Jive plus GMD

Ilias Moysidis and Jing Ma

Fred Hutchinson Cancer Research Center

Abstract

1 Methodology

1.1 General Matrix Decomposition (GMD)

In this section we will describe the GMD problem for a matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$. Let $\mathbf{Q} = \mathbf{\Delta}^{-1}$ and $\mathbf{R} = \mathbf{\Sigma}^{-1}$, where $\mathbf{\Delta} \in \mathbb{R}^{n \times n}, \mathbf{\Sigma} \in \mathbb{R}^{p \times p}$ and $\mathbf{\Delta}, \mathbf{\Sigma} \succ \mathbf{0}$. We define the matrix norm $\|\mathbf{X}\|_{\mathbf{Q},\mathbf{R}} = \sqrt{\operatorname{tr}(\mathbf{QXRX}^\intercal)}$ which we call $\mathbf{Q}, \mathbf{R}$-norm. The GMD optimization problem is the best rank-s approximation to the data with respect to the $\mathbf{Q}, \mathbf{R}$-norm:

$$\begin{aligned}
&\text{minimize } \|\mathbf{X} - \mathbf{UDV}^\intercal\|^2_{\mathbf{Q},\mathbf{R}}, \\
&\text{subject to } \mathbf{U}^\intercal\mathbf{QU} = \mathbf{I}_s,\ \mathbf{V}^\intercal\mathbf{RV} = \mathbf{I}_s,\ \mathbf{D} \text{ is diagonal and } \mathbf{D} \succ \mathbf{0}.
\end{aligned}$$

Let $\|\cdot\|_F$ denote the Frobenius norm. The aforementioned problem can be recast as a classic singular value decomposition (SVD) optimization problem, since it is equivalent to

$$\text{minimize } \|\mathbf{Q}^{1/2}\mathbf{XR}^{1/2} - (\mathbf{Q}^{1/2}\mathbf{U})\mathbf{D}(\mathbf{R}^{1/2}\mathbf{V})^\intercal\|^2_F,$$

subject to $(\mathbf{Q}^{1/2}\mathbf{U})^\intercal(\mathbf{Q}^{1/2}\mathbf{U}) = \mathbf{I}_s$, $(\mathbf{R}^{1/2}\mathbf{V})^\intercal(\mathbf{R}^{1/2}\mathbf{V}) = \mathbf{I}_s$, $\mathbf{D}$ is diagonal and $\mathbf{D} \succ \mathbf{0}$.

Therefore, the GMD problem for $\mathbf{X}$ is equivalent to the SVD problem for $\mathbf{Q}^{1/2}\mathbf{X}\mathbf{R}^{1/2}$. If $\mathbf{U}^*, \mathbf{V}^*, \mathbf{D}^*$ is the solution of the SVD for $\mathbf{Q}^{1/2}\mathbf{X}\mathbf{R}^{1/2}$, then the solution of the equivalent GMD is $\mathbf{U} = \mathbf{Q}^{-1/2}\mathbf{U}^*, \mathbf{V} = \mathbf{R}^{-1/2}\mathbf{V}^*, \mathbf{D} = \mathbf{D}^*$.

1.2 Model

Let $\mathbf{X}_k \in \mathbb{R}^{n \times p_k}$ be the data matrix of dataset k, for $k = 1, \ldots, K$ and define $p = \sum_{k=1}^K p_k$. Without loss of generality we assume that these data matrices are normalized, i.e.

$$\sum_{i=1}^n X_{k,ij} = 0 \text{ and } \|\mathbf{X}_k\|_{\mathbf{Q},\mathbf{R}_k} = 1$$

for all k, j.

Let $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ be the covariance matrix between the samples and $\mathbf{\Delta}_k \in \mathbb{R}^{p_k \times p_k}$ be the covariance between the variables of the k-th category. Define the precision matrices $\mathbf{Q} = \mathbf{\Sigma}^{-1}$, $\mathbf{R}_k = \mathbf{\Delta}_k$ and the block diagonal matrices

$$\mathbf{\Delta} = \text{block.diag}(\mathbf{\Delta}_1, \ldots, \mathbf{\Delta}_K) \quad \text{and} \quad \mathbf{R} = \text{block.diag}(\mathbf{R}_1, \ldots, \mathbf{R}_K).$$

We begin by defining the joint structure matrix. Let $\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{V}^\intercal$, where:

1. $\mathbf{D} \in \mathbb{R}^{r \times r}$ is a positive definite, diagonal matrix,

2. $\mathbf{U} \in \mathbb{R}^{n \times r}$ satisfies $\mathbf{U}^\intercal\mathbf{Q}\mathbf{U} = \mathbf{I}_r$, and

3. $\mathbf{V} \in \mathbb{R}^{p \times r}$ satisfies $\mathbf{V}^\intercal\mathbf{R}\mathbf{V} = \mathbf{I}_r$.

If we write the right-singular vector matrix of the joint structure as

$$\mathbf{V} = (\mathbf{V}_1^\intercal \ldots \mathbf{V}_K^\intercal)^\intercal, \quad \mathbf{V}_k \in \mathbb{R}^{p_k \times r}$$

we can write the joint structure in the form $\mathbf{J} = (\mathbf{J}_1 \dots \mathbf{J}_K)$, where $\mathbf{J}_k = \mathbf{U}\mathbf{D}\mathbf{V}_k^\intercal$.

Let $\mathbf{A}_k = \mathbf{U}_k\mathbf{D}_k\mathbf{W}_k^\intercal$ be the individual structure matrix, where:

1. $\mathbf{D}_k \in \mathbb{R}^{r_k \times r_k}$ is a positive definite, diagonal matrix,
2. $\mathbf{U}_k \in \mathbb{R}^{n \times r_k}$ satisfies $\mathbf{U}_k^\intercal\mathbf{Q}\mathbf{U}_k = \mathbf{I}_{r_k}$, and
3. $\mathbf{W}_k \in \mathbb{R}^{p_k \times r}$ satisfies $\mathbf{W}_k^\intercal\mathbf{R}_k\mathbf{W}_k = \mathbf{I}_{r_k}$.

We define the concatenated individual structure matrix $\mathbf{A} = (\mathbf{A}_1 \dots \mathbf{A}_K)$.

Finally, we define the matrix-normal noise matrices $\mathbf{E}_k \sim \mathcal{MN}_{n\times p_k}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{\Delta}_k)$ such that $\mathbf{E}_i \perp\!\!\!\perp \mathbf{E}_j$ for $i \neq j$. If $\mathbf{E} = (\mathbf{E}_1 \dots \mathbf{E}_K)$ is the concatenated noise matrix, then $\mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{\Delta})$.

The basic assumption of this paper is that the observed data follow the model

$$\mathbf{X}_k = \mathbf{J}_k + \mathbf{A}_k + \mathbf{E}_k, \quad k = 1, \dots, K.$$

For identifiability reasons, we demand that the columns of joint and individual structure are orthogonal, i.e.

$$\mathbf{J}^\intercal\mathbf{Q}\mathbf{A} = \mathbf{0}.$$

This assumption is not restrictive since for every pair of matrices $\mathbf{J}, \mathbf{A}$ there exist unique $\tilde{\mathbf{J}}, \tilde{\mathbf{A}}$ such that

$$\mathbf{J} + \mathbf{A} = \tilde{\mathbf{J}} + \tilde{\mathbf{A}} \quad \text{and} \quad \tilde{\mathbf{J}}^\intercal\mathbf{Q}\tilde{\mathbf{A}} = \mathbf{0}.$$

2 Algorithms for estimating J and A

2.1 GJIVE

Suppose that we know the matrices $\mathbf{X}, \mathbf{A}$. Then, $\mathbf{J}$ is the best rank-r approximation of $\mathbf{X} - \mathbf{A}$. Similarly, suppose that we know the matrices $\mathbf{X}_k, \mathbf{J}_k$. Then, $\mathbf{A}_k$ is the best rank-r_k approximation of $\mathbf{X}_k - \mathbf{J}_k$.

To estimate the matrices $\mathbf{J}, \mathbf{A}$ we use the following algorithm:

1. Set $\mathbf{A} = \mathbf{0}$

2. Solve

$$\begin{aligned}&\text{minimize } \|\mathbf{X} - \mathbf{A} - \mathbf{U}\mathbf{D}\mathbf{V}^\intercal\|_{\mathbf{Q},\mathbf{R}}\\&\text{subject to } \mathbf{U}^\intercal\mathbf{Q}\mathbf{U} = \mathbf{I}_r, \quad \mathbf{V}^\intercal\mathbf{R}\mathbf{V} = \mathbf{I}_r, \quad \mathbf{D} \succ \mathbf{0}, \text{ diagonal matrix.}\end{aligned}$$

3. Set $\mathbf{J}_k = \mathbf{U}\mathbf{D}\mathbf{V}_k^\intercal$ and solve

$$\begin{aligned}&\text{minimize } \|\mathbf{X}_k - \mathbf{J}_k - \mathbf{U}_k\mathbf{D}_k\mathbf{W}_k^\intercal\|_{\mathbf{Q},\mathbf{R}_k}\\&\text{subject to } \mathbf{U}_k^\intercal\mathbf{Q}\mathbf{U}_k = \mathbf{I}_{r_k}, \quad \mathbf{W}_k^\intercal\mathbf{R}_k\mathbf{W}_k = \mathbf{I}_{r_k}, \quad \mathbf{D}_k \succ \mathbf{0}, \text{ diagonal matrix.}\end{aligned}$$

4. Set $\mathbf{A} = (\mathbf{U}_1\mathbf{D}_1\mathbf{W}_1^\intercal \dots \mathbf{U}_K\mathbf{D}_K\mathbf{W}_K^\intercal)$.

5. Repeat steps 2-4 until convergence.

2.2 Decorelatted JIVE

To estimate the matrices $\mathbf{J}, \mathbf{A}$ we use the following algorithm:

1. Set $\mathbf{A} = \mathbf{0}$

2. Solve

$$\begin{aligned}&\text{minimize } \|\mathbf{Q}^{1/2}(\mathbf{X} - \mathbf{A})\mathbf{R}^{1/2} - \mathbf{U}\mathbf{D}\mathbf{V}^\intercal\|_F\\&\text{subject to } \mathbf{U}^\intercal\mathbf{U} = \mathbf{I}_r, \quad \mathbf{V}^\intercal\mathbf{V} = \mathbf{I}_r, \quad \mathbf{D} \succ \mathbf{0}, \text{ diagonal matrix.}\end{aligned}$$

3. Set $\mathbf{J}_k = \mathbf{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}_k^\intercal\mathbf{\Delta}_k^{1/2}$ and solve

$$\begin{aligned}&\text{minimize } \|\mathbf{Q}^{1/2}(\mathbf{X}_k - \mathbf{J}_k)\mathbf{R}_k^{1/2} - \mathbf{U}_k\mathbf{D}_k\mathbf{W}_k^\intercal\|_F\\&\text{subject to } \mathbf{U}_k^\intercal\mathbf{U}_k = \mathbf{I}_{r_k}, \quad \mathbf{W}_k^\intercal\mathbf{W}_k = \mathbf{I}_{r_k}, \quad \mathbf{D}_k \succ \mathbf{0}, \text{ diagonal matrix.}\end{aligned}$$

4. Set $\mathbf{A} = (\boldsymbol{\Sigma}^{1/2}\mathbf{U}_1\mathbf{D}_1\mathbf{W}_1^\intercal\boldsymbol{\Delta}_1^{1/2} \dots \boldsymbol{\Sigma}^{1/2}\mathbf{U}_K\mathbf{D}_K\mathbf{W}_K^\intercal\boldsymbol{\Delta}_K^{1/2})$.

5. Repeat steps 2-4 until convergence.

2.3 GAJIVE

Suppose we have $\mathbf{X}_1 \in \mathbb{R}^{n\times p_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{n\times p_2}$ of rank r_1 and r_2 respectively. The joint structure is the linear subspace

$$\mathrm{col}(\mathbf{X}_1) \cap \mathrm{col}(\mathbf{X}_2)$$

If $\mathrm{GMD}_{r_i}(\mathbf{X}_i, \mathbf{Q}, \mathbf{R}_i) = (\mathbf{U}_i, \boldsymbol{\Sigma}_i, \mathbf{V}_i)$, $i = 1, 2$, then

$$\mathrm{col}(\mathbf{X}_1) \cap \mathrm{col}(\mathbf{X}_2) = \mathrm{col}(\mathbf{U}_1) \cap \mathrm{col}(\mathbf{U}_2)$$

Define $\tilde{\mathbf{U}} = (\mathbf{U}_1, \mathbf{U}_2)$ and $s = \min(r_1, r_2)$. Let $\mathrm{GMD}_s(\tilde{\mathbf{U}}, \mathbf{Q}, \mathbf{I}_s) = (\mathbf{H}, \mathbf{D}, \mathbf{W})$. The principal angles $\theta_1, \dots, \theta_s$ in increasing order are

$$\arccos(d_1^2 - 1), \dots, \arccos(d_s^2 - 1).$$

Suppose $\theta_1 = \dots = \theta_l = 0$. Then the first l columns of $\mathbf{H}$ form an orthonormal basis $\mathbf{G}$ for the joint structure. Define the projection matrix $\mathbf{P} = \mathbf{G}\mathbf{G}^\intercal\mathbf{Q}$. Then

$$\mathbf{J} = \mathbf{P}(\mathbf{X}_1, \mathbf{X}_2), \quad \mathbf{A} = (\mathbf{I} - \mathbf{P})(\mathbf{X}_1, \mathbf{X}_2).$$

2.4 SUM-GMD

Let $\mathbf{X} = (\mathbf{X}_1 \dots \mathbf{X}_K) \in \mathbb{R}^{n\times p}$. We only assume a joint structure $\mathbf{J}$ for $\mathbf{X}$. Define $s = \sum_{k=1}^K r_k$. To estimate $\mathbf{J}$ we solve

$$\begin{aligned}&\text{minimize } \|\mathbf{X} - \mathbf{U}\mathbf{D}\mathbf{V}^\intercal\|_F \\ &\text{subject to } \mathbf{U}^\intercal\mathbf{Q}\mathbf{U} = \mathbf{I}_r, \quad \mathbf{V}^\intercal\mathbf{R}\mathbf{V} = \mathbf{I}_s, \quad \mathbf{D} \succ \mathbf{0}, \text{ diagonal matrix.}\end{aligned}$$

3 Related PCA Model

Define $\mathbf{M} = \mathbf{J} + \mathbf{A}$. Then,

$$\mathbf{Y} := (\mathbf{X} - \mathbf{M})\mathbf{R}^{1/2} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{I}_p),$$
$$\mathbf{Z} := \mathbf{Q}^{1/2}(\mathbf{X} - \mathbf{M}) \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{I}_n, \mathbf{\Delta})$$

If $\mathbf{Y}_j$ denotes the j-th column of $\mathbf{Y}$ and $\mathbf{Z}_i$ denotes the i-th row of $\mathbf{Z}$, then

$$\mathbf{Y}_1, \ldots, \mathbf{Y}_p \overset{i.i.d.}{\sim} \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma}),$$
$$\mathbf{Z}_1, \ldots, \mathbf{Z}_n \overset{i.i.d.}{\sim} \mathcal{N}_p(\mathbf{0}, \mathbf{\Delta}).$$

Suppose $\mathbf{A} \in \mathbb{R}^{n\times r}$, $\mathbf{A}_i \in \mathbb{R}^{p_i\times r_i}$ are full rank matrices and $\mathbf{G} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r)$, $\mathbf{G}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{r_i})$. Suppose furthermore $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_n)$, $\mathbf{E}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{p_i})$ and define $\mathbf{Y} = \mathbf{AG} + \mathbf{E}$, $\mathbf{Z}_i = \mathbf{A}_i\mathbf{G}_i + \mathbf{E}_i$. Then

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{AA}^\intercal + \sigma^2\mathbf{I}_n) \quad \text{and} \quad \mathbf{Z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{A}_i\mathbf{A}_i^\intercal + \sigma^2\mathbf{I}_{p_i})$$

The maximum likelihood estimator of $\mathbf{A}$ is

$$\hat{\mathbf{A}} = \hat{\mathbf{U}}(\hat{\mathbf{\Lambda}} - \hat{\sigma}^2\mathbf{I})^{1/2}\hat{\mathbf{S}}^\intercal,$$

where $(\hat{\mathbf{U}}, \hat{\mathbf{\Lambda}})$ is the eigenvalue decomposition of the sample covariance matrix $\hat{\mathbf{\Sigma}}$, and $\mathbf{S}$ is an arbitrary orthonormal matrix. Similarly for $\mathbf{A}_i$.

To connect the two models, we have to assume

$$\mathbf{X}_i \sim \mathcal{MN}_{n\times p_i}(\mathbf{UDV}_i^\intercal + \mathbf{U}_i\mathbf{D}_i\mathbf{W}_i^\intercal, \mathbf{ULU}^\intercal, \mathbf{U}_i\mathbf{L}_i\mathbf{U}_i^\intercal)$$

4 Matrix Normal PCA model

Let the full rank matrices $\mathbf{A} \in \mathbb{R}^{n\times r}$, $\mathbf{B}_i \in \mathbb{R}^{p_i\times r_i}$ and the random variables $\mathbf{Y}_i \sim \mathcal{MN}_{r\times p_i}(\mathbf{0}, \mathbf{I}_r, \mathbf{I}_{p_i})$, $\mathbf{Z}_i \sim \mathcal{MN}_{n\times r_i}(\mathbf{0}, \mathbf{I}_n, \mathbf{I}_{r_i})$, $\mathbf{E}_i \sim \mathcal{MN}_{n\times p_i}(\mathbf{0}, \sigma^2\mathbf{I}_n, \sigma^2\mathbf{I}_{p_i})$ and suppose

$$\mathbf{X}_i = \mathbf{AY}_i + \mathbf{Z}_i\mathbf{B}_i^\intercal + \mathbf{E}_i.$$

Then

$$\mathbf{X}_i \sim \mathcal{MN}_{n\times p_i}(\mathbf{0}, \mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n, \mathbf{B}_i\mathbf{B}_i^\intercal + \sigma^2\mathbf{I}_{p_i}), \quad i = 1, \ldots, K.$$

Proof. Define $\mathbf{B} = \operatorname{block.diag}(\mathbf{B}_1, \ldots, \mathbf{B}_K)$. The negative log-likelihood up to a constant is

$$\begin{aligned} l(\mathbf{A}, \mathbf{B}, \sigma|\mathbf{X}) = \operatorname{tr}&\left[\left(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p\right)^{-1}\mathbf{X}^\intercal\left(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n\right)^{-1}\mathbf{X}\right] \\ &+ n\log\det\left(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p\right) + p\log\det\left(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n\right). \end{aligned}$$

The partial derivatives are

$$\begin{aligned} \frac{\partial l(\mathbf{A}, \mathbf{B}, \sigma|\mathbf{X})}{\partial \mathbf{A}} &= -2(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-1}\mathbf{X}(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-1}\mathbf{X}^\intercal(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-1}\mathbf{A} \\ &\quad + 2p(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-1}\mathbf{A} \\ \frac{\partial l(\mathbf{A}, \mathbf{B}, \sigma|\mathbf{X})}{\partial \mathbf{B}} &= -2(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-1}\mathbf{X}^\intercal(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-1}\mathbf{X}(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-1}\mathbf{B} \\ &\quad + 2n(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-1}\mathbf{B}, \\ \frac{\partial l(\mathbf{A}, \mathbf{B}, \sigma|\mathbf{X})}{\partial \sigma} &= 2\sigma\operatorname{tr}\left[(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-2}\mathbf{X}^\intercal(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-1}\mathbf{X}\right] \\ &\quad + 2\sigma\operatorname{tr}\left[(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-2}\mathbf{X}(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-1}\mathbf{X}^\intercal\right] \\ &\quad + 2n\sigma\operatorname{tr}\left[(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-1}\right] + 2p\sigma\operatorname{tr}\left[(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-1}\right] \end{aligned}$$

One class of stationary points are given by the equations

$$\begin{aligned} \mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n - \frac{1}{p}\mathbf{X}(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-1}\mathbf{X}^\intercal &= \mathbf{0}, \\ \mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p - \frac{1}{n}\mathbf{X}^\intercal(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-1}\mathbf{X} &= \mathbf{0}, \\ n\operatorname{tr}\left[(\mathbf{B}\mathbf{B}^\intercal + \sigma^2\mathbf{I}_p)^{-1}\right] + p\operatorname{tr}\left[(\mathbf{A}\mathbf{A}^\intercal + \sigma^2\mathbf{I}_n)^{-1}\right] &= \mathbf{0}. \end{aligned} \tag{4.1}$$

Suppose $\mathbf{A} = \mathbf{U}_\mathbf{A}\mathbf{D}_\mathbf{A}\mathbf{V}_\mathbf{A}^\intercal$, $\mathbf{B} = \mathbf{U}_\mathbf{B}\mathbf{D}_\mathbf{B}\mathbf{V}_\mathbf{B}^\intercal$ are the singular value decompositions of $\mathbf{A}$ and $\mathbf{B}$, such that $\mathbf{U}_\mathbf{A}^\intercal\mathbf{U}_\mathbf{A} = \mathbf{U}_\mathbf{A}\mathbf{U}_\mathbf{A}^\intercal = \mathbf{I}_n$, $\mathbf{U}_\mathbf{B}^\intercal\mathbf{U}_\mathbf{B} = \mathbf{U}_\mathbf{B}\mathbf{U}_\mathbf{B}^\intercal = \mathbf{I}_p$ and $\mathbf{D}_\mathbf{A}$, $\mathbf{D}_\mathbf{B}$ are rectangular diagonal matrices. Define the diagonal matrices $\mathbf{\Lambda_A} = \mathbf{D_A D_A^\intercal} + \sigma^2 \mathbf{I}_n$, $\mathbf{\Lambda_B} = \mathbf{D_B D_B^\intercal} + \sigma^2 \mathbf{I}_p$. Then (4.1) becomes

$$
\begin{aligned}
\mathbf{U_A \Lambda_A U_A^\intercal} - \frac{1}{p}\mathbf{X U_B \Lambda_B^{-1} U_B^\intercal X^\intercal} &= \mathbf{0}, \\
\mathbf{U_B \Lambda_B U_B^\intercal} - \frac{1}{n}\mathbf{X^\intercal U_A \Lambda_A^{-1} U_A^\intercal X} &= \mathbf{0}, \\
n \operatorname{tr}(\mathbf{\Lambda_B^{-1}}) + p \operatorname{tr}(\mathbf{\Lambda_A^{-1}}) &= \mathbf{0}.
\end{aligned} \tag{4.2}
$$

Let $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\sigma}^2)$ be the solution of (4.2). Define the

$$
\begin{aligned}
\hat{\mathbf{\Sigma}} = \hat{\mathbf{A}}\hat{\mathbf{A}}^\intercal, \quad \tilde{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}} + \hat{\sigma}^2 \mathbf{I}_n, \\
\hat{\mathbf{\Delta}} = \hat{\mathbf{B}}\hat{\mathbf{B}}^\intercal, \quad \tilde{\mathbf{\Delta}} = \hat{\mathbf{\Delta}} + \hat{\sigma}^2 \mathbf{I}_p.
\end{aligned}
$$

From (4.2) we see that in order to find $\hat{\mathbf{A}}, \hat{\mathbf{B}}$ we first need to find matrices $\tilde{\mathbf{\Sigma}}, \tilde{\mathbf{\Delta}}$ such that

$$
\begin{aligned}
\tilde{\mathbf{\Sigma}} &= \frac{1}{p}\mathbf{X}\tilde{\mathbf{\Delta}}^{-1}\mathbf{X}^\intercal, \\
\tilde{\mathbf{\Delta}} &= \frac{1}{n}\mathbf{X}^\intercal\tilde{\mathbf{\Sigma}}^{-1}\mathbf{X}.
\end{aligned}
$$

We then perform eigenvalue decomposition on $\tilde{\mathbf{\Sigma}}, \tilde{\mathbf{\Delta}}$ to find $(\hat{\mathbf{U}}_\mathbf{A}, \hat{\mathbf{\Lambda}}_\mathbf{A}), (\hat{\mathbf{U}}_\mathbf{B}, \hat{\mathbf{\Lambda}}_\mathbf{B})$. After that we find $\hat{\sigma}^2$ by solving

$$
n \operatorname{tr}\left(\hat{\mathbf{\Lambda}}_\mathbf{B}^{-1}\right) + p \operatorname{tr}\left(\hat{\mathbf{\Lambda}}_\mathbf{A}^{-1}\right) = \mathbf{0}.
$$

Finally,

$$
\begin{aligned}
\hat{\mathbf{\Delta}}_\mathbf{A} &= \hat{\mathbf{\Lambda}}_\mathbf{A}^{1/2} - \hat{\sigma}^2 \mathbf{I}_n, \\
\hat{\mathbf{\Delta}}_\mathbf{B} &= \hat{\mathbf{\Lambda}}_\mathbf{B}^{1/2} - \hat{\sigma}^2 \mathbf{I}_p.
\end{aligned}
$$

$\square$

5 Simulation

5.1 Constructing row and column covariance matrices

For $i, j \in \mathbb{N} \cup \{0\}$, we define the following metrics:

$$\begin{aligned}
d(i,j) &= |i-j|, \\
d_1(i,j) &= \frac{d(i,j)}{1+d(i,j)}, \\
d_2(i,j) &= \log(1+d(i,j)), \\
d_3(i,j) &= \begin{cases} \max\{i,j\}, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}.
\end{aligned}$$

We choose $\rho \in (0,1)$ and define the row covariance matrix $\mathbf{\Sigma} = (\Sigma_{ij})$ such that

$$\Sigma_{ij} = \rho^{d(i,j)}.$$

For each category k, we define the column covariance matrix $\mathbf{\Delta}_k = (\Delta_{k,ij})$ such that

$$\Delta_{k,ij} = \rho^{d_{k \bmod 3}(i,j)}.$$

5.2 Constructing joint and individual structure

The goal of this section is to construct the joint structure $\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{V}^\intercal$ and the individual structures $\mathbf{A}_k = \mathbf{U}_k\mathbf{D}_k\mathbf{V}_k^\intercal$, $k = 1, \ldots, K$. We start by generating a random orthonormal basis

$$\mathbf{B} = \begin{pmatrix} \mathbf{e}_1 & \ldots & \mathbf{e}_n \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

We construct the left-singular matrices such that

$$\mathbf{U} = \mathbf{\Sigma}^{1/2}\begin{pmatrix} \mathbf{e}_1 & \ldots & \mathbf{e}_r \end{pmatrix},$$

$$
\begin{aligned}
\mathbf{U}_1 &= \mathbf{\Sigma}^{1/2}\begin{pmatrix} \mathbf{e}_{r+1} & \dots & \mathbf{e}_{r+r_1} \end{pmatrix},\\
&\vdots\\
\mathbf{U}_K &= \mathbf{\Sigma}^{1/2}\begin{pmatrix} \mathbf{e}_{r+r_1+\ldots+r_{K-1}+1} & \dots & \mathbf{e}_{r+r_1+\ldots+r_K} \end{pmatrix}.
\end{aligned}
$$

This way we make sure that

$$\mathbf{U}^\intercal\mathbf{Q}\mathbf{U} = \mathbf{I}_r, \quad \mathbf{U}_k^\intercal\mathbf{Q}\mathbf{U}_k = \mathbf{I}_{r_k}.$$

Furthermore, we guarante that the intersection of the column spaces of $\mathbf{J}$ and $\mathbf{A}_1, \ldots, \mathbf{A}_K$ is the zero vector since

$$\mathbf{J}^\intercal\mathbf{Q}\mathbf{A}_k = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{J} \perp \mathbf{A}_k, \quad k = 1, \ldots, K.$$

This means that the individual structures are truly individual because they do not share a common subspace with the joint space. In addition, we make sure that the intersection of the column spaces of $\mathbf{A}_1, \ldots, \mathbf{A}_K$ is the zero vector since

$$\mathbf{A}_i^\intercal\mathbf{Q}\mathbf{A}_j = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{A}_j \perp \mathbf{A}_i, \quad i \neq j$$

This means that there is no common structure inside the individual structures.

Let $\mathbf{\Delta}$ be a block diagonal matrix with blocks $\mathbf{\Delta}_1, \ldots, \mathbf{\Delta}_K$. We construct the right-singular matrices $\mathbf{V}, \mathbf{V}_1, \ldots, \mathbf{V}_K$ such that

$$
\begin{aligned}
\mathbf{V} &= \mathbf{\Delta}^{1/2}\mathbf{C},\\
\mathbf{V}_1 &= \mathbf{\Delta}^{1/2}\mathbf{C}_1,\\
&\vdots\\
\mathbf{V}_K &= \mathbf{\Delta}^{1/2}\mathbf{C}_K,
\end{aligned}
$$

where $\mathbf{C}, \mathbf{C}_1, \ldots, \mathbf{C}_K$ are orthonormal matrices. This way we make sure that

$$\mathbf{V}^\intercal\mathbf{R}\mathbf{V} = \mathbf{I}_r, \quad \mathbf{V}_k^\intercal\mathbf{R}_k\mathbf{V}_k = \mathbf{I}_{r_k}, \quad k = 1, \ldots, K,$$

where $\mathbf{R} = \mathbf{\Delta}^1$ and $\mathbf{R}_k = \mathbf{\Delta}_k^{-1}$.

Finally, we randomly choose numbers $\lambda_1, \ldots, \lambda_{r+r_1+\ldots+r_K}$ such that $1 \leq \lambda_i \leq 20$ and

$$\begin{aligned}
&\lambda_1 \geq \ldots \geq \lambda_r, \\
&\lambda_{r+1} \geq \ldots \geq \lambda_{r+r_1}, \\
&\vdots \\
&\lambda_{r+r_1+\ldots r_{K-1}+1} \geq \ldots \geq \lambda_{r+r_1+\ldots+r_K}.
\end{aligned}$$

With these number we construct the matrices

$$\begin{aligned}
\mathbf{D} &= \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_r \end{pmatrix}, \\
\mathbf{D}_1 &= \begin{pmatrix} \lambda_{r+1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_{r+r_1} \end{pmatrix}, \\
&\vdots \\
\mathbf{D}_K &= \begin{pmatrix} \lambda_{r+r_1+\ldots+r_{k-1}+1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_{r+r_1+\ldots+r_K} \end{pmatrix}.
\end{aligned}$$

Let $s \leq \min\{n, p\}$. We first describe how to construct a matrix $\mathbf{U} \in \mathbb{R}^{n \times s}$ such that $\mathbf{U}^\intercal \mathbf{Q} \mathbf{U} = \mathbf{I}_s$ for $\mathbf{Q} \in \mathbb{R}^{n \times n}$, $\mathbf{Q} \succ \mathbf{0}$. Choose $\mathbf{e}_1, \ldots, \mathbf{e}_s \in \mathbb{R}^n$ such that $\mathbf{e}_i^\intercal \mathbf{e}_j = \delta_{ij}$, where δ_{ij} is the Kroneker delta. There are many ways to construct such vectors. An easy way is to find a known orthonormal basis of $\mathbb{R}^s$ and then attach $n - s$ zeros at the end of each vector. A second approach in the same line of though would be to choose a known basis of $\mathbb{R}^n$ and simply remove $n - s$ vectors from it. Another way is to arbitrarily choose s linearly independent vectors of $\mathbb{R}^n$ and then apply the Gram-Schmidt process on them. The second step of this process is to define $\mathbf{U} = \mathbf{Q}^{-1/2}(\mathbf{e}_1 \ \ldots \ \mathbf{e}_s)$. Then,

$$(\mathbf{e}_1 \ \ldots \ \mathbf{e}_s)^\intercal (\mathbf{e}_1 \ \ldots \ \mathbf{e}_s) = \mathbf{I}_s \Leftrightarrow (\mathbf{Q}^{1/2}\mathbf{U})^\intercal (\mathbf{Q}^{1/2}\mathbf{U}) = \mathbf{I}_s \Leftrightarrow \mathbf{U}^\intercal \mathbf{Q} \mathbf{U} = \mathbf{I}_s.$$

Similarly, we construct a matrix $\mathbf{V} \in \mathbb{R}^{p \times s}$ such that $\mathbf{V}^\intercal \mathbf{R} \mathbf{V} = \mathbf{I}_s$. Finally, we simply define a diagonal positive definite matrix $\mathbf{D} \in \mathbb{R}^{s \times s}$. We define the joint structure to be

$$\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{V}^\intercal.$$

Let $s_k \leq \min\{n, p_k\}$. Similarly like before, we construct matrices $\mathbf{U}_k \in \mathbb{R}^{n \times s_k}, \mathbf{D}_k \in \mathbb{R}^{s_k \times s_k}, \mathbf{W}_k \in \mathbb{R}^{p_k \times s_k}$ such that $\mathbf{U}_k^\intercal \mathbf{Q} \mathbf{U}_k = \mathbf{I}_{s_k}$, $\mathbf{W}_k^\intercal \mathbf{R}_k \mathbf{W}_k = \mathbf{I}_{s_k}$ and $\mathbf{D}_k$ is a positive definite diagonal matrix. We define the individual structure to be

$$\mathbf{A}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{W}_k^\intercal.$$

We repeat this process K times to construct the concatenated matrix $\mathbf{A} = (\mathbf{A}_1 \ \ldots \ \mathbf{A}_k)$. To ensure identifiability we require that $\mathbf{J}^T \mathbf{Q} \mathbf{A}_k = \mathbf{0}$ and $\mathbf{A}_i^\intercal \mathbf{Q} \mathbf{A}_j = \mathbf{0}$ for $i \neq j$.

In order to ensure identifiability we need to generate $\mathbf{U}$ and $\mathbf{U}_k$ by choosing different vectors from the same basis. A necessary condition for this to happen is $n \geq r + \sum_k r_k$.

Having constructed the common and individual structure, we define the observed data matrix to be

$$\mathbf{X} = \mathbf{J} + \mathbf{A} + \mathbf{E},$$

where $\mathbf{E}$ is drawn from a matrix normal distribution $MN_{n,p}(\mathbf{0}, \mathbf{Q}^{-1}, \mathbf{R}^{-1})$.

6 Principal angles between subspaces of $\mathbb{R}^n$

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and positive definite matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ define the inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{Q}} = \mathbf{x}^{\intercal}\mathbf{Q}\mathbf{y}.$$

Let L, M be subspaces in $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{\mathbf{Q}})$. Without loss of generality we assume

$$\dim L = l \leq m = \dim M$$

The principal angles between L and M, $0 \leq \theta_1 \leq \theta_2 \leq \ldots \leq \theta_l \leq \pi/2$ are defined by

$$\cos\theta_i = \frac{\langle \mathbf{x}_i, \mathbf{y}_i \rangle_{\mathbf{Q}}}{\|\mathbf{x}_i\|_{\mathbf{Q}} \|\mathbf{y}_i\|_{\mathbf{Q}}} = \operatorname{argmax} \left\{ \frac{\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{Q}}}{\|\mathbf{x}\|_{\mathbf{Q}} \|\mathbf{y}\|_{\mathbf{Q}}} : \begin{array}{ll} \mathbf{x} \in L, & \mathbf{x} \perp \mathbf{x}_k, \\ \mathbf{y} \in M, & \mathbf{y} \perp \mathbf{y}_k, \end{array} \; k = 1 \ldots, i-1 \right\}.$$

Note that

$$\theta_1 = \ldots = \theta_k = 0 < \theta_{k+1} \quad \Leftrightarrow \quad \dim L \cap M = k > 0.$$

Lemma 1. *Let the columns of $\mathbf{G}_L \in \mathbb{R}^{n\times l}$ and $\mathbf{G}_M \in \mathbb{R}^{n\times m}$ be orthonormal bases for L and M respectively, and let*

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_l \geq 0$$

be the singular values of $\mathbf{G}_L^{\intercal}\mathbf{Q}\mathbf{G}_M$. Then

$$\cos\theta_i = \sigma_i, \quad i = 1, \ldots, l,$$

and

$$\sigma_1 = \ldots = \sigma_k = 1 > \sigma_{k+1} \quad \Leftrightarrow \quad \dim L \cap M = k > 0.$$

Proof. Define

$$\mathbf{s} = \mathbf{G}_L \mathbf{y}, \quad \mathbf{t} = \mathbf{G}_M \mathbf{z}.$$

Then,

$$\mathbf{s}^\intercal \mathbf{Q} \mathbf{s} = \mathbf{y}^\intercal \mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_L \mathbf{y} = \mathbf{y}^\intercal \mathbf{y},$$
$$\mathbf{t}^\intercal \mathbf{Q} \mathbf{t} = \mathbf{z}^\intercal \mathbf{G}_M^\intercal \mathbf{Q} \mathbf{G}_M \mathbf{z} = \mathbf{z}^\intercal \mathbf{z}.$$

Therefore, the problem

$$\mathbf{y}_k^\intercal \mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M \mathbf{z}_k = \operatorname{argmax} \left\{ \mathbf{y}^\intercal \mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M \mathbf{z} : \begin{array}{c} \mathbf{y} \in \mathbb{R}^l, \mathbf{z} \in \mathbb{R}^m, \\ \|\mathbf{y}\|_2 = \|\mathbf{z}\|_2 = 1, \quad i = 1, \ldots, k-1 \\ \mathbf{y}^\intercal \mathbf{y}_i = \mathbf{z}^\intercal \mathbf{z}_i = 0, \end{array} \right\}$$

can be rewritten as

$$\mathbf{s}_k^\intercal \mathbf{Q} \mathbf{t}_k = \operatorname{argmax} \left\{ \mathbf{s}^\intercal \mathbf{Q} \mathbf{t} : \begin{array}{c} \mathbf{s} \in L, \mathbf{t} \in M, \\ \|\mathbf{s}\|_{\mathbf{Q}} = \|\mathbf{t}\|_{\mathbf{Q}} = 1, \quad i = 1, \ldots, k-1 \\ \mathbf{s}^\intercal \mathbf{Q} \mathbf{s}_i = \mathbf{t}^\intercal \mathbf{Q} \mathbf{t}_i = 0, \end{array} \right\}.$$

□

Lemma 1 shows that the principal angle problem of L, M in $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{\mathbf{Q}})$ is equivalent to the problem of singular value decomposition of $\mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M$ in $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_2)$

Theorem 1. *Let the columns of* $\mathbf{G}_L \in \mathbb{R}^{n \times l}$ *and* $\mathbf{G}_M \in \mathbb{R}^{n \times m}$ *be orthonormal bases for* L *and* M *respectively, and denote*

$$\mathbf{G} = (\mathbf{G}_L, \mathbf{G}_M).$$

Let the singular values of $\mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M$ *be*

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_l \geq 0.$$

Then the singular values of $\mathrm{GMD}(\mathbf{G}, \mathbf{Q}, \mathbf{I}_{l+m})$ *in decreasing order are*

$$\sqrt{1+\sigma_1}, \ldots, \sqrt{1+\sigma_l}, \underbrace{1, \ldots, 1}_{m-l}, \sqrt{1-\sigma_l}, \ldots, \sqrt{1-\sigma_1}.$$

Proof. Since

$$\mathbf{G}^\intercal \mathbf{Q} \mathbf{G} = \begin{pmatrix} \mathbf{I}_l & \mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M \\ \mathbf{G}_M^\intercal \mathbf{Q} \mathbf{G}_L & \mathbf{I}_m \end{pmatrix} = \mathbf{I}_{l+m} + \begin{pmatrix} \mathbf{0} & \mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M \\ \mathbf{G}_M^\intercal \mathbf{Q} \mathbf{G}_L & \mathbf{0} \end{pmatrix}$$

By Lemma 1 there exist $\mathbf{U} \in \mathbb{R}^{l \times l}$, $\mathbf{\Sigma} \in \mathbb{R}^{l \times l}$ and $\mathbf{V} \in \mathbf{R}^{m \times l}$ such that

$$\mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\intercal$$

and

$$\mathbf{U}^\intercal \mathbf{U} = \mathbf{V}^\intercal \mathbf{V} = \mathbf{I}_l, \quad \mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_l).$$

We can write the SVD in its complete form by including the singular vectors that correspond to the zero singular values, i.e.

$$\mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M = \mathbf{U} (\mathbf{\Sigma} \quad \mathbf{0}_{l \times (m-l)}) \begin{pmatrix} \mathbf{V}^\intercal \\ \tilde{\mathbf{V}}^\intercal \end{pmatrix},$$

where $(\mathbf{V} \quad \tilde{\mathbf{V}})$ is an orthonormal basis of $\mathbf{R}^m$. Define

$$\mathbf{W} = \begin{pmatrix} \frac{1}{\sqrt{2}} \mathbf{U} & \mathbf{0} & -\frac{1}{\sqrt{2}} \mathbf{U} \\ \frac{1}{\sqrt{2}} \mathbf{V} & \tilde{\mathbf{V}} & \frac{1}{\sqrt{2}} \mathbf{V} \end{pmatrix}.$$

Then

$$\begin{pmatrix} \mathbf{0} & \mathbf{G}_L^\intercal \mathbf{Q} \mathbf{G}_M \\ \mathbf{G}_M^\intercal \mathbf{Q} \mathbf{G}_L & \mathbf{0} \end{pmatrix} = \mathbf{W} \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{\Sigma} \end{pmatrix} \mathbf{W}^\intercal \quad \text{and} \quad \mathbf{W}^\intercal \mathbf{W} = \mathbf{W} \mathbf{W}^\intercal = \mathbf{I}_{l+m}.$$

Therefore,

$$\mathbf{G}^\intercal\mathbf{Q}\mathbf{G} = \mathbf{W}\begin{pmatrix} \mathbf{I}_l + \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-l} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_l - \mathbf{\Sigma} \end{pmatrix}\mathbf{W}^\intercal.$$

Suppose the generalized matrix decomposition of $\mathbf{G}$ is $\mathbf{X}\mathbf{D}\mathbf{Y}^\intercal$, where

$$\begin{aligned} &\mathbf{X} \in \mathbb{R}^{n\times(l+m)}, \quad \mathbf{X}^\intercal\mathbf{Q}\mathbf{X} = \mathbf{I}_{l+m}, \\ &\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_{l+m}), \\ &\mathbf{Y} \in \mathbb{R}^{(l+m)\times(l+m)}, \quad \mathbf{Y}^\intercal\mathbf{Y} = \mathbf{Y}\mathbf{Y}^\intercal = \mathbf{I}_{l+m}. \end{aligned}$$

Then

$$\mathbf{G}^\intercal\mathbf{Q}\mathbf{G} = \mathbf{Y}\mathbf{D}\mathbf{X}^\intercal\mathbf{Q}\mathbf{X}\mathbf{D}\mathbf{Y}^\intercal = \mathbf{Y}\mathbf{D}^2\mathbf{Y}^\intercal,$$

which implies that

$$d_i^2 = \begin{cases} 1 + \sigma_i, & 1 \le i \le l \\ 1, & l+1 \le i \le m \\ 1 - \sigma_i, & m+1 \le l \le m+l \end{cases}$$

□

Theorem 1 shows that the principal angle problem of L, M in $(\mathbb{R}^n, \langle\cdot,\cdot\rangle_{\mathbf{Q}})$ is equivalent to the GMD$(\mathbf{G}, \mathbf{Q}, \mathbf{I}_{l+m})$

Let $\mathbf{X}_1 \in \mathbb{R}^{n\times p_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{n\times p_2}$ of rank r_1 and r_2 respectively. Without loss of generality we assume $r_1 \le r_2$. Let the generalized matrix decomposition of $\mathbf{X}_k$ be

$$\mathbf{X}_k = \mathbf{U}_k\mathbf{D}_k\mathbf{V}_k^\intercal, \quad k = 1, 2,$$

where $\mathbf{U}_k^\intercal\mathbf{Q}\mathbf{U}_k = \mathbf{I}_{r_k}$ and $\mathbf{V}_k^\intercal\mathbf{R}_k\mathbf{V}_k = \mathbf{I}_{r_k}$.

Lemma 2. *Let* $\mathbf{X} \in \mathbb{R}^{n\times p}$ *and suppose* $\mathrm{GMD}_r(\mathbf{X}, \mathbf{Q}, \mathbf{R}) = (\mathbf{U}, \mathbf{\Sigma}, \mathbf{V})$*. Then the column space of* $\mathbf{X}$ *is equal to the column space of* $\mathbf{U}$*.*

Proof. Define $\mathbf{S} = (\mathbf{\Sigma} \quad \mathbf{0}) \in \mathbb{R}^{r\times p}$. There exists matrix $\mathbf{V}^*$ such that $\tilde{\mathbf{V}}^\intercal \mathbf{R} \tilde{\mathbf{V}} = \mathbf{I}_p$, where $\tilde{\mathbf{V}} = (\mathbf{V} \quad \mathbf{V}^*) \in \mathbb{R}^{p\times p}$.

The inverse of $\tilde{\mathbf{V}}$ exists. Indeed,

$$\mathbf{I}_p = (\mathbf{R}^{1/2}\tilde{\mathbf{V}})^\intercal(\mathbf{R}^{1/2}\tilde{\mathbf{V}}) \Rightarrow (\mathbf{R}^{1/2}\tilde{\mathbf{V}})^{-1} = \tilde{\mathbf{V}}^\intercal \mathbf{R}^{1/2} \Rightarrow \tilde{\mathbf{V}}^{-1} = \tilde{\mathbf{V}}^\intercal \mathbf{R}$$

It is a know fact that for $\mathbf{A}, \mathbf{B}$ such that $\mathbf{B}^{-1}$ exists we have

$$\mathrm{col}(\mathbf{AB}) = \mathrm{col}(\mathbf{A}).$$

Thus,

$$\mathrm{col}(\mathbf{X}) = \mathrm{col}(\mathbf{US}\tilde{\mathbf{V}}^\intercal) = \mathrm{col}(\mathbf{US}) = \mathrm{col}(\mathbf{U\Sigma}) = \mathrm{col}(\mathbf{U}).$$

□

Lemma 2 implies that the column space of the joint structure is equal to

$$\mathrm{col}(\mathbf{X}_1) \cap \mathrm{col}(\mathbf{X}_2) = \mathrm{col}(\mathbf{U}_1) \cap \mathrm{col}(\mathbf{U}_2).$$

Define $\tilde{\mathbf{U}} = (\mathbf{U}_1, \mathbf{U}_2)$. Let $\mathbf{W\Lambda Y}^\intercal$ be the r_1-rank generalized matrix decomposition of $\tilde{\mathbf{U}}$, such that $\mathbf{W}^\intercal\mathbf{QW} = \mathbf{I}_{r_1}$ and $\mathbf{Y}^\intercal\mathbf{Y} = \mathbf{I}_{r_1}$. By Theorem 1 the singular values of $\mathrm{GMD}(\tilde{\mathbf{U}}, \mathbf{Q}, \mathbf{I}_{r_1})$ in decreasing order are

$$\sqrt{1+\cos\theta_1}, \ldots, \sqrt{1+\cos\theta_{r_1}}.$$

Define $\lambda_i = \sqrt{1+\cos\theta_i}$. Thus, the principal angles of $\mathbf{U}_1$ and $\mathbf{U}_2$ in increasing order are

$$0 \le \theta_1 = \arccos(\lambda_1^2 - 1) \le \ldots \le \theta_{r_1} = \arccos(\lambda_{r_1}^2 - 1) \le \frac{\pi}{2}.$$

Let $\mathbf{W} = (\mathbf{w}_1 \dots \mathbf{w}_{r_1})$. Suppose $\theta_1 = \dots = \theta_s = 0$. Then $\tilde{\mathbf{W}} = (\mathbf{w}_1 \dots \mathbf{w}_s)$ is the orthonormal basis of the joint structure. Define the projection matrix $\mathbf{P} = \tilde{\mathbf{W}}\tilde{\mathbf{W}}^\intercal\mathbf{Q}$. The joint structure and the individual structure of $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ are given by $\mathbf{PX}$ and $(\mathbf{I} - \mathbf{P})\mathbf{X}$ respectively.

7 Metrics

To compare the methods we are using two different metrics. Let $\mathbf{J}, \mathbf{A}$ be the true joint and individual structure matrices, and $\hat{\mathbf{J}}, \hat{\mathbf{A}}$ be the estimated joint and individual structure matrices. To compare the estimation error we use

$$\frac{\|\mathbf{J} - \hat{\mathbf{J}}\|_{\mathbf{Q},\mathbf{R}}}{\|\mathbf{J}\|_{\mathbf{Q},\mathbf{R}}} + \frac{\|\mathbf{A} - \hat{\mathbf{A}}\|_{\mathbf{Q},\mathbf{R}}}{\|\mathbf{A}\|_{\mathbf{Q},\mathbf{R}}}.$$

The covariance matrices $\mathbf{Q}, \mathbf{R}$ are chosen so that they are in accordance with the assumptions for the estimator. For example if we are dealing with the JIVE method we use $\mathbf{Q} = \mathbf{I}_n, \mathbf{R} = \mathbf{I}_p$. On the other hand if we are dealing with the iterative JIVE estimators we use the given non-trivial $\mathbf{Q}, \mathbf{R}$.

The second metric that we use has to do with subspace recovery. Suppose $\mathbf{J} = \mathbf{UDV}^\intercal$ and $\hat{\mathbf{J}} = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{V}}^\intercal$. The projection onto the true joint column space is given by $\mathbf{P} = \mathbf{UU}^\intercal\mathbf{Q}$ and onto the estimated joint column space is given by $\hat{\mathbf{P}} = \hat{\mathbf{U}}\hat{\mathbf{U}}^\intercal\mathbf{Q}$. The subspace recovery error is given by

$$\|\mathbf{P} - \hat{\mathbf{P}}\|_F.$$

Again $\mathbf{Q}$ is chosen in accordance with the assumptions about the space and it can be different between the true and the estimated space.

8 Metric for variation

Let $\mathbf{J}$ and $\mathbf{A}$ be defined as above. Then,

$$\begin{aligned}
\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}} &= \operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{R}\mathbf{X}^\intercal) \\
&= \operatorname{tr}[\mathbf{Q}(\mathbf{J}+\mathbf{A})\mathbf{R}(\mathbf{J}+\mathbf{A})^\intercal] \\
&= \operatorname{tr}(\mathbf{Q}\mathbf{J}\mathbf{R}\mathbf{J}^\intercal) + 2\operatorname{tr}(\mathbf{Q}\mathbf{J}\mathbf{R}\mathbf{A}^\intercal) + \operatorname{tr}(\mathbf{Q}\mathbf{A}\mathbf{R}\mathbf{A}^\intercal).
\end{aligned}$$

By assumption, $\mathbf{J}^\intercal\mathbf{Q}\mathbf{A} = \mathbf{0}$. Therefore,

$$\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}} = \operatorname{tr}(\mathbf{Q}\mathbf{J}\mathbf{R}\mathbf{J}^\intercal) + \operatorname{tr}(\mathbf{Q}\mathbf{A}\mathbf{R}\mathbf{A}^\intercal) = \|\mathbf{J}\|^2_{\mathbf{Q},\mathbf{R}} + \|\mathbf{A}\|^2_{\mathbf{Q},\mathbf{R}}.$$

Let the solution of GMD for $\mathbf{J}$ be $\mathbf{U}\mathbf{D}\mathbf{V}^\intercal$. Then,

$$\|\mathbf{J}\|^2_{\mathbf{Q},\mathbf{R}} = \operatorname{tr}(\mathbf{Q}\mathbf{U}\mathbf{D}\mathbf{V}^\intercal\mathbf{R}\mathbf{V}\mathbf{D}\mathbf{U}^\intercal) = \operatorname{tr}(\mathbf{Q}\mathbf{U}\mathbf{D}^2\mathbf{U}^\intercal) = \operatorname{tr}(\mathbf{D}^2\mathbf{U}^\intercal\mathbf{Q}\mathbf{U}) = \operatorname{tr}(\mathbf{D}^2).$$

Writing $\mathbf{A}$ in its concatenated form $(\mathbf{A}_1 \ \ldots \ \mathbf{A}_K)$ we get,

$$\|\mathbf{A}\|^2_{\mathbf{Q},\mathbf{R}} = \operatorname{tr}(\mathbf{Q}\mathbf{A}\mathbf{R}\mathbf{A}^\intercal) = \sum_{k=1}^{K} \operatorname{tr}(\mathbf{Q}\mathbf{A}_k\mathbf{R}_k\mathbf{A}_k^\intercal) = \sum_{k=1}^{K} \|\mathbf{A}_k\|^2_{\mathbf{Q},\mathbf{R}_k}.$$

Let the solution of GMD for $\mathbf{A}_k$ be $\mathbf{U}_k\mathbf{D}_k\mathbf{W}_k^\intercal$. Then,

$$\begin{aligned}
\|\mathbf{A}_k\|^2_{\mathbf{Q},\mathbf{R}_k} &= \operatorname{tr}(\mathbf{Q}\mathbf{U}_k\mathbf{D}_k\mathbf{W}_k^\intercal\mathbf{R}_k\mathbf{W}_k\mathbf{D}_k\mathbf{U}_k^\intercal) \\
&= \operatorname{tr}(\mathbf{Q}\mathbf{U}_k\mathbf{D}_k^2\mathbf{U}_k^\intercal) \\
&= \operatorname{tr}(\mathbf{D}_k^2\mathbf{U}_k^\intercal\mathbf{Q}\mathbf{U}_k) \\
&= \operatorname{tr}(\mathbf{D}_k^2).
\end{aligned}$$

Therefore,

$$\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}} = \operatorname{tr}(\mathbf{D}^2) + \sum_{k=1}^{K} \operatorname{tr}(\mathbf{D}_k^2).$$

From the above, we deduce that the quantities $\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}}$, $\|\mathbf{J}\|^2_{\mathbf{Q},\mathbf{R}}$ and $\|\mathbf{A}\|^2_{\mathbf{Q},\mathbf{R}}$ express a form of variation. In particular, $\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}}$ expresses the total squared variation of the generalized joint matrix decomposition. If we write $\mathbf{X}$ and $\mathbf{J}$ in their concatenated forms $(\mathbf{X}_1 \ \ldots \ \mathbf{X}_K)$ and $(\mathbf{J}_1 \ \ldots \ \mathbf{J}_K)^\intercal$ respectively, we get

$$\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}} = \operatorname{tr}(\mathbf{QXRX}^\intercal) = \sum_{k=1}^{K} \operatorname{tr}(\mathbf{QX}_k\mathbf{R}_k\mathbf{X}_k^\intercal) = \sum_{k=1}^{K} \|\mathbf{X}_k\|^2_{\mathbf{Q},\mathbf{R}_k},$$
$$\|\mathbf{J}\|^2_{\mathbf{Q},\mathbf{R}} = \operatorname{tr}(\mathbf{QJRJ}^\intercal) = \sum_{k=1}^{K} \operatorname{tr}(\mathbf{QJ}_k\mathbf{R}_k\mathbf{J}_k^\intercal) = \sum_{k=1}^{K} \|\mathbf{J}_k\|^2_{\mathbf{Q},\mathbf{R}_k}$$

Hence,

$$\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}} = \sum_{k=1}^{K} \|\mathbf{X}_k\|^2_{\mathbf{Q},\mathbf{R}_k} = \sum_{k=1}^{K} \left(\|\mathbf{J}_k\|^2_{\mathbf{Q},\mathbf{R}_k} + \|\mathbf{A}_k\|^2_{\mathbf{Q},\mathbf{R}_k}\right).$$

From this we conclude that $\|\mathbf{X}_k\|^2_{\mathbf{Q},\mathbf{R}_k}$ expresses the squared variation of the k-th dataset. Finally, since

$$\sum_{k=1}^{K} \|\mathbf{J}_k\|^2_{\mathbf{Q},\mathbf{R}_k} = \operatorname{tr}(\mathbf{D}^2) \quad \text{and} \quad \sum_{k=1}^{K} \|\mathbf{A}_k\|^2_{\mathbf{Q},\mathbf{R}_k} = \sum_{k=1}^{K} \operatorname{tr}(\mathbf{D}_k^2),$$

we can further decompose the squared variation of $\|\mathbf{X}_k\|^2_{\mathbf{Q},\mathbf{R}_k}$ into the squared variation $\|\mathbf{J}_k\|^2_{\mathbf{Q},\mathbf{R}_k}$ explained by the joint component and the squared variation $\|\mathbf{A}_k\|^2_{\mathbf{Q},\mathbf{R}_k}$ explained by the individual component. In summary, the statistics that we are interested in are

$$\frac{\|\mathbf{J}_k\|^2_{\mathbf{Q},\mathbf{R}_k}}{\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}}} \quad \text{and} \quad \frac{\|\mathbf{A}_k\|^2_{\mathbf{Q},\mathbf{R}_k}}{\|\mathbf{X}\|^2_{\mathbf{Q},\mathbf{R}}}.$$

9 Supplementary

Theorem 2. *1) If* $\mathbf{U}^\intercal\mathbf{Q}\mathbf{U}_i = \mathbf{0}$*, then* $\mathbf{J}^\intercal\mathbf{Q}\mathbf{A}_i = \mathbf{0}$*. 2) For* $i \neq j$*, if* $\mathbf{U}_i^\intercal\mathbf{Q}\mathbf{U}_j = \mathbf{0}$*, then* $\mathbf{U}_i^\intercal\mathbf{Q}\mathbf{U}_j = \mathbf{0}$*.*

Proof. We will only prove 1 because 2 is proved similarly.

$$\mathbf{J}^\intercal\mathbf{Q}\mathbf{A}_i = \mathbf{V}\mathbf{D}\mathbf{U}^\intercal\mathbf{Q}\mathbf{U}_i\mathbf{D}_i\mathbf{W}_i^\intercal = \mathbf{0}.$$

□

Corollary 1. *Let* $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ *be an orthonormal basis of* $\mathbb{R}^n$ *and* $\mathbf{U} = \mathbf{Q}^{-1/2}(\mathbf{e}_{l_1} \ldots \mathbf{e}_{l_s})$*,* $\mathbf{U}_i = \mathbf{Q}^{-1/2}(\mathbf{e}_{m_1} \ldots \mathbf{e}_{m_{r_i}})$*,* $\mathbf{U}_j = \mathbf{Q}^{-1/2}(\mathbf{e}_{h_1} \ldots \mathbf{e}_{h_{r_j}})$*. If*

1. $\{l_1, \ldots, l_s\} \cap \{m_1, \ldots, m_{r_i}\} = \emptyset$*, then* $\mathbf{J}^\intercal\mathbf{Q}\mathbf{A}_i = \mathbf{0}$*,*
2. $\{m_1, \ldots, m_{r_i}\} \cap \{h_1, \ldots, h_{r_j}\} = \emptyset$*, then* $\mathbf{U}_i^\intercal\mathbf{Q}\mathbf{U}_j = \mathbf{0}$*.*

Proof. We will only prove 1. By Theorem 2, if $\mathbf{U}^\intercal\mathbf{Q}\mathbf{U}_i = \mathbf{0}$, then $\mathbf{J}^\intercal\mathbf{Q}\mathbf{A}_i = \mathbf{0}$. If $\{l_1, \ldots, l_s\} \cap \{m_1, \ldots, m_{r_i}\} = \emptyset$, then

$$\mathbf{U}^\intercal\mathbf{Q}\mathbf{U}_i = \begin{pmatrix} \mathbf{e}_{l_1}^\intercal \\ \vdots \\ \mathbf{e}_{l_s}^\intercal \end{pmatrix} (\mathbf{e}_{m_1} \ldots \mathbf{e}_{m_{r_i}}) = \begin{pmatrix} \mathbf{e}_{l_1}^\intercal\mathbf{e}_{m_1} & \ldots & \mathbf{e}_{l_1}^\intercal\mathbf{e}_{m_{r_i}} \\ \vdots & \ddots & \vdots \\ \mathbf{e}_{l_s}\mathbf{e}_{m_1} & \ldots & \mathbf{e}_{l_s}\mathbf{e}_{m_{r_i}} \end{pmatrix} = \mathbf{0}.$$

□

Theorem 3. *Assume that* $\bigcap_{k=1}^K \mathrm{col}(\mathbf{X}_k) \neq \{\mathbf{0}\}$*. Then, there exist* $\mathbf{J}_k, \mathbf{A}_k$ *such that*

1. $\mathbf{X}_k = \mathbf{J}_k + \mathbf{A}_k$*,*
2. $\mathrm{col}(\mathbf{J}_k) = \mathrm{col}(\mathbf{J})$*,*
3. $\mathrm{col}(\mathbf{J}_k) \perp \mathrm{col}(\mathbf{A}_k)$*,*
4. $\bigcap_{k=1}^K \mathrm{col}(\mathbf{A}_k) = \{\mathbf{0}\}$*.*

Proof. 1. Define

$$\mathrm{col}(\mathbf{J}) = \bigcap_{k=1}^{K} \mathrm{col}(\mathbf{X}_k) \subset \mathrm{col}(\mathbf{X}_k) \subset \mathbb{R}^n.$$

Then, there exists $\mathbf{V} \in \mathbf{R}^{n\times r}$ such that $\mathbf{V}^\intercal \mathbf{V} = \mathbf{I}_r$ and $\mathrm{col}(\mathbf{J}) = \mathrm{span}(V)$. Define the projection matrix $\mathbf{P} = \mathbf{V}\mathbf{V}^\intercal \in \mathbb{R}^{n\times n}$. Then, we can write

$$\mathbf{X}_k = \mathbf{P}\mathbf{X}_k + (\mathbf{I} - \mathbf{P})\mathbf{X}_k.$$

Define $\mathbf{J}_k = \mathbf{P}\mathbf{X}_k$ and $\mathbf{A}_k = (\mathbf{I} - \mathbf{P})\mathbf{X}_k$. Then,

$$\begin{aligned}
\mathrm{col}(\mathbf{X}_k) &= (\mathrm{col}(\mathbf{X}_k) \cap \mathrm{ran}(\mathbf{P})) \oplus (\mathrm{col}(\mathbf{X}_k) \cap \mathrm{ran}(\mathbf{I} - \mathbf{P})) \\
&= \mathrm{col}(\mathbf{J}) \oplus (\mathrm{col}(\mathbf{X}_k) \cap \mathrm{ran}(\mathbf{I} - \mathbf{P})) \\
&.
\end{aligned}$$

Thus, $\mathrm{col}(\mathbf{J}_k) = \mathrm{col}(\mathbf{J})$ and $\mathrm{col}(\mathbf{A}_k) = \mathrm{col}(\mathbf{X}_k) \cap \mathrm{ran}(\mathbf{I} - \mathbf{P})$.

In addition,

$$\mathrm{col}(\mathbf{J}_k) = \mathrm{ran}(\mathbf{P}) \text{ and } \mathrm{col}(\mathbf{A}_k) \subset \mathrm{ran}(\mathbf{I} - \mathbf{P}) \Rightarrow \mathrm{col}(\mathbf{J}_k) \perp \mathrm{col}(\mathbf{A}_k).$$

Finally,

$$\begin{aligned}
\bigcap_{k=1}^{K} (\mathrm{col}(\mathbf{X}_k) \cap \mathrm{ran}(\mathbf{I} - \mathbf{P})) &= \left(\bigcap_{k=1}^{K} \mathrm{col}(\mathbf{X}_k) \cap \mathrm{ran}(\mathbf{I} - \mathbf{P}) \right) \\
&= \mathrm{ran}(\mathbf{P}) \cap \mathrm{ran}(\mathbf{I} - \mathbf{P}) \\
&= \{\mathbf{0}\}.
\end{aligned}$$

$\square$

References