\text{vec}(\mathcal{Y}) = [\boldsymbol{X}\boldsymbol{B} \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathbf{1_R} + [\mathcal{E}' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathbf{1_R} + \text{vec}(\mathcal{E}'),$$

where $\odot$ is the column-wise Kronecker product. Since $\mathcal{E}'$ is independent with $\mathcal{E}$ and $\text{cov}(\text{vec}(\mathcal{E})) = \boldsymbol{I}_{d^3}$, we only need to calculate $\text{cov}([\mathcal{E}' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathbf{1_R})$. Note that

$$[\mathcal{E}' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathbf{1_R} = \begin{bmatrix} (\mathcal{E}_1' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3)\mathbf{1_R} \\ \cdots \\ (\mathcal{E}_d' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3)\mathbf{1_R} \end{bmatrix}, \quad \text{and} \quad \mathcal{E}_i' \perp \mathcal{E}_j', i \neq j \in [d],$$

where $\mathcal{E}_i' \in \mathbb{R}^{1\times R}$ refers to the i-th row of $\mathcal{E}'$. Therefore, we know that $\text{cov}([\mathcal{E}' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathbf{1_R})$ is block-wise diagonal with diagonal elements $\text{cov}((\mathcal{E}_i' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3)\mathbf{1_R}), i \in [d]$. Also notice that

$$(\mathcal{E}_i' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3)\mathbf{1_R} = \sum_{k=1}^{R} \mathcal{E}_{ik}' \otimes \boldsymbol{A}_{2k} \otimes \boldsymbol{A}_{3k} = [\boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathcal{E}_i'^T.$$

Therefore, we have

$$\text{cov}([\mathcal{E}_i' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathbf{1_R}) = \text{cov}([\boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathcal{E}_i'^T) = [\boldsymbol{A}_2 \odot \boldsymbol{A}_3]\Sigma[\boldsymbol{A}_2 \odot \boldsymbol{A}_3]^T,$$

and thus the whole covariance matrix $\text{cov}(\text{vec}(\mathcal{Y}))$ is

$$\begin{aligned} \text{cov}(\text{vec}(\mathcal{Y})) &= \text{cov}(\text{vec}(\mathcal{E})) + \text{cov}([\mathcal{E}' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathbf{1_R}) \\ &= \boldsymbol{I}_{d^3} + \begin{bmatrix} [\boldsymbol{A}_2 \odot \boldsymbol{A}_3]\Sigma[\boldsymbol{A}_2 \odot \boldsymbol{A}_3]^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & [\boldsymbol{A}_2 \odot \boldsymbol{A}_3]\Sigma[\boldsymbol{A}_2 \odot \boldsymbol{A}_3]^T \end{bmatrix} \end{aligned}$$

Claim: The second part of the covariance can not be written as normal tensor ensemble $\mathcal{N}(0, \Sigma_1, \Sigma_2, \Sigma_3)$, where $\Sigma_1 = \boldsymbol{I}_d, \Sigma_2 = \boldsymbol{A}_2\boldsymbol{A}_2^T, \Sigma_3 = \boldsymbol{A}_3\boldsymbol{A}_3^T$.

Proof of the Claim: We prove the claim from two aspects.

Technical Reason: If second part $[\![\mathcal{E}', \boldsymbol{A}_2, \boldsymbol{A}_3]\!]$ can be formulated in a normal tensor ensemble, then the covariance $\text{cov}([\mathcal{E}' \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathbf{1_R})$ can be written in the form $\Sigma_1 \otimes \Sigma_2 \otimes \Sigma_3$. Obviously, we have $\Sigma_1 = \boldsymbol{I}_d$. Then, we only need to verify $(\boldsymbol{A}_2 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_3)^T = (\boldsymbol{A}_2\boldsymbol{A}_2^T) \otimes (\boldsymbol{A}_3\boldsymbol{A}_3^T)$. However, it is not true.

Let $\tilde{\boldsymbol{A}}_{2i}, \tilde{\boldsymbol{A}}_{3i}$ denote the rows of the matrix $\boldsymbol{A}_2, \boldsymbol{A}_3$. Then we have

$$\boldsymbol{A}_2 \odot \boldsymbol{A}_3 = \begin{bmatrix} \tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{31} \\ \vdots \\ \tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{3d} \\ \tilde{\boldsymbol{A}}_{22} * \tilde{\boldsymbol{A}}_{31} \\ \vdots \\ \tilde{\boldsymbol{A}}_{2d} * \tilde{\boldsymbol{A}}_{3d} \end{bmatrix}_{d^2 \times R}$$

where $*$ denote the Hadamard product, and

$$\begin{aligned}(\boldsymbol{A}_2 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_3)^T &= \begin{bmatrix} [\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{31}][\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{31}]^T & \cdots & [\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{31}][\tilde{\boldsymbol{A}}_{2d} * \tilde{\boldsymbol{A}}_{3d}]^T \\ [\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{32}][\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{31}]^T & \cdots & [\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{32}][\tilde{\boldsymbol{A}}_{2d} * \tilde{\boldsymbol{A}}_{3d}]^T \\ \vdots & \vdots & \vdots \\ [\tilde{\boldsymbol{A}}_{2d} * \tilde{\boldsymbol{A}}_{3d}][\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{31}]^T & \cdots & [\tilde{\boldsymbol{A}}_{2d} * \tilde{\boldsymbol{A}}_{3d}][\tilde{\boldsymbol{A}}_{2d} * \tilde{\boldsymbol{A}}_{3d}]^T \end{bmatrix} \\ &= \begin{bmatrix} [\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{21}][\tilde{\boldsymbol{A}}_{31} * \tilde{\boldsymbol{A}}_{31}]^T & \cdots & [\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{2d}][\tilde{\boldsymbol{A}}_{31} * \tilde{\boldsymbol{A}}_{3d}]^T \\ [\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{21}][\tilde{\boldsymbol{A}}_{32} * \tilde{\boldsymbol{A}}_{31}]^T & \cdots & [\tilde{\boldsymbol{A}}_{21} * \tilde{\boldsymbol{A}}_{2d}][\tilde{\boldsymbol{A}}_{32} * \tilde{\boldsymbol{A}}_{3d}]^T \\ \vdots & \vdots & \vdots \\ [\tilde{\boldsymbol{A}}_{2d} * \tilde{\boldsymbol{A}}_{21}][\tilde{\boldsymbol{A}}_{3d} * \tilde{\boldsymbol{A}}_{31}]^T & \cdots & [\tilde{\boldsymbol{A}}_{2d} * \tilde{\boldsymbol{A}}_{2d}][\tilde{\boldsymbol{A}}_{3d} * \tilde{\boldsymbol{A}}_{3d}]^T \end{bmatrix},\end{aligned}$$

where the second equation follows $(a * b)(c * d)^T = \sum_i a_i b_i c_i d_i = (a * c)(b * d)^T$ for row vectors a, b, c, d. On the other hand, we have

$$\boldsymbol{A}_2\boldsymbol{A}_2^T = \begin{bmatrix} \tilde{\boldsymbol{A}}_{21}\tilde{\boldsymbol{A}}_{21}^T & \cdots & \tilde{\boldsymbol{A}}_{21}\tilde{\boldsymbol{A}}_{2d}^T \\ \tilde{\boldsymbol{A}}_{22}\tilde{\boldsymbol{A}}_{21}^T & \cdots & \tilde{\boldsymbol{A}}_{22}\tilde{\boldsymbol{A}}_{2d}^T \\ \vdots & \vdots & \vdots \\ \tilde{\boldsymbol{A}}_{2d}\tilde{\boldsymbol{A}}_{21}^T & \cdots & \tilde{\boldsymbol{A}}_{2d}\tilde{\boldsymbol{A}}_{2d}^T \end{bmatrix},$$

and similar to $\boldsymbol{A}_3\boldsymbol{A}_3^T$. Therefore, consider constants r, s, v, w, we have

$$\left[\boldsymbol{A}_2\boldsymbol{A}_2^T \otimes \boldsymbol{A}_3\boldsymbol{A}_3^T\right]_{d(r-1)+v, d(s-1)+w} = \tilde{\boldsymbol{A}}_{2r}\tilde{\boldsymbol{A}}_{2s}^T\tilde{\boldsymbol{A}}_{3v}\tilde{\boldsymbol{A}}_{3w}^T,$$

and

$$\left[(\boldsymbol{A}_2 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_3)^T\right]_{d(r-1)+v, d(s-1)+w} = (\tilde{\boldsymbol{A}}_{2r} * \tilde{\boldsymbol{A}}_{2s})(\tilde{\boldsymbol{A}}_{3v} * \tilde{\boldsymbol{A}}_{3w})^T.$$

Note that $\tilde{\boldsymbol{A}}_{2r}\tilde{\boldsymbol{A}}_{2s}^T\tilde{\boldsymbol{A}}_{3v}\tilde{\boldsymbol{A}}_{3w}^T$ is the summation of R^2 terms while $(\tilde{\boldsymbol{A}}_{2r} * \tilde{\boldsymbol{A}}_{2s})(\tilde{\boldsymbol{A}}_{3v} * \tilde{\boldsymbol{A}}_{3w})^T$ is the summation of R terms which are included in the R^2 terms. Therefore, we claim that $(\boldsymbol{A}_2 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_3)^T \neq (\boldsymbol{A}_2\boldsymbol{A}_2^T) \otimes (\boldsymbol{A}_3\boldsymbol{A}_3^T)$.

Model reason: Any tensor $\mathcal{Y}$ in the class of normal tensor $\mathcal{N}(0, \Sigma_1, \Sigma_2, \Sigma_3)$ should be written as

$$\mathcal{Y} = \mathcal{Z} \times_1 \Sigma_1^{1/2} \times_2 \Sigma_2^{1/2} \times_3 \Sigma_3^{1/2},$$

where $\mathcal{Z}$ has i.i.d. standard normal entries, and the kronecker structure comes from the vectorization

$$\text{Cov}(\text{vec}(\mathcal{Y})) = \text{Cov}\left[\Sigma_1^{1/2} \otimes \Sigma_2^{1/2} \otimes \Sigma_3^{1/2}\text{vec}(\mathcal{Z})\right].$$

Since $\text{Cov}(\text{vec}(\mathcal{Z})) = \boldsymbol{I}_{R^3}$, the covariance of vectorized $\mathcal{Y}$ is

$$\text{Cov}(\text{vec}(\mathcal{Y})) = \left[\Sigma_1^{1/2} \otimes \Sigma_2^{1/2} \otimes \Sigma_3^{1/2}\right]\left[\Sigma_1^{1/2} \otimes \Sigma_2^{1/2} \otimes \Sigma_3^{1/2}\right]^T = [\Sigma_1 \otimes \Sigma_2 \otimes \Sigma_3],$$

where the last equation follows the mixed-product property of Kronecker product.

Back to our case, the tensor $[\![\mathcal{E}', \boldsymbol{A}_2, \boldsymbol{A}_3]\!]$ can be re-written in a tucker product form

$$\mathcal{Y}' = [\![\mathcal{E}', \boldsymbol{A}_2, \boldsymbol{A}_3]\!] = \mathcal{D} \times_1 \mathcal{E}' \times_2 \boldsymbol{A}_2 \times \boldsymbol{A}_3,$$

where $\mathcal{D}$ is a super-diagonal tensor with only $\mathcal{D}_{iii} = 1$ and others remain 0. Therefore, $\mathcal{D} \times_1 \mathcal{E}'$ is not a tensor with i.i.d. standard normal entries since most of the entries remain 0. Therefore, we have

$$\text{Cov}(\text{vec}(\mathcal{Y}')) = \text{Cov}\left[\boldsymbol{I}_d \otimes \boldsymbol{A}_2 \otimes \boldsymbol{A}_3\text{vec}(\mathcal{D} \times_1 \mathcal{E}')\right] = [\boldsymbol{I}_d \otimes \boldsymbol{A}_2 \otimes \boldsymbol{A}_3]\, \boldsymbol{H}\, [\boldsymbol{I}_d \otimes \boldsymbol{A}_2 \otimes \boldsymbol{A}_3]^T,$$

where $\boldsymbol{H}$ is a diagonal matrix indicates the non-zero entries in $\text{Cov}(\text{vec}(\mathcal{D} \times_1 \mathcal{E}'))$. Due to the selection matrix $\boldsymbol{H}$, we can not represent tensor $\mathcal{Y}'$ in the normal tensor ensemble.

$\square$

References