matrix case:
Y = PQ:
P deterministic (orthonornomal) matrix ,
Q i.i.d noise entry.

Y=[S1, S2]
S1: row covariance —> P*P^T
S2: column covariance —> I

Thought about SupCP

Jiaxin Hu

May 21, 2021

Y=[D; noise, A2, A3];

T = [D; I, A2, A3] — deteriministic (r-by-d2-by-d3)
A = T\times_1 noise factor
—> A[1,,]=linear combination of matrix slices
A[2,,.]-linear

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,

no: X,
E: i.i.d;
A1 = (i.i.d noise)

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

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 $\mathcal{N}(0, \Sigma)$.

Cov(Y) = [S1, S2, S3]
where S1=I_{d-by-d}
S2 = A2*A2^T; S3= A3* A3^T

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

$$(\mathcal{E}'_i \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3)\mathbf{1_R} = \sum_{k=1}^{R} \mathcal{E}'_{ik} \otimes \boldsymbol{A}_{2k} \otimes \boldsymbol{A}_{3k} = [\boldsymbol{A}_2 \odot \boldsymbol{A}_3]\mathcal{E}'^T_i.$$

Therefore, we have

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

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

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

assume = I

2 SupCP performance

Write the second part in terms of normal tensor ensemble.
Y = MVN(0, Sigma_1, Sigma_2, Sigma_3), where Sigma_k is mode-k covariance.

In this note, we consider the Gaussian data and all the matrices are full rank. Let $\mathcal{Y} \in \mathbb{R}^{d\times d\times d}$, $\boldsymbol{X}_k \in \mathbb{R}^{d\times p}, k \in [3]$. Consider the tucker rank $\boldsymbol{r} = (r, r, r)$ and CP rank R. The dimension of $\boldsymbol{M}_k$ can be obtained by the context.

What is Sigma_k, i=1,2,3 in this case? Dimension? Intuition?

model 3: [B1X1, B2X2; B3X3]

vs. model 1: [A1, A2, A3].
provide {A1,A2,A3}~{in distribution same} [B1(X1^TX1)^{-1}X1, ….].

A1: d-by-r matrix, orthonormal column emsemble
B1X1: d-by-r matrix, orthonormal column emsemble (provided p>r)

2.1 Without supervision

Recall the STD and SupCP models without the supervision.

$$\begin{aligned} STD &: \quad \mathcal{Y} = \mathcal{C} \times \{\boldsymbol{M}_1, \boldsymbol{M}_2, \boldsymbol{M}_3\} + \mathcal{E} \\ SupCP &: \quad \mathcal{Y} = [\![\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3]\!] + \mathcal{E}. \end{aligned}$$

The fitted value vec($\mathcal{Y}$) for STD and SupCP lie in

$$\begin{aligned} \mathcal{P}_{STD} &= \{C(\boldsymbol{M}_1 \otimes \boldsymbol{M}_2 \otimes \boldsymbol{M}_3) \mid \boldsymbol{M}_k \in \mathbb{R}^{d\times r}, \boldsymbol{M}_k^T \boldsymbol{M}_k = \boldsymbol{I}_r\}, \\ \mathcal{P}_{SupCP} &= \{C(\boldsymbol{A}_1 \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3) \mid \boldsymbol{A}_k \in \mathbb{R}^{d\times R}\}, \end{aligned}$$

respectively, where $C(\boldsymbol{X})$ refers to the column space of the matrix $\boldsymbol{X}$. Note that rank$(\boldsymbol{M}_1 \otimes \boldsymbol{M}_2 \otimes \boldsymbol{M}_3) = r^3$ and rank$(\boldsymbol{A}_1 \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_3) = R$.

1. If the true signal is generated from STD, the fitted value for STD model is

$$\text{vec}(\hat{\mathcal{Y}}_{STD}) \in C(\hat{\boldsymbol{M}}_1 \otimes \hat{\boldsymbol{M}}_2 \otimes \hat{\boldsymbol{M}}_3), \quad \text{for some} \quad \hat{\boldsymbol{M}}_k^T \hat{\boldsymbol{M}}_k = \boldsymbol{I}_r.$$

   If $R \leq r^3$, the space $\mathcal{P}_{SupCP}$ may not cover the best estimation from the true model, vec$(\hat{\mathcal{Y}}_{STD})$. Because vec$(\hat{\mathcal{Y}}_{STD})$ is a combination of r^3 bases of $\mathbb{R}^3$, and the $\hat{\mathcal{Y}}_{SupCP} \in \mathcal{P}_{SupCP}$ is a combination of R bases of $\mathbb{R}^3$.

   If $R > r^3$, we can expect the space $\mathcal{P}_{SupCP}$ may cover the best estimation vec$(\hat{\mathcal{Y}}_{STD})$.

   See the following figures for numerical results.

References

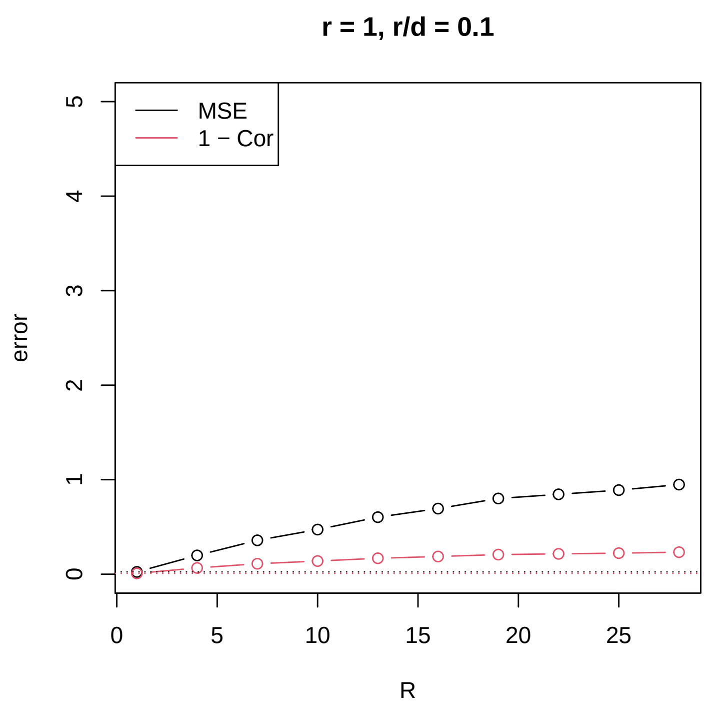

Figure 1: Error of CP model when data is generated from Tucker model. The dashed lines are errors of the STD estimation. Here we consider $\boldsymbol{r} = (1, 1, 1)$ with $d = 10$.

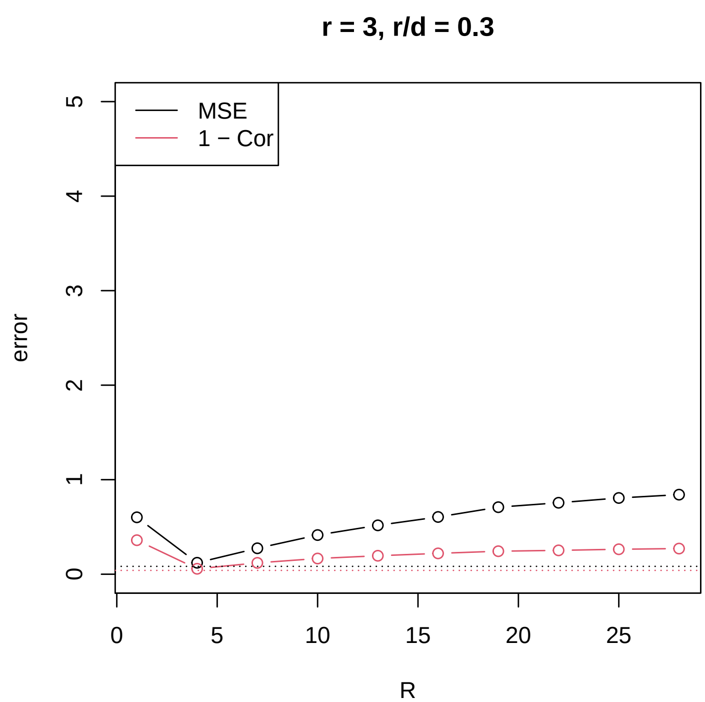

Figure 2: Error of CP model when data is generated from Tucker model. The dashed lines are errors of the STD estimation. Here we consider $\boldsymbol{r} = (3, 3, 3)$ with $d = 10$.

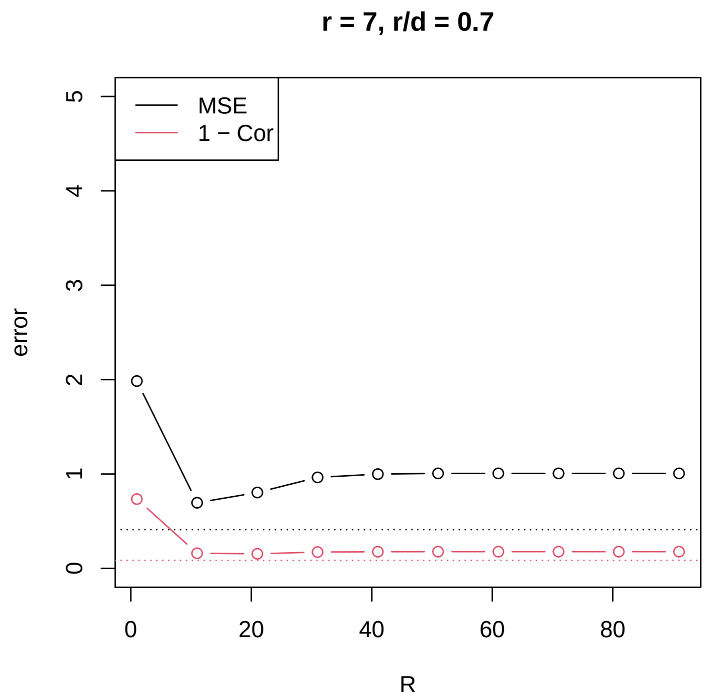

Figure 3: Error of CP model when data is generated from Tucker model. The dashed lines are errors of the STD estimation. Here we consider $\boldsymbol{r} = (7, 7, 7)$ with $d = 10$.

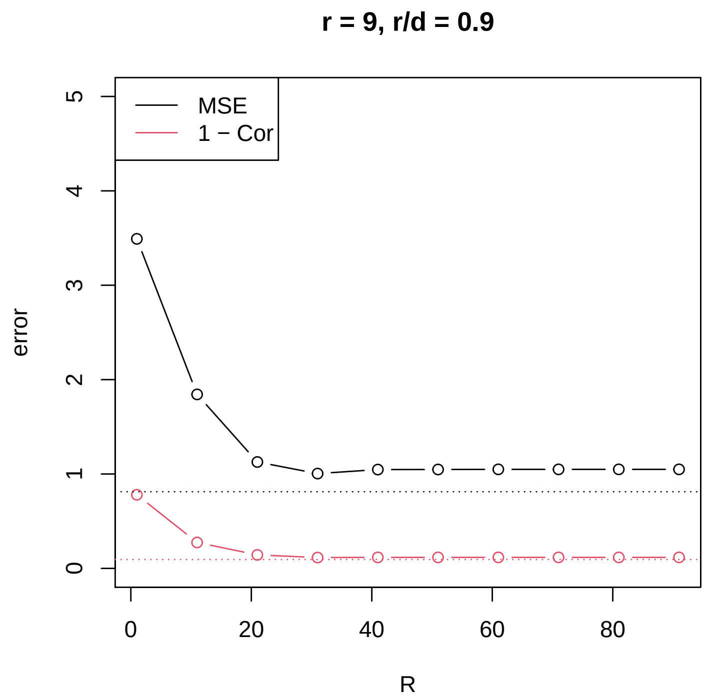

Figure 4: Error of CP model when data is generated from Tucker model. The dashed lines are errors of the STD estimation. Here we consider $\boldsymbol{r} = (9, 9, 9)$ with $d = 10$.