Integrative decomposition of multi-source data by identifying partially-joint score subspaces

SeoWon Choi\textsuperscript{a}, Sungkyu Jung\textsuperscript{a,1}

\textsuperscript{a}Department of Statistics, Seoul National University, Seoul, Republic of Korea

Abstract

Analysis of multi-source dataset, where data on the same objects are collected from multiple sources, is of rising importance in many fields, most notably in multi-omics biology. A novel framework and algorithms for integrative decomposition of such multi-source data are proposed to identify and sort out common factor scores in terms of whether the scores are relevant to all data sources (fully joint), to some data sources (partially joint), or to a single data source. The key difference between the proposed method and existing approaches is that raw source-wise factor score subspaces are utilized in the identification of the partially-joint block-wise association structure.

To identify common score subspaces, which may be partially joint to some of data sources, from noisy observations, the proposed algorithm sequentially computes one-dimensional flag means among source-wise score subspaces, then collects the subspaces that are close to the mean. The proposed decomposition boasts fast computational speed, and is superior in identifying the true partially-joint association structure and recovering the joint loading and score subspaces than competing approaches. The proposed decomposition is applied to a blood cancer multi-omics data set, containing measurements from three data sources. Our method identifies a latent score, partially joint to the drug panel and methylation profile data sources but not relevant to RNA sequencing profiles, which helps discovering hidden clusters in the data.

Keywords: Multi-block data, Factor model, Principal angles, Data integration, Dimension reduction

Email addresses: chsw1@snu.ac.kr (SeoWon Choi), sungkyu@snu.ac.kr (Sungkyu Jung)

\textsuperscript{1}This work was supported by Samsung Science and Technology Foundation under Project Number SSTF-BA2002-03.
1. Introduction

In various fields of science and technology, there is a growing interest in analyzing multi-source data in an integrative way. By the multi-source data, we mean data obtained for multiple groups of variables on the same set of subjects. Each group of variables is observed from a common source, and form a data block. A prominent example of multi-source data is modern multi-omics data that include gene expressions, RNA sequencing, mutations, epigenetic markers or metagenomic materials (Subramanian et al., 2020). The recent development of high-throughput technologies enables us to extract these sources of information comprehensively from a given preparation of cancer/normal tissue samples (Reuter et al., 2015; Norris et al., 2017).

One of the main challenges in analyzing multi-omics data is that data blocks come from distinct measurements of different sources. For example, our motivating data set consists of three data blocks, each from a drug response panel, genome-wide DNA methylation profiles and RNA sequencing profiles (Dietrich et al., 2018); see Section 6 for a detailed description of the data. Separately analyzing each data block hinders the assessment of inter-relations among different data blocks. To capture the potentially joint association structures in these multi block data, the linked component model has been oftentimes used (Smilde et al., 2003; VanDeun et al., 2009; Lock et al., 2013).

Suppose that there are $K$ data blocks, $X_k \in \mathbb{R}^{p_k \times n}$, for $k = 1, \ldots, K$, observed for $n$ subjects. The subjects in the data are common and matched, i.e., the measurements for the $i$th subject appear in the $i$th column of each data matrix. Assuming zero mean, each data block is decomposed into

$$X_k = Z_k + E_k = U_k V_k^T + E_k, \quad k = 1, \ldots, K,$$

where the low-rank “signal” matrix $Z_k$ is factored into a loading matrix $U_k$ and a score matrix $V_k$, which is perturbed by “noise” matrix $E_k$. The linked component model further assumes that two or more data blocks can potentially share a common score component (VanDeun et al., 2009). An extreme example is that all scores are common to all data blocks, that is, $V_1 = \cdots = V_k$, as in Smilde et al. (2003). The models considered in Lock et al. (2013) and Feng et al. (2018) allow some scores to be common to all data blocks, explaining the joint variation among all data blocks, and some
scores to be specific to each data blocks, explaining the individual variations in a single data block.

Following Gaynanova and Li (2019) and Gao et al. (2021), we in addition allow partially-joint scores that are shared across multiple, but not necessarily all, data blocks. An illustrative example of such a model, for $K = 3$ blocks of data, is

$$(X_1^T, X_2^T, X_3^T)^T = UW^T + E,$$

where $U$ is the loading matrix, $W_{(1,2,3)}$ is a matrix of scores that affect all data blocks, and $W_{(1,2)}$ is a partially-joint score matrix, affecting only the first two data blocks, but not the third. The scores in $W_{3}$ are specific to the third data block. Our goal is to delineate such an association structure from a multi-source data.

In this paper, we develop a novel framework and estimation strategy for integrative decomposition of multi-source data, by identifying scores that are fully joint to all data sources, partially joint to some, or specific to a single data source. The framework utilizes the signal score subspace $V_k$, a rank $r_k$ subspace of $\mathbb{R}^n$, spanned by the rows of the signal matrix $Z_k$ in (1). The rationale for using $[V_k]$ is straightforward: If a common score of rank $r$ is shared by the first two data blocks, then the intersection of $[V_1]$ and $[V_2]$ is non-empty, and is a dimension $r$ subspace. To capture and summarize the partially-joint block-wise association structure from the score subspaces $[V_k]$, we define the collection of tuples

$$S := \mathcal{S}([V_k]) = \{(S, r(S)) : S \subset \mathcal{K}\},$$

in which a non-empty subset $S$ of $\mathcal{K} = \{1, \ldots, K\}$ denotes a specific pattern of block-wise association, and $r(S)$ denotes the dimension of common scores corresponding to each association pattern $S$. Note that a pattern $S = \{1, 2\}$ (for instance) represents that there are common scores of dimension $r(S)$ partially joint to the first and second data sources. See Section 2 for a detailed construction of $\mathcal{S}$, and Section 4 for conditions to guarantee the uniqueness of $\mathcal{S}$.

To identify the partially-joint score structure $\mathcal{S}$ from noisy observations, we propose to compute the flag mean (Draper et al., 2014) of signal score subspaces $[V_k]$’s, and to test whether the mean is indeed close to each signal
score subspace. If a mean \( w \) is “close” to \([V_1]\) and \([V_2]\), but not to other subspaces, then it becomes a basis of (estimated) \( W_{(1,2)} \). We use a tunable parameter to determine the closeness between two subspaces. The algorithm, detailed in Section 3, is quick in decomposing multi-source data sets, and boasts a superior performance in identifying \( \mathcal{G} \) and in the estimation of the subspaces spanned by the loading matrices and by the common score matrices.

Recently, there has been a growing interest in integrative decomposition of multi-source data (Lock et al., 2013; Li and Jung, 2017; Feng et al., 2018; Li and Gaynanova, 2018; Gaynanova and Li, 2019; Park and Lock, 2020; Gao et al., 2021; Lock et al., 2022). Among these, Gaynanova and Li (2019) and Gao et al. (2021) also considered modeling partially-joint association structures. However, these authors focused on the loading matrix \( U \), and exploited the source-wise sparse structure of the matrix \( U \), as seen in (2). On the contrary, we explicitly utilize the signal score matrices \( V_k \) in (1) in identifying partially-joint scores of (2). The loading matrix is naturally obtained as a subsequent step in our proposal. Our approach extends the angle-based joint and individual variation identification of Feng et al. (2018), in which partially joint variations were not considered. In Section 5, we numerically confirm that our proposal finds the true association structure more accurately than the aforementioned decomposition methods.

The rest of article is organized as follows. In Section 2, we formally present our integrative decomposition framework and define the partially joint structure \( \mathcal{G} \), followed by the proposed decomposition algorithm and tuning parameter selection procedure in Section 3. In Section 4, conditions to guarantee unique decompositions are discussed. Section 5 is devoted to numerical illustrations and comparisons to existing methods. In Section 6, we demonstrate the use of the proposed method in an analysis of a blood cancer multi-omics data set associated with drug responses, and reveal that the proposal detects a latent signal pattern, partially joint across two, but not all, data sources, which is not easily seen by a separate analysis of individual data blocks. Proofs, technical lemmas, examples and extended numerical results are given in the supplementary material. R codes and data to reproduce the numerical results in this article can be found at https://github.com/sungkyujung/psi.
2. Statistical framework

2.1. Partially-joint structures

Consider a set of row-centered data matrices \( X_k \in \mathbb{R}^{p_k \times n} \) for \( k = 1, \ldots, K \), where the \( i \)th column of each data matrix corresponds to the same (\( i \)th) subject. Let \( p = \sum_k p_k \). We assume that each \( X_k \) is additively decomposed into the rank \( r_k \) true signal block \( Z_k \) and random errors \( E_k \) so that \( X_k = Z_k + E_k \).

We consider a model in which the concatenated signal matrix \( Z = (Z_1^T, \ldots, Z_K^T)^T \) is decomposed into

\[
Z = UW^T,
\]

where each column of \( W \) (corresponding to a latent score) has Euclidean norm 1, but not necessarily orthonormal. We require both \( U \) and \( W \) to have full column rank \( r \), and \( r < \min(n, p) \).

Our goal is to find a block-wise association structure among \( Z_k \)'s, in which a latent score (a column of \( W \)) is linked to multiple, but not necessarily all of, \( Z_k \)'s. This “partial-jointness” of common scores is imposed by assuming block-sparsity on \( U \). An example of such model for \( K = 3 \) blocks of data sets is

\[
\begin{pmatrix}
Z_1 \\
Z_2 \\
Z_3
\end{pmatrix} = \begin{pmatrix}
U_{(1),1} & U_{(1),2} & 0 \\
U_{(2),1} & U_{(2),2} & 0 \\
U_{(3),1} & 0 & U_{(3),3}
\end{pmatrix}
\begin{pmatrix}
W_1 \\
W_2 \\
W_3
\end{pmatrix}^T,
\]

where the blocks \( U_{(1),3}, U_{(2),3} \) and \( U_{(3),2} \) are exactly zero. In this case, the scores in \( W_1 \) are shared by all signal blocks \( Z_1, Z_2 \) and \( Z_3 \) (thus are related to all data sources), whereas \( W_2 \) is shared only by \( Z_1 \) and \( Z_2 \) (that is, partially joint to data sources), and \( W_3 \) is specific to \( Z_3 \).

Each block-wise association pattern can be represented by the set of data-source indices, or an index-set, a subset of \( \{1, \ldots, K\} \), denoted generally by \( S \). For each index-set \( S \), the number of columns of \( W \) associated with \( S \) is called the rank of \( S \) and is denoted by \( r(S) \). We summarize the block-sparsity via a partially-joint structure, which is the collection of pairs \((S, r(S))\) (with \( r(S) > 0 \)). In our example, if the number of columns of \( W_1, W_2 \) and \( W_3 \) are 2, 1 and 1, respectively, then the partially-joint structure \( \mathcal{S} \) corresponding to (3) is \(((\{1, 2, 3\}, 2), (\{1, 2\}, 1), (\{3\}, 1))\).

As can be seen in the example (3) of the decomposition \( Z = UW^T \), for each index-set \( S \), the columns of \( W \) associated with \( S \) are concatenate into
an $n \times r(S)$ matrix. These blocks of $W$ are called partially-joint scores. The columns of $U$ associated with $S$ are also concatenated into a $p \times r(S)$ matrix, which can be split into block matrices of $p_k \times r(S)$ corresponding to each $Z_k$. We call these block matrices partially-joint loadings. If $k \notin S$, then the partially-joint loading matrix corresponding to $Z_k$ is zero. In our example, $W_1, W_2$ and $W_3$ are partially-joint scores and $U^{(1,1)}, U^{(2,1)}, U^{(3,1)}, U^{(2,1)}$ and $U^{(2,2)}$ are partially-joint loadings.

For given $Z_k$’s, our aim is to identify partially-joint loadings $U$ and scores $W$. Only with the block-sparsity constraint, there may be multiple candidates of both $U$ and $W$. In the next subsection, we present our strategy of identifying $U$ and $W$ by finding the column space of each $W_i$.

Remark 1. The decomposition of $Z_k$’s into the partially-joint loading and score matrices is not unique. If instead of $W_i$ (and its corresponding loading matrix $U^{(k),i}$), one chooses $W_i' = W_i R$, for an orthogonal matrix $R$, then the corresponding loading matrix becomes $U'_{(k),i} = U_{(k),i} R$. Nevertheless, if $U_{(k),i}$ is zero, then $U_{(k),i} R$ is zero for any orthogonal $R$, so the block-sparsity of $U$ is invariant to the choices of basis.

2.2. Partially-joint score subspace and index-set ordering

We now define the partially-joint score subspace $[W_i]$, a linear subspace spanned by the columns of a partially-joint score $W_i$, in a constructive manner. The definition depends mainly on the two concepts: the signal score subspace of signal block $Z_k$ and the ordering of index-sets.

Let the true signal matrix be further decomposed into $Z_k = U_k V_k$, as in the factor analysis, where $V_k \in \mathbb{R}^{n \times r_k}$ is the factor score matrix satisfying $V_k^T V_k = I_{r_k}$. We call the column space of $V_k$, denoted by $[V_k]$, as the signal score subspace for $Z_k$. Note that $[V_k]$ is a rank-$r_k$ subspace of $\mathbb{R}^n$, spanned by the columns of $V_k$.

A geometric implication of using the signal score subspaces $[V_k]$’s in defining partially-joint scores $W_i$’s is as follows: Take an index set $S = \{1, 2\} \subset \{1, \ldots, K\}$ as an example. If $[V_1]$ and $[V_2]$ overlap, then the corresponding signal blocks $Z_1$ and $Z_2$ share a common score, say $W_{\text{cand}}$, represented by the subspace $[W_{\text{cand}}] := [V_1] \cap [V_2]$. Such a common score $W_{\text{cand}}$ is at least partially-joint to $Z_1$ and $Z_2$ (and may be related to all or some of $Z_3, \ldots, Z_K$). As we have assumed that signal blocks $Z_k$’s share the corresponding partially-joint score, the intersections $\cap_{k \in S} [V_k]$ of signal score subspaces play an important role, as a building block, in identifying partially-joint score subspaces $[W_i]$. 

6
However, one of the challenges of such an approach is that index-sets, as subsets of \( \{1, \ldots, K\} \), have a partially-ordered relation with respect to set-inclusion. As two index-sets can be either nested or intersected, the construction of a partially-joint score subspace corresponding to an index-set may affect other partially-joint score subspaces. Thus we suggest a plausible way of constructing partially-joint score subspaces by specifying an order among index-sets.

We give an ordering to index-sets, sorted by the number of elements in descending order. The first index-set is \( S_1 = \{1, \ldots, K\} \), the largest set (corresponding the fully-joint score subspace), and each of the next \( K \) index-sets \( S_2, \ldots, S_{K+1} \) contains exactly \( K - 1 \) indices and so on. (There are \( 2^K - 1 \) different index-sets.) Note that index-set ordering is not unique for given \( K \), since there can be permutations among index-sets of the same size. Conditions to guarantee invariance of the decomposition with respect to different choices of indexing will be discussed in Section 4.

**Example 2.1.** For \( K = 3 \), \( S_1 = \{1, 2, 3\} \) as appeared in the top row of Fig. 1 and the next three sets \( (S_2, S_3, S_4) \) are given by the next row of the figure. That is, \( S_2, S_3, \) and \( S_4 \) are \( \{1, 2\}, \{2, 3\}, \{1, 3\} \), respectively. Likewise, \( S_5 = \{1\}, S_6 = \{2\}, S_7 = \{3\} \). Elements in \( S_i \) stands for the indices of data blocks that potentially possess a common score. Such an ordering is not unique, and one may set, e.g., \( S_6 = \{3\}, S_7 = \{2\} \).

Using the signal score subspaces \( [V_k] \)'s and a pre-specified index-set ordering, we now define partially-joint score subspaces.
Definition 1. Suppose that data matrices $X_k = Z_k + E_k \in \mathbb{R}^{p_k \times n}$ for $k = 1, \ldots, K$ with true signal block $Z_k$ are given, in which the data matrices are matched, i.e., the $j$th row of each $X_k$ represents the measurements for the $j$th individual. With an index-set ordering $S_1, \ldots, S_{2^K - 1}$, a set of corresponding partially-joint score subspace $[W_i]$’s for $i = 1, \ldots, 2^K - 1$ are defined sequentially:

$$[W_1] := \bigcap_{k \in S_1} [V_k],$$
$$[W_2] := P_1^\perp \left( \bigcap_{k \in S_1} [V_k] \right),$$
$$\vdots$$
$$[W_i] := \left( \bigcap_{j : j < i, S_j \cap S_i \neq \emptyset} P_j^\perp \right) \left( \bigcap_{k \in S_i} [V_k] \right).$$

Here, $P_j^\perp$ is the projection transformation of $\mathbb{R}^n$ onto the orthogonal complement of $[W_j]$. The notation $\bigcap_{j \in J} P_j^\perp ([A])$, for a set $J \subset \{1, \ldots, 2^K - 1\}$ and a subspace $[A]$, stands for the repeated applications of $P_j^\perp$ on $[A]$, where $P_j^\perp$ are applied one by one by the increasing order of $j \in J$.

Note that depending on $\{[V_k]\}$, a $[W_i]$ may be $\{0\}$. (As a convention, $\text{rank}\{0\} = 0$.) The block-wise association structure $\mathcal{S}$ for the matched data $X_k$’s is then the collection of $(S_i, r(S_i))$ for which $r(S_i) = \text{rank}([W_i]) > 0$.

Example 2.2. Suppose $K = 3$ and the index-set ordering is given as in Example 2.1, with $r(S_1) = r(S_2) = r(S_4) = r(S_7) = 1$, and $r(S_i) = 0$ otherwise. Partially-joint score subspaces are obtained as $W_1 = \bigcap_{k \in S_1} [V_k]$, $W_2 = P_1^\perp \left( \bigcap_{k \in S_2} [V_k] \right)$, $W_4 = P_2^\perp \circ P_1^\perp \left( \bigcap_{k \in S_4} [V_k] \right)$ and $W_7 = P_4^\perp \circ P_1^\perp \left( \bigcap_{k \in S_7} [V_k] \right)$.

Partially-joint score subspaces have the following basic properties.

1. The partially-joint score subspaces involved with each $Z_k$ can be used to restore the signal score subspace. That is,

$$[V_k] = +_{i \in \{k \in S_i\}} [W_i],$$

as shown in Lemma A.3 in the supplementary material. Here the notation ‘+’ means the sum of subspaces.

2. For any $i \neq j$, $[W_i]$ and $[W_j]$ do not overlap, i.e., $[W_i] \cap [W_j] = \{0\}$, but they are not necessarily orthogonal. However, if both $[W_i]$ and $[W_j]$ are related to a common data source, then they are orthogonal, as shown in the following lemma.
Lemma 2.1. For $i, j \in \{1, \ldots, 2^K - 1\}$ and $S_i \cap S_j \neq \emptyset$, $[W_i] \perp [W_j]$.

Lemma 2.1 ensures that all partially-joint score subspaces relevant to the $k$th data block, $\{[W_i] : k \in S_i, i = 1, \ldots, 2^K - 1\}$, are orthogonal to each other. Thus, the subspaces in the right hand side of (4) are indeed orthogonal to each other.

2.3. Partially-joint score and loading

Given $S$, choose an orthogonal basis $W_i \in \mathbb{R}^{n \times r(S_i)}$ of $[W_i]$, which becomes a partially joint score. Let $W(k)$ be the column-wise concatenation of $W_i$'s for $i \in \{i : k \in S_i\}$. That is, $W(k)$ collects the partially-joint scores corresponding to the $k$th data source. By Lemma 2.1, the matrix $W(k)$ is an orthogonal matrix. The partially-joint loadings corresponding to $Z_k$ are then given by

$$U(k) = Z_k W(k),$$

where $U(k)$ is the column-wise concatenation of $U_i$'s for $i \in \{i : k \in S_i\}$. Indeed we can easily check that $U(k) W_T(k) = Z_k W(k) W_T(k) = Z_k$, since $W(k) W_T(k)$ is the projection matrix onto $[V_k]$ by the property (1) above. The partially-joint loading $U(k,i)$ for $i \in \{i : k \in S_i\}$ is the $(k,i)$th block of $U$. For $i \not\in \{i : k \in S_i\}, U(k,i) = 0$.

3. Estimation

In practice, the signal $Z_k$ and error $E_k$ of row-centered data matrices $X_k \in \mathbb{R}^{n \times n}$ for $k = 1, \ldots, K$ are unknown. We assume that the ranks $r_k$ of the signal matrices are pre-determined and extract the signal matrix using a low-rank approximation of $X_k$, and write the rank $r_k$ approximation of $X_k$ by $\hat{Z}_k$. The basis of the (empirical) signal score subspace $\hat{V}_k$ is given either by the right singular vectors of $\hat{Z}_k$ from or by any off-the-shelf factor model estimates. We use the singular value decomposition (SVD) for our numerical examples.

An overview of the estimation algorithm is as follows: (1) on the outer loop, we iterate through all index-sets $S_i$ on a given ordering, and (2) on the inner loop, we obtain the partially-joint score subspace estimate $[\hat{W}_i]$ from $\{[\hat{V}_k]\}_{k \in S_i}$, identifying the basis of $[\hat{W}_i]$ one-by-one.

A major challenge in doing so is that we only have the sample versions of signal blocks, $\hat{Z}_k$, and thus the available sample signal score subspaces $[\hat{V}_k]$
is a perturbed version of the true signal score subspaces \([V_k]\). That is, even if \([V_k]\)’s overlap (i.e., \(\cap_{k \in S_i}[V_k] \neq \{0\}\)) the sample version may not overlap: \(\cap_{k \in S_i}[\hat{V}_k] = \{0\}\). Thus, there is a need to give a slack on identifying the “intersection” of \([V_k]\)’s, accounting for random perturbations in \([\hat{V}_k]\). We propose to use principal angles between subspaces for such identification, further developed in Section 3.1.

In Section 3.1, we propose a procedure to estimate a basis matrix \(\hat{W}_i\), which gives the partially-joint score subspace estimate \(\hat{W}_i\) at each \(i\)th step. The rank of the corresponding \(S_i\) is set as \(\hat{r}(S_i) = \text{rank}(\hat{W}_i)\). Finishing the iteration over all \(S_i\)’s, an estimated partially-joint structure \(\hat{S} = \{(S_i, \hat{r}(S_i))\}\) and a set of corresponding \(\hat{W}_i\)’s are obtained. In Section 3.2, we then discuss the estimation of corresponding partially-joint loading matrices from \(\hat{S}\) and \(\hat{W}_i\)’s.

### 3.1. Partially-joint Score Subspace Estimation

Our strategy for estimating partially-joint score subspace \(\hat{W}_i\), for an index-set \(S_i\), is by collecting one-dimensional bases, that lie “close” to each and every score subspaces in \(\{[\hat{V}_k]\}_{k \in S_i}\).

For the notion of closeness, the principal angle between subspaces is used, and we will determine that two subspaces are close if the principal angle is smaller than a threshold \(\lambda \in [0, \pi/2)\). For now, \(\lambda\) is treated as a predetermined tuning parameter, and a data-driven choice of \(\lambda\) will be discussed in Section 3.3.

For a given \(\lambda \in [0, \pi/2)\), we estimate \(\hat{W}_i\) sequentially for \(i = 1, \ldots, 2^K - 1\), using the estimation algorithm presented below. At the \(i\)th stage, we begin with \(W_i = \emptyset\), to which the identified one-dimensional bases of the partially-joint score subspace are added. Algorithm 1 summarizes the proposed partially-joint structure identification procedure, using steps (a)—(c).

**Step (a):** As a candidate that may be included in \(\hat{W}_i\), compute the mean direction \(\hat{w}\) among \(\{[\hat{V}_k]\}_{k \in S_i}\), provided that \([\hat{V}_k] \neq \{0\}\) for all \(k \in S_i\). The mean direction minimizes the sum of the squares of the subspace distances between a candidate \(w\) and subspaces \([\hat{V}_k]\) for \(k \in S_i\), and is

\[
\hat{w} = \arg\min_w \sum_{k \in S_i} \left(\langle w, [\hat{V}_k]\rangle\right)^2, 
\]

(5)
Algorithm 1: Partially-joint Structure Identification

input: $\hat{V}_1, \ldots, \hat{V}_K, S_1, \ldots, S_{2^K-1}, \lambda$

for $i = 1, 2, \ldots, 2^K - 1$ do

Set $W_i = \emptyset$;

while dim($\hat{V}_k$) > 0 for all $k \in S_i$ do

(a) Compute the mean direction $\hat{w}$ of $\{\hat{V}_k\}_{k \in S_i}$, see (5);

if the condition (6) is satisfied then

(b) Let $W_i \leftarrow W_i \cup \{\hat{w}\}$;

(c) Update $\hat{V}_k \leftarrow \hat{V}_k,\text{trunc}$ for each $k \in S_i$

else

break;

end

end

Let $\hat{r}(S_i) = |W_i|$, write $\hat{W}_i$ for the $n \times \hat{r}(S_i)$ matrix consisting of elements in $W_i$ and record $(\hat{S}_i, \hat{r}(S_i), \hat{W}_i)$ in $\hat{S}$;

end

Result: $\hat{S}$

where $d([w], [B]) = 1/\sqrt{2} \cdot \|ww^T - BB^T\|_F$ is the Frobenius-norm distance between subspaces $[w]$ and $[B]$ (Ye and Lim, 2016). We chose the Frobenius norm, since for any choice of the basis $\hat{V}_k$ for $[\hat{V}_k]$,

$$\sum_{k \in S_i} d([w], [\hat{V}_k])^2 = |S_i| - w^T \left( \sum_{k \in S_i} \hat{V}_k \hat{V}_k^T \right) w = |S_i| - w^T (HH^T) w,$$

where $H$ is the matrix given by the column-wise binding of $\hat{V}_k$’s, and $\hat{w}$ is the first left singular vector of $H$. We mention in passing that $[\hat{w}]$ is the one-dimensional flag mean of subspaces $\{[\hat{V}_k]\}_{k \in S_i}$ (Draper et al., 2014).

Step (b): Check whether all of the signal score subspaces $[\hat{V}_k]$ are not too dispersed from $\hat{w}$ in (5). For this, given the prespecified $\lambda$, we check whether the principal angle between the mean direction and each of $[\hat{V}_k]$ is at most $\lambda$, i.e.,

$$\theta([\hat{w}], [\hat{V}_k]) < \lambda, \quad \text{for all } k \in S_i. \tag{6}$$

Here, the principal angle $\theta([w], [B]) := \arcsin(d([w], [B])) \in [0, \pi/2]$ is the
acute angle formed by the vector \( w \) and the subspace \([B]\) [Björck and Golub, 1973]. If the condition (6) is not satisfied, then skip the following and move to the next stage for \( S_{i+1} \). If (6) is satisfied, then \( W_i \) is updated to \( W_i \leftarrow W_i \cup \{\hat{w}\} \).

Step (c): Deflate each \([\hat{V}_k]\) for \( k \in S_i \), so that the next mean direction is orthogonal to previous ones. In particular, we “peel” the basis \( \hat{w} \), that has been added to \( W_i \), from each of \([\hat{V}_k]\), \( k \in S_i \). Since \( \hat{w} \) is not exactly in \([\hat{V}_k]\), the one-dimensional subspace closest to \( \hat{w} \) is removed. Specifically, let \([\hat{V}_{k,\text{trunc}}]\) be the orthogonal complement of \( P_{\hat{V}_k} \hat{w} \) within \([\hat{V}_k]\). (Note that \( \dim([\hat{V}_{k,\text{trunc}}]) = \dim([\hat{V}_k]) - 1 \).) We update \([\hat{V}_k]\) by \([\hat{V}_k] \leftarrow [\hat{V}_{k,\text{trunc}}]\), for each \( k \in S_i \). If any \([\hat{V}_k]\) for \( k \in S_i \) becomes \( \{0\} \), then move to the next stage for \( S_{i+1} \). Otherwise, move back to step (a).

We give an illustrative example for the three steps.

**Example 3.1.** Suppose \( n = 3 \) and \( K = 3 \). Assume that we are at the stage \( S_i = \{1, 2\} \) and \([\hat{V}_1]\) has been deflated (in the previous stages) to be of dimension 1, while \([\hat{V}_2]\) is of dimension 2, as in Fig. 3. There, \([\hat{V}_1]\) is generated by \((\cos \pi/6, 0, \sin \pi/6)^T\), and \([\hat{V}_2]\) by \((1, 0, 0)^T\) and \((0, 1, 0)^T\). In Step (a), the mean direction is \( \hat{w} = (\cos \pi/12, 0, \sin \pi/12)^T \). In Step (b), for \( k = 1, 2 \), \( \theta([\hat{w}, [\hat{V}_k]]) = \pi/12 \), (6) is satisfied for any tuning parameter \( \lambda > \pi/12 \). Thus we say that \([\hat{V}_1]\) and \([\hat{V}_2]\) share a partially-joint score subspace, and \( \hat{w} \) is included as a basis of \([W_i]\). In step (c), \([\hat{V}_2]\) is updated to \([\hat{V}_2] \leftarrow [\hat{V}_{2,\text{trunc}}] = \text{span}((0, 1, 0)^T)\), and \([\hat{V}_1]\) becomes \( \{0\} \). Since there is no more scores left to exploit, we move onto the next stage for \( S_{i+1} = \{2, 3\} \).

For a singleton set \( S_i = \{k\} \), the mean direction \( \hat{w} \) is any unit vector in (the deflated) \([\hat{V}_k]\), and the condition (6) is always satisfied. Thus, for this case, \( \hat{W}_i = \hat{V}_k \), in place of steps (a)—(c) above.

It can be seen that the estimated partially-joint score matrices \( \hat{W}_i \) consist of orthogonal columns, and \( \hat{W}_i^T \hat{W}_i = I_{r(S_i)} \), because the deflated subspaces in Step (c) above are orthogonal to \( \hat{w} \), and the next mean direction lied in the union of the deflated subspaces.

3.2. Partially-joint Loading Matrix Estimation

Given the partially-joint structure estimate \( \hat{S} = \{(S_i, \hat{r}(S_i)) : i \in I_K\} \) and the corresponding partially-joint score subspace estimates \( \hat{W}_i \)'s, we ob-
Figure 2: A figurative description for computing partially-joint score subspace $[\hat{W}_i]$, for $S_i = \{1, 2\}$. The two-dimensional subspace $[\hat{V}_2]$ is depicted as a disk. $w_{\{1,2\}}$ is the mean direction and $p_2$ stands for the projection of $w_{\{1,2\}}$ onto $[\hat{V}_2]$.

Obtain the estimated partially-joint loading matrix $\hat{U}$.

Let $\hat{Z} = (\hat{Z}_1^T, \ldots, \hat{Z}_K^T)^T \in \mathbb{R}^{p \times n}$, where $p = \sum_{k=1}^{K} p_k$. Denote the column-wise concatenation of $\hat{W}_i$ as $\hat{W} \in \mathbb{R}^{n \times \hat{r}}$, where $\hat{r} = \sum_{i \in I_K} \hat{r}(S_i)$.

We estimate the partially-joint loading matrix from the optimization problem

$$\hat{U} = \arg\min_{U \in \mathbb{R}^{p \times \hat{r}}} \| \hat{Z} - U \cdot \hat{W}^T \|_F^2.$$  \hspace{1cm} (7)

with a constraint that $U$ in (7) has the block-wise sparsity structure that corresponds to the estimated partially-joint structure $\hat{\mathcal{S}}$. For example, given $K = 2$ and $\hat{\mathcal{S}} = \{\{(1,2), 2\}, \{(1), 1\}\}$, $U$ has a block sparsity structure

$$p_1 \begin{bmatrix} U_{(1),1} & U_{(1),2} \\ \vdots & \vdots \end{bmatrix} p_2 \begin{bmatrix} U_{(2),1} \\ 0 \end{bmatrix}$$

with the number of columns of $(U_{(1),1}^T, U_{(2),1}^T)^T$ and $U_{(1),2}$ are two and one, respectively.

With the block-wise sparse constraint imposed, the objective function (7) for $U$ can be written separately for each data block, i.e.,

$$\| \hat{Z} - U \cdot \hat{W}^T \|_F^2 = \sum_{k=1}^{K} \| \hat{Z}_k - U_{(k)} \hat{W}_{(k)}^T \|_F^2.$$  \hspace{1cm} (8)
Here, $U_{(k)}$ and $\hat{W}_{(k)}$ are the column-wise concatenation of each $U_{(k),i}$'s and $\hat{W}_i$'s, respectively, for $i \in \{i : k \in S_i \text{ and } \hat{r}(S_i) > 0\}$. The minimizer of (8) is

$$\hat{U}_{(k)} = \tilde{Z}_k \hat{W}_{(k)} (\hat{W}_{(k)}^T \hat{W}_{(k)})^{-1},$$

and $\hat{U}_{(k),i}$ for $i \in J_{(k)}$ are obtained by disjoining $\hat{U}_{(k)}$. By the block-wise sparse structure, set $\hat{U}_{(k),i} = 0$ if $k \notin S_i$.

### 3.3. Tuning Parameter Selection

The partially-joint structure identification, proposed in Section 3.1, depends heavily on the tuning parameter $\lambda \in [0, \pi/2)$. If $\lambda$ is too small, then all scores are identified as individual scores, specific to each data blocks. If $\lambda$ is too large, then individual and partially-joint scores may be falsely identified as fully-joint scores.

We use data splitting to select the value of tuning parameter $\lambda \in [0, \pi/2)$. For a single instance of data splitting, randomly split $n$ samples of $X = [X_1^T, \ldots, X_K^T]^T$ into two groups of equal proportions, the training set $X_{tr} = [X_{tr,1}^T, \ldots, X_{tr,K}^T]^T$ and the test set $X_{test} = [X_{test,1}^T, \ldots, X_{test,K}^T]^T$.

Given the signal rank $r_k$ of each $X_k$, we then extract the training signal matrices $\tilde{Z}_{tr,k}$ for $k = 1, \ldots, K$ using the rank $r_k$ approximation of $X_{tr,k}$. For each $\lambda$ on the tuning parameter grid, we identify the partially-joint structure from $\tilde{Z}_{tr,k}$'s, and obtain the partially-joint score $\hat{W}_{tr,\lambda}$ and the partially-joint loading matrix $\hat{U}_{tr,\lambda}$, as discussed in Sections 3.1 and 3.2.

To assess the degrees to which the estimates are generalized to the test set, we first evaluate the score matrix for the test set, given by the loading matrix estimates $\hat{U}_{tr,\lambda}$ from the training data. The test score matrix $\hat{W}_{test,\lambda}$ is defined as the minimizer $\hat{W}_{test,\lambda} \in \mathbb{R}^{n_{test} \times \hat{r}}$ of

$$\|X_{test} - \hat{U}_{tr,\lambda} W^T\|_F^2$$

subject to $W^T W = I_{\hat{r}}$. The test score matrix is then $\hat{W}_{test,\lambda} = P_{\lambda} Q_{\lambda}^T$, where $P_{\lambda}$ and $Q_{\lambda}$ are left and right singular vector matrices of $X_{test}^T \hat{U}_{tr,\lambda} = P_{\lambda} \Sigma_{\lambda} Q_{\lambda}^T$ using the Eckart-Young theorem (Eckart and Young, 1936). We remark that the constraint $W^T W = I_{\hat{r}}$ assumes orthogonality among the scores, which may not be satisfied. Other options include imposing $W^T W = \hat{W}_{tr,\lambda}^T \hat{W}_{tr,\lambda}$, thereby requiring the correlations in the test scores to be the same as those in the training scores. Due to high variability of $\hat{W}_{tr,\lambda}^T \hat{W}_{tr,\lambda}$ over random splitting, this approach generally leads to unstable choice of $\lambda$. 

14
Let $\tilde{\lambda}$ be the value of $\lambda$ for which the minimum of an empirical risk is attained. The empirical risk, defined for $\lambda \in [0, \pi/2]$, is

$$\text{Risk}(\lambda) = \sum_{k=1}^{K} \frac{\|X_{\text{test},k} - \hat{U}_{\text{tr},\lambda,(k)} \hat{W}_{\text{test},\lambda}^T\|^2_F}{\|X_{\text{test},k}\|^2_F},$$  \hspace{1cm} (9)$$

where $\hat{U}_{\text{tr},\lambda,(k)}$ is the $k$th row block of $\hat{U}_{\text{tr},\lambda}$. A similar form was used in Gaynanova and Li (2019). The corresponding partially-joint structure is $\hat{S}_{\text{tr}}(\tilde{\lambda})$.

Note that the tuned parameter $\tilde{\lambda}$ is optimal in the sense that it minimizes (9). However, as $\tilde{\lambda}$ is tuned using only a half of the sample (with greater variability of $\hat{w}$ and $[\hat{V}_i]$), directly using $\tilde{\lambda}$ as the angle threshold (6) for the whole data (with doubled sample size) generally results in identifying too many common scores. On the other hand, the estimated structure $\hat{S}_{\text{tr}}(\tilde{\lambda})$ corresponding to the tuned parameter $\tilde{\lambda}$ does not heavily depend on the sample size. Therefore, we use the chosen structure $\hat{S}_{\text{tr}}(\tilde{\lambda})$, rather than the parameter $\tilde{\lambda}$, in determining the angle threshold used in (6) for the whole data. Let $\hat{Z}_k$ be the rank $r_k$ approximation of the whole data $X_k$. Again on the grid of $\lambda$’s, we obtain the partially-joint structure $S(\lambda)$ from $\hat{Z}_k$’s. Then we choose the best value $\hat{\lambda}$ that minimize $d(\hat{S}_{\text{tr}}(\hat{\lambda}), S(\lambda))$, where $d$ is a measure of dissimilarity between two structures defined below.

### 3.3.1. Measure of Dissimilarity

In generalizing $\hat{S}_{\text{tr}}(\hat{\lambda})$ to the whole data, we compute dissimilarity measure between $\hat{S}_{\text{tr}}(\hat{\lambda})$ and $S(\lambda)$ for each candidate $\lambda$. For this, we represent each element $(S, r(S)) \in S$ by $r(S)$ copies of a binary vector $a_S \{0, 1\}^K$ where the $i$th element of $a_S$ is 1 if $i \in S$, 0 otherwise. For example, for $K = 2$, $(S, r(S)) = (\{1, 2\}, 2)$ is represented by \{(1, 1), (1, 1)\}. Collecting all such binary vectors for each element of $S$, as an instance $\mathcal{S} = ((\{1, 2\}, 2), (\{1\}, 1))$ is represented by the multi-set \{(1, 1), (1, 1), (1, 0)\}.

We devise a measure that swiftly captures dissimilarity between two multi-sets, under a situation that we need to compute dissimilarity as many time as the number of $\hat{\lambda}$’s on the grid for each instance of splitting. Let $d_H(a, b)$ be the number of elements between column vectors $a, b \in \{0, 1\}^K$, i.e. Hamming distance, which we will use as a ground distance. The dissimilarity measure $d(\mathcal{S}, \mathcal{S}')$ between two partially-joint structure $\mathcal{S}, \mathcal{S}'$ is
defined as follows
\[
d(\mathcal{S}, \mathcal{S}') = \sum_{a \in A \setminus B} d_H^2(a, B \setminus A) + \sum_{b \in B \setminus A} d_H^2(b, A \setminus B),
\]
where \(A\) and \(B\) are the multi-sets corresponding to \(\mathcal{S}\) and \(\mathcal{S}'\), respectively. See Section B.1 of the supplementary material for an exemplary description. Such a distance between multi-sets has been appeared in the literature. In fact, \(d\) is a special version of the matching distance between multi-sets defined in Bolt et al. (2022).

Remark 2. It should be noted that \(d\) is not a metric, but a semi-metric. The conditions (1) \(d(\mathcal{A}, \mathcal{B}) = 0\) if and only if \(\mathcal{A} = \mathcal{B}\) and (2) \(d(\mathcal{A}, \mathcal{B}) = d(\mathcal{B}, \mathcal{A})\) hold, but (3) the triangular inequality does not hold in general.

4. Theory

Given an ordering of index-sets, defined in Section 2.2 and the signal blocks \(Z = \{Z_k : k = 1, \ldots, K\}\), the partially-joint structure \(\mathcal{S}(Z) = \{(S_i, r(S_i), [W_i]) : i = 1, \ldots, 2^K - 1\}\) is uniquely determined. (We abuse the notation \(\mathcal{S}(Z)\) to include the partially-joint score subspaces \([W_i]\)'s, for this section.) Unfortunately, for different orderings given to index-sets the ranks \(r(S_i)\) and the partially-joint score subspaces \([W_i]\) may be different. In this section, we introduce conditions on relations among \([V_k]\)'s for \(\mathcal{S}(Z)\) to be uniquely determined regardless of the choice of the orders of index-sets provided that they obey the partially-ordered relation with respect to set-inclusion.

Let \(\mathcal{I}_K = \{1, \ldots, 2^K - 1\}\). For each \(l = 1, \ldots, K\), let \(\mathcal{J}_l = \{i \in \mathcal{I}_K : |S_i| = l\}\) be the set of all indices with size \(l\). For example, \(\mathcal{J}_l\) corresponds to the \(l\)th row of Figure 1. For \(l = 1, \ldots, K - 1\), we set \([I_i] = +_{i \in \{i : |S_i| > l\}} (\cap_{k \in S_i} [V_k]) = +_{i \in \{i : |S_i| > l\}} [W_i]\), and the projection onto \([I_i] \perp\) in \(\mathbb{R}^n\) is denoted by \(P_i^\perp\). Note that when evaluating the rank for \([W_i]\) for \(i \in \mathcal{J}_l\), the definition of \([W_i]\) in Definition 1 utilizes the deflated score subspaces orthogonal to \([I_i]\), using the given order of \(S_i\)'s among \(i \in \mathcal{J}_l\).

Let \(\mathcal{K} = \{1, \ldots, K\}\). We say \(\{[V_k]\}_{k \in \mathcal{K}}\) to be relatively independent if, for every \(l = 1, \ldots, K - 1\) and \(i \in \mathcal{J}_l\), \(P_i^\perp (\cap_{k \in S_i} [V_k])\) is linearly independent to \([C_{l-i}] := +_{j \in \mathcal{J}_{l-i}} (P_j^\perp (\cap_{k \in S_i} [V_k]))\). In words, \(\{[V_k]\}_{k \in \mathcal{K}}\) is relatively independent, if for each and every layer \(\mathcal{J}_l\), each deflated subspace is linearly independent to \(\mathcal{J}_{l-i}\).
independent to $[C_{l,-i}]$, which is the sum of the other deflated subspaces in the layer $J_l$.

If $P_{li}^\perp(\cap_{k \in S_i}[V_k])$ is orthogonal to $[C_{l,-i}]$ for every $l = 1, \ldots, K - 1$ and $i \in J_l$, then $\{[V_k]\}_{k \in K}$ is said to be relatively orthogonal. We immediately check that relative orthogonality implies relative independence.

**Theorem 1.** Given matched data matrices $X_k = Z_k + E_k \in \mathbb{R}^{p_k \times n}$ for $k = 1, \ldots, K$ with true signal $Z_k$ and error $E_k$, if $\{[V_k]\}_{k \in K}$ is relatively independent, then, regardless of the ordering of index-sets, there exists a unique set of pairs $\{(S_i, r(S_i))\}_{i \in I_K}$.

Under only the relative independence condition, the determination of the partially-joint score subspaces $[W_i]$ corresponding to $S_i$, $i \in I_K$, may not be unique, and depends on the ordering of index-sets (see Examples A.3 and A.4 in the supplementary material). To ensure uniqueness of $[W_i]$’s, we require a rather strong assumption. We say that $\{[V_k]\}_{k \in K}$ is absolutely orthogonal, if (1) $\{[V_k]\}_{k \in K}$ is relatively orthogonal, and (2) for each $l = 1, \ldots, K - 1$ and for every $i \in J_l$,

$$P_{li}^\perp(\cap_{k \in S_i}[V_k]) = P_{li}^\perp(\cap_{k \in S_i}[V_k]),$$

where $[J_i] = +_{j \in J_{i \succ l}}(\cap_{k \in S_j}[V_k]) = +_{j \in J_{i \succ l}}[W_j]$ and $J_{i \succ l} = \{j : |S_j| > l, S_i \cap S_j \neq \phi\}$.

**Theorem 2.** Given matched data matrices $X_k = Z_k + E_k \in \mathbb{R}^{p_k \times n}$ for $k = 1, \ldots, K$ with true signal $Z_k$ and error $E_k$, if $\{[V_k]\}_{k \in K}$ is absolutely orthogonal, then partially-joint score subspaces $[W_i]$ for $i \in I_K$ are uniquely determined.

The uniqueness of each partially-joint loading subspace $[U_{(k),i}]$ is deduced from the uniqueness of $[W_i]$’s.

**Corollary 1.** Given matched data matrices $X_k = Z_k + E_k \in \mathbb{R}^{p_k \times n}$ for $k = 1, \ldots, K$ with true signal $Z_k$ and error $E_k$, if $\{[V_k]\}_{k \in K}$ is absolutely orthogonal, then each partially-joint loading subspace $[U_{(k),i}]$ is uniquely determined for $k = 1, \ldots, K$ and $i \in I_{(k)}$.

We provide proofs of theorems in this section and examples for relative independence and absolute orthogonality in Section A of the supplementary material.
5. Simulation Study

5.1. Example Dataset Generation

In the simulation study, we use the following data generation setting for numerically analyzing the performance of our proposal. Throughout, we use $K = 3$ blocks of data sets, in which the association structures are given by the ranks of index-sets.

First, we set a pre-determined rank $r(S_i)$ for each index-set $S_i$ for $i = 1, \ldots, 2^K - 1$. The partially-joint score matrix $W_i \in \mathbb{R}^{n \times r(S_i)}$ for each $S_i$ is given by the orthonormal basis of a random $n \times r(S_i)$ matrix consisting of independently standard normal variates.

The generic partially-joint loading matrix $U_{(k), i} \in \mathbb{R}^{p_k \times r(S_i)}$ for $i = 1, \ldots, 2^K - 1$ and $k = 1, \ldots, K$ is a column-wise concatenation of randomly generated vectors $u_{k,i,j} \in \mathbb{R}^{p_k}$, $j = 1, \ldots, r(S_i)$. Each entry of $u_{k,i,j}$ is generated following $\mathcal{N}(0, \sigma_{i,j}^2)$ independently. The magnitude of signal $\sigma_{i,j}^2$'s depend on the simulation settings, and are summarized as $\sigma_M^2 = \{(\sigma_{i,j}^2) : i = 1, \ldots, 2^K - 1, j = 1, \ldots, r(S_i), r(S_i) > 0\}$.

We then derive the generic signal matrix $Z_k$ for $k = 1, \ldots, K$ by $Z_k = \sum_{i=1}^{2^k-1} U_{(k), i} W_i^T$. The concatenation of generic loading matrices is denoted $U$, and is of size $\sum_{k=1}^{K} p_k \times \sum_{i=1}^{2^k-1} r(S_i)$.

The error matrix $E_k$ is generated element-wisely, such that $e_{k,i,j} \sim \mathcal{N}(0, \sigma^2)$ independently for $i = 1, \ldots, p_k$ and $j = 1, \ldots, n$. The magnitude of error $\sigma^2$ is set as the reciprocal of the signal-to-noise ratio (SNR), $\sigma^2 = 1/(\text{SNR})$.

We use the following six models.

1. (Individuals) $\mathcal{G} = \{(\{1\}, 2), (\{2\}, 2), (\{3\}, 2)\}$, $\sigma_M^2 = \{(1.4, 0.8), (1.3, 0.7), (1.2, 0.6)\}$

2. (Fully joint) $\mathcal{G} = \{(\{1, 2, 3\}, 2)\}$, $\sigma_M^2 = \{(1.0, 0.9)\}$

3. (Circular, partially joint) $\mathcal{G} = \{(\{1, 2\}, 2), (\{1, 3\}, 2), (\{2, 3\}, 2)\}$, $\sigma_M^2 = \{(1.4, 0.8), (1.3, 0.7), (1.2, 0.6)\}$

4. (Mix of fully joint and individuals) $\mathcal{G} = \{(\{1, 2, 3\}, 2), (\{1\}, 2), (\{2\}, 2), (\{3\}, 2)\}$, $\sigma_M^2 = \{(1.5, 0.8), (1.4, 0.7), (1.3, 0.6), (1.2, 0.5)\}$

5. (Fully joint and partially joint) $\mathcal{G} = \{(\{1, 2, 3\}, 2), (\{1, 2\}, 2), (\{1, 3\}, 2), (\{2, 3\}, 2)\}$, $\sigma_M^2 = \{(1.5, 0.8), (1.4, 0.7), (1.3, 0.6), (1.2, 0.5)\}$
6. (All possible combinations) $\mathcal{S} = \{(1, 2, 3), 2, (1, 2), 2, (1, 3), 2, (2, 3), 2, (1, 2), (2, 2), (3, 2), (2, 3), 2, 2, (1, 3), 2, (1, 2, 3), 2, (1, 5, 0.5), (1.4, 0.4), (1.3, 0.3), (1.2, 0.2)\}$

$\sigma_M^2 = \{(1.8, 0.8), (1.7, 0.7), (1.6, 0.6), (1.5, 0.5), (1.4, 0.4), (1.3, 0.3), (1.2, 0.2)\}$

5.2. Results on Comparative Study

In this subsection, we numerically compare the performance of our proposal to other competitors, including SLIDE (Gaynanova and Li, 2019), COBS (Gao et al., 2021), AJIVE (Feng et al., 2018), and JIVE (Lock et al., 2013). From the estimates of each method, the partially-joint structure $\mathcal{S}$, as well as the concatenated partially-joint score matrix $\hat{W}$ and loading matrix $\hat{U}$ can be extracted. See Section B.2 of the supplementary material for a brief review of these methods. Our proposal will be called the method of partially-joint structure identification, or PSI for short.

To assess the efficacy of finding the true partially-joint structure $\mathcal{S}$ and proper loading and score matrices, we use the following measures for PSI and other four methods.

1. **Accuracy**: The rate of finding the true partially-joint structure, $\mathbb{P}(\mathcal{S} = \hat{\mathcal{S}})$.

2. **Mean relative squared error (RSE)**: The relative size of the reconstruction error for each $Z_k$, 

$$\frac{1}{K} \sum_{k=1}^{K} \frac{\|Z_k - \hat{U}_k \hat{W}_k^T\|_F^2}{\|Z_k\|_F^2}.$$ 

3. $\overline{\theta}(U, \hat{U})$ for Partially-joint loading matrix $\hat{U}$: We measure the difference between $U$ and $\hat{U}$ as follows. We denote the principal angles between $U_{k,i}$ and $\hat{U}_{(k)}$ by $\theta_{U_{k,i,j}}$ for $k = 1, \ldots, K$, $i = 1, \ldots, 2^K - 1$ and $j = 1, \ldots, r(S_i)$. We report the average $\overline{\theta}(U, \hat{U})$ of all $\theta_{U_{k,i,j}}$’s.

4. $\overline{\theta}(W, \hat{W})$ for Partially-joint score $\hat{W}$: We measure the difference between $W$ and $\hat{W}$ as follows. We denote the principal angle between $w_{i,j}$ and $\hat{W}$ by $\theta_{W_{i,j}}$ for $i = 1, \ldots, 2^K - 1$ and $j = 1, \ldots, r(S_i)$. We report the average $\overline{\theta}(W, \hat{W})$ of all $\theta_{W_{i,j}}$’s as.

The simulation was conducted on different values of SNR (10 and 5) for the example models 1 to 6 with $n = 200$ and $p_1 = p_2 = p_3 = 200$. For
each model, we randomly generate a true partially-joint loading matrix $U$ which is then fixed throughout repetitions. For each SNR value, we repeat the simulation 100 times.

We performed a comparative study of all 5 methods over these datasets. The initial ranks were given to each method as follows. For PSI and AJIVE, we have given the initial ranks for each $X_k$ by their own true rank. For example, in Model 3, $r_k = 4$ for each $k$. SLIDE and JIVE require the concatenated data matrix $X$ as input, thus we have used the whole data matrix for these methods. (We also have used reduced-rank matrices, with the ranks given by the true model, as input for SLIDE and JIVE, but the results turned out to be worse.) For COBS, we use the concatenated $X$ and set the number of components as the sum of each index-set’s true rank (for example, $6 = 2 + 2 + 2$ in Model 3). Accuracy of finding the true structure and the average (and standard deviation) of RSE, $\bar{\theta}(U, \hat{U})$ and $\bar{\theta}(W, \hat{W})$, over 100 repetitions, are reported in Tables 1 and 2.

In Model 1, in which case there are only individual scores, PSI and AJIVE have identified the true partially-joint structure for almost all instances, but the performance of PSI is slightly better. SLIDE is comparable in terms of Accuracy and RSE and but sometimes provides large deviations in estimating the partially-joint loadings $U$. One possible reason is that SLIDE estimates partially-joint scores by enforcing $W^TW = I_{\hat{r} \times \hat{r}}$, whereas PSI and AJIVE allow non-orthogonality of these scores between different index-sets.

In Model 2, in which case there are only joint score, all methods but JIVE find the true structure quite well for large enough SNRs, and PSI, SLIDE and AJIVE have estimated almost identical partially-joint score subspaces.

In Model 3, in which partially-joint scores are entangled in a cyclic structure, PSI and SLIDE show good performance in identifying the true structures. On the other hand, other methods have completely failed to identify the partially joint structure, as expected. We observe that AJIVE and JIVE yield surprisingly nice figures in $\bar{\theta}(U, \hat{U})$ and $\bar{\theta}(W, \hat{W})$, and it may call an explanation. For Model 3, AJIVE and JIVE estimate a fully-joint structure rather than the (true) partially-joint structure. In spite of this inaccurate identification of structures, the estimation of $U$ and $W$ for AJIVE and JIVE becomes similar to applying an SVD for the concatenated $X$, which sometimes yields a good performance.

Simulation results for Models 4–6 are contained in Table 2. Model 4 represents the case where both joint and individual scores are present. Both
Table 1: Comparative Study on Model 1 to 3. The unit for Accuracy is percent.

| Model | SNR | Measure | PSI | SLIDE | bcv | COBS | AJIVE | JIVE |
|-------|-----|---------|-----|-------|-----|------|-------|------|
| 1     | 15  | Accuracy | 100 | 100   | 30  | 96   | 91    |      |
|       |     | RSE      | 0.24 (0.01) | 0.24 (0.01) | 0.41 (0.04) | 0.24 (0.02) | 0.26 (0.03) |      |
|       |     | $\bar{\theta}(U, \hat{U})$ | 15.99 (0.47) | 50.99 (1.59) | 54.84 (5.77) | 16.04 (0.59) | 16 (0.48) |      |
|       |     | $\bar{\theta}(W, \hat{W})$ | 21.18 (0.39) | 21.18 (0.4) | 23.97 (0.84) | 21.45 (1.27) | 21.22 (0.4) |      |
| 2     | 15  | Accuracy | 100 | 100   | 100 | 100  | 100   | 100  |
|       |     | RSE      | 0.38 (0.01) | 0.39 (0.03) | 0.56 (0.04) | 0.38 (0.03) | 0.42 (0.09) |      |
|       |     | $\bar{\theta}(U, \hat{U})$ | 19.93 (0.66) | 49.32 (10.35) | 60.77 (4.87) | 19.99 (0.74) | 19.03 (0.68) |      |
|       |     | $\bar{\theta}(W, \hat{W})$ | 25.94 (0.59) | 27.22 (3.38) | 31.4 (1.99) | 26.15 (1.1) | 25.99 (0.59) |      |
| 3     | 15  | Accuracy | 100 | 100   | 0   | 0    | 0     | 0    |
|       |     | RSE      | 0.16 (0) | 0.16 (0.01) | 0.35 (0.02) | 0.27 (0.01) | 0.48 (0.13) |      |
|       |     | $\bar{\theta}(U, \hat{U})$ | 15.34 (0.38) | 29.99 (3.98) | 29.16 (3.14) | 16.01 (0.45) | 14.88 (0.39) |      |
|       |     | $\bar{\theta}(W, \hat{W})$ | 15.56 (0.35) | 15.23 (0.33) | 17.08 (0.5) | 16 (0.57) | 15.06 (0.4) |      |
| 4     | 15  | Accuracy | 100 | 100   | 0   | 0    | 0     | 0    |
|       |     | RSE      | 0.25 (0.01) | 0.24 (0.01) | 0.52 (0.04) | 0.4 (0.01) | 0.74 (0.16) |      |
|       |     | $\bar{\theta}(U, \hat{U})$ | 18.91 (0.44) | 32.74 (4.79) | 41.14 (4.44) | 20.3 (0.68) | 18.34 (0.47) |      |
|       |     | $\bar{\theta}(W, \hat{W})$ | 19.28 (0.43) | 18.79 (0.39) | 23.25 (1.31) | 20.67 (1.05) | 18.69 (0.43) |      |

SLIDE and AJIVE showed superior performance in identifying the true structure in this case, and PSI was as competent in estimating score subspaces as AJIVE. Models 5 and 6 are most complicated with all types of associations (joint, partially-joint and individual) are present. For these cases both PSI and SLIDE show superior performances than other methods. (For Model 6, the algorithm for COBS did not converge.) Overall, the proposed PSI performs best in terms of the accuracy in estimating the true structure, for all models considered, and is comparable to the competitors in terms of RSE and estimating the loading and score subspaces. Among the competitors, we find SLIDE’s overall performance is among the best. While JIVE surprisingly
shows great performance in estimating the loading and score subspaces, we emphasize that JIVE fails in identifying the block-wise association structures, measured by the accuracy, for almost all models.

Table 2: Comparative Study on Model 4 to 6. The unit for Accuracy is percent.

| Model | SNR | Measure | PSI 100 | SLIDE bcv 100 | COBS 15 | AJIVE 100 | JIVE 98 |
|-------|-----|---------|---------|--------------|--------|-----------|--------|
| 4     | 15  | Accuracy | 100     | 15           | 100    | 98        |
|       |     | RSE      | 0.18 (0.01) | 0.17 (0.01) | 0.33 (0.03) | 0.17 (0) | 0.19 (0.01) |
|       |     | θ(U, ̂U) | 15.11 (0.31) | 28.76 (6.35) | 35.5 (4.08) | 15.09 (0.31) | 15.27 (0.31) |
|       |     | θ(W, ̂W) | 18.97 (0.38) | 18.72 (0.37) | 21.65 (0.86) | 18.84 (0.37) | 18.78 (0.38) |
| 10    |     | Accuracy | 100     | 35           | 53     | 100       | 69     |
|       |     | RSE      | 0.28 (0.01) | 0.28 (0.01) | 0.47 (0.03) | 0.28 (0.01) | 0.31 (0.03) |
|       |     | θ(U, ̂U) | 18.82 (0.45) | 32.05 (5.6)  | 43.66 (3.55) | 18.79 (0.45) | 18.87 (0.47) |
|       |     | θ(W, ̂W) | 23.28 (0.51) | 23.13 (1.18) | 28.58 (1.73) | 23.1 (0.49)  | 22.99 (0.49) |
| 5     |     | Accuracy | 24      | 5            | 4      | 100       | 15     |
|       |     | RSE      | 0.75 (0.08) | 0.71 (0.07)  | 0.76 (0.03) | 0.64 (0.02) | 0.89 (0.18) |
|       |     | θ(U, ̂U) | 28.46 (0.95) | 55.1 (10.3)  | 58.55 (4.85) | 27.56 (0.66) | 27.22 (0.7)  |
|       |     | θ(W, ̂W) | 32.87 (0.87) | 42.97 (7.98) | 48.74 (3.46) | 33.04 (0.8)  | 32.36 (0.83) |
| 5     |     | Accuracy | 5       | 100          | 0      | 0         | 0      |
|       |     | RSE      | 0.14 (0)  | 0.14 (0)     | 0.3 (0.02)  | 0.2 (0.01)  | 0.47 (0.08) |
|       |     | θ(U, ̂U) | 14.78 (0.29) | 21.89 (4.68) | 26.81 (1.32) | 15.09 (0.32) | 14.12 (0.4)  |
|       |     | θ(W, ̂W) | 14.93 (0.3)  | 14.31 (0.29) | 16.06 (0.38) | 14.92 (0.38) | 13.77 (0.38) |
| 10    |     | Accuracy | 100     | 99           | 0      | 0         | 0      |
|       |     | RSE      | 0.23 (0.01) | 0.22 (0.01)  | 0.46 (0.02) | 0.31 (0.01) | 0.73 (0.1)  |
|       |     | θ(U, ̂U) | 18.3 (0.41)  | 26.21 (3.95) | 36.22 (2.49) | 28.78 (0.44) | 17.47 (0.51) |
|       |     | θ(W, ̂W) | 18.52 (0.35) | 17.67 (0.33) | 22.04 (1.3)  | 18.85 (0.62) | 17.03 (0.43) |
| 5     |     | Accuracy | 9       | 0            | 0      | 0         | 0      |
|       |     | RSE      | 0.75 (0.11) | 0.68 (0.05)  | 0.75 (0.03) | 0.66 (0.03) | 1.58 (0.18) |
|       |     | θ(U, ̂U) | 28.64 (1.14) | 50.51 (6.92) | 57.92 (3.66) | 27.97 (0.73) | 25.28 (0.78) |
|       |     | θ(W, ̂W) | 26.18 (1.1)  | 29.69 (2.18) | 39.35 (2.21) | 29.23 (1.44) | 24.73 (0.7)  |
| 6     |     | Accuracy | 100     | 96           | 0      | 0         | 0      |
|       |     | RSE      | 0.15 (0)  | 0.15 (0)     | -       | 0.17 (0.01) | 0.19 (0.02) |
|       |     | θ(U, ̂U) | 14.44 (0.26) | 34.21 (2.7)  | -       | 14.3 (0.27) | 14.35 (0.42) |
|       |     | θ(W, ̂W) | 17.5 (0.3)  | 16.99 (0.27) | -       | 17.5 (0.3)  | 17.09 (0.29) |
| 10    |     | Accuracy | 100     | 73           | -      | 0 (0)     | 0 (0) |
|       |     | RSE      | 0.24 (0)  | 0.24 (0.01)  | -       | 0.27 (0.01) | 0.36 (0.05) |
|       |     | θ(U, ̂U) | 17.92 (0.29) | 35.78 (3.15) | -       | 17.82 (0.36) | 17.63 (0.53) |
|       |     | θ(W, ̂W) | 21.53 (0.31) | 21.53 (1.62) | -       | 21.33 (0.38) | 20.77 (0.4)  |
| 5     |     | Accuracy | 18      | 0            | -      | 0         | 0      |
|       |     | RSE      | 0.66 (0.07) | 0.63 (0.03)  | -       | 0.65 (0.02) | 1.14 (0.11) |
|       |     | θ(U, ̂U) | 27.35 (0.77) | 52.93 (6.03) | -       | 27.18 (0.55) | 25.24 (0.69) |
|       |     | θ(W, ̂W) | 30.16 (0.97) | 37.01 (4.35) | -       | 31.72 (0.82) | 28.54 (0.65) |

We also have compared the computation time needed by each method. Our method, PSI, took remarkably small computation times in all situations. For each of six models, PSI took 0.8—9.1 seconds in computation, compared to the time ranges 240—940 seconds for SLIDE, 10—73 seconds for COBS (when converged) and 6.3—417 seconds for JIVE. AJIVE took about 12—30 seconds. Table B.5 in the supplementary material lists the computation time for our method and other four methods.

22
5.3. Results on Imbalanced Signal Strength between Joint and Individual Components

Next, we consider cases where the signal strengths of joint components and individual components are grossly imbalanced. Consider a new model, with $K = 3$ and $p_1 = p_2 = p_3 = 100$, whose index-set ordering is $S_1 = \{1, 2, 3\}$, $S_5 = \{1\}$, $S_6 = \{2\}$ and $S_7 = \{3\}$. We set inherent joint rank $r(S_1) = 10$ and individual ranks $r(S_5) = r(S_6) = r(S_7) = 10$. Other index-sets have zero ranks, $r(S_2) = r(S_3) = r(S_4) = 0$. Throughout, $n = 200$.

For the case in which there are larger variations in the joint component than the individual components, we set $\sigma^2_{1,j} \gg \sigma^2_{i,j}$ for $i = 5, 6, 7$ and all $j$’s. In the opposite case, we give larger variations in the individual components than the joint component, $\sigma^2_{i,j} \gg \sigma^2_{1,j}$ for $i = 5, 6, 7$ and all $j$’s. Details are given in Section B.5 of the supplementary materials.

We carried out comparative simulations on both cases at SNR levels $\infty$ and 10. Not only the success rate of finding true structure but also the numbers of estimated joint and individual components were evaluated. In Table 3 in which case the joint component has larger variations, PSI and JIVE showed good performances to other methods. SLIDE failed to identify (weak) individual structures. AJIVE showed a compatible performance at SNR 10. When the individual components have larger variations, in Table 4, AJIVE is the best to detect (weak) joint structure, boasting the capability of Wedin bound simulation. SLIDE and JIVE also show compatible performances.

6. Real Data Analysis

In this section, we apply the proposed PSI to a blood cancer multi-omics data set linked to a drug response panel [Dietrich et al., 2018], available in the European Genome-Phenome Archive under accession number EGAS0000100174. We have chosen to use 121 cases diagnosed with chronic
Table 4: The case where the individual components have larger variations than the joint components.

| SNR | Measure | PSI  | SLIDE bcv | COBS  | AJIVE | JIVE |
|-----|---------|------|-----------|-------|-------|------|
| ∞   | Accuracy| 100  | 0         | 0     | 0     | 0    |
|     | Joint   | 10 (0)| 10.47 (0.73) | 38.95 (1.06) | 3.8 (1.83) | 14.42 (4.05) |
|     | Individual | 30 (0) | 25.81 (1.41) | 0.09 (0.29) | 48.6 (5.48) | 27.91 (1.17) |
| 10  | Accuracy| 0    | 0         | 0     | 100   | 0    |
|     | Joint   | 1.56 (2.15) | 10.79 (1.27) | 29.61 (2.89) | 10 (0) | 13.34 (3.66) |
|     | Individual | 46.84 (9.15) | 23.37 (1.76) | 3.45 (1.5) | 30 (0) | 27.09 (1.19) |

lymphocytic leukemia (CLL). The drug response panel \(X_{\text{Drug}}\) records the \textit{ex vivo} cell viabilities at a series of 5 concentrations, for each of 62 drugs that target onco-related pathways or are used widely in clinical practice. This multi-omics data set consists of the genome-wide DNA methylation profiles and the RNA sequencing profiles. The top 5,000 most variable CpG sites were selected \(X_{\text{Meth}}\) from the 450K illumina assay DNA methylation profile. As for the RNA sequencing profile, we selected the top 5,000 gene expressions with the largest stabilized variances \(X_{\text{Exp}}\) from the high-throughput sequencing (HTS) assay. In summary, we have three blocks of data sets \(X_{\text{Drug}} \in \mathbb{R}^{310 \times 121}, X_{\text{Meth}} \in \mathbb{R}^{5000 \times 121}\) and \(X_{\text{Exp}} \in \mathbb{R}^{5000 \times 121}\).

We repeat the estimation process over 100 repetitions (of data splitting) and select the mode structure, that is, the estimated partially-joint structure that appears the most out of 100 repetitions. The estimated mode structure is

\[
\hat{\mathcal{G}} = \{(\{\text{Drug, Meth}\}, 1), (\{\text{Drug}\}, 4), (\{\text{Meth}\}, 41), (\{\text{Exp}\}, 3)\},
\]

which means that PSI detected the index-set \{Drug, Meth\} of rank 1, and no fully-joint score or other forms of partially-joint scores were detected. This structure appeared 53 times out of 100 repetitions. The index-set \{Drug, Meth\} of rank 1 stands for the existence of one-dimensional latent score, partially-joint for Drug and Meth data sets (but not for Exp data set).

PSI showed better performance in computation time over other methods. In an Intel ® Xeon ® CPU E5-2640 v4 @ 2.4GHz system, it took 9.38 seconds per single data splitting instance (54.56 seconds for 100 instances with 40 cores). In comparison, the SLIDE with bcr method detected the same structure as our method, but it took almost 9 hours on the same machine. The COBS yielded an eccentric result, giving the fully-joint score of rank 50, taking about 14 minutes.
PSI was also robust over different choices of initial ranks. As a preprocessing, we set the ranks of the signal matrices using the principal component analysis which accounts for cumulative proportion of variances at 30%, 40%, 50% and 60%. Resulting signal ranks for Drug, Meth and Exp datasets are (2, 6, 1), (2, 15, 1), (3, 28, 2) and (5, 42, 3), respectively. For all cases, PSI detected the same index-set \{Drug, Meth\} of rank 1.

The analysis by PSI for the patient sample multi-block data reveals a particular pattern in the latent partially-joint score, which cannot be identified by applying, e.g. principal component analysis to each data block. To support this claim, we plot the reconstructed matrices $\hat{U}_2^T \hat{W}_2$ for the identified partially-joint and individual parts of the data in Fig. 3. Both the samples and the variables of $X_{\text{Drug}}$ are ordered by a hierarchical biclustering applied to the partially-joint component $\hat{U}_{\text{(Drug),2}}^T \hat{W}_{\text{2}}$ of $X_{\text{Drug}}$, where $\hat{W}_2 = \hat{W}_{\{\text{Drug}, \text{Meth}\}}$. The matrix $Z_{\text{Drug},s_2} := \hat{U}_{\text{(Drug),2}}^T \hat{W}_{\text{2}}$ is shown in the top left part of Fig. 3. The variables of $X_{\text{Meth}}$ are similarly ordered.

Focusing on the partially-joint scores corresponding to \{Drug, Meth\}, the samples are clustered into two distinct subgroups. These subgroups are shown in Fig. 4(a), and are denoted by groups $\alpha$ and $\beta$. There, it can be seen that the variables in $X_{\text{Drug}}$ and $X_{\text{Meth}}$ show a contrasting pattern according to the two subgroups $\alpha$ and $\beta$. Comparing $Z_{\text{Drug},s_2}$ (Fig. 4(a)) with the whole $X_{\text{Drug}}$ (Fig. 4(b)), we observe that the subgroups identified above are hidden in $X_{\text{Drug}}$. Moreover, the subgroups $\alpha$ and $\beta$ are well-separated in the partially-joint score $\hat{W}_{\{\text{Drug}, \text{Meth}\}}$ of $Z_{\text{Drug},s_2}$ (Fig. 4(c)), while it is hard to find any subgroup in the principal component scores of whole $X_{\text{Drug}}$ (Fig. 4(d)). The same conclusion can be made by inspecting the component of $X_{\text{Meth}}$ corresponding to the partially-joint score $\hat{W}_{\{\text{Drug}, \text{Meth}\}}$, denoted $Z_{\text{Meth},s_2}$, and the whole $X_{\text{Meth}}$ in Fig. 4(e) and (f). Thus we observed that PSI gives a more effective measure of finding inherent subgroups in a multi-omics data set than a separate application of the principal component analysis on each of the data matrix.

To find indicators that best explain the subgroups $\alpha$ and $\beta$, we conducted the Fisher exact test simultaneously on 59 gene mutations or chromosome defects of each patients, available as an ancillary information. The p-values from each Fisher exact test were adjusted by the Benjamini-Hochberg (BH) method. The smaller p-values indicate stronger differences of each mutation or chromosome defect between the subgroup $\alpha$ and $\beta$. We found that the immunoglobulin heavy chain variable (IGHV) region mutation status
Figure 3: The reconstructed components according to the estimated partially-joint structure. The values are all normalized for each column, then truncated to line on $[-2, 2]$. has the most associated relation with the subgroups at BH-adjusted p-value $1.026 \times 10^{-13}$. We also present the $2 \times 2$ table of the IGHV status and the subgroups, see Tables C.1 and C.2 in the supplementary material. We postulate IGHV mutation status gives a substantial explanation for the subgrouping of CLL patients, that is, wild type matches to the subgroup $\alpha$ and mutation type to the subgroup $\beta$. Survival analyses on overall survival rates was also conducted, and we found statistically significant differences in survival between the two subgroups $\alpha$ and $\beta$ as shown in Fig. 5.

Again in Fig. 4(a), the variables in $Z_{\text{Drug},S_2}$ can be clustered into subgroups [a] and [b] showing a contrasting response pattern to the subgroups $\alpha$ and $\beta$ (variables showing weak responses were excluded as subgroup [c]). The subgroup [a] shows higher viability for $\beta$ (IGHV mutated) than $\alpha$ (IGHV wild type) and vice versa for [b]. Table C.3 (in the supplementary material) presents the list of prominent drugs that have appeared in subgroups [a] and [b] at least 4 times out of 5 concentrations. For the subgroup [a], the table lists a number of inhibitor drugs that target the B cell receptor (BCR) components, such as Bruton’s tyrosine kinase (BTK; spebrutinib, ibrutinib), phosphatidylinositol 3-kinase (PI3K; idelalisib, duvelisib) and spleen tyro-
sine kinase (STK; tamatinib, PRT062607 HCL), AKT inhibitor (MK-2206) or SRC inhibitors (dasatinib) targets signal transduction pathways that promotes survival and growth of B cell lymphocytes. Unexpected encounter with HSP90 inhibitor (AT13387 or Onalespib) may be related to the stability of lymphocyte-specific SRC family kinases (Mshaik et al., 2021). The appearance of CHK inhibitors (PF 477736, AZD7762, CCT241533) may be linked to repairing mechanisms of DNA damages at G2 phase, known to be associated with WEE1 kinase and the AKT/PKB pathway (Zhang and Hunter, 2014). For the subgroup [b], the appearance of mTOR inhibitor (everolimus) may suggest that mTOR pathway and shows different drug sensitivities to the BCR component, despite the fact that it is on the downstream of AKR/PKB
Figure 5: The difference in overall survival between the subgroups $\alpha$ and $\beta$ is displayed on a Kaplan plot with p-value 0.0028 from the log-rank test.

pathway. The role of IGHV in this implication requires further investigation. BCL2 inhibitor (navitoclax) and rotenone might be related to the role of mitochondria in apoptosis (Wang and Youle, 2014).

Supplementary Materials

The supplementary materials contain the proofs of Lemma 2.1, Theorems 1 and 2, and Corollary 1, a brief review on competing methods, detailed model settings used for simulations, additional numerical results and auxiliary information on real data analysis, including the table of drugs selected by the proposed method.

References

Björck, A., Golub, G., 1973. Numerical methods for computing angles between linear subspaces. Mathematics of Computation 27(123), 579–594.

Bolt, G., Lunag’omez, S., Nemeth, C., 2022. Distances for comparing multisets and sequences ArXiv:2206.08858v1.

Dietrich, S., Oleś, M., Lu, J., Sellner, L., et al., 2018. Drug-perturbation-based stratification of blood cancer. Journal of Clinical Investigation 128(1), 427–445.
Draper, B., Kirvy, M., Marks, J., Marrinan, T., Peterson, C., 2014. A flag representation for finite collections of subspaces of mixed dimensions. Linear Algebra and its Applications 451, 15–32.

Eckart, C., Young, G., 1936. The approximation of one matrix by another of lower rank. Psychometrika 1, 211–218.

Feng, Q., Jiang, M., Hannig, J., Marron, J., 2018. Angle-based joint and individual variation explained. Journal of Multivariate Analysis, 241–265.

Gao, X., Lee, S., Jung, S., 2021. Covariate-driven factorization by thresholding for multi-block data. Biometrics 77(3), 1011–1023.

Gaynanova, I., Li, G., 2019. Structural learning and integrative decomposition of multi-view data. Biometrics 75, 1121–1132.

Li, G., Gaynanova, I., 2018. A general framework for association analysis of heterogeneous data. The Annals of Applied Statistics 12(3), 1700–1726.

Li, G., Jung, S., 2017. Incorporating covariates into integrated factor analysis of multi-view data. Biometrics 72, 1433–1442.

Lock, E., Hoadley, K., Marron, J., Nobel, A., 2013. Joint and individual variation explained (jive) for integrated analysis of multiple data types. Annals of Application Statistics 7(1), 523–542.

Lock, E.F., Park, J.Y., Hoadley, K.A., 2022. Bidimensional linked matrix factorization for pan-omics pan-cancer analysis. The annals of applied statistics 16, 193.

Mshaik, R., Simonet, J., Georgievski, A., et al., 2021. Hsp90 inhibitor nvp-bep800 affects stability of src kinases and growth of t-cell and b-cell acute lymphoblastic leukemias. Blood Cancer Journal 11(61). URL: https://www.nature.com/articles/s41408-021-00450-2 doi:10.1038/s41408-021-00450-2

Norris, J.L., Farrow, M.A., Gutierrez, D.B., et al., 2017. Integrated, high-throughput, multiomics platform enables data-driven construction of cellular responses and reveals global drug mechanisms of action. Journal of Proteome Research 16(3), 1364–1375. doi:https://doi.org/10.1021/acs.jproteome.6b01004.
Park, J.Y., Lock, E.F., 2020. Integrative factorization of bidimensionally linked matrices. Biometrics 76, 61–74.

Reuter, J.A., Spacek, D., Snyder, M.P., 2015. High-throughput sequencing technologies. Molecular Cell 58(4), 586–597. doi:10.1016/j.molcel.2015.05.004.

Smilde, A.K., Westerhuis, J.A., de Jong, S., 2003. A framework for sequential multiblock component methods. Journal of Chemometrics 17, 323–337.

Subramanian, I., Verma, S., Kumar, S., Jere, A., Anamika, K., 2020. Mult-omics data integration, interpretation, and its application. Bioinformatics and Biology Insights URL: https://journals.sagepub.com/doi/full/10.1177/1177932219899051, doi:10.1177/1177932219899051.

VanDeun, K., Smilde, A.K., van der Werf, M.J., Kiers, H.A., VanMechelen, I., 2009. A structured overview of simultaneous component based data integration. BMC Bioinformatics 10, 246.

Wang, C., Youle, R.J., 2014. The role of mitochondria in apoptosis. Annu Rev Genet. 43, 95–118.

Ye, K., Lim, L.H., 2016. Schubert varieties and distances between subspaces of different dimensions. SIAM Journal of Matrix Analysis and Applications 37(3), 1176–1197.

Zhang, Y., Hunter, T., 2014. Roles of chk1 in cell biology and cancer therapy. International Journal of Cancer 134(5), 1013–1023.
SUPPLEMENTARY MATERIAL

A. Proofs

A.1. Some basic facts for proof

As it is the most elementary fact, we begin by recapitulating Lemma 2 and show its proof.

**Lemma 2.** For \(i, j \in \mathcal{I}_K = \{1, \ldots, 2^K - 1\}\) and \(S_i \cap S_j \neq \emptyset\), \([W_i] \perp [W_j]\).

**Proof.** From the definition of the partially-joint score subspace, \([W_i]\), we immediately find that

\[ [W_i] \perp [W_j], \quad i, j \in \mathcal{I}_K \text{ and } S_j \cap S_i \neq \emptyset \]

by the range-kernel complementarity property of the vector space projection transformation.

We denote \(\mathcal{I}_{<i} = \{j : j < i, S_j \cap S_i \neq \emptyset\}\). Hereafter \(\mathcal{N}(T)\) is the null space of a linear transformation \(T\) of \(\mathbb{R}^n\) and \(\mathcal{R}(T)\) its range space.

**Lemma A.1.** For \(i \in \mathcal{I}_K\), we have \(\mathcal{N}(\bigcap_{k \in \mathcal{I}_{<i}} P_k^\perp) = \bigoplus_{k \in \mathcal{I}_{<i}} [W_k]\).

**Proof.** Let \(v \in \bigoplus_{k \in \mathcal{I}_{<i}} [W_k]\). Then there exists a unique \(\{v_k\}_{k \in \mathcal{I}_{<i}}\) with \(v_k \in [W_k]\) such that the sum of all \(v_k\) is \(v\). For each \(j_1 \in \mathcal{I}_{<i}\) and \(v_{j_1} \in [W_{j_1}]\), it can be easily checked that \((\bigcap_{k \in \mathcal{I}_{<j_1}} P_k^\perp)(v_{j_1}) = \{0\}\) and then \((\bigcap_{k \in \mathcal{I}_{<j_1}} P_k^\perp)(v_{j_1}) = \{0\}\) follows. Thus \((\bigcap_{k \in \mathcal{I}_{<i}} P_k^\perp)(v) = \{0\}\) and \(\mathcal{N}(\bigcap_{k \in \mathcal{I}_{<i}} P_k^\perp) \supset \bigoplus_{k \in \mathcal{I}_{<i}} [W_k]\).

Conversely, let \(v' \notin \bigoplus_{k \in \mathcal{I}_{<i}} [W_k]\). Then there exists a unique \(\{v_k\}_{k \in \mathcal{I}_{<i}}\) with \(v_k \in [W_k]\) and non-zero \(a \in (\bigoplus_{k \in \mathcal{I}_{<i}} [W_k])^\perp\) such that \(v'\) is the sum of all \(v_k\) and \(a\). For each \(j_1 \in \mathcal{I}_{<i}\) and \(v_{j_1} \in [W_{j_1}]\), we have \((\bigcap_{k \in \mathcal{I}_{<j_1}} P_k^\perp)(v_{j_1}) = \{0\}\) as before. Since \(a \perp [W_{k'}]\) for any \(k' \in \mathcal{I}_{<i}\), we also have \((\bigcap_{k \in \mathcal{I}_{<i}} P_k^\perp)(a) \neq 0\). Thus we have \((\bigcap_{k \in \mathcal{I}_{<i}} P_k^\perp)(v') \neq \{0\}\) and \(\mathcal{N}(\bigcap_{k \in \mathcal{I}_{<i}} P_k^\perp) \subset \bigoplus_{k \in \mathcal{I}_{<i}} [W_k]\). \(\square\)

**Lemma A.2.** For \(i \in \mathcal{I}_K\), we have \(\bigoplus_{k \in \mathcal{I}_{<i}} [W_k] = +_{k \in \mathcal{I}_{<i}} (\cap_{k' \in S_k}[V_{k'}])\).

**Proof.** We give a proof by induction on \(k\). If \(k = 1\), there is nothing to prove.
If \(k = 2\), \([W_1] = \cap_{k' \in S_1}[V_{k'}]\), so the statement is true for \(k = 1, 2\).

For any \(k \geq 3\), suppose the statement holds, that is,

\[ \bigoplus_{j \in \mathcal{I}_{<m}} [W_j] = +_{j \in \mathcal{I}_{<m}} (\cap_{k' \in S_j}[V_{k'}]) \]

31
for all $1 \leq m \leq k$. Let $\tilde{k}$ be the largest element in $I_{<k+1}$. We denote $P = \bigcirc_{k' \in I_{<k}} P_{k'}^{+}$ and $P^{\perp}$, the projection onto $\mathcal{N}(P)$ of $\mathbb{R}^n$. Then, we have

$$+_{j \in I_{<k+1}} (\cap_{k' \in S_j} [V_{k'}]) = +_{j \in I_{<k}} [W_j] + \cap_{k' \in S_k} [V_{k'}]$$

$$= +_{j \in I_{<k}} [W_j] + P(\cap_{k' \in S_k} [V_{k'}]) + P^{\perp}(\cap_{k' \in S_k} [V_{k'}])$$

Indeed, $P^{\perp}(\cap_{k' \in S_k} [V_{k'}]) \subset \mathcal{N}(P)$ and $\mathcal{N}(P) = \bigoplus_{k' \in I_{<k}} [W_{k'}]$ by the previous lemma. Thus

$$+_{j \in I_{<k+1}} (\cap_{k' \in S_j} [V_{k'}]) = +_{j \in I_{<k}} [W_j] + \cap_{k' \in S_k} [V_{k'}]$$

$$= +_{j \in I_{<k}} [W_j] + [W_{\tilde{k}}]$$

$$= +_{j \in I_{<k+1}} [W_j].$$

Therefore the statement holds for any $k \geq 3$ and the proof is completed. □

**Lemma A.3.** For $k = 1, \ldots, K$, $+_{i \in \{k \in S_i\}} [W_{S_i}] = [V_k]$.

**Proof.** In Lemma A.2, we set $i$ such that $S_i = \{k\}$. Then $I_{<i} = \{j : k \in S_j, |S_j| > 1\}$ and $+_{j \in I_{<1}} [W_j] = +_{j \in I_{<1}} (\cap_{k' \in S_j} [V_{k'}])$. The proof goes similarly as in Lemma A.2. Let $\tilde{i}$ be the largest element in $I_{<i}$. We denote $P = \bigcirc_{k' \in I_{<i}} P^{+}_{k'}$ and $P^{\perp}$, the projection onto $\mathcal{N}(P)$ of $\mathbb{R}^n$.

$$+_{j \in I_{<i} \cup S_i} (\cap_{k' \in S_j} [V_{k'}]) = +_{j \in I_{<i}} [W_j] + \cap_{k' \in S_i} [V_{k'}]$$

$$= +_{j \in I_{<i}} [W_j] + P(\cap_{k' \in S_i} [V_{k'}]) + P^{\perp}(\cap_{k' \in S_i} [V_{k'}])$$

$$= +_{j \in I_{<i}} [W_j] + P(\cap_{k' \in S_i} [V_{k'}]) + P^{\perp}(\cap_{k' \in S_i} [V_{k'}])$$

$$= +_{j \in I_{<i}} [W_j] + [W_i]$$

$$= +_{j \in I_{<i} \cup S_i} [W_j].$$

By the set inclusion-exclusion principle, as $I_{<i} \cup S_i$ involves every index-set that contains $k$, it is immediate that $+_{j \in I_{<i} \cup S_i} (\cap_{k' \in S_j} [V_{k'}]) = [V_{\tilde{k}]}]$. □

**A.2. Proof of Theorem 1**

**Proof.** We claim that $\text{rank}([W_i])$ is uniquely determined. By Sylvester’s law of nullity, we have

$$\text{rank}([W_i]) = \text{rank}(\cap_{k \in S_i} [V_k]) - \text{rank}(\mathcal{N}(\bigcirc_{j \in I_{<i}} P^{\perp}) \cap (\cap_{k \in S_i} [V_k])),$$
where $\mathcal{I}_{<i} = \{ j : j < i, S_j \cap S_i \neq \phi \}$ as before. By lemma $A.1$ and $A.2$ Sylvester’s law of nullity for our theorem is restated as

$$\text{rank}([W_i]) = \text{rank}(\cap_{k \in S_i}[V_k]) - \text{rank}((+_{k \in \mathcal{I}_{<i}} (\cap_{k' \in S_k}[V_{k'}])) \cap (\cap_{k \in S_i}[V_k])).$$

We want to make this expression in a more explicit form. For that, we suggest the following assertions.

Let $l = |S_i|$. We set $[I_l] = +_{k \in (k : |S_k| > l)} (\cap_{k' \in S_k}[V_{k'}])$. The projection onto $[I_l]$ in $\mathbb{R}^n$ is denoted by $P_{I_l}$ and the projection onto $[I_l]^{\perp}$ is denoted by $P_{I_l}^{\perp}$. Moreover, we consider an index set $\mathcal{J}_{i,l} = \mathcal{I}_{<i} \cap \{ k : |S_k| > l \}$ and let $[J_l] = +_{k \in \mathcal{J}_{i,l}} (\cap_{k' \in S_k}[V_{k'}])$. The projection onto $[J_l]$ in $\mathbb{R}^n$ is denoted by $P_{J_l}$ and the projection onto $[J_l]^{\perp}$ by $P_{J_l}^{\perp}$. Finally, we define $\mathcal{J}_{i,l} = \mathcal{I}_{<i} \cap \{ k : |S_k| = l \}$.

**Lemma A.4.** $P_{J_l}^{\perp} = P_{I_l}^{\perp} \circ P_{J_l}^{\perp}$.

**Proof.** Trivial from the fact $\mathcal{N}(P_{J_l}^{\perp}) \subset \mathcal{N}(P_{I_l}^{\perp})$. \qed

**Lemma A.5.** If $v_j \in P_{J_l}^{\perp}(\cap_{k' \in S_j}[V_{k'}])$ with $j \in \mathcal{J}_{i,l} \setminus \mathcal{J}_{i,l}$, then $v_j \in (\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_m}[V_{k'}])$ for some $m < i$, $m \neq j$ such that $S_m \cap S_i = \{0\}$ and not for any $m < i$, $m \neq j$ such that $S_m \cap S_i \neq \{0\}$.

**Proof.** Consider the cases

1. $v_j \in (\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_m}[V_{k'}])$ for some $m < i$, $m \neq j$ such that $S_m \cap S_i \neq \{0\}$ but not for any $S_m \cap S_i = \{0\}$,

2. $(\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_m}[V_{k'}]) = \{0\}$ for only $m \in \mathcal{J}_{i,l}$

In case (1), $v_j$ becomes automatically an element of $\cap_{k' \in S_i}[V_{k'}]$ such that $S_j \subset S_i$ and $S_m \subset S_i$. Then $v_j \in [I_l]$ since $S_i \cap S_i \neq \phi$ and $|S_i| > |S_i|$, and this is a contradiction. In case (2), $v_j \notin \mathcal{R}(P_{I_l})$ and this is contradict to the assumption. \qed

**Proposition A.1.** If $[V_k]_{k \in \mathcal{K}}$ is relatively independent, then $+_{k \in \mathcal{J}_{i,l}} P_{J_l}^{\perp}(\cap_{k' \in S_k}[V_{k'}])$ and $P_{J_l}^{\perp}(\cap_{k \in S_i}[V_k])$ are independent.

**Proof.** We want to show that the relative independence of $[V_k]_{k \in \mathcal{K}}$ is violated if $+_{k \in \mathcal{J}_{i,l}} P_{J_l}^{\perp}(\cap_{k' \in S_k}[V_{k'}])$ and $P_{J_l}^{\perp}(\cap_{k \in S_i}[V_k])$ are linearly dependent. Suppose that there is nonzero $v \in +_{k \in \mathcal{J}_{i,l}} P_{J_l}^{\perp}(\cap_{k' \in S_k}[V_{k'}]) \cap P_{J_l}^{\perp}(\cap_{k \in S_i}[V_k])$ with $l = |S_i|$. We are going to show that $+_{k \in \mathcal{J}_{i,l}} P_{J_l}^{\perp}(\cap_{k' \in S_k}[V_{k'}]) \cap P_{J_l}^{\perp}(\cap_{k \in S_i}[V_k])$ is not $\{0\}$ in the following cases.
First we exclude the case where \( P_{i_l}^1(\cap_{k' \in S_l}[V_{k'}]) = \{0\} \). If so we have \( \cap_{k' \in S_l}[V_{k'}] = \oplus_{k'' \in \{k'':S_l \subseteq S_{k''}\}}[W_{k''}] = +_{k'' \in \{k'':S_l \subseteq S_{k''}\}}[V_{k''}] \), then \( \cap_{k' \in S_l}[V_{k'}] \) itself is in \([I_i] = \mathcal{N}(P_{i_l}^1)\).

Next, under the assumption that \( P_{i_l}^1(\cap_{k' \in S_l}[V_{k'}]) \neq \{0\} \), we run through the following situations. Now on \( J_{i,l}' = \{k : |S_k| = l, k \in \mathcal{I}_{<i}, P_{i_l}^1(S_k) \neq \{0\}\} \).

(i) \(|J_{i,l}| = 0\) : There is nothing to prove.

(ii) \(|J_{i,l}| = 1\) and \(|J_{i,l}'| = 0\) :

We will show that this case is vacuous. Suppose there exist nonzero \( v \in P_{J_{i,l}}^1(\cap_{k' \in S_i}[V_{k'}]) \cap P_{J_{i,l}}^1(\cap_{k' \in S_k}[V_{k'}]) \) with \( k \in J_{i,l} \). As \( P_{J_{i,l}}^1(\cap_{k' \in S_i}[V_{k'}]) = \{0\} \), we can deduce that \( P_{J_{i,l}}^1(v) = 0 \).

We pick a vector \( u \) in \( \cap_{k' \in S_j}[V_{k'}] \) such that \( S_i \subseteq S_j \) and \( S_j \subseteq S_j \). As \( \cap_{k' \in S_i}[V_{k'}] \subseteq \cap_{k' \in S_j}[V_{k'}] \) and \( \cap_{k' \in S_j}[V_{k'}] \subseteq \cap_{k' \in S_k}[V_{k'}] \), we have \( u \in (\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}]) \). As \( S_i \cap S_j \neq \emptyset \), we observe that \( \cap_{k' \in S_j}[V_{k'}] \cap [I_i] = \mathcal{N}(P_{J_{i,l}}^1) \).

\[
\begin{align*}
u \in & P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \\
& \subset P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap P_{J_i}^1((\cap_{k' \in S_k}[V_{k'}])).
\end{align*}
\]

Let \( w = u + v \). Since \( u \in P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \) with respect to \( S_k \) and the same for \( S_i \), we find that \( w \in (\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}]) \), and then, \( w \in [I_i] = \mathcal{N}(P_{J_{i,l}}^1) \). But this forces \( P_{J_{i,l}}^1(w) = v \) and \( v \) to be zero and leads to vacuity.

(iii) \(|J_{i,l}| = 1\) and \(|J_{i,l}'| = 1\) :

Suppose there is nonzero \( v \in P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \) with \( k \in J_{i,l} \). From Lemma \ref{A.4}, we have \( P_{J_{i,l}}^1(v) \in P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \). By the same argument, \( P_{J_{i,l}}^1(v) \in P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \). Then there is a nonzero \( P_{J_{i,l}}^1(v) \in P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \).

(iv) \(|J_{i,l}| \geq 2\), \(|J_{i,l}'| = 0\) and \( v \in P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \) with some \( k \in J_{i,l} \) : the same as case (ii).

(v) \(|J_{i,l}| \geq 2\), \(|J_{i,l}'| = 0\) and \( v \notin P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \) with any of \( k \notin J_{i,l} \) :

We will show that this case is a generalization of case (ii) and also vacuous. As \( v \in +_{k \in J_{i,l}} P_{J_{i,l}}^1((\cap_{k' \in S_i}[V_{k'}]) \cap (\cap_{k' \in S_k}[V_{k'}])) \), we express

\[
v = \sum_{j \in J_{i,l}} a_j v_j
\]

34
for each \( a_j \in \mathbb{R} \) (at least two of them are nonzero) and \( v_j \in P_{I_i}^\perp(\cap_{k' \in S_j}[V_{k'}]) \).

By Lemma A.5 for each \( v_j \) for \( j \in J_{i,t} \), we can find \( S_{m,j} \) such that \( v_j \in S_{m,j} \) and \( m < i, m \neq j, S_m \cap S_i = \{0\} \). And this also implies that there exists certain \( S_{t,j} \) and \( v_j \in S_{t,j} \) such that \( S_j \subset S_{t,j} \) and \( S_m,j \subset S_{t,j} \). Here we point out that as \( S_{t,j} \cap S_i \neq \phi \), since \( S_j \cap S_i \neq \phi \). Then all \( \cap_{k' \in S_{t,j}}[V_{k'}] \) for \( j \in J_{i,t} \) and their linear combinations are subsets of \([I_i]\).

This leads to the conclusion \( v \in [I_i] \) and shows the vacuity of this case.

(vi) \( |J_{i,l}| \geq 2, |J'_{i,l}| \geq 1 \) and \( v \in P_{I_i}^\perp(\cap_{k' \in S_k}[V_{k'}]) \) with some \( k \in J_{i,l} \):
If \( k \in J'_{i,l} \), then the same as the case (iii). If \( k \not\in J'_{i,l} \), then the same as the case (ii).

(vii) \( |J_{i,l}| \geq 2, |J'_{i,l}| \geq 1 \) and \( v \not\in P_{I_i}^\perp(\cap_{k' \in S_k}[V_{k'}]) \) with any of \( k \in J_{i,l} \):
As \( v \in +_{k \in J_{i,l}} P_{I_i}^\perp(\cap_{k' \in S_k}[V_{k'}]) \), we express
\[
v = \sum_{j \in J_{i,l}} a_j v_j
\]
for each \( a_j \in \mathbb{R} \) (at least two of them are nonzero) and \( v_j \in P_{I_i}^\perp(\cap_{k' \in S_j}[V_{k'}]) \).

Note that if \( a_{j'} = 0 \) for all \( j' \in J'_{i,l} \), then this case is essentially the same as case (v), so we only consider the situation at least one \( a_{j'} \neq 0 \) for \( j' \in J'_{i,l} \).

By Lemma A.5 and its consequences in case (v), for each \( j \in J_{i,l} \setminus J'_{i,l} \), there exists \( S_{t,j} \) and \( v_j \in S_{t,j} \) such that \( S_j \subset S_{t,j} \) and \( S_m,j \subset S_{t,j} \). As previously discussed, \( \cap_{k' \in S_{t,j}}[V_{k'}] \) for \( j \in J_{i,l} \setminus J'_{i,l} \) is a subset of \([I_i]\). So we rule out the terms involving \( j \in J_{i,l} \setminus J'_{i,l} \) and then
\[
P_{I_i}^\perp(v) = \sum_{j' \in J'_{i,l}} a_{j'} P_{I_i}^\perp(v_{j'}).
\]
Since each \( \cap_{k' \in S_j}[V_{k'}] \neq \{0\} \) for \( j \in J'_{i,l} \) is a subset of \([I_i]\) and at least one \( a_{j'} \neq 0 \), we deduce that \( P_{I_i}^\perp(v) \) is non-zero. Therefore, the relative independence is violated.

\[\square\]
Now the last term in the RHS of our law of nullity is re-expressed as

\[
(+_{k\in\mathcal{I}_c}(\cap_{k'\in S_k}[V_{k'}])) \cap (\cap_{k\in S_k}[V_k]) =((+_{k\in\mathcal{J}_{c,<l}}(\cap_{k'\in S_k}[V_{k'}])) \oplus (+_{k\in\mathcal{J}_{c,l}} P_{j,l}^{\perp}(\cap_{k'\in S_k}[V_{k'}]))) \cap (\cap_{k\in S_k}[V_k])
\]

(A.1)

In Proposition A.1, we have observed that \(+_{k\in\mathcal{J}_{c,l}} P_{j,l}^{\perp}(\cap_{k'\in S_k}[V_{k'}])\) and \(P_{j,l}^{\perp}(\cap_{k\in S_k}[V_k])\) are independent. As \(P_{j,l}(\cap_{k\in S_k}[V_k]) \subset +_{k\in\mathcal{J}_{c,<l}}(\cap_{k'\in S_k}[V_{k'}])\), the term (A.1) becomes

\[
(+_{k\in\mathcal{I}_c}(\cap_{k'\in S_k}[V_{k'}])) \cap (\cap_{k\in S_k}[V_k]) = (+_{k\in\mathcal{J}_{c,<l}}(\cap_{k'\in S_k}[V_{k'}]) \cap (P_{j,l}(\cap_{k\in S_k}[V_k]))).
\]

Finally, we have demonstrated

\[
\text{rank}([W_i]) = \text{rank}(\cap_{k\in S_k}[V_k]) - \text{rank} \left( (+_{k\in\mathcal{I}_c}(\cap_{k'\in S_k}[V_{k'}]) \cap (\cap_{k\in S_k}[V_k]) \right)
\]

\[
= \text{rank}(\cap_{k\in S_k}[V_k])
\]

\[
- \text{rank} \left( (+_{k\in\mathcal{J}_{c,<l}}(\cap_{k'\in S_k}[V_{k'}]) \cap (P_{j,l}(\cap_{k\in S_k}[V_k])) \cap (\cap_{k\in S_k}[V_k])) \right)
\]

\[
= \text{rank}(\cap_{k\in S_k}[V_k]) - \text{rank}(P_{j,l}(\cap_{k\in S_k}[V_k])).
\]

It is notable that the determination of \(\text{rank}([W_i])\) depends only on \(\mathcal{J}_{i,l}\), that is, the set of indices \(j\) such that \(|S_j| > |S_i|\) and \(S_j \cap S_i \neq \emptyset\). In other words, it does not depend on any index-sets of the same size as \(S_i\) and their orderings.

As partially-joint score subspaces \([W_i]\) are constructed recursively and the determination of each \([W_i]\)’s rank only depends on \(\mathcal{J}_{i,l}\), not on the ordering index-sets of size \(l = |S_i|\), we conclude that the set of pairs \(\{(S_i, r(S_i)) : i \in \mathcal{I}_K\}\) is unique.

\[\square\]

A.3. Proof of Theorem 2

Proof. We give a proof recursively on \(l \in K\). If \(l = K\), there only exists \([W_1] = \cap_{k'\in S}[V_{k'}]\) for \(S_1 = K\), so there is nothing to prove. If \(l = K - 1\), by absolute orthogonality, all \(P_{l,l}^{\perp}(\cap_{k'\in S}[V_{k'}])\) are orthogonal each other for \(i = \{2, \ldots, K + 1\}\). Therefore, in determining each \([W_i]\) for \(i = \{2, \ldots, K + 1\}\), other \([W_t]\) for \(t \in \mathcal{J}_{i,2} = \mathcal{I}_{<l} \cap \{i' : |S_{i'}| = 2\}\) does not affect on the construction of \([W_i]\).

For any \(l \leq K - 2\), suppose the statement holds, that is, a partially-joint score subspace \([W_j]\) such that \(|S_j| = l' > l\) is uniquely determined only by
[W_{j'}] for S_{j'} > S_j. For each index i ∈ J_i = \{i' : |S_i| = l\} (regardless of ordering), we have P_h(∩_{k ∈ S_i}[V_{k'}]) = P_{\hat{h}}(∩_{k ∈ S_i}[V_{k'}]) by absolute orthogonality. Then among all indices i' ∈ J_i, P_{i'}^+(∩_{k ∈ S_{j'}}[V_{k'}]) are orthogonal each other.

Suppose an ordering on the set of all index-sets of size l is given, denoted by (S_{i_1}, \ldots, S_{i_n}) with h = K_C. For i_1, [W_{i_1}] is just determined as P_{i_1}^+(∩_{k ∈ S_{i_1}}[V_{k'}]). Next for i_2, as [W_{i_1}] and P_{i_2}^+(∩_{k ∈ S_{i_2}}[V_{k'}]) are orthogonal, we check that P_{i_1}^+ \circ P_{i_2}^+ (∩_{k ∈ S_{i_2}}[V_{k'}]) = P_{i_2}^+(∩_{k ∈ S_{i_2}}[V_{k'}]). Thus [W_{i_2}] is determined regardless of [W_{i_1}]. In recursive manner, for i ∈ \{i_3, \ldots, i_h\}, [W_i] is determined regardless of all [W_{i'}] for i' ∈ J_{i,i}, or in other word, is uniquely determined as P_{i_i}^+(∩_{k ∈ S_i}[V_{k'}]) depending only on [W_{j'}]s for S_{j'} > l.

\[\square\]

A.4. Proof of Corollary 1

Proof. From the discussions of Section 2, we know that [W_{(k)}] is indeed the score subspace of Z_k for each k = 1, \ldots, K. Thus, given unique [W_i]'s for i ∈ I_{(k)} from Theorem 2, the subspace \[U_{(k),i}\] generated by U_{(k),i} is unique. \[\square\]

A.5. Examples

The following two examples presents the cases where relative independence is satisfied and not:

**Example A.1.** Let K = 3, n = 4 and

\[
V_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 0 \end{pmatrix}, \quad V_3 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix}
\]

then \[J_1 = [(1, 0, 0, 0)^T]\] and \[P_{i_1}^+[V_1] = [(0, 1, 0, 0)^T], P_{i_1}^+[V_2] = [(0, 1/\sqrt{2}, 1/\sqrt{2}, 0)^T]\] and \[P_{i_1}^+[V_3] = [(0, 0, 1/\sqrt{2}, 1/\sqrt{2})^T]\] are linearly independent. Thus \{[V_1], [V_2], [V_3]\} is relatively independent.

**Example A.2.** Let K = 3, n = 4 and

\[
V_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 0 \end{pmatrix}, \quad V_3 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix}
\]

As \[P_{i_1}^+[V_3]∩(P_{i_1}^+[V_1]+P_{i_1}^+[V_2]) = [(0, 0, 1, 0)^T]\] is not empty, thus \{[V_1], [V_2], [V_3]\} is not relatively independent.

37
We present the following examples that support Theorem 1, that relative independence indeed guarantee the uniqueness of the partially-joint structure \( \mathcal{S} = \{(S_i, r(S_i)) : i \in I_K, r(S_i) > 0 \} \).

Example A.3 (cont’d from Example A.1). Under the ordering \( S_1 = \{1, 2, 3\}, S_5 = \{1\}, S_6 = \{2\}, S_7 = \{3\} \), we obtain \( [W_1] = [(1, 0, 0, 0)^T], [W_5] = [(0, 1, 0, 0)^T], [W_6] = [(0, 0, 1, 0)^T] \) and \( [W_7] = [(0, 0, 0, 1)^T] \). Then we have
\[
\mathcal{S} = \{(\{1, 2, 3\}, 1), (\{1\}, 1), (\{2\}, 1), (\{3\}, 1)\}.
\]

On the other hand, in the case \( S_1 = \{1, 2, 3\}, S_5 = \{2\}, S_6 = \{1\}, S_7 = \{3\} \), we obtain \( [W_1] = [(1, 0, 0, 0)^T], [W_5] = [(0, 1, 0, 0)^T], [W_6] = [(0, -1/\sqrt{2}, 1/\sqrt{2}, 0)^T] \) and \( [W_7] = [(0, 0, 0, 1)^T] \). The partially-joint structure is still the same as above.

Example A.4 (cont’d from Example A.2). Under the ordering \( S_1 = \{1, 2, 3\}, S_5 = \{1\}, S_6 = \{2\}, S_7 = \{3\} \), we obtain \( [W_1] = [(1, 0, 0, 0)^T], [W_5] = [(0, 1, 0, 0)^T], [W_6] = [(0, 0, 1, 0)^T] \) and \( [W_7] = [(0, 0, 0, 1)^T] \). The partially-joint structure is
\[
\mathcal{S} = \{(\{1, 2, 3\}, 1), (\{1\}, 1), (\{2\}, 1), (\{3\}, 1)\}.
\]

On the other hand, in the case \( S_1 = \{1, 2, 3\}, S_5 = \{3\}, S_6 = \{2\}, S_7 = \{1\} \), we obtain \( [W_1] = [(1, 0, 0, 0)^T], [W_5] = [(0, 1, 0, 0)^T], [W_6] = [(0, 0, 1/\sqrt{2}, 1/\sqrt{2})^T] \), \( [W_7] = [(0, 0, 1/\sqrt{2}, -1/\sqrt{2})^T] \) and \( [W_7] = \{0\} \). This time, the partially-joint structure is
\[
\mathcal{S} = \{(\{1, 2, 3\}, 1), (\{1\}, 0), (\{2\}, 1), (\{3\}, 2)\}.
\]

We also present the following examples that support Theorem 2, that absolute orthogonality guarantee the uniqueness of the partially-joint score subspaces.

Example A.5. Let \( K = 4 \), \( n = 7 \) and
\[
V_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}^T
\]
\[
V_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}^T
\]
\[
V_3 = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 & 0 \end{pmatrix}^T
\]
\[
V_4 = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 & 0 \end{pmatrix}^T
\]
This example satisfies absolute orthogonality. Between under two orderings $(S_{11} = \{3, 4\}, S_{12} = \{1\}, S_{13} = \{2\})$ and $(S_{11} = \{3, 4\}, S_{12} = \{2\}, S_{13} = \{1\})$, the determinations of $[W_{12}]$ and $[W_{13}]$ are the same.

Example A.6. Let $K = 4$, $n = 7$ and

\[
V_1 = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0 \right)^T
\]
\[
V_2 = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 1, 0 \right)^T
\]
\[
V_3 = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, 0 \right)^T
\]
\[
V_4 = \left( \frac{1}{\sqrt{2}}, 0, 0, 0, 1, 0 \right)^T.
\]

This example satisfies relative orthogonality, but not absolute orthogonality. Between under two orderings $(S_{11} = \{3, 4\}, S_{12} = \{1\}, S_{13} = \{2\})$ and $(S_{11} = \{3, 4\}, S_{12} = \{2\}, S_{13} = \{1\})$, the determinations of $[W_{12}]$ and $[W_{13}]$ are not the same because $[V_1]$ and $[V_2]$ are not orthogonal and $[W_{11}] = (1/\sqrt{2}, 1/\sqrt{2}, 0, 0, 0, 0)^T$ does not have an effect on the determination of $[W_{12}]$ and $[W_{13}]$ by definition.

B. Additional Information on Simulation Study

B.1. Example of the Measure of Dissimilarity between Two Partially-Joint Structures

For the comparison between two partially-joint structure, we devised a following concept of the dissimilarity measure between two partially-joint structures.

We turn a partially-joint structure into the corresponding binary structure matrix. For example, let $K = 3$ and partially-joint structure be

\[
\mathcal{G} = \{(1, 2, 3, 2, W_1), (1, 2, 1, W_2), (1, 3, 1, W_3), (2, 3, 1, W_4), (3, 1, W_7)\}.
\]

Then we have the corresponding binary structure matrix

\[
T = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 1
\end{pmatrix}.
\]
Consider two binary structure matrices $T_1 \in \{0, 1\}^{K \times m_1}$ and $T_2 \in \{0, 1\}^{K \times m_2}$.

Discarding all the identical columns, we denote the remaining columns $\tilde{T}_1$ and $\tilde{T}_2$. For example, from

$$T_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

we obtain

$$\tilde{T}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \tilde{T}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

For each remaining column of $\hat{T}_1$ (or $\hat{T}_2$), find the closest column of $\hat{T}_2$ (or $\hat{T}_1$) in the Hamming distance. The dissimilarity measure between $\hat{T}_1$ and $\hat{T}_2$ is the sum of the squares of all these Hamming distance.

In the example above, for $(1, 1, 0)^T$ of $\tilde{T}_1$, the closest column of $\tilde{T}_2$ is $(1, 1, 1)^T$ and the difference is 1. For $(1, 1, 1)^T$ of $\tilde{T}_2$, the difference between $(1, 1, 0)^T$ is 1 and for $(0, 1, 1)^T$ of $\tilde{T}_2$, it is 2. The dissimilarity measure is then $1^2 + 1^2 + 2^2 = 6$.

### B.2. Review on Methodology of Other Methods

We briefly review the methodology of AJIVE [Feng et al., 2018], SLIDE [Gaynanova and Li, 2019], COBS [Gao et al., 2021] and JIVE [Lock et al., 2013].

**AJIVE** In AJIVE, each signal matrix $Z_k \in \mathbb{R}^{p_k \times n}$ is regarded as a sum of joint structure $J_k$ and individual structure $I_k$ for $k = 1, \ldots, K$. A joint structure $J_k$ is viewed as the score subspace $[V_M] \in \mathbb{R}^n$ shared by all $Z_i$’s.

AJIVE extracts each estimated signal matrix $\hat{Z}_k$ from dataset $X_k$ and $\hat{Z}_k$ is of initial rank estimate $r_k$. The estimated shared joint component $[\hat{V}_M]$ is obtained as a flag mean among score subspaces of $\hat{Z}_k$’s, or $[\hat{V}_1], \ldots, [\hat{V}_k]$, in a sense of the projection Frobenius norm distance as our method. The rank $r_J$ of $[\hat{V}_M]$ (called the joint rank) is estimated using the simulated distribution of the largest singular value of the concatenated matrix of random directions and that of Wedin bounds:

1. If the largest squared singular value of the column concatenation matrix $\hat{V}$ of $\hat{V}_k$’s is larger than the 5th percentile of the simulated distribution
of the largest squared singular value of the concatenation matrix of random orthogonal matrices of the same size as $\hat{V}_k$'s (or random direction bound), then $[\hat{V}_M]$ is not generated by noise in 95 percent of probability.

(2) If there are $\hat{r}_J$ squared singular values of $\hat{V}$ are larger than the 95th percentile of the simulated distribution of Wedin bounds, then the first $\hat{r}_J$ right singular vectors are used as the basis for the estimated joint score subspace $[\hat{V}_M]$.

The estimated joint structure $\hat{J}_k$ is a projection of the dataset $X_k$ onto the estimated joint score subspace $[\hat{V}_M]$, that is, $\hat{J}_k = X_k \hat{V}_M \hat{V}_M^T$. Each estimated individual structure $\hat{I}_k$ is obtained as $X_k \cdot (I - \hat{V}_M \hat{V}_M^T)$. The row spaces of each estimated individual structure $\hat{I}_k$ is orthogonal to $[\hat{V}_M]$. There is no guarantee that individual structures are mutually orthogonal. The joint score matrix is just defined as $\hat{V}_M$ and the corresponding joint loading matrix for $k$th data source is a regression of $\hat{J}_k$ on $\hat{V}_M \hat{V}_M^T$, computed as $\hat{J}_k \cdot \hat{V}_M$. The individual loading and score matrices are obtained from SVDs of $\hat{I}_k$'s.

SLIDE SLIDE identifies the partially-joint structure with the penalized matrix factorization, that is, 

$$(\tilde{U}, \tilde{V}) = \arg \min_{U,V} \sum_{k=1}^{K} \frac{1}{2} \|X_k - UV^T\|^2_F + \lambda \sum_{j=1}^{r} \|U_{kj}\|_2 \quad \text{s.t.} \quad V^TV = I,$$

where $U_{kj}$ is the $j$th column of the loading matrix $U_k$ of $X_k$ and $r$ is the number of all possible sparsity patterns. After computing $\tilde{U}$ and $\tilde{V}$ with an iterative algorithm, the corresponding structure matrix $\tilde{T}$ is obtained from the sparse structure of $\tilde{U}$. Note that the concept of the structure matrix $\tilde{T}$ here is identical to the partially-joint structure matrix of ours in Section B.1. Even though this optimization problem is nonconvex and there is no guarantee about convergence to the global optimum, authors reported that a local solution can be obtained heuristically by initializing $V$ with the left singular matrix of concatenated $X$ that works well in the simulation.

Then SLIDE estimate the loading and score matrices, $\hat{U}$ and $\hat{V}$, for the structure $\hat{T}$ by solving the following optimization problem with an iterative algorithm,

$$(\hat{U}, \hat{V}) = \arg \min_{U,V} \|X - UV^T\|^2_F \quad \text{s.t.} \quad V^TV = I,$$
with the constraint that the loading \( U \) has the same sparsity structure as \( \hat{T} \).

In model validation, SLIDE adapt the block cross validation (BCV) procedure to select the best structure \( \hat{T}_{\text{best}} \). BCV splits rows and columns of each dataset \( X_k \) into submatrices \( X_k^{11}, X_k^{12}, \ldots, X_k^{21} \). Then it holds out a set of submatrices \( X_{ij}^{11}, X_{ij}^{12}, \ldots, X_{ij}^{K} \) of the same sub-block position in each dataset \( X_k \) and evaluate the prediction error on \( X_{ij}^{11} \)'s. Given a set of structure candidates \( \hat{T}_1, \hat{T}_2, \ldots \), we select one that minimizes the error across all folds.

**JIVE** Like in AJIVE, JIVE decompose each signal matrix \( X_i \) as a sum of joint structure \( J_i \) and individual structure \( I_i \), for \( i = 1, \ldots, K \). After defining \( R \) to be a row concatenation of \( R_i = X_i - J_i - I_i \), JIVE estimate both joint and individual structures by minimizing \( \|R\|_F^2 \) under the given ranks. An alternating iterative algorithm is implemented for the estimation finding individual component with given joint component at one step and vice versa at another step. The estimated joint structure is identical to the first \( r_J \) terms in the SVD of \( X \) with individual components removed and the estimated individual structures to the first \( r_i \) terms in the SVD of \( X_i \) with the joint component removed. The selection of \( r_J \) and \( r_i \)'s are validated using the permutation test.

**COBS** COBS iteratively estimates a sequence of loading vectors, \( u_i \) for \( i = 1, \ldots, r \) for given \( r \), while updating the data matrix \( X = [X_1, \ldots, X_k] \in \mathbb{R}^{n \times \sum p_k} \). The algorithm starts with \( X^{[0]} = X \). At the \( i \)-th step, with the current data matrix \( X^{[i-1]} \), the \( i \)-th loading vector \( u_i \) is estimated solving the following maximization problem, that is,

\[
\tilde{u}_i = \max_u \| (X^{[i-1]})^T u \|_2^2 \quad \text{s.t.} \quad u^T u = 1.
\]

As each \( u_i \) is equipped with block structure \( u_i = \left( u_{(1)}^T \ldots u_{(K)}^T \right)^T \), we can give sparsity at two levels of thresholding, one for block-wise sparsity and the other for overall sparsity in estimating \( \tilde{u}_i \). The tuning parameters \( \alpha_v \in [0, 1] \) and \( \lambda_v \geq 0 \) control the two levels of sparsity respectively, and \( \tilde{u} \) is thresholded as a normalized solution of

\[
\min_x \frac{1}{2} \| x - \tilde{u}_i \|^2 + \gamma_1 \| x \|_1 + \gamma_2 \| x \|_{2,1},
\]

where \( \gamma_1 = \alpha_v \lambda_v \) and \( \gamma_2 = (1 - \alpha_v) \lambda_v \). Then the score vector \( v_i \) is estimated as the empirical BLUP and dataset \( X^{[i-1]} \) is updated as \( X^{[i]} = X^{[i-1]} - \tilde{u}_i \hat{v}_T \).
B.3. Computation time comparison

Table B.5: Time comparison among our method and other four methods (SLIDE, COBS, AJIVE and JIVE). The computation time reported. The unit is second(s). Averages and standard deviations are over 10 replications.

| Model | SNR | PSI | SLIDE bcr | COBS | AJIVE | JIVE |
|-------|-----|-----|-----------|------|-------|------|
| 1     | 15  | 0.78 (0.1) | 283.61 (8.37) | 11.36 (0.19) | 12.01 (0.13) | 6.32 (0.15) |
|       | 10  | 0.77 (0.06) | 281.73 (5.77) | 10.9 (0.13) | 12.06 (0.14) | 148.55 (74.98) |
|       | 5   | 0.76 (0.02) | 362.66 (3.67) | 11.33 (0.09) | 12.01 (0.1) | 108.64 (0.57) |
| 2     | 15  | 0.95 (0.15) | 283.91 (6.74) | 3.85 (0.15) | 12.35 (0.36) | 19.83 (0.5) |
|       | 10  | 0.86 (0.03) | 332.45 (6.2) | 3.54 (0.15) | 12.04 (0.14) | 24.44 (0.25) |
|       | 5   | 0.87 (0.08) | 361.34 (9.2) | 3.09 (0.07) | 12.1 (0.06) | 51.79 (1.21) |
| 3     | 15  | 1.73 (0.13) | 265.59 (5.26) | 11.83 (0.48) | 11.92 (0.16) | 56.79 (1.09) |
|       | 10  | 1.62 (0.09) | 287.99 (5.2) | 10.07 (0.18) | 12.01 (0.41) | 148.38 (3.41) |
|       | 5   | 1.52 (0.08) | 398.44 (6.75) | 11.6 (0.19) | 12.06 (0.27) | 65.95 (1.82) |
| 4     | 15  | 1.48 (0.12) | 237.01 (4.61) | 14.52 (0.36) | 12.57 (0.97) | 67.55 (1.47) |
|       | 10  | 1.39 (0.04) | 280.57 (4.04) | 16.39 (0.15) | 12.04 (0.12) | 28.72 (0.52) |
|       | 5   | 1.41 (0.09) | 360.8 (4.95) | 15.02 (0.12) | 12 (0.09) | 37.87 (0.16) |
| 5     | 15  | 2.45 (0.13) | 240.77 (5.24) | 15.65 (0.18) | 12.01 (0.19) | 100.54 (2.14) |
|       | 10  | 2.36 (0.06) | 252.43 (5.94) | 16.2 (0.2) | 11.95 (0.12) | 70.98 (8.43) |
|       | 5   | 2.36 (0.11) | 359.94 (5.37) | 15.47 (0.14) | 11.92 (0.12) | 108.68 (15.49) |
| 6     | 15  | 9.07 (0.68) | 399.56 (8.64) | 70.99 (3.57) | 32.79 (1.91) | 439.35 (16.24) |
|       | 10  | 7.86 (0.96) | 770.04 (13.9) | 72.64 (2.43) | 32.77 (1.99) | 297.46 (9.04) |
|       | 5   | 8.84 (0.61) | 940.57 (37.64) | 73.77 (3.87) | 32.79 (2.24) | 417.45 (63.1) |

B.4. Results on Tuning Parameter Selection

We report the performances of our tuning parameter selection procedure of Section 3.3 using the six models given in Section 5. Since the tuning parameter \( \lambda \) represents a threshold for principal angles, the candidates for \( \lambda \) are given by \( \lambda = 0^\circ, 1^\circ, \ldots, 90^\circ \). For each value of \( \lambda \), we evaluated the empirical risk. As discussed in Section 3.3, we take the parameter \( \tilde{\lambda} \) that gives the smallest empirical risk and also compare \( d(\hat{S}_{tr}(\tilde{\lambda}), \hat{S}(\lambda)) \) as a function of \( \lambda \). We also present \( d(\hat{S}(\lambda), \hat{S}) \), which reflects how much the estimated structure differs from the true structure on each value of angle threshold, under the situation where the true structure (‘oracle’) is known.

First, consider SNR = 10. The empirical risk is minimized at an interval of \( \lambda \)'s, and for any \( \lambda \) in the interval, the corresponding structure \( \hat{S} \) matches the true \( S \); the valley bottoms of empirical risk (solid line) are posited inside those of the measure of dissimilarity (dashed line) with value zero, as seen in Fig. B.6. For each model in the figure, solid line (empirical risk) shows a
similar shape as dashed line (the measure of dissimilarity), which implies that empirical risk well reflects the difference between the estimated structure and the true structure.

Consider next the signal-to-noise ratio is small, SNR = 2. Unfortunately, the empirical risk is minimized at smaller values of $\lambda$ than desired; see Fig. B.7. Unlike Fig. B.6, solid line (empirical risk) shows a far different shape than dashed line (the measure of dissimilarity), except for Model 1, which implies that empirical risks fail to detect the true structure. This is due to the lower value of SNR, with which the magnitude of noise overwhelms that of signal. As the score vectors of each dataset have almost random directions in low SNRs, there is a tendency that signals from a partially-joint (and fully-joint) scores are counted separately as if they belong to individual data blocks.
Figure B.7: The values of empirical risk (solid) and $d(\hat{\Theta}_{tr}(\lambda), \hat{\Theta}(\lambda))$ (dotted), $d(\hat{\Theta}(\lambda), \Theta)$, (dashed).
B.5. Simulation Settings for Section 5.4

In the unbalance of signal strength between joint and individual component settings, we set $n = 200$ and $K = 3$.

In the first case, we set the elements of $\sigma^2_M$ as

1. Joint ($S_1$) : $(15, 14.5, \ldots, 5.5)$,
2. Individual 1 ($S_2$) : $(0.150, 0.141, 0.132, \ldots, 0.069)$,
3. Individual 2 ($S_3$) : $(0.147, 0.138, 0.129, \ldots, 0.066)$,
4. Individual 3 ($S_4$) : $(0.144, 0.135, 0.126, \ldots, 0.063)$,

so the strength of joint signals are about 100 times stronger than those of individual signals.

In the second case, we set the elements of $\sigma^2_M$ as

1. Joint ($S_1$) : $(0.15, 0.145, \ldots, 0.055)$,
2. Individual 1 ($S_2$) : $(15, 14.1, 13.2, \ldots, 6.9)$,
3. Individual 2 ($S_3$) : $(14.7, 13.8, 12.9, \ldots, 6.6)$,
4. Individual 3 ($S_4$) : $(14.4, 13.5, 12.6, \ldots, 6.3)$,

so the strength of individual signals are about 100 times stronger than those of joint signals.
C. Additional Information on Real Data Analysis

Table C.1: Fisher’s exact tests between gene mutations/chromosome defects and the CLL subgroups, α and β were conducted simultaneously. The p-values were adjusted with the FDR (Benjamini-Hochberg method) and the top 5 lowest FDR (BH method) adjusted p-values were listed.

| Gene mutation/Chromosome defect | adjusted p-value |
|---------------------------------|------------------|
| IGHV                            | $1.036 \times 10^{-13}$ |
| MED12                           | 0.173            |
| del17p13                        | 0.174            |
| del13q14                        | 0.178            |
| TP53                            | 0.184            |

Table C.2: The association between IGHV mutation status and the CLL subgroups with an FDR (BH method) adjusted p-value $1.036 \times 10^{-13}$ from the Fisher’s exact test. Nine missing data for IGHV mutation status were excluded.

| Mutation \ Subgroup | $\alpha$ | $\beta$ |
|---------------------|----------|---------|
| IGHV Wild type      | 49       | 8       |
| IGHV Mutated        | 7        | 48      |
Table C.3: The most appeared drugs in the subgroups [a] and [b] with appearances at least four times out of 5 concentrations.

| Drug name          | Target pathway                                      | Appearances |
|--------------------|-----------------------------------------------------|-------------|
| spebrutinib        | BTK                                                 | 5           |
| idelalisib         | PI3K delta                                          | 5           |
| duvelisib          | PI3K gamma, PI3K delta                              | 5           |
| tamatinib          | SYK                                                | 5           |
| dasatinib          | ABL1, KIT, LYN, PDGFRA, PDGFRB, SRC                 | 5           |
| PF 477736          | CHK1, CHK2                                          | 5           |
| MK-2206            | AKT1/2 (PKB)                                        | 5           |
|                   |                                                     |             |
| selumetinib        | MEK1/2                                              | 4           |
| PRT062607 HCL      | SYK                                                | 4           |
| AZD7762            | CHK1/2                                              | 4           |
| CCT241533          | CHK2                                                | 4           |
| TAE684             | ALK                                                | 4           |
| MK-1775            | WEE1                                                | 4           |
| AT133387           | HSP90                                               | 4           |

| Drug name          | Target pathway                                      | Appearances |
|--------------------|-----------------------------------------------------|-------------|
| everolimus         | mTOR                                               | 5           |
| thapsigargin       | SERCA                                              | 5           |
| orlistat           | LPL                                                | 5           |
| rotenone           | Electron transport chain in mitochondria            | 5           |
| alatinib           | EGFR, ERBB2                                         | 4           |
| fludarabine        | Purine analogue                                     | 4           |
| navitoclax         | BCL2, BCL-XL, BCL-W                                 | 4           |