Non-iterative Joint and Individual Variation Explained

Qing Feng, Jan Hannig and J.S.Marron

Department of Statistics and Operations Research

The University of North Carolina at Chapel Hill

Abstract

Integrative analysis of disparate data blocks measured on a common set of experimental subjects is one major challenge in modern data analysis. This data structure naturally motivates the simultaneous exploration of the joint and individual variation within each data block resulting in new insights. For instance, there is a strong desire to integrate the multiple genomic data sets in The Cancer Genome Atlas (TCGA) to characterize the common and also the unique aspects of cancer genetics and cell biology for each source. In this paper we introduce Non-iterative Joint and Individual Variation Explained (Non-iterative JIVE), capturing both joint and individual variation within each data block. This is a major improvement over earlier approaches to this challenges in terms of both a new conceptual understanding and a fast linear algebra computation. An important mathematical contribution is the use of principal angles between appropriate subspaces and perturbation theory in the segmentation of joint an individual variation. Furthermore, this makes our method robust against the heterogeneity among data blocks, without a need for normalization. An application to gene expression and copy number data reveals different behaviors of each type of signal in characterizing tumor types. Another application to mortality data reveals interesting historical lessons.

keywords: Data Integration, Variation decomposition, Singular Value Decomposition, Principal Angel, Perturbation theory, Heterogeneity.

1 Introduction

A major challenge in modern data analysis is data integration, combining diverse information from disparate data sets measured on a common set of experimental subjects. A unified and insightful understanding of the set of data blocks is expected from simultaneously exploring the joint variation representing the inter-block associations and the individual variation specific to each block. This requires rigorous definitions of each type of variation together with a method to obtain identifiable decomposition to a set of heterogeneous data blocks.

Lock et al. (2013) proposed a model frame and method (JIVE) to decompose each data block into three matrices. These model different types of variation, including a low-rank approximation of the joint variation across the blocks, low-rank approximations of the individual variation for each data block, and residual noise. Drawbacks of that approach include a slow, iterative algorithm, and a need for arbitrary normalization of the data sets. The iterative computation raises further concern about convergence and even identifiability. G. Zhou and Mandic (2015) formulated JIVE decomposition as a quadratic optimization problem with restrictions to ensure identifiability. This proposed algorithm, although showing a major improvement in efficiency, still requires iterations and a tuning parameter to control the number of joint components.

Here a simultaneous solution to these problems is proposed. The key insight is to describe the JIVE components using row space (assuming columns are the data objects). This focuses the methodology on signatures (linear combinations) of samples, e.g. patients, which gives straightforward definitions of the components and thus provides identifiability. Segmentation of the variation into joint and individual components is done using a *principal angle analysis* based on a perturbation theory. This approach eliminates the selection of a tuning parameter which simplify the analysis. A further benefit of this new approach is that data normalizations (e.g. to handle non data scaling, and differing number of features) is no longer needed.

Other methods that simultaneously decompose joint variation patterns and individual variation patterns have also been developed. Westerhuis et al. (1998) discusses two types of methods. One main type extends the traditional Principal Component Analysis (PCA), such

as Consensus PCA and Hierarchical PCA first introduced by Wold (1987); Wold et al. (1996). An overview of extended PCA methods is discussed in Smilde et al. (2003). This type of method computes the block scores, block loadings, global loadings and global scores based on an iterative procedure. The other main type of method are extensions of Partial Least Squares (PLS) (Wold, 1985) or Canonical Correlation Analysis (CCA) (Hotelling, 1936) that seek associated patterns between the two data blocks by maximizing covariance/correlation. For example, Wold et al. (1996) introduced multi-block PLS and hierarchical PLS (HPLS) and Trygg and Wold (2003) proposed *O2-PLS* to better reconstruct joint signals by removing structured individual variation.

A connection between extended PCA and extended PLS methods is discussed in Hanafi et al. (2011). Both types of methods provide an integrative analysis by taking the inter-block associations into account. These papers make the recommendations to use normalization to address potential scale heterogeneity, including normalizing by the Frobenius norm, or the largest singular value of each data block etc. However, there are no consistent criteria for normalization and some of these methods have convergence problems. An important point is that none of these approaches provides simultaneous decomposition highlighting joint and individual modes of variation with the goal of contrasting these to reveal new insights.

1.1 Practical Motivation

Simultaneous variation decomposition has been useful in many practical application, e.g. genomic cancer research. For example, Lock and Dunson (2013), Kühnle (2011), Mo et al. (2013) performed integrative clustering on multiple sources to reveal novel and consistent subtypes based on understanding of joint and individual variation. Other types of application include analysis of multi-source metabolomic data (Kuligowski et al., 2015), extraction of commuting patterns in railway networks (Jere et al., 2014), recognition of braincomputer interface (Zhang et al., 2015) and etc.

Our primary motivating data example is The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013), which has disparate genomic data types. Integration of these is fundamental for studying cancer on a molecular level. As a concrete example, gene expression and copy number aberrations are considered here for a set of 815 breast cancer tumor samples. For each tumor

sample, there are measurements of 16615 gene expression features and 24174 copy number features. The two data sources have different dimensions and very different scalings.

The tumor samples are classified into five molecular subtypes: Basal-like, HER2, Luminal A, Luminal B and Normal-like. An integrative analysis of gene expression and copy number targets the association between the two types of features that jointly classify the subtypes. In addition, the features to individually identify the subtypes of each source can be highlighted to give new insights.

1.2 Toy Example

A clear view of the challenges brought by multiple disparate data blocks, in particular the heterogeneity among data blocks is given here using a toy example. This toy example has two 100×100 data blocks, X and Y, with patterns corresponding to joint and individual structures. Figure 1 shows colormap views of matrices, with the value of each matrix entry colored according to the color bar at the bottom of each panel. A careful look at the color bar scaling shows the values are almost 4 orders of magnitude larger for the top matrices. Each column of these matrices is regarded as a common data object and each row is considered as one feature. Each of the two raw data matrices (X and Y in the left column) is the sum of joint, individual and noise components shown in the other columns.

The joint signal for both blocks is constant across columns, thus having the same rank one row space, but different matrix entries. The X joint signal has the value 25000 in the first 25 rows and -25000 in the second 25 rows, with 0 in the remaining rows. The Y joint signal has -5 in the last 25 rows and 5 in the second to last 25 rows and 0 elsewhere. The X individual signal partitions the columns into five groups, each of size 20. Those columns have -5000, 5000, -5000, 5000, 0 as their respective values. The individual signal for Y partitions the columns into four groups, each of size 25. Those columns have values -2, -1, 1, 2, respectively. Note that these two individual row spaces are quite different. The noise matrices are standard Gaussian random matrices (scaled by 5000 for X).

Note the toy data exhibit very strong scale heterogeneity across blocks. For simplicity of presentation, the important signals are represented in the mean structure. Therefore we will work with the non-centered data in the following analyses.

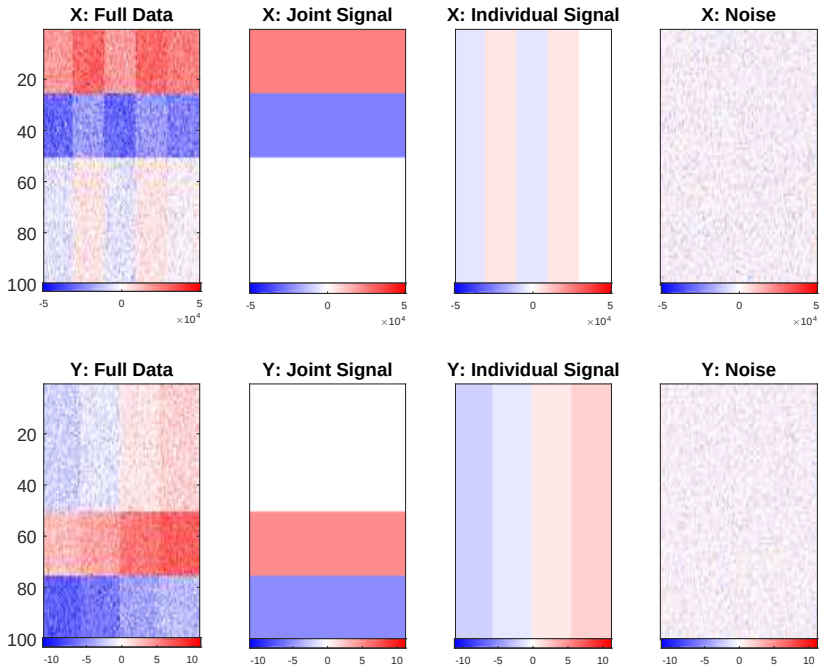

Figure 1: Data blocks X (top) and Y (bottom) in the toy example. The first column of figures present the observed data matrices with each type of signal and noise matrices depicted in the remaining columns. Color bar at the bottom of each panel. These structures are challenging to capture using conventional methods due to different orders of magnitude.

A naive integrative analysis can be done by concatenating X and Y on columns and performing a singular value decomposition on this concatenated matrix. Figure 2 shows the results for 3 choices of rank. The rank 1 approximation reasonably captures the joint variation in each block, because the concatenated matrix has essentially the constant vector as first right SVD eigenvector. The rank 2 approximation in the center shows the columns grouped by 5, clearly driven by the X individual component, because the much larger X values drive the analysis. One might hope that the Y individual component would show up in the rank 3 approximation. However, because the noise in the X matrix is so large, a noise component from X dominates the Y signal, so the important latter component completely disappears. PLS might be used to address the magnitude difference in this examples and capture the signal components. However, PLS fails to indicate which component is joint among blocks and which component is individual for each block.

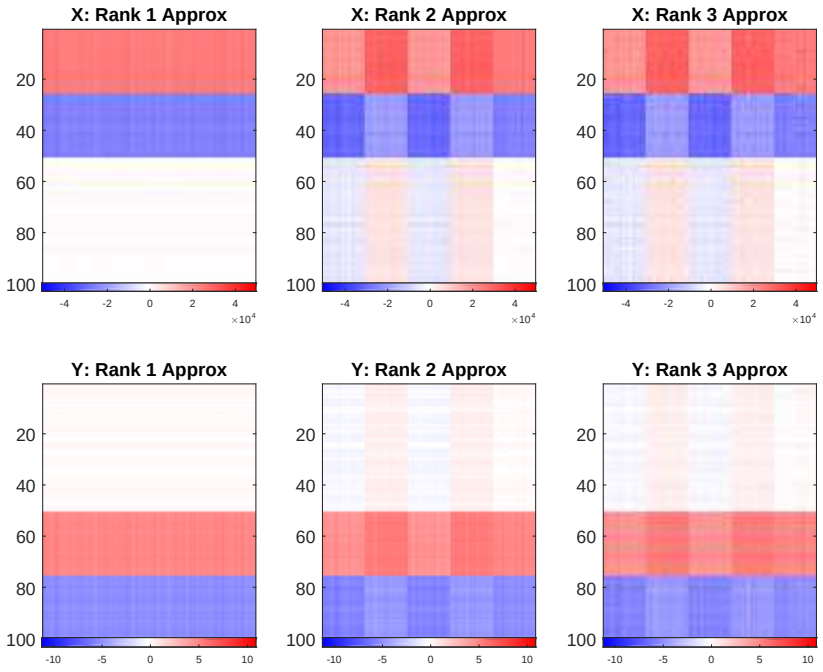

Figure 2: Shows the concatenation SVD approximation of each block for rank 1 (left), 2 (center) and 3 (right). Although block X has a relatively accurate approximation when the rank is chosen as 2 or 3, the individual pattern in block Y has never been captured due to the heterogeneity between X and Y.

The left column of Figure 3 shows the JIVE approximation of each data block which well captures the signal variations within both X and Y. What's more, JIVE distinguishes the types of variation by providing approximations of both joint and individual signal, depicted in the remaining columns. This example highlights a major strength of the proposed version of JIVE: no data normalization is needed for correct handling of disparate data, even when they differ by orders of magnitude as here. As shown in this example conventional methods require normalization which can be very challenging. Data normalization e.g. scaling by Frobenius norm will properly address the magnitude difference in this example. But simple variance re-scaling is clearly inadequate when the number of features in the two data sets are widely different such as the genomic data sets in TCGA described in Section 1.1

The rest of this paper is organized as follows. Section 2 describes the proposed method. Results of application to a TCGA breast cancer data set and a mortality data set are

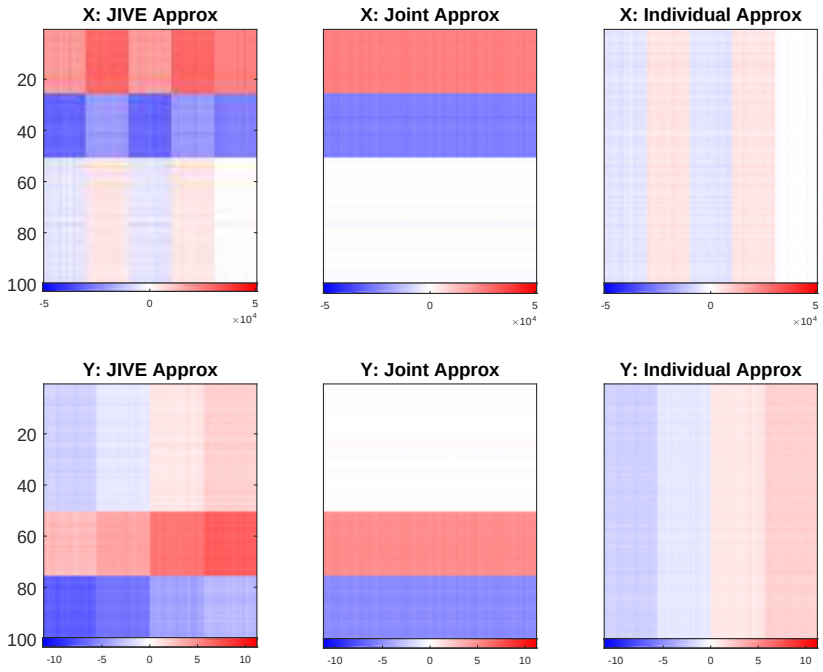

Figure 3: JIVE method approximation of the data blocks X and Y in the toy example are shown in the first column of figures, with the joint and individual signal matrices depicted in the remaining columns. Both quite diverse types of variations are well captured for each data block by new proposed JIVE.

presented in Section 3.

2 Proposed Method

In this section the details of the new proposed method are discussed. We start from a two-block data set and provide a population model for this case in Section 2.1. The theoretical foundations based on matrix perturbation theory from linear algebra (Stewart, 1990) are given in Section 2.2. These theoretical results motivate our estimation approach which is proposed in Section 2.3.

2.1 Population Model

Matrices X $(d_X \times n)$ and Y $(d_Y \times n)$ are the two data blocks for study. The columns are regarded as data objects, one for each experimental subject, while rows are considered as features. X and Y therefore have the same number of columns and perhaps different row dimensions.

Both X and Y are modeled as low rank signals A_X, A_Y perturbed by additive noise matrices E_X, E_Y. Each low rank signal is the summation of two matrices containing joint and individual variation, denoted as J_X and I_X, J_Y and I_Y respectively:

$$X = A_X + E_X = J_X + I_X + E_X$$

$$Y = A_Y + E_Y = J_Y + I_Y + E_Y$$

Our approach focuses on data objects, e.g. patient *signatures*, which are determined by the *row patterns* (essentially right basis vectors of appropriate SVDs) living in the row space, $\mathbb{R}^n$. Therefore, the joint structure matrices are defined as sharing the same row space

$$row(J_X) = row(J_Y) = row(J),$$

while the individual spaces are individual in the sense that the intersection of row spaces is the zero vector space, i.e.

$$row(I_X) \cap row(I_Y) = \{\vec{0}\}.$$

To ensure an identifiable variation decomposition, orthogonality between the row spaces of matrices containing joint and individual variation is assumed. In particular, $row(J) \perp row(I_X)$ and $row(J) \perp row(I_Y)$. Note that orthogonality between I_X and I_Y is *not* assumed. Under these assumptions, the model is identifiable in the sense:

Theorem 1 *Given matrices A_X and A_Y, there are unique matrices J_X, J_Y, I_X and I_Y so that:*

- $A_X = J_X + I_X$, $A_Y = J_Y + I_Y$
- $row(J_X) = row(J_Y) = row(J)$

- $row(J) \perp row(I_X)$, $row(J) \perp row(I_Y)$
- $row(I_X) \cap row(I_Y) = \{\vec{0}\}$

The proof is provided in the Appendix. Our model follows the JIVE idea proposed in Lock et al. (2013) but provides much fast computation, enhanced understanding and a clearer mathematical framework. Lock et al. (2013) imposed the rank constraints on the joint structure i.e. $rank(J_X) = rank(J_Y)$ but did not clearly formulate the definition of a common row space. Furthermore, the orthogonality between joint and individual structure was imposed on matrices instead of row spaces i.e. $J_X A_X^T = \mathbf{0}$, $J_Y A_Y^T = \mathbf{0}$. This did not highlight the role of row spaces in defining variation structure.

The additive noise matrices are assumed to follow an isotropic error model where the energy of projection is invariant to the direction in both row and column spaces. Important examples include the multivariate standard normal distribution and the multivariate student distribution.

The singular values of each noise matrix are assumed to be smaller than the smallest singular values of each signal to give identifiability. The assumption on the noise distribution here is much less strong than the classical i.i.d. Gaussian random matrix. Another important point is that this spherical error assumption only comes into play when determining the number of joint components. Other than that, the estimation approach given in Section 2.3 reconstructs each signal matrix based on SVD and thus is quite robust against the error distribution.

2.2 Theoretical Foundations

The main challenge is segmentation of the joint and individual variation in the presence of noise which individually perturbs each signal. The estimates of the joint spaces $row(J_X)$ and $row(J_Y)$, while expected to be similar, are no longer exactly the same due to noise. Let $\tilde{A}_X$ and $\tilde{A}_Y$ be approximations of A_X and A_Y, respectively. If some subspaces of $\tilde{A}_X$ and $\tilde{A}_Y$ have a very small angle, they can be considered as estimates of the common row space under different perturbations. Application of the results of the *Generalized* $\sin\theta$ *Theorem* (Wedin, 1972) is proposed to decide when the two subspaces are close enough to be regarded as

estimates of the joint row space. Based on this theorem, the number of joint components can be determined resulting in an appropriate segmentation.

For this, a notion of distance between two subspaces is defined. For consistency with the Generalized $\sin\theta$ Theorem, this is defined as the difference of projection matrices under the Euclidean norm. Let $\mathcal{Q}$, $\mathcal{Q}^\dagger$ be l dimensional subspaces of $\mathbb{R}^n$. The distance between the two subspaces is $\|P_{\mathcal{Q}} - P_{\mathcal{Q}^\dagger}\|$ (Stewart, 1990) and can also be written as

$$\begin{aligned}\rho(\mathcal{Q}, \mathcal{Q}^\dagger) &= \|(I - P_{\mathcal{Q}})P_{\mathcal{Q}^\dagger}\| \\ &= \|(I - P_{\mathcal{Q}^\dagger})P_{\mathcal{Q}}\|\end{aligned}$$

An insightful understanding of this defined distance $\rho(\mathcal{Q}, \mathcal{Q}^\dagger)$ comes from a principal angle analysis (Jordan, 1875; Hotelling, 1936) of the subspaces $\mathcal{Q}$ and $\mathcal{Q}^\dagger$. Denote the principal angles between $\mathcal{Q}$ and $\mathcal{Q}^\dagger$ as $\Theta(\mathcal{Q}, \mathcal{Q}^\dagger) = \{\theta_1, \cdots, \theta_l\}$ with $\theta_1 \geq \theta_2 \cdots \geq \theta_l$. The distance ρ is equal to the sine of the maximal principal angle i.e. $\sin\theta_1$. This suggests that the largest principal angle between two subspaces can indicate their closeness i.e. distance. Take $\mathcal{Q}$ as a theoretical subspace and $\mathcal{Q}^\dagger$ as its perturbed subspace. Under a slight perturbation, the largest principal angle between $\mathcal{Q}$ and $\mathcal{Q}^\dagger$ is expected to be small.

The generalized $\sin\theta$ theorem provides a bound for the sine of the largest principal angle between a subspace and its perturbation, e.g. the subspaces $\mathcal{Q}$ and $\mathcal{Q}^\dagger$. This bound quantifies how the theoretical subspace $\mathcal{Q}$ is affected by noise. Denote the SVDs of $\tilde{A}_X$ and $\tilde{A}_Y$ by $\tilde{U}_X\tilde{\Sigma}_X\tilde{V}_X^T$ and $\tilde{U}_Y\tilde{\Sigma}_Y\tilde{V}_Y^T$. The largest principal angle between the row space of A_X, A_Y and their estimates $\tilde{A}_X$, $\tilde{A}_Y$, denoted θ_X and θ_Y respectively, are bounded based on the generalized $\sin\theta$ theorem. In particular,

Theorem 2 (The Generalized $\sin\theta$ Theorem; Wedin 1972) *Let θ_X and θ_Y be the largest principal angle for each subspace of signal and its estimate. The sines of θ_X and θ_Y are bounded*

$$\sin\theta_X \leq \frac{max(\|E_X\tilde{V}_X\|, \|E_X^T\tilde{U}_X\|)}{\sigma_{min}(\tilde{\Sigma}_X)}, \tag{1}$$

$$\sin\theta_Y \leq \frac{max(\|E_Y\tilde{V}_Y\|, \|E_Y^T\tilde{U}_Y\|)}{\sigma_{min}(\tilde{\Sigma}_Y)}, \tag{2}$$

where $\sigma_{min}(\tilde{\Sigma}_X)$, $\sigma_{min}(\tilde{\Sigma}_Y)$ are the smallest singular values of $\tilde{A}_X$ and $\tilde{A}_Y$.

This bound measures how far the perturbed space can be away from the theoretical one. The deviation is bounded by the maximal value of noise energy on column and row spaces and also the smallest signal singular values. This is consistent with the intuition that a largest principal angle is small when the signal is strong and perturbations are weak. Besides, taking noise energy on both column and row spaces resolves the issue brought by high dimensionality.

Following the population model, the initial estimates of the signal row spaces contain two type of subspaces

$$row(\tilde{A}_X) = row(\tilde{J}_X) + row(\tilde{I}_X)$$

$$row(\tilde{A}_Y) = row(\tilde{J}_Y) + row(\tilde{I}_Y)$$

in which $row(\tilde{J}_X)$ and $row(\tilde{J}_Y)$ are the individually perturbed row spaces of $row(J)$; $row(\tilde{I}_X)$ and $row(\tilde{I}_Y)$ are the row spaces of the estimated individual components orthogonal to $row(\tilde{J}_X)$ and $row(\tilde{J}_Y)$ respectively. Considering this, the principal angles between $row(\tilde{J}_X)$ and $row(\tilde{J}_Y)$ tend to be close to zero and smaller than the principal angles between $row(\tilde{I}_X)$ and $row(\tilde{I}_Y)$. Therefore, segmentation into joint and individual can be done by a principal angle analysis of subspaces $row(\tilde{A}_X)$ and $row(\tilde{A}_Y)$. The largest principal angle between joint components should be bounded in terms of the perturbation bounds for individual deviation from the common subspace i.e. θ_X and θ_Y given in the generalized $\sin\theta$ theorem. In particular,

Lemma 1 *Let θ be the largest principal angle between two subspaces that are each perturbation of the common row space within $row(A_X)$ and $row(A_Y)$. That angle is bounded by*

$$\theta \leq \theta_X + \theta_Y$$

in which θ_X and θ_Y are the angles given in Theorem 2.

The proof is provided in the Appendix.

Notice that the bound in Theorem 2 is applicable but cannot be directly used for data analysis since the error matrices E_X and E_Y are not observable. As the error matrices are assumed to be isotropic, we propose to use bootstrap to re-sample noisy directions from the

residuals of the low rank approximations. The Euclidean norm of these error related terms can thus be estimated by projecting observed data onto the subspace spanned by re-sampled directions. This bootstrap based method can also provide confidence intervals of these perturbation bounds. Taking the bound as a threshold, the joint and individual structures within the row spaces of $\tilde{A}_X$ and $\tilde{A}_Y$ can be segmented. More details of estimating the perturbation bound will be discussed in Section 2.3.

2.3 Estimation Approach

The algorithm uses SVD as a building block to find an estimate of the targeted decomposition . A three-step algorithm is outlined below.

1. Obtain an initial estimate of the signal row space of each data block by thresholding the singular values.

2. Extract the joint row space from the signal row spaces using the bound of Lemma 2.

3. Decompose each data matrix into joint and individual variation matrices using projections onto the row space in Step 2.

As a basic illustration for each step we use the toy example described in Section 1. Details for each step appear in the following subsections.

2.3.1 Signal Space Initial Extraction

Even though the signal components A_X and A_Y are low rank, data matrices X and Y are usually of full rank due to corruption by noise. SVD works as a signal extraction device in this step, keeping components with singular values greater than selected thresholds individually for each data block. These thresholds are selected using a multi-scale perspective. For example, by finding relatively big jumps in a Scree plot. Figure 4 shows the Scree plots of each data block for the toy example in Section 1. Both plots suggest selection of rank as 2 since the first two components stand out while the rest of the singular values decay slowly showing no clear jump.

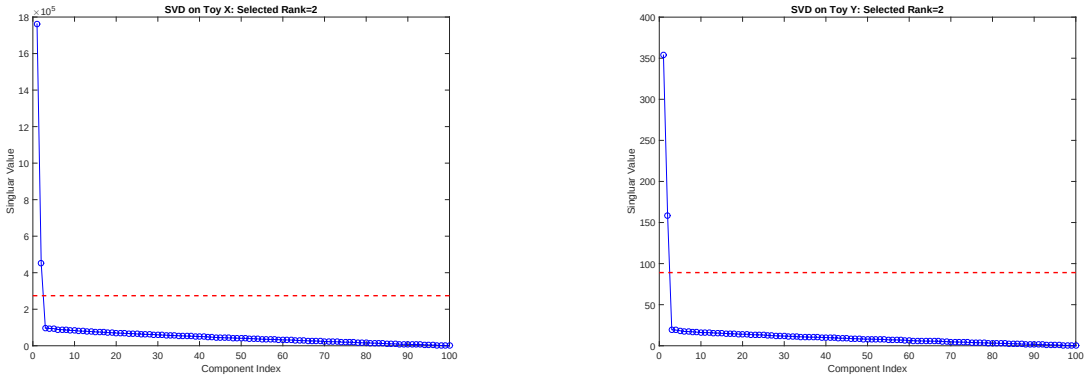

Figure 4: Scree plots for the toy data sets X (left) and Y (right). Both plots display the singular values associated with a component in descending order versus the index of the component. The two components with singular values above the dotted red threshold line are regarded as the initial signal components in the first step of JIVE.

Let $\tilde{r}_X$, $\tilde{r}_Y$ be the initial estimates of the signal ranks r_X and r_Y. In the toy example $\tilde{r}_X = \tilde{r}_Y = 2$. Each data block has a low rank approximation, $\tilde{A}_X$ and $\tilde{A}_Y$, as the initial estimates of the signal matrices A_X and A_Y

$$\tilde{A}_X = \tilde{U}_X \tilde{\Sigma}_X \tilde{V}_X^T$$

$$\tilde{A}_Y = \tilde{U}_Y \tilde{\Sigma}_Y \tilde{V}_Y^T$$

where $\tilde{U}_X$ keeps the left singular vectors corresponding to the largest $\tilde{r}_X$ singular values and $\tilde{U}_Y$ is similarly defined. The initial estimates of the signal row spaces, denoted as $row(\tilde{A}_X)$ and $row(\tilde{A}_Y)$, are spanned by the right singular vectors in $\tilde{V}_X$ and $\tilde{V}_Y$ respectively.

2.3.2 Row Space Segmentation

The segmentation into joint and individual subspaces is based on studying principal angles between the initial estimate of signal row spaces and the perturbation bound derived in Lemma 2.

The principal angles between $row(\tilde{A}_X)$ and $row(\tilde{A}_Y)$ are computed by performing SVD

on a concatenation of their right singular vector matrices, (Miao and Ben-Israel, 1992) i.e.

$$M \triangleq \begin{bmatrix} \tilde{V}_X^T \\ \tilde{V}_Y^T \end{bmatrix} = U_M \Sigma_M V_M^T$$

where the singular values determine the principal angles, $\Theta(row(\tilde{A}_X), row(\tilde{A}_Y)) = \{\theta_1, \cdots, \theta_l\}$ as

$$\theta_i = \arccos(1 - (\sigma_{M,i})^2), \ i = 1, \cdots, l = min(\tilde{r}_X, \tilde{r}_Y).$$

Figure 5 depicts the principal angles of the concatenated right singular vector matrices for the toy example. The associated principal angles between the initially estimated signal row spaces are labeled next to the first two component as 2.42 and 74.82 (degree).

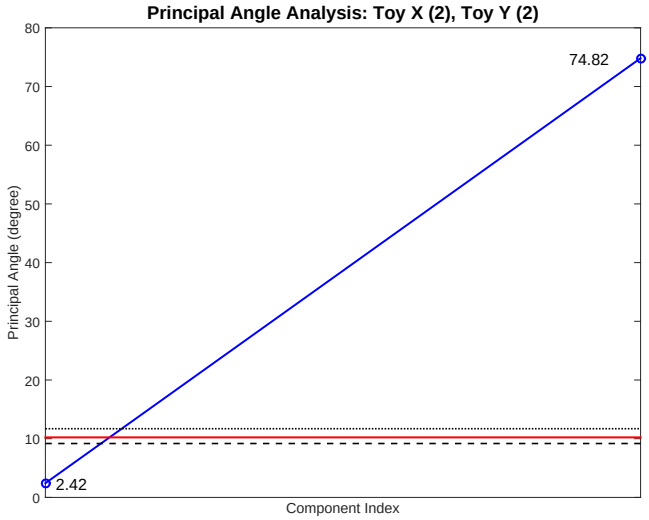

Figure 5: Principal angles between the initial estimates of signal row spaces. The bound for the largest angle is 10.43, suggesting the existence of one joint component. This correctly captures the underlying structure of this toy example.

Besides, given a left singular vector in U_M i.e. u_M, a pair of principal vectors in each subspace can be constructed by projecting $\tilde{V}_X$ and $\tilde{V}_Y$ onto the vector u_M. Denote u_M as the concatenation of $[u_{M,X}; u_{M,Y}]$. Note that the length of $u_{M,X}$ is equal to the number of columns of $\tilde{V}_X$ and similarly for the other part. The principal vectors in each subspace can be written as $\tilde{V}_X u_{M,X}$ and $\tilde{V}_Y u_{M,Y}$ respectively. The angle between the pair of principal vectors is equal to the principal angle computed from the singular value corresponding to

u_M. The right singular vector in V_M points in the same direction as the sum of principal vector pairs of each subspace.

This SVD decomposition can be understood as a tool sorting the pairs of directions within the two subspaces in increasing order of the angle between each pair. When the corresponding principal angle is smaller than the perturbation bound θ, the pair of principal vectors can be considered as pointing in the same joint direction under different perturbations. The joint matrices within each data block are defined to have the same row space. To assure such definition, the right singular vector can be taken as an estimate of the theoretical joint direction. Assume there are $\hat{r}_J$ principal angles smaller than the bound θ. The first $\hat{r}_J$ right singular vectors are used as the spanning set of the estimate of $row(J)$ i.e. $row(\hat{J})$.

As mentioned in Section 2.2, the bound θ requires the estimation of terms $\|E_X\tilde{V}_X\|$, $\|E_X^T\tilde{U}_X\|$, $\|E_Y\tilde{V}_Y\|$ and $\|E_Y^T\tilde{U}_Y\|$. These terms are the measurements of energies of noise matrices projected onto the signal column and row spaces. Since an isotropic error model is assumed, the energy of noise matrices on arbitrary directions are supposed to be equal. Thus, we propose to resample noisy directions from the residuals of the low rank approximations. Given a subspace spanned by the resampled directions, the observed data blocks are projected onto this subspace and the Euclidean norm is calculated. A typical number of resamples is 1000. The quantiles of this distribution provide a simulated confidence interval of the perturbation bound. Typically median is chosen as the estimate of angle bound for exploratory analysis. For certain cases that a tight angle bound is desired, the 5^{th} quantile can by used as a conservative threshold for finding joint components. Given the threshold, the joint and individual structures within the row spaces of $\tilde{A}_X$ and $\tilde{A}_Y$ can be segmented.

The estimated perturbation bound is 10.43 (degrees) for the toy example. This bound together with its confidence interval suggests that the number of joint components should be taken to be 1. The corresponding right singular vector is taken as the estimate of the orthonormal basis of the joint space.

2.3.3 Final Decomposition

Based on the estimate of the joint row space, matrices containing joint variation in each data block can be reconstructed by projecting X and Y onto this estimated space. Define the

matrix $\hat{V}_J$ as $[v_{M,1}, \cdots, v_{M,\hat{r}_J}]$, where $v_{M,i}$ is the i^{th} column in the matrix V_M. The projection matrix is

$$P_J = \hat{V}_J(\hat{V}_J^T\hat{V}_J^{-1}\hat{V}_J^T \tag{3}$$

and the estimates of joint structure of X and Y are

$$\hat{J}_X = XP_J \tag{4}$$

$$\hat{J}_Y = YP_J \tag{5}$$

The row space of joint structure is orthogonal to the row spaces of each individual structure. Therefore, the original data blocks are projected to the orthogonal space of $row(\hat{J})$. The projection matrix onto the orthogonal space of $row(\hat{J})$ is $P_J^\perp = I - P_J$ and the projections of each data block are denoted as $X^\perp$, $Y^\perp$ respectively i.e.

$$X^\perp = XP_J^\perp \tag{6}$$

$$Y^\perp = YP_J^\perp \tag{7}$$

Finally we threshold this projection by performing SVD on $X^\perp$, $Y^\perp$. The components with singular values larger than the first thresholds from Section 2.3.1 are kept as the individual components, denoted as $\hat{I}_X$ and $\hat{I}_Y$. The remaining components of each SVD are regarded as an estimate of the noise matrices.

By taking a union of the estimated row spaces of each type of variation, the estimated signal row spaces are

$$row(\hat{A}_X) = row(\hat{J}) + row(\hat{I}_X)$$

$$row(\hat{A}_Y) = row(\hat{J}) + row(\hat{I}_Y)$$

with rank $\hat{r}_X = \hat{r}_J + \hat{r}_{IX}$, $\hat{r}_Y = \hat{r}_J + \hat{r}_{IY}$ respectively.

Due to the adjustment of directions of joint components, these final estimates of signal row spaces may be different from those obtained in the initial signal extraction step. Note that even the estimates of rank $\hat{r}_X$ and $\hat{r}_Y$ might also differ from the initial estimates $\tilde{r}_X$, $\tilde{r}_Y$.

2.4 Post JIVE Analysis

Given the variation decompositions of each data block, several types of post JIVE analyses are available for exploring the joint and individual sample variation patterns. The estimates of joint variation within each data block can be represented by SVD

$$\hat{J}_X = \hat{U}_J^X \hat{\Sigma}_J^X \hat{V}_J^X$$

$$\hat{J}_Y = \hat{U}_J^Y \hat{\Sigma}_J^Y \hat{V}_J^Y$$

in which $\hat{V}_J^X$, $\hat{V}_J^Y$ are the $\hat{r}_J \times n$ spanning basis matrices of the estimated common joint row space $row(\hat{J})$. Note that the singular values $\hat{\Sigma}_J^X$, $\hat{\Sigma}_J^Y$ can be completely different, since they are driven by the sample variation pattern and can reflect very different amounts of variation between the blocks. The loading matrices $\hat{U}_J^X$ $(d_X \times \hat{r}_J)$, $\hat{U}_J^Y$ $(d_Y \times \hat{r}_J)$ respectively specify two distinct column spaces.

There are three important matrix representations of the information in the joint space, with differing uses in post JIVE analyses.

1. *Full Matrix Representation.* For applications where the original features are the main focus (such as finding driving genes) the full matrix representations $\hat{J}_X$ $(d_X \times n)$ and $\hat{J}_Y$ $(d_Y \times n)$ must be used.

2. *Block Specific Score (BSS).* For applications where the relationships between subjects are the main focus (such as discrimination between subtypes) large computational gains are available for using the much low dimensional representations $\hat{\Sigma}_J^X \hat{V}_J^X$ $(\hat{r}_J \times n)$ and $\hat{\Sigma}_J^Y \hat{V}_J^Y$ $(\hat{r}_J \times n)$. This results in no loss of information when rotation invariant methods are used.

3. *Common Normalized Score (CNS).* When it is desirable to study the component of joint behavior that is separate from within block variations, the analysis should focus on a common basis of $row(\hat{J})$, $\hat{V}_J$ $(\hat{r}_J \times n)$ from Section 2.3.3.

The relationship between BSS and CNS is analogous to that of the traditional covariance (i.e PLS) and correlation (i.e CCA) analysis.

The individual variation within blocks can be similarly analyzed resulting in both BSS and CNS analyses for the individual components. When original features are important, the full matrix

$$\hat{I}_X = \hat{U}_I^X \hat{\Sigma}_I^X \hat{V}_I^X$$

$$\hat{I}_Y = \hat{U}_I^Y \hat{\Sigma}_I^Y \hat{V}_I^Y$$

with dimension $d_X \times n$ and $d_Y \times n$ are available. Otherwise large computational savings are available from the BSS version $\hat{\Sigma}_I^X \hat{V}_I^X$ $(\hat{r}_{IX} \times n)$and $\hat{\Sigma}_I^Y \hat{V}_I^Y$ $(\hat{r}_{IY} \times n)$. For studying scale free behaviors, use the *Individual Normalized Score (INS)* $\hat{V}_I^Y$ $(\hat{r}_{IX} \times n)$ and $\hat{V}_I^Y$ $(\hat{r}_{IY} \times n)$.

3 Data Analysis

In this section, we apply Non-iterative JIVE to two real data sets, TCGA breast cancer (Hu et al., 2015) and Spanish mortality as analyzed in (Marron and Alonso, 2014). Detailed analyses are given in Section 3.1 and Section 3.2.

3.1 TCGA Data

A prominent goal of modern cancer research, of which TCGA is a major resource, is the combination of biological insights from multiple types of measurements made on common subjects. JIVE is a powerful new tool for gaining such insights. Here gene expression and copy number features measured on a common set of breast cancer samples are taken as an example. Genetic subtypes have proven to be fundamental to precision medicine, so insights about these will be used to assess the performance of JIVE.

We perform JIVE for several subsets of the data, including all tumors, HER2 and Luminal, and Luminal alone. Table 1 states the variation explained by JIVE decomposition for each subset. As shown in the table, most of the copy number variation (about 80%) is joint with gene expression for all of these subsets. On the other hand, the gene expression data contains a much larger percentage of individual variation (about 60%) that differs from copy number. This observation is consistent with expected biology because copy number

variation tends to generate variation in gene expression, while there are many other sources of variation that also drive gene expression.

Data source	Comparison	Joint	Individual
All Tumors	Gene expression	35%	65%
	Copy number	80%	20%
HER2 & Luminal	Gene expression	34%	66%
	Copy number	73%	27%
Luminal Only	Gene expression	41%	59%
	Copy number	81%	19%

Table 1: Percentage of variation explained by joint BSS structure, individual BSS structure for gene expression and copy number data. Shows that copy number variation mainly associates with gene expression, but gene expression is more diverse as expected.

Additional biological insights come from post analysis of these JIVE decompositions. Subtype differences are explored by performing classifications on both joint and individual variation. This was done using both the CNS/INS and the BSS data representations. Results are similar so only CNS results are shown here. This gives a straightforward joint analysis because it is based on the common set of joint scores. The classification directions are obtained by Distance Weight Discrimination(DWD) (Marron et al., 2007) which is useful because of the high dimensional nature of these data. Class differences are quantified by DiProPerm hypothesis tests (Wei et al., 2015) based on 100 permutations. Strength of the evidence is usually measured by permutation p-values. However, in this case most p-values are zero. Thus a more interpretable measure of strength of evidences is to provide DiProPerm z-scores. We also report the area under the ROC curve (AUC) (Hanley and McNeil, 1982), to show the classification accuracy.

Figure 6 presents the results of classification analysis of joint variation within the three data subsets. Each panel shows a separation of subtypes by projecting the CNS of joint structure onto the DWD discrimination direction. The dots are a jitter plot of the data, using colors and symbols to distinguish the subtypes. Each symbol is a data point whose

horizontal coordinate is the value and vertical coordinate is the height based on data ordering. The curves are Gaussian kernel density estimates i.e. smoothed histograms, which show the distribution of the subtypes.

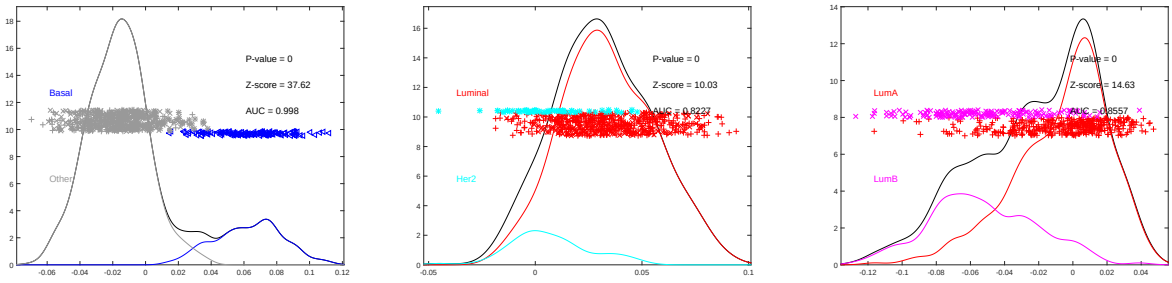

Figure 6: One dimensional projection of joint structures onto the DWD discriminant direction. Basal-like vs. the other tumor subtypes (left), HER2 vs. Luminal(center) and Luminal A vs. B (right). A strong separation is apparent between Basal and the other tumor subtypes, while there is more overlap for the other two classifications. This contrast indicates different discriminatory power of joint variations between these different subsets of gene expression and copy number.

The left plot of Figure 6 presents a clear visual separation between Basal-like and other tumor subtypes. The high value z-score of 37.6 and AUC also suggest a strongly significant class difference. The middle plot visualizes the discrimination between HER2 and Luminal. Although the z-score of 10.0 from DiProPerm indicates a significant difference, the visual separation is not as large. The separation between Luminal A and B, depicted in the right plot, is still not as strong as the Basal-like vs. the other tumor subtypes but has stronger evidence than HER2 and Luminal suggested by the DiProPerm z-score of 14.6. The contrast of separations in Figure 6 indicates the distinct discriminant powers of the joint signals within different data subsets. The joint signal between gene expression and copy number shows strong power for distinguishing Basal-like from the other tumor subtypes but is not quite as powerful for the other two class comparisons. This contrast is consistent with the known biological fact that the Basal-like subtype has much stronger copy number variations than the Her2 subtype.

A similar study is conducted for the individual variation within gene expression and copy number, which reveals a contrast with the joint variation. Table 2 gives the DiProPerm

z-scores and AUCs for the INSs of each individual variation. Differing from the joint variation, the individual variations within copy number do not have power for distinguishing Basal-like from Other tumors, and Luminal A from B. The table shows that the DiProPerm z-scores are not significant and the AUCs are almost equivalent to random guessing(around 0.5). For these two class comparisons, the individual variation within gene expression still present substantial discriminant power but much weaker than the joint variation. The insignificant separation of individual variation within copy number and the dramatic decrease in discriminant power of individual variation within gene expression suggest that the class differences are mostly explained by the joint variation between gene expression and copy number. Besides, in view of the fact shown in Table 1 that gene expression contains a large proportion of individual variation, this is a strong indicator that the individual structure of gene expression may be driven by some additional biological components. A further investigation could be a clustering analysis of these individual variations to identity new subtypes which might lead to better treatments.

The discrimination between HER2 and Luminal tells a different story. The individual variations within both gene expression and copy number present significant discriminatory power; in particular, the individual gene expression has an even better classification than its joint variation. This suggests that copy number features may not work jointly with gene expression features to distinguish HER2 and Luminal.

Data source	Data Type	Z-score (P-value)	AUC
All Tumors	Gene expression	9.61 (0)	0.7829
	Copy number	1.1 (0.145)	0.5663
HER2 & Luminal	Gene expression	20.64 (0)	0.9643
	Copy number	9.37 (0)	0.7551
Luminal Only	Gene expression	11.77 (0)	0.8052
	Copy number	0.67 (0.267)	0.5704

Table 2: Z-scores and AUC of individual structure in classifying different pairwise classes. Except HER2 versus Luminal, the other two comparisons indicate a less significant discrimination in the individual variation.

A further understanding of these genomic sources can be obtained by looking at the loading plots given by each classification. In particular, we have identified a set of gene expression features associated with a set of copy number features that work together to separate the compared classes.

3.2 Spanish Mortality Data

A quite different data set from the Human Mortality Database is studied here, which is consist of both male and female Spanish people. This data set further demonstrates the advantage of JIVE in gaining insights. For each gender data block, there is a matrix of *mortality*, defined as the number of people who died divided by the total, for a given age group and year. Because mortality varies by several orders of magnitude, the $\log_{10}$ of the mortality is studied here. Each row represents an age group from 0 to 95, and each column represents a year between 1908 and 2002. In order to associate the historical events with the variations of mortality, columns (i.e. mortality as a function of age) are considered as the common set of data objects of each gender block. Marron and Alonso (2014) performed analysis on the male block and showed interesting interpretations related to Spanish history. Here we are looking for a deeper analysis which integrates both males and females by exploring joint and individual variation patterns.

Non-iterative JIVE is applied to the two gender blocks centered by subtracting the mean of each age group, since the mean structure contains essential variation information. The most interesting JIVE analysis comes from 3 male and 3 female components. The resulting JIVE gives 2 joint components and 1 of each individual component. Since the loading matrices provide important information of the effect of different age groups, BSS analysis together with loading matrices is most informative here.

Figure 7 shows a view of the first joint components for the males (left) and females (right) that is very different from the heat map views used in Section 1.1. While these components are matrices, additional insights come from plotting the rows of the matrices as curves over year (top) and the columns as curves over age (bottom). The curves over year (top) are colored using a heat color scheme, indexing age (black = 0 through red = 40 to yellow = 95 as shown in the vertical color bar on the bottom left). The curves over

age (bottom) are colored using a rainbow color scheme (magenta = 1908 through green = 1960 to red = 2002, shown in the horizontal color bar in the top) and use the vertical axis as domain with horizontal axis as range to highlight the fact that these are column vectors. Additional visual cues to the matrix structure are the horizontal rainbow color bar in the top panel, showing that year indexes columns of the data matrix and the vertical heat color bar (bottom) showing that age indexes rows of the component matrix. Because this is a single component, i.e. a rank one approximation of the data, each curve is a multiple of a single eigenvector. The corresponding coefficients are shown on the right. In conventional PCA/SVD terminology, the upper BSS coefficients are called *loadings*, and are in fact the entries of the left eigenvectors (colored using the heat color scale on the bottom). Similarly, the lower coefficients are called *scores* and are the entries of the right eigenvectors, colored using the rainbow bar shown in the top.

The scores plots together with the rows as curves plots in Figure 7 indicate a dramatic improvement in mortality after the 1950s for both males and females. The scores plots are bimodal indicating rapid overall improvement in mortality around the the 1950s. This is also visible in the rows as curves plot. Thus the first mode of joint variation is driven by overall improvement in mortality. In addition to the overall improvement, the rows as curves and scores plot also show the major mortality events, the global flu pandemic of 1918 and the Spanish Civil war in the late 1930s. The loading plots together with the columns as curves plots present the different impacts of this common variation on different age groups for males and females. The loadings plot for males suggests the improvement in mortality is gradually increasing from older towards younger age groups. In contrast, the female block has a bimodal kernel density estimate of the loadings. This shows that female of child bearing age have received large benefits from improving health care. This effect is similarly visible from comparing the female versus male columns as curves.

The second BSS components of joint variation within each gender are similarly visualized in Figure 8. This common variation reflects the contrast between the years around 1950 and the years around 1980 which can be told from the curves in the left top and the colors in the right bottom subplots in both male and female panel. In the scores plot, the green circles, seen on the left end, represent the years around 1950 when the automobile penetration

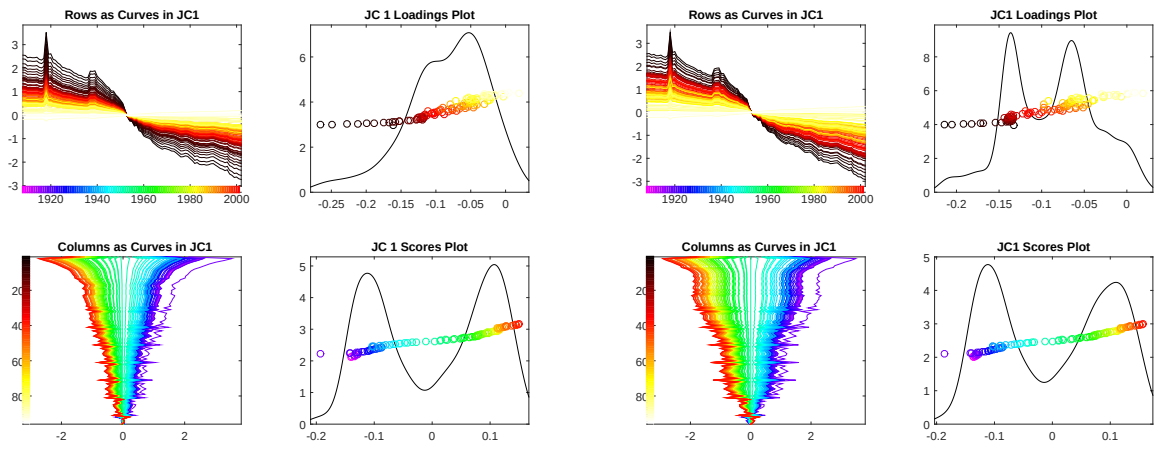

Figure 7: The first BSS joint components of male (left panel) and female (right panel) contain the common modes of variation caused by the overall improvement across different age groups, as can be seen from the scores plots in the right bottom of each panel. The dramatic decrease happened around the 1950s shown in the column projection plot. The decrease degrees vary from age groups.

started. And the orange to red circles on the right end correspond to years around 1980, after seat belt legislation was first introduced in Spain. These modes of variation can be interpreted as the increase in fatalities caused by automobiles and later improvements in safety such as seat belts and safer roads. The upper left loadings plot of males shows that these automobile events had a stronger influence on the 20-45 males in terms of both larger values and a second peak in the kernel density estimate. Although this contrast can also be seen in the loadings plot of females, it is not as strong as for the male block. The JC2 loadings plots show an interesting outlier, the babies of age zero. We speculate this shows an effect in improvement of post-natal care that coincidently happened around the same time.

Another interesting result comes from the studying first individual components (IC1) of males and females, shown in Figure 9. In the scores plot of males (left), the blue circles stand out from the rest, corresponding to the years of the Spanish civil war when a significant spike can be seen in the rows as curves plot. Young to middle age groups are affected more than the others which can be found in the loadings plot and columns as curves plot. Such year variation pattern, however, cannot be detected in the individual variation component

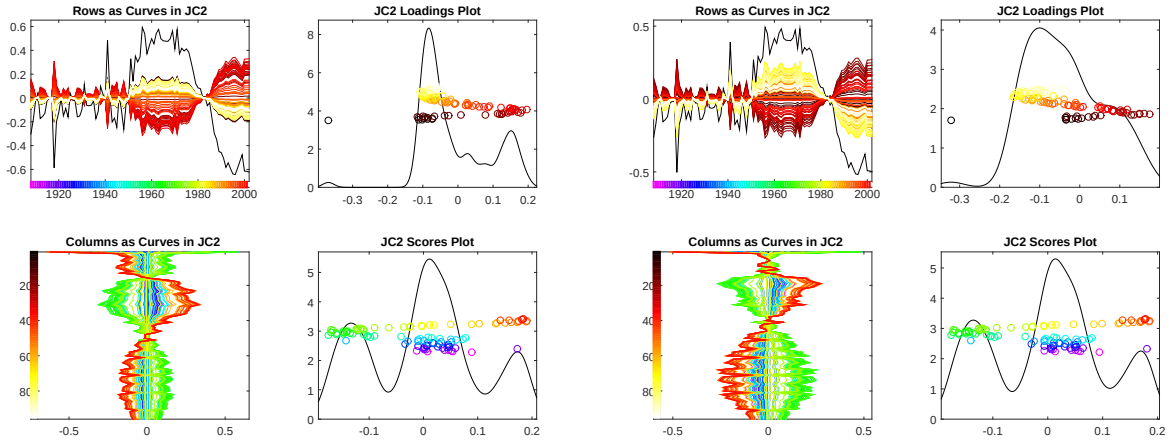

Figure 8: The second joint components of male (left) and female (right) contain the common modes of variation driven by the increase in fatalities caused by automobile penetration and later improvement due to safety improvements. This can be seen from the scores plots in the right bottom. The loadings plots show that this automobile event exerted a significantly stronger impact on the 20-45 males.

of females. The columns as curves plot on the lower left suggest some type of 5-year age rounding effect, which is seen to occur mostly during the earlier years as indicated both in the rows as curves plot and the colors of the peaks in the columns as curves plot. Note that the plot scales show that the individual female effects are much smaller in magnitude than the male effects.

Acknowledgements

This research was supported in part by the National Science Foundation under Grant No. 1016441 and 1512945.

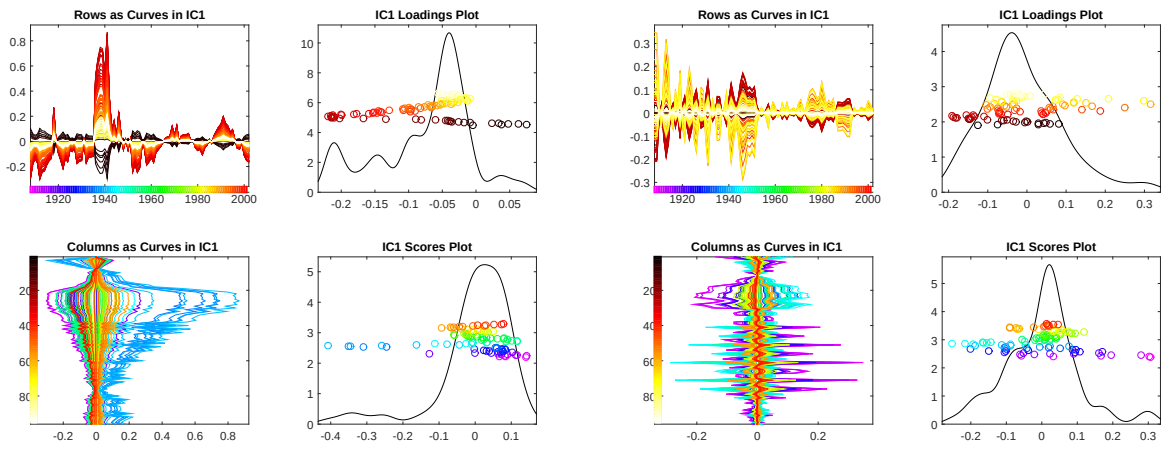

Figure 9: The individual component of male (left) contains the variation driven by Spanish civil war which can be seen from the blue circles on the right end of right bottom plot. The Spanish civil war mainly affected the young to middle age male.

4 Appendix

Theorem 3 (Existence of the decomposition) *Let A be the matrix concatenation of the signal matrices A_X, A_Y. There exist matrices J_X, J_Y, I_X and I_Y such that*

$$A = \begin{bmatrix} A_X \\ A_Y \end{bmatrix} = \begin{bmatrix} J_X \\ J_Y \end{bmatrix} + \begin{bmatrix} I_X \\ I_Y \end{bmatrix}$$

which satisfy the following constraints:

1. $row(J_X) = row(J_Y) \triangleq row(J)$
2. $row(I_X) \cap row(I_Y) = \{\vec{0}\}$
3. $row(J) \perp row(I_X)$ *and* $row(J) \perp row(I_Y)$

Proof 1 *Define a space $row(J)$ as the intersection of row spaces of A_X and A_Y, $row(A_X)$ and $row(A_Y)$.*

$$row(J) = row(A_X) \cap row(A_Y)$$

Then, define spaces $row(I_X)$ and $row(I_Y)$ orthogonal to $row(J)$ by performing the operation

$$row(I_X) = row(A_X) - row(J)$$

$$row(I_Y) = row(A_Y) - row(J)$$

Therefore, the three created spaces $row(J)$, $row(I_X)$ *and* $row(I_Y)$ *satisfy the three constraints in the theorem. Let* P_J *be the projection matrix onto the row space* $row(J)$*. Define*

$$J_X = A_X P_J, J_Y = A_Y P_J$$

Furthermore, let P_{IX}, P_{IY} *be projection matrices onto the row spaces* $row(I_X)$, $row(I_Y)$ *respectively. Define*

$$I_X = A_X P_{IX}, I_Y = A_Y P_{IY}$$

Based on the constraints of the three row spaces, projection matrices therefore satisfy

$$P_{AX} = P_J + P_{IX}, P_J \perp P_{IX}$$

$$P_{AY} = P_J + P_{IY}, P_J \perp P_{IY}$$

where P_{AX}, P_{AY} *are respectively projection matrices onto row spaces of* A_X *and* A_Y*. Hence, there exist matrices* J_X, J_Y, I_X *and* I_Y *such that* $A = \begin{bmatrix} A_X \\ A_Y \end{bmatrix} = \begin{bmatrix} J_X \\ J_Y \end{bmatrix} + \begin{bmatrix} I_X \\ I_Y \end{bmatrix}$ *simultaneously satisfying the stated constraints.*

Theorem 4 (uniqueness) *The matrices* J_X, J_Y, I_X *and* I_Y *in Theorem 3 are uniquely defined.*

Proof 2 *Assume there exist another three different row spaces* $row(J)^*$, $row(I_X)^*$ *and* $row(I_Y)^*$ *satisfying the constraints that*

$$row(J)^* + row(I_X)^* = row(A_X)$$

$$row(J)^* + row(I_Y)^* = row(A_Y)$$

$$row(J)^* \perp row(I_X)^*, row(J)^* \perp row(I_X)^*$$

$$row(I_X)^* \cap row(I_Y)^* = \{\vec{0}\}$$

Since, for $\forall \alpha \in row(J)^*$, $\alpha \in row(A_X)$ *and* $\alpha \in row(A_Y)$ *indicating* $\alpha \in row(A_X) \cap row(A_Y) = row(J)$, $row(J)^* \subsetneq row(J)$.

Assume a vector $\beta \in row(J) - row(J)^*$, then $\beta \in row(I_X)^*$ and $\beta \in row(I_Y)^*$. Thus, $row(I_X)^* \cap row(I_Y)^* \neq \{\vec{0}\}$ which contradicts the constraints.

Due to the uniqueness of the row spaces of each matrix, J_X, J_Y, I_X and I_Y in Theorem 3 are uniquely defined.

Lemma 2 *Let θ be the largest principal angle between two subspaces that are results of individually perturbed common row space. Such angle is bounded by*

$$\theta \leq \theta_X + \theta_Y$$

in which θ_X and θ_Y are the angles given in Theorem 2.

Proof 3 *Let $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ be the projection matrices onto the individually perturbed joint row spaces. And let P be the projection matrices onto the common joint row space J. Thus, we have*

$$\begin{aligned}
\sin\theta &= \|(I - P_{\mathcal{X}})P_{\mathcal{Y}}\| && (8)\\
&= \|(I - P_{\mathcal{X}})(I - P + P)P_{\mathcal{Y}}\| && (9)\\
&\leq \|(I - P_{\mathcal{X}})(I - P)P_{\mathcal{Y}}\| + \|(I - P_{\mathcal{X}})PP_{\mathcal{Y}}\| && (10)\\
&= \|(I - P_{\mathcal{X}})(I - P)(I - P)P_{\mathcal{Y}}\| + \|(I - P_{\mathcal{X}})PPP_{\mathcal{Y}}\| && (11)\\
&\leq \|(I - P_{\mathcal{X}})(I - P)\|\|(I - P)P_{\mathcal{Y}}\| + \|(I - P_{\mathcal{X}})P\|\|PP_{\mathcal{Y}}\| && (12)
\end{aligned}$$

in which $\|(I - P_{\mathcal{X}})P\| = \sin\theta_X$, $\|(I - P_{\mathcal{Y}})(I - P)\| = \cos\theta_X$, $\|(I - P_{\mathcal{Y}})P\| = \sin\theta_Y$ and $\|(I - P_{\mathcal{Y}})(I - P)\| = \cos\theta_Y$. There,

$$\sin\theta \leq \cos\theta_X \sin\theta_Y + \sin\theta_X \cos\theta_Y = \sin(\theta_X + \theta_Y)$$

and we have $\theta \leq \theta_X + \theta_Y$.

References

A. Cichocki Y. Zhang G. Zhou and D. Mandic. Group component analysis from multi-block data: Common and individual feature extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 2015.

Mohamed Hanafi, Achim Kohler, and El-Mostafa Qannari. Connections between multiple co-inertia analysis and consensus principal component analysis. *Chemometrics and intelligent laboratory systems*, 106(1):37–40, 2011.

James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, pages 321–377, 1936.

Yi-Juan Hu, Wei Sun, Jung-Ying Tzeng, and Charles M Perou. Proper use of allele-specific expression improves statistical power for cis-eqtl mapping with rna-seq data. *Journal of the American Statistical Association*, (just-accepted), 2015.

Shashank Jere, Justin Dauwels, Muhammad Tayyab Asif, Nikola Mitro Vie, Andrzej Cichocki, and Patrick Jaillet. Extracting commuting patterns in railway networks through matrix decompositions. In *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, pages 541–546. IEEE, 2014.

Camille Jordan. Essai sur la géométrie à n dimensions. *Bulletin de la Société mathématique de France*, 3:103–174, 1875.

Oliver Kühnle. *Integration of multiple high-throughput data-types in cancer research*. PhD thesis, 2011.

Julia Kuligowski, David Pérez-Guaita, Ángel Sánchez-Illana, Zacarías León-González, Miguel de la Guardia, Máximo Vento, Eric F Lock, and Guillermo Quintás. Analysis of multi-source metabolomic data using joint and individual variation explained (jive). *Analyst*, 2015.

Eric F Lock and David B Dunson. Bayesian consensus clustering. *Bioinformatics*, page btt425, 2013.

Eric F Lock, Katherine A Hoadley, JS Marron, and Andrew B Nobel. Joint and individual variation explained (jive) for integrated analysis of multiple data types. *The annals of applied statistics*, 7(1):523, 2013.

J Steve Marron and Andrés M Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 2014.

JS Marron, Michael J Todd, and Jeongyoun Ahn. Distance-weighted discrimination. *Journal of the American Statistical Association*, 102(480):1267–1271, 2007.

Jianming Miao and Adi Ben-Israel. On principal angles between subspaces in R^n. *Linear algebra and its applications*, 171:81–98, 1992.

Qianxing Mo, Sijian Wang, Venkatraman E Seshan, Adam B Olshen, Nikolaus Schultz, Chris Sander, R Scott Powers, Marc Ladanyi, and Ronglai Shen. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proceedings of the National Academy of Sciences*, 110(11):4245–4250, 2013.

Age K Smilde, Johan A Westerhuis, and Sijmen de Jong. A framework for sequential multiblock component methods. *Journal of chemometrics*, 17(6):323–337, 2003.

Gilbert W Stewart. Matrix perturbation theory. 1990.

Johan Trygg and Svante Wold. O2-pls, a two-block (x±y) latent variable regression (lvr) method with an integral osc® lter^2. *J. chemometrics*, 17:53–64, 2003.

Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972.

Susan Wei, Chihoon Lee, Lindsay Wichers, and JS Marron. Direction-projection-permutation for high dimensional hypothesis tests. *Journal of Computational and Graphical Statistics*, (just-accepted), 2015.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M Stuart, Cancer Genome Atlas Research Network, et al. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Johan A Westerhuis, Theodora Kourti, and John F MacGregor. Analysis of multiblock and hierarchical pca and pls models. *Journal of chemometrics*, 12(5):301–321, 1998.

Herman Wold. Partial least squares. *Encyclopedia of statistical sciences*, 1985.

Svante Wold. *PLS modeling with latent variables in two or more dimensions*. 1987.

Svante Wold, Nouna Kettaneh, and Kjell Tjessem. Hierarchical multiblock pls and pc models for easier model interpretation and as an alternative to variable selection. *Journal of chemometrics*, 10(5-6):463–482, 1996.

Yu Zhang, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Ssvep recognition using common feature analysis in brain–computer interface. *Journal of neuroscience methods*, 244:8–15, 2015.