

Effect of Data Transformation on Residue

Hyuk Cho Inderjit S. Dhillon

Department of Computer Sciences
The University of Texas at Austin
Austin, TX, 78712, USA

11th October, 2007

Technical Report # TR-07-55

Abstract

Recently, Aguilar-Ruiz [2005] considers a data matrix containing both scaling and shifting factors and shows that the mean squared residue [Cheng and Church, 2000], called RESIDUE(II) in this paper, is useful to discover shifting patterns, but not appropriate to find scaling patterns. This finding draws our attention on the weakness of RESIDUE(II) measure and the need of new approaches to discover both scaling and shifting patterns in the considered matrix. To resolve the weakness of RESIDUE(II) in finding scaling patterns, we propose a simple remedy that still uses the same residue measure. The main idea is to remove hidden scaling factors in the considered data matrix by taking a specific data transformation. We investigate various data transformations including no transformation, double centering, mean centering, standard deviation normalization, and Z-score transformation. Further, we apply these data transformations to row/column dimension of data matrix models with different global/local scaling and global/local shifting factors. First, we characterize the properties of the data transformations on different data matrix models, including six Euclidean co-clustering schemes in Bregman co-clustering algorithms [Banerjee et al., 2007] and other existing data models in the literature. In particular, we formally analyze the effect of each data transformation on the two residues [Cho et al., 2004], here called RESIDUE(I) and RESIDUE(II), respectively. Then, we apply all the data transformations to publicly available human cancer gene expression datasets and empirically validate the analysis results by using the minimum sum squared residue co-clustering (MSSRCC) algorithms [Cho et al., 2004]. In conclusion, through column standard deviation normalization or column Z-score transformation, we are able to overcome the shortcoming of RESIDUE(II) in finding scaling patterns and discover both scaling and shifting patterns.

1 Introduction

Hartigan [1972]'s pioneering work, named *direct clustering*, stimulated a vast amount of research on co-clustering algorithms. Co-clustering aims at identifying homogeneous local patterns, each of which consists of a subset of rows and a subset of columns in a given two dimensional matrix. In particular, co-clustering has attracted genomic researchers, because its idea is compatible with our understanding of cellular processes, where a subset of genes are coregulated under a certain experimental conditions, but to behave almost independently under other conditions [Ben-Dor

et al., 2003]. For a survey on the application of co-clustering to biological data analysis see the recent paper by Madeira and Oliveira [2004].

Cheng and Church [2000] are considered to be the first to apply co-clustering, called also *biclustering*, to gene expression data. They proposed a greedy search heuristic that generates biclusters, one at a time, which satisfy a certain homogeneity constraint, called *mean squared residue*. Since then, several similar approaches that make use of the residue have been proposed to enhance the work of Cheng and Church [2000]. Recently, Cho et al. [2004] developed two minimum sum squared co-clustering (MSSRCC) algorithms: one objective function is based on the partitioning model proposed by Hartigan [1972] and the other one is based on the squared residue formulated by Cheng and Church [2000]. The residues are defined in equations (2.1) and (2.2) in the next chapter.

The mean squared residue measure, which utilizes RESIDUE(II), has been popularly used, however little research on its detail properties has been done. Recently, Aguilar-Ruiz [2005] models a data matrix containing both scaling and shifting factors and formally proves that the mean squared residue measure depends on the scaling variance in the considered data matrix. This finding issues the weakness of the residue measure and the need of new approaches to discover both scaling and shifting patterns. Motivated by the data model and the approach in Aguilar-Ruiz [2005], we propose a simple remedy to address its weakness in finding scaling patterns, still using the same residue measure. We suggest to take a specific data transformation before computing the residue in order to handle hidden scaling factors. We consider no transformation, double centering, column/row mean centering, column/row standard deviation normalization, and column/row Z-score transformation. The details of these data transformations are explained in Section 3. Then, we apply each data transformation to rows or columns of data matrix models derived from different combinations of global/local scaling and global/local shifting factors. We analyze and compare the effect of data transformations of several data matrix models on the two residues [Cho et al., 2004], called RESIDUE(I) and RESIDUE(II). Furthermore, using MSSRCC, we empirically demonstrate the advantage of the data transformations with publicly available human cancer microarrays, including Colon cancer, Leukemia, Lung Cancer, and Mixed-Lineage Leukemia (MLL). Both analysis and experimental results reveal that column standard deviation normalization and column Z-score transformation are the effective for both RESIDUE(I) and RESIDUE(II).

The rest of this paper is organized as follows: In Section 2 we introduce some definitions and facts used in this paper. We describe the considered data transformations in Section 3. Then, we formally analyze the effects of data transformations and summarize the analysis results in Section 4. We discuss the experimental results with human cancer gene expression datasets in Section 5. Finally, the paper is concluded with some remark.

2 Formulation

We adapt the following definitions in Aguilar-Ruiz [2005], Cheng and Church [2000], and Cho et al. [2004] to fit for our context.

2.1 Co-cluster

DEFINITION 1 (data matrix). A data matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, whose (i, j) -th element is denoted by a_{ij} , is defined as follows:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

For example, a microarray can be defined with two finite sets, the set of genes and the set of experimental conditions. Note that Aguilar-Ruiz [2005] describes the microarray whose rows represent experimental condition and columns represent genes. However, in this paper, we will consider a microarray which consists of examples of genes in rows and attributes as experimental conditions in columns.

DEFINITION 2 (co-cluster). Let $I \subseteq \{1, 2, \dots, m\}$ denote the set of indices of the rows in a row cluster and $J \subseteq \{1, 2, \dots, n\}$ denote the set of indices of the columns in a column cluster. A submatrix of \mathbf{A} induced by the index sets I and J is called a co-cluster and denoted as $\mathbf{A}_{IJ} \in \mathbb{R}^{|I| \times |J|}$, where $|I|$ and $|J|$ denote the cardinality of index set I and index set J , respectively.

2.2 Sum squared residue

In order to evaluate the coherence of such a co-cluster, we define the two *residues*, RESIDUE(I) and RESIDUE(II), of an element a_{ij} in the co-cluster determined by index sets I and J as below.

DEFINITION 3 (residues). RESIDUE(I) is defined as

$$h_{ij} = a_{ij} - a_{IJ} \tag{2.1}$$

and RESIDUE(II) is defined as

$$h_{ij} = a_{ij} - a_{iJ} - a_{IJ} + a_{IJ}, \tag{2.2}$$

where the mean of the entries in row i whose column indices are in J is computed by $a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{ij}$, the mean of the entries in column j whose row indices are in I by $a_{IJ} = \frac{1}{|I|} \sum_{i \in I} a_{ij}$, and the mean of all the entries in the co-cluster whose row and column indices are in I and J by $a_{IJ} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} a_{ij}$.

DEFINITION 4 (sum squared residue (SSR)). Let $\mathbf{H}_{IJ} \in \mathbb{R}^{|I| \times |J|}$ be the residue matrix whose entries are described by one of the two residues, RESIDUE(I) in equation (2.1) and RESIDUE(II) in equation (2.2). Then, the sum squared residue of \mathbf{H}_{IJ} is defined as

$$SSR = \|\mathbf{H}_{IJ}\|^2 = \sum_{i \in I, j \in J} h_{ij}^2, \tag{2.3}$$

where the norm $\|\mathbf{X}\|$ denotes the Frobenius norm of matrix \mathbf{X} , i.e., $\|\mathbf{X}\|^2 = \sum_{i,j} x_{ij}^2$.

2.3 Patterns

We generalize the definition of data patterns in Aguilar-Ruiz [2005] so that data patterns contain all possible scaling and shifting factors.

DEFINITION 5 (global/local scaling (gsc/lsc) and global/local shifting (gsh/lsh) pattern). A bicluster shows both a shifting and a scaling pattern when it follows the expression

$$a_{ij} = \pi_i \times \alpha_j + \beta_j, \quad (2.4)$$

where π_i is the base value for row (e.g., gene) i , α_j is the scaling factor for column (e.g., experimental condition) j , and β_j is the shifting factor for column (e.g., experimental condition) j . Combining scaling and shifting patterns results in the following four patterns: global scaling (gsc) and global shifting pattern (gsh) when $a_{ij} = \pi_i \times \alpha + \beta$; global scaling (gsc) and local shifting pattern (lsh) when $a_{ij} = \pi_i \times \alpha + \beta_j$; local scaling (lsc) and global shifting pattern (gsh) when $a_{ij} = \pi_i \times \alpha_j + \beta$; and local scaling (lsc) and local shifting pattern (lsh) when $a_{ij} = \pi_i \times \alpha_j + \beta_j$.

We borrow the concepts of “local” and “global” scaling and shifting from Cheng and Church [2000], Cho et al. [2004], and Aguilar-Ruiz [2005]. Notice that the pattern defined in equation (2.4) is in the form of the linear (or affine) transformation. The shifting factor, α_j corresponds to the multiplicative component (or slope) and β_j corresponds to the additive component (or intercept) of the linear transformation. Madeira and Oliveira [2004] introduced other specific patterns to model a perfect co-cluster with constant rows, columns, and coherent values as follows: $a_{ij} = \pi + \gamma_i$, $a_{ij} = \pi + \beta_j$, and $a_{ij} = \pi + \gamma_i + \beta_j$, where π is the typical value within the co-cluster A_{IJ} , γ_i is the adjustment for row $i \in I$, and β_j is the adjustment of column $j \in J$. However, in this paper, we consider the data matrix that includes both shifting and scaling patterns denoted in equation (2.4).

2.4 Facts in data matrix \mathbf{A}

We use the following facts in a data matrix \mathbf{A} , throughout the paper.

Means. The mean of the base values of \mathbf{A} is defined by $\mu_\pi = \frac{1}{m} \sum_{i=1}^m \pi_i$, the mean of the scaling factors by $\mu_\alpha = \frac{1}{n} \sum_{j=1}^n \alpha_j$ for lsc and $\mu_\alpha = \alpha$ for gsc, and the mean of the shifting factors by $\mu_\beta = \frac{1}{n} \sum_{j=1}^n \beta_j$ for lsh and $\mu_\beta = \beta$ for gsh. Also, the mean of row i is computed by $a_{i.} = \frac{1}{n} \sum_{j=1}^n a_{ij} = \pi_i \alpha + \beta$ for gsc and lsh, $a_{i.} = \pi_i \alpha + \mu_\beta$ for gsc and lsh, $a_{i.} = \pi_i \mu_\alpha + \beta$ for lsc and gsh, and $a_{i.} = \pi_i \mu_\alpha + \mu_\beta$ for lsc and lsh. The mean of column j is obtained by $a_{.j} = \frac{1}{m} \sum_{i=1}^m a_{ij} = \mu_\pi \alpha + \beta$ for gsc and gsh, $a_{.j} = \mu_\pi \alpha + \beta_j$ for gsc and lsh, $a_{.j} = \mu_\pi \alpha_j + \beta$ for lsc and gsh, and $a_{.j} = \mu_\pi \alpha_j + \beta_j$ for lsc and lsh. The mean of all the elements is computed by $a_{..} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n a_{ij} = \mu_\pi \alpha + \beta$ for gsc and gsh, $a_{..} = \mu_\pi \alpha + \mu_\beta$ for gsc and lsh, $a_{..} = \mu_\pi \mu_\alpha + \beta$ for lsc and gsh, and $a_{..} = \mu_\pi \mu_\alpha + \mu_\beta$ for lsc and lsh.

Variance. The variance of the base values is defined as $\sigma_\pi^2 = \frac{1}{m} \sum_{i=1}^m (\pi_i - \mu_\pi)^2$, the variance of the scaling factors as $\sigma_\alpha^2 = \frac{1}{n} \sum_{j=1}^n (\alpha_j - \mu_\alpha)^2$, and the variance of the shifting factors as $\sigma_\beta^2 = \frac{1}{n} \sum_{j=1}^n (\beta_j - \mu_\beta)^2$. Let a and b be real constants and X and Y be discrete random variables, then the variances of X and Y , denoted as σ_X^2 and σ_Y^2 , respectively, satisfy the properties:

1. $\sigma_a^2 = \sigma_b^2 = 0$;
2. $\sigma_{(aX+b)}^2 = a^2 \sigma_X^2$;
3. $\sigma_{(aX+bY)}^2 = a^2 \sigma_X^2 + b^2 \sigma_Y^2$.

Therefore, the variance of row i is computed as $\sigma_{i.}^2 = 0$ for gsc and gsh, $\sigma_{i.}^2 = \sigma_\beta^2$ for gsc and lsh, $\sigma_{i.}^2 = \pi_i^2 \sigma_\alpha^2$ for lsc and gsh, and $\sigma_{i.}^2 = \pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2$ for lsc and lsh. The variance of column j is computed as $\sigma_{.j}^2 = \sigma_\pi^2 \alpha^2$ for gsc and $\sigma_{.j}^2 = \sigma_\pi^2 \alpha_j^2$ for lsc.

Note that neither global nor local shifting factor affects the column variance, $\sigma_{.j}$, since variance of a constant shifting factor added to each column is 0. For the same reason, neither row mean

centering nor column mean centering affects the row variance, σ_i , and the column variance, σ_j , respectively.

3 Data Transformations

Raw data values have a limitation that raw values do not disclose how they vary from the central tendency of the distribution. Therefore, transformation of the raw data is considered one of the most important steps for various data mining processes since the variance of a variable will determine its importance in a given model [Sánchez et al., 1994]. Recently, Wouters et al. [2003] emphasize the importance of appropriate weighting in the analysis of microarray data, where they compare different transformation methods as the building blocks of three multivariate projection methods. In addition, Kluger et al. [2003] incorporate different normalizations of genes and conditions in the hope of discarding the irrelevant constant background noise. An early study on preprocessing can be found in Harshman and Lundy [1984]. Recently, Bro and Smilde [2003] and Smilde et al. [2004] discuss a number of important features of the common preprocessing steps of centering and scaling, focussing on two-way bilinear data analysis and generalizing the results to multiway data analysis.

In this study we investigate the following data transformations and the details are explained in the sequel. Further, we summarize in Table 1 the data transformations of the data matrix whose entry a_{ij} is defined in equation (2.4).

No transformation (NT). No centering or scaling is taken. In other words, $a'_{ij} = a_{ij}$, *i.e.*, the raw matrix is directly input to MSSRCC.

Double centering (DC.) Double centering is defined as

$$a'_{ij} = a_{ij} - a_{i.} - a_{.j} + a_{..}$$

for $i = 1, \dots, m$ and $j = 1, \dots, n$. Through DC, each entry of a data matrix \mathbf{A} becomes $a'_{ij} = (\pi_i - \mu_\pi)(\alpha_j - \mu_\alpha)$. Note that we have $a'_{i.} = a'_{.j} = 0$ and consequently $a'_{..} = 0$, since DC transforms the data matrix to have both row means and column means to be 0. DC has become a classic technique for background correction. After this centering, the data matrix is projected on a hyperplane that runs through the origin and is orthogonal to the all-ones vector. It leads to a reduction by one of the rank of the original matrix [Wouters et al., 2003].

Column/row mean centering (MC). Column mean centering is defined as

$$a'_{ij} = a_{ij} - a_{.j}$$

for $i = 1, \dots, m$ and $j = 1, \dots, n$. Row mean centering is defined similarly with $\mu_{i.}$. Through column MC, each entry becomes $a'_{ij} = \pi_i\alpha_j + \beta_j - \mu_{.j}$. Therefore, row mean, column mean, and whole mean become $a'_{i.} = \pi_i\mu_\alpha + \mu_\beta - a_{..}$, $a'_{.j} = \mu_\pi\alpha_j + \beta_j - a_{.j}$, and $a'_{..} = \mu_\pi\mu_\alpha + \mu_\beta - a_{..}$, respectively. Through row MC, each entry becomes $a'_{ij} = \pi_i\alpha_j + \beta_j - a_{i.}$. Similarly, row mean, column mean, and whole mean become $a'_{i.} = \pi_i\mu_\alpha + \mu_\beta - a_{i.}$, $a'_{.j} = \mu_\pi\alpha_j + \beta_j - a_{..}$, and $a'_{..} = \mu_\pi\mu_\alpha + \mu_\beta - a_{..}$, respectively.

Column/row standard deviation normalization (SDN). Column standard deviation normalization is defined as

$$a'_{ij} = \frac{a_{ij}}{\sigma_{.j}}$$

for $i = 1, \dots, m$ and $j = 1, \dots, n$. Row standard deviation normalization is defined similarly with $a_{i.}$ and $\sigma_{i.}^2$. Through column SDN each column has a unit variance and through row SDN each row has a unit variance..

Table 1: Data transformation

Type	Target	Transformed data matrix
NT	n/a	$\begin{pmatrix} \pi_1\alpha_1 + \beta_1 & \pi_1\alpha_2 + \beta_2 & \cdots & \pi_1\alpha_n + \beta_n \\ \pi_2\alpha_1 + \beta_1 & \pi_2\alpha_2 + \beta_2 & \cdots & \pi_2\alpha_n + \beta_n \\ \vdots & \vdots & \vdots & \vdots \\ \pi_m\alpha_1 + \beta_1 & \pi_m\alpha_2 + \beta_2 & \cdots & \pi_m\alpha_n + \beta_n \end{pmatrix}$
DC	n/a	$\begin{pmatrix} a_{11} - a_{1.} - a_{.1} + a_{..} & a_{12} - a_{1.} - a_{.2} + a_{..} & \cdots & a_{1n} - a_{1.} - a_{.n} + a_{..} \\ a_{21} - a_{2.} - a_{.1} + a_{..} & a_{22} - a_{2.} - a_{.2} + a_{..} & \cdots & a_{2n} - a_{2.} - a_{.n} + a_{..} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} - a_{m.} - a_{.1} + a_{..} & a_{m2} - a_{m.} - a_{.2} + a_{..} & \cdots & a_{mn} - a_{m.} - a_{.n} + a_{..} \end{pmatrix}$
MC	Row	$\begin{pmatrix} \pi_1\alpha_1 + \beta_1 - a_{1.} & \pi_1\alpha_2 + \beta_2 - a_{1.} & \cdots & \pi_1\alpha_n + \beta_n - a_{1.} \\ \pi_2\alpha_1 + \beta_1 - a_{2.} & \pi_2\alpha_2 + \beta_2 - a_{2.} & \cdots & \pi_2\alpha_n + \beta_n - a_{2.} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_m\alpha_1 + \beta_1 - a_{m.} & \pi_m\alpha_2 + \beta_2 - a_{m.} & \cdots & \pi_m\alpha_n + \beta_n - a_{m.} \end{pmatrix}$
	Col	$\begin{pmatrix} \pi_1\alpha_1 + \beta_1 - a_{.1} & \pi_1\alpha_2 + \beta_2 - a_{.2} & \cdots & \pi_1\alpha_n + \beta_n - a_{.n} \\ \pi_2\alpha_1 + \beta_1 - a_{.1} & \pi_2\alpha_2 + \beta_2 - a_{.2} & \cdots & \pi_2\alpha_n + \beta_n - a_{.n} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_m\alpha_1 + \beta_1 - a_{.1} & \pi_m\alpha_2 + \beta_2 - a_{.2} & \cdots & \pi_m\alpha_n + \beta_n - a_{.n} \end{pmatrix}$
SDN	Row	$\begin{pmatrix} \frac{1}{\sigma_{1.}} (\pi_1\alpha_1 + \beta_1) & \frac{1}{\sigma_{1.}} (\pi_1\alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{1.}} (\pi_1\alpha_n + \beta_n) \\ \frac{1}{\sigma_{2.}} (\pi_2\alpha_1 + \beta_1) & \frac{1}{\sigma_{2.}} (\pi_2\alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{2.}} (\pi_2\alpha_n + \beta_n) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma_{m.}} (\pi_m\alpha_1 + \beta_1) & \frac{1}{\sigma_{m.}} (\pi_m\alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{m.}} (\pi_m\alpha_n + \beta_n) \end{pmatrix}$
	Col	$\begin{pmatrix} \frac{1}{\sigma_{.1}} (\pi_1\alpha_1 + \beta_1) & \frac{1}{\sigma_{.2}} (\pi_1\alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{.n}} (\pi_1\alpha_n + \beta_n) \\ \frac{1}{\sigma_{.1}} (\pi_2\alpha_1 + \beta_1) & \frac{1}{\sigma_{.2}} (\pi_2\alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{.n}} (\pi_2\alpha_n + \beta_n) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma_{.1}} (\pi_m\alpha_1 + \beta_1) & \frac{1}{\sigma_{.2}} (\pi_m\alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{.n}} (\pi_m\alpha_n + \beta_n) \end{pmatrix}$
ZT	Row	$\begin{pmatrix} \frac{1}{\sigma_{1.}} (\pi_1\alpha_1 + \beta_1 - a_{1.}) & \frac{1}{\sigma_{1.}} (\pi_1\alpha_2 + \beta_2 - a_{1.}) & \cdots & \frac{1}{\sigma_{1.}} (\pi_1\alpha_n + \beta_n - a_{1.}) \\ \frac{1}{\sigma_{2.}} (\pi_2\alpha_1 + \beta_1 - a_{2.}) & \frac{1}{\sigma_{2.}} (\pi_2\alpha_2 + \beta_2 - a_{2.}) & \cdots & \frac{1}{\sigma_{2.}} (\pi_2\alpha_n + \beta_n - a_{2.}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma_{m.}} (\pi_m\alpha_1 + \beta_1 - a_{m.}) & \frac{1}{\sigma_{m.}} (\pi_m\alpha_2 + \beta_2 - a_{m.}) & \cdots & \frac{1}{\sigma_{m.}} (\pi_m\alpha_n + \beta_n - a_{m.}) \end{pmatrix}$
	Col	$\begin{pmatrix} \frac{1}{\sigma_{.1}} (\pi_1\alpha_1 + \beta_1 - a_{.1}) & \frac{1}{\sigma_{.2}} (\pi_1\alpha_2 + \beta_2 - a_{.2}) & \cdots & \frac{1}{\sigma_{.n}} (\pi_1\alpha_n + \beta_n - a_{.n}) \\ \frac{1}{\sigma_{.1}} (\pi_2\alpha_1 + \beta_1 - a_{.1}) & \frac{1}{\sigma_{.2}} (\pi_2\alpha_2 + \beta_2 - a_{.2}) & \cdots & \frac{1}{\sigma_{.n}} (\pi_2\alpha_n + \beta_n - a_{.n}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma_{.1}} (\pi_m\alpha_1 + \beta_1 - a_{.1}) & \frac{1}{\sigma_{.2}} (\pi_m\alpha_2 + \beta_2 - a_{.2}) & \cdots & \frac{1}{\sigma_{.n}} (\pi_m\alpha_n + \beta_n - a_{.n}) \end{pmatrix}$

Abbreviations: DC – Double Centering; MC – Mean Centering; NT – No Transformation; SDN – Standard Deviation Normalization; and ZT – Z-score Transformation.

Column/row Z-score transformation (ZT). Column standardization is defined as

$$a'_{ij} = \frac{a_{ij} - a_{.j}}{\sigma_{.j}}$$

for $i = 1, \dots, m$ and $j = 1, \dots, n$. Row standardization is defined similarly with $a_{i.}$ and σ_i^2 . It is also called “autoscaling”, where the measurements are scaled so that each column/row has a zero mean and a unit variance [Kowalski and Bender, 1972]. Through ZT, the relative variation in intensity is emphasized, since ZT is a linear transformation, which keeps the relative positions of observations and the shape of the original distribution. In fact, column ZT is the combination of column MC and column SDN. Similarly, row ZT is the combination of row MC and row SDN.

4 Analysis

In this section, we analyze the effect of the data transformations on the residues defined in equations (2.1) and (2.2). In reality, rows and columns in a co-cluster are not necessary to be consecutive. However, for brevity we consider the following co-cluster, \mathbf{A}_{IJ} , whose entries consist of first $|I|$ rows and first $|J|$ columns in \mathbf{A} as

$$\mathbf{A}_{IJ} = \begin{pmatrix} \pi_1\alpha_1 + \beta_1 & \pi_1\alpha_2 + \beta_2 & \cdots & \pi_1\alpha_{|J|} + \beta_{|J|} \\ \pi_2\alpha_1 + \beta_1 & \pi_2\alpha_2 + \beta_2 & \cdots & \pi_2\alpha_{|J|} + \beta_{|J|} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_{|I|}\alpha_1 + \beta_1 & \pi_{|I|}\alpha_2 + \beta_2 & \cdots & \pi_{|I|}\alpha_{|J|} + \beta_{|J|} \end{pmatrix}.$$

Because of space limitation, we focus on analyzing RESIDUE(II) for the three data transformations including NT, column SDN, and column ZT, which clearly demonstrate the effect of the specific data transformation. However, for comparison purpose, we provide the results of RESIDUE(I) and RESIDUE(II) for all the data transformations in Tables 5 and 9, respectively, each of which is classified based on target dimension, scaling type, and shifting type. The residues summarized in Tables 5 and 9 for the other remaining data transformations can be easily obtained, following the similar steps in this section (see Appendix).

4.1 No transformation (NT)

(i, j) -th entry of row $i \in I$ and column $j \in J$ of co-cluster \mathbf{A}_{IJ} is described as $a_{ij} = \pi_i\alpha_j + \beta_j$. Then, the mean of the base values of \mathbf{A}_{IJ} is computed by $\mu_{\pi_I} = \frac{1}{|I|} \sum_{i \in I} \pi_i$. and the mean of the scaling factors by $\mu_{\alpha_J} = \frac{1}{|J|} \sum_{j \in J} \alpha_j$, and the mean of the shifting factors by $\mu_{\beta_J} = \frac{1}{|J|} \sum_{j \in J} \beta_j$. Also, the mean of row i is obtained by $a_{iJ} = \pi_i\mu_{\alpha_J} + \mu_{\beta_J}$, the mean of column j by $a_{Ij} = \mu_{\pi_I}\alpha_j + \beta_j$, and the mean of all the elements by $a_{IJ} = \mu_{\pi_I}\mu_{\alpha_J} + \mu_{\beta_J}$. Using these values, we obtain RESIDUE(II), $h_{ij} = (\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$. Consequently, the sum squared residue (SSR) can be computed as

$$\begin{aligned} SSR &= \|\mathbf{H}_{IJ}\|^2 \\ &= \sum_{i \in I, j \in J} h_{ij}^2 \\ &= \sum_{i \in I, j \in J} (\pi_i - \mu_{\pi_I})^2 (\alpha_j - \mu_{\alpha_J})^2 \\ &= |I||J|\sigma_{\pi_I}^2\sigma_{\alpha_J}^2, \end{aligned} \tag{4.1}$$

Table 2: Transformed data matrix

Type	Target	Transformed data matrix
SDN	Col	$\begin{pmatrix} \frac{1}{\sigma_{i,1}} (\pi_1 \alpha_1 + \beta_1) & \frac{1}{\sigma_{i,2}} (\pi_1 \alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{i,n}} (\pi_1 \alpha_n + \beta_n) \\ \frac{1}{\sigma_{i,1}} (\pi_2 \alpha_1 + \beta_1) & \frac{1}{\sigma_{i,2}} (\pi_2 \alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{i,n}} (\pi_2 \alpha_n + \beta_n) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma_{i,1}} (\pi_m \alpha_1 + \beta_1) & \frac{1}{\sigma_{i,2}} (\pi_m \alpha_2 + \beta_2) & \cdots & \frac{1}{\sigma_{i,n}} (\pi_m \alpha_n + \beta_n) \end{pmatrix}$
		$= \frac{1}{\sigma_\pi} \begin{pmatrix} \pi_1 + \frac{\beta_1}{\alpha_1} & \pi_1 + \frac{\beta_2}{\alpha_2} & \cdots & \pi_1 + \frac{\beta_n}{\alpha_n} \\ \pi_2 + \frac{\beta_1}{\alpha_1} & \pi_2 + \frac{\beta_2}{\alpha_2} & \cdots & \pi_2 + \frac{\beta_n}{\alpha_n} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_m + \frac{\beta_1}{\alpha_1} & \pi_m + \frac{\beta_2}{\alpha_2} & \cdots & \pi_m + \frac{\beta_n}{\alpha_n} \end{pmatrix}$
ZT	Col	$\begin{pmatrix} \frac{1}{\sigma_{i,1}} (\pi_1 \alpha_1 + \beta_1 - a_{i,1}) & \frac{1}{\sigma_{i,2}} (\pi_1 \alpha_2 + \beta_2 - a_{i,2}) & \cdots & \frac{1}{\sigma_{i,n}} (\pi_1 \alpha_n + \beta_n - a_{i,n}) \\ \frac{1}{\sigma_{i,1}} (\pi_2 \alpha_1 + \beta_1 - a_{i,1}) & \frac{1}{\sigma_{i,2}} (\pi_2 \alpha_2 + \beta_2 - a_{i,2}) & \cdots & \frac{1}{\sigma_{i,n}} (\pi_2 \alpha_n + \beta_n - a_{i,n}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma_{i,1}} (\pi_m \alpha_1 + \beta_1 - a_{i,1}) & \frac{1}{\sigma_{i,2}} (\pi_m \alpha_2 + \beta_2 - a_{i,2}) & \cdots & \frac{1}{\sigma_{i,n}} (\pi_m \alpha_n + \beta_n - a_{i,n}) \end{pmatrix}$
		$= \frac{1}{\sigma_\pi} \begin{pmatrix} \pi_1 - \mu_\pi & \pi_1 - \mu_\pi & \cdots & \pi_1 - \mu_\pi \\ \pi_2 - \mu_\pi & \pi_2 - \mu_\pi & \cdots & \pi_2 - \mu_\pi \\ \vdots & \vdots & \vdots & \vdots \\ \pi_m - \mu_\pi & \pi_m - \mu_\pi & \cdots & \pi_m - \mu_\pi \end{pmatrix}$

Abbreviations: SDN – Standard Deviation Normalization; ZT – Z-score Transformation.

where $\sigma_{\pi_I}^2 = \frac{1}{|I|} \sum_{i \in I} (\pi_i - \mu_{\pi_I})^2$ and $\sigma_{\alpha_J}^2 = \frac{1}{|J|} \sum_{j \in J} (\alpha_j - \mu_{\alpha_J})^2$.

In fact, SSR shown in equation (4.1) is a revisit of Theorems in Aguilar-Ruiz [2005], where mean squared residue was used with no data transformation. Equation (4.1) shows that SSR is dependent on both the variance of base values and the variance of scaling factors, but independent from shifting factors. Accordingly, any shifting operations such as DC and MC to the given data matrix should not contribute to RESIDUE(II) (see RESIDUE(II) values for NT, DC, and MC in Table 9).

4.2 Column standard deviation normalization (SDN)

Originally, data matrix \mathbf{A} contains both local scaling and local shifting factors. Through column SDN, it is transformed to have the constant global scaling factor, *i.e.*, 1, and the local shifting factors, *i.e.*, $\frac{\beta_j}{\alpha_j}$ (see Table 2). To be more specific, (i, j) -th entry is transformed as $a_{ij} = \frac{1}{\sigma_j} (\pi_i \alpha_j + \beta_j) = \frac{1}{\sigma_j} \left(\pi_i + \frac{\beta_j}{\alpha_j} \right)$. Then, row mean, column mean, and whole mean of co-cluster \mathbf{A}_{IJ} are computed by $a_{iJ} = \frac{1}{|J|} \sum_{j \in J} \frac{1}{\sigma_j} (\pi_i \alpha_j + \beta_j)$, $a_{Ij} = \frac{1}{|I|} \sum_{i \in I} \frac{1}{\sigma_j} (\pi_i \alpha_j + \beta_j)$, and $a_{IJ} = \frac{1}{|I||J|} \sum_{i \in I} \sum_{j \in J} \frac{1}{\sigma_j} (\pi_i \alpha_j + \beta_j)$, respectively. Therefore, using RESIDUE(II), we can capture the perfect co-cluster, *i.e.*, zero RESIDUE(II), for all the possible combinations of global/local scaling and global/local shifting patterns of a data matrix. Furthermore, RESIDUE(II) captures the coherent patterns noted in Cho et al. [2004] as $x e^T + e y^T$, where $e = [11 \cdots 1]^T$, and x and y are arbitrary column vectors. However, using RESIDUE(I), we may not capture the perfect co-cluster as summarized in Table 5.

4.3 Column Z-score transformation (ZT)

As illustrated in Table 2, through column ZT, data matrix \mathbf{A} is transformed data matrix to have the constant global scaling factor, *i.e.*, 1, and the constant global shifting factor, *i.e.*, $-\mu_{\pi}$. To be more specific, (i, j) -th entry is transformed as $a_{ij} = \frac{1}{\sigma_j} (\pi_i \alpha_j + \beta_j - a_{.j}) = \frac{1}{\sigma_j} (\pi_i - \mu_{\pi})$. Then, row mean of co-cluster \mathbf{A}_{IJ} is obtained by $a_{iJ} = \frac{1}{\sigma_j} (\mu_{\pi_i} - \mu_{\pi}) = a_{ij}$, and column mean and whole mean by $a_{Ij} = \frac{1}{\sigma_j} (\mu_{\pi_I} - \mu_{\pi}) = a_{IJ}$. Like column SDN, we obtain zero REDIDUE(II) for all the possible combinations of scaling and shifting patterns (see Table 9). However, using MSSRCC with RESIDUE(I), we may not capture the perfect co-clusters because SSR is still dependent on the variance of base values.

5 Experimental results

In this section, we empirically show the effect of data transformations on the four publicly available human cancer microarray datasets including Colon cancer [Alon et al., 1999], Leukemia [Golub et al., 1999], Lung cancer [Armstrong et al., 2002], and MLL [Armstrong et al., 2002]. For MSSRCC, we fix $\tau = 10^{-3} \|\mathbf{A}\|^2$ for batch updates and $\tau = 10^{-6} \|\mathbf{A}\|^2$ for local search steps. With spectral initialization and local search strategy, MSSRCC generates 100×2 or 100×3 co-clusters within 20 seconds for all the datasets.

Data preprocessing. Since we use publicly available gene expression matrices, they have already been preprocessed in various ways using image analysis, expression quantization, normalization, and screening out. Therefore, the numerical values in each dataset are very different; the number of genes are thousands, while the number of samples are relatively small, often less than a hundred;

Table 3: Description of microarray datasets used in our experiments.

	Colon	Leukemia	Lung	MLL
# original genes	2000	7129	12533	12582
# samples	62	72	181	72
# sample classes	2	2	2	3
Sample class names	Normal(20) Tumor(42)	ALL(47) AML(25)	ADCA(150) MPM(31)	ALL(24) AML(25) MLL(23)
$ max/min $	15	5	5	5
$ max - min $	500	500	600	5500
# remaining genes	1096	3571	2401	2474

Abbreviations: ALL – Acute Lymphoblastic Leukemia; AML – Acute Myeloid Leukemia; ADCA – Adenocarcinoma; MPM – Malignant Pleural Mesothelioma; and MLL – Mixed-Lineage Leukemia. The number after each sample class name denotes the number of samples in the sample class.

and many of the genes are noisy and redundant. These make feature selection an important issue. However, instead of utilizing sophisticated feature selection algorithms, we apply the following simple preprocessing steps usually adopted in microarray experiments (Bø and Jonassen [2002], Dettling and Bühlmann [2002], Dudoit and Fridlyand [2002]) to detect differential expression: we filter out genes whose relative deviation ($|max/min|$) or absolute deviation ($|max - min|$) is less than predefined values, where max and min refer respectively to the maximum and minimum expression levels for a particular gene across all samples. Further, the gene expression values in Colon dataset were transformed by taking the base-10 logarithm. Details of each dataset are summarized in Table 3.

Tissue sample clustering evaluation measure. To evaluate the performance of sample clusterings, we quantify tissue sample clustering performance using the following clustering accuracy measure:

$$Accuracy(\%) = \frac{1}{T} \left(\sum_{i=1}^l t_i \right) \times 100,$$

where T denotes the total number of samples, l denotes the number of sample clusters, and t_i denotes the numbers of the samples correctly clustered into a sample class i . We first form a confusion matrix where (i, j) -th entry gives the number of samples in cluster i that belong to the true class j . Each t_i is a diagonal element of the corresponding confusion matrix whose cluster labels are permuted so that sum of diagonal elements is maximized.

Performance comparison. Figure 1 illustrates the average performance of tissue sample accuracy using MSSRCC with RESIDUE(I). The accuracy values are varied with different data transformations, because RESIDUE(I) is affected by different factors as summarized in Table 5. For example, through column ZT, RESIDUE(I) is independent from both scaling and shifting factors and hence generates the best accuracy values ((a), (b), and (d)). Through DC or column MC, RESIDUE(I) is not affected by shifting factors, but still affected by scaling factors. DC generates the second best accuracy values ((a), (c), and (d)) and both DC and column MC gives similar performance ((a) and (d)). Furthermore, there are three interesting cases in Figure 1: (1) column SDN and column ZT for Leukemia dataset ((b)); (2) column SDN and DC for both Leukemia and MLL datasets ((b) and (d)); and (3) ZT and column MC for Leukemia, Lung, and MLL datasets

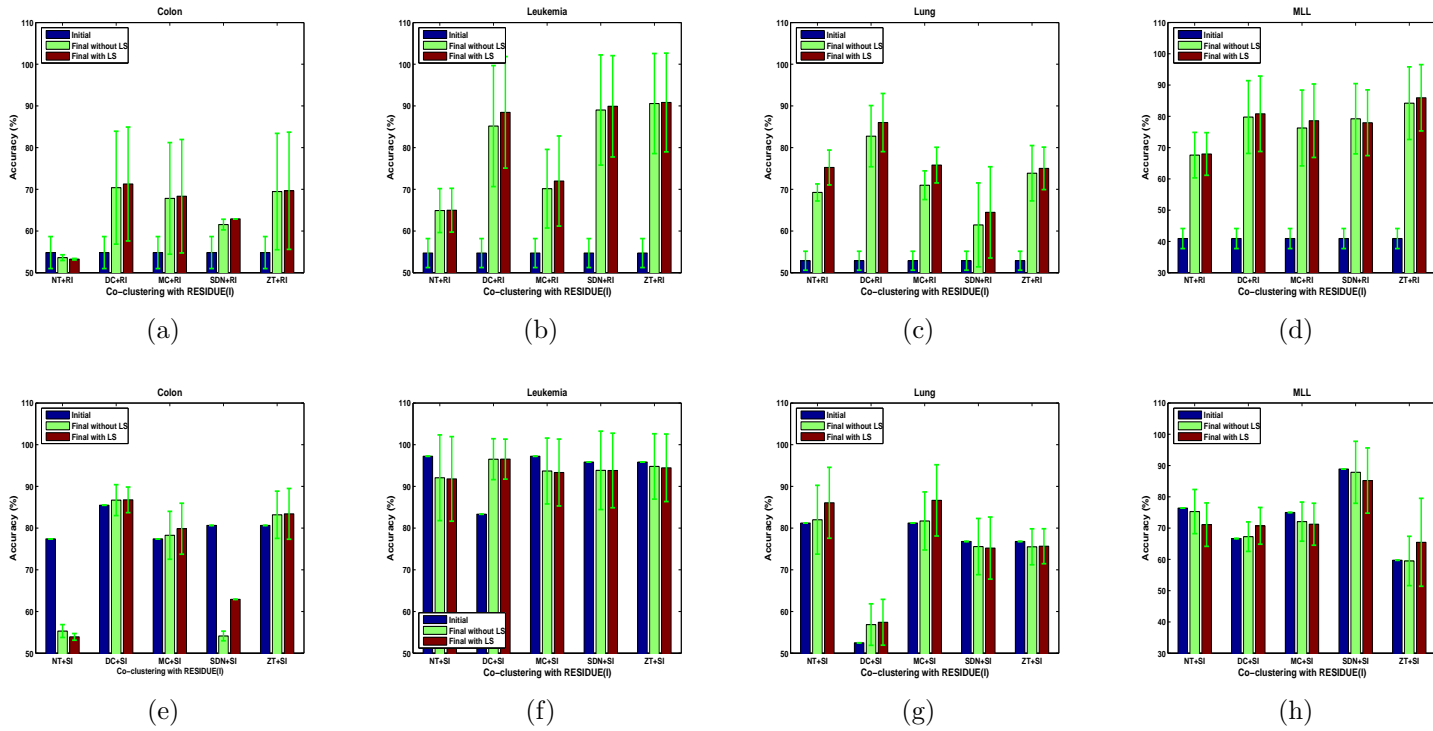


Figure 1: Average tissue sample clustering accuracy using MSSRCC with RESIDUE(I). The accuracy values are averaged over 1 to 100 gene clusters. (a)-(d) are averaged over 10 random runs and (e)-(h) are obtained with deterministic spectral initialization. Abbreviations: RI – Random Initialization; SI – Spectral Initialization; NT – No Transformation; DC – Double Centering; MC – (column) Mean Centering; SDN – (column) Standard Deviation Normalization; ZT – (column) Z-score Transformation; and LS – Local Search.

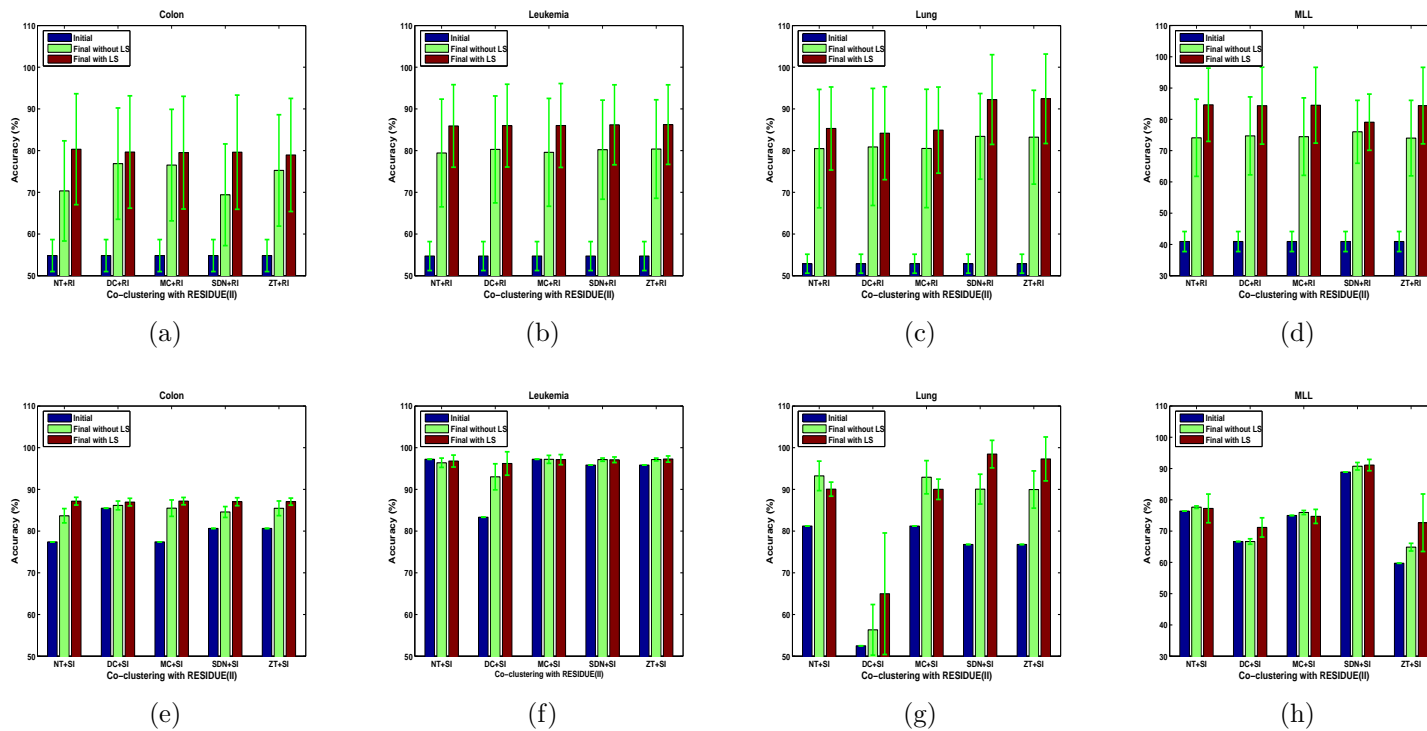


Figure 2: Average tissue sample clustering accuracy using MSSRCC with RESIDUE(II). The accuracy values are averaged over 1 to 100 gene clusters. (a)-(d) are averaged over 10 random runs and (e)-(h) are obtained with deterministic spectral initialization. Abbreviations: RI – Random Initialization; SI – Spectral Initialization; NT – No Transformation; DC – Double Centering; MC – (column) Mean Centering; SDN – (column) Standard Deviation Normalization; ZT – (column) Z-score Transformation; and LS – Local Search.

((f), (g), and (h)). We interpret these experimental results based on RESIDUE(I) in Table 5 as follows: (1) and (2) may satisfy that the variance of $\frac{\beta_j}{\alpha_j}$ (with column SDN) is negligible, i.e., $\frac{\beta_j}{\alpha_j} \approx \frac{1}{|J|} \sum_{j \in J} \frac{\beta_j}{\alpha_j}$ for each of the two datasets; (3) may be the case that both $\beta_j \approx \mu_{\beta_j}$ (with ZT) and $\alpha_j \approx \mu_{\alpha_j}$ (with column MC).

Figure 2 elucidates that the experimental results with RESIDUE(II) are consistent with the analytic results in Table 9. For example, NT, DC, and MC with RI ((a)-(d)) and NT and MC with SI ((e)-(h)) result in almost identical accuracy performance, since RESIDUE(II) with NT, DC, or MC is $(\pi_i - \mu_{\pi_i})(\alpha_j - \mu_{\alpha_j})$ as shown in Table 9. Furthermore, as analyzed in the previous section, both column SDN and column ZT help MSSRCC with RESIDUE(II) to capture perfect co-clusters. Therefore, they are supposed to generate similar performance and Figure 2 clearly verifies the consistency of the experimental results.

Overall performance of NT with RESIDUE(I) is not better than that of other data transformations ((a)-(d) in Figure 1). DC has been widely used in many applications. For example, Torgerson [1952] used the double-centering formula in metric multi-dimensional scaling (MDS), Gower [1966] required this data transformation step in principal coordinate analysis (PCoA), and Lewi [1976] also utilized the same transformation in spectral map analysis (SMA). Other applications of DC can be found in Kluger et al. [2003], Lewi [1989], Sánchez et al. [1994]. The experimental results show that DC is effective with RESIDUE(I) ((a)-(d) in Figure 1), but relatively less effective with RESIDUE(II) (Figure 2) because RESIDUE(II) with NT, DC, and MC are same. Seasholtz and Kowalski [1992] showed that MC may not be an optimal preprocessing method under certain conditions and also Sánchez et al. [1994] reported that the results from MC was worse than those from NT. The performance of MC with RESIDUE(I) is almost identical to or a little bit better than that of NT, but not better than that of either SDN or ZT for most cases.

It is worthy mentioning that spectral initialization plays a significant role in improving initial as well as final accuracy values for all the cases. In addition, the standard deviation lines clearly show the trend that spectral initialization results in more stable accuracy than random initialization. Furthermore, local search strategy takes important part in further improvement of quality of co-clustering.

6 Conclusion and remark

Aguilar-Ruiz [2005] issues the need of a new metric to discover both scaling and shifting patterns, showing that RESIDUE(II) can discover any shifted patterns but may not capture some scaled patterns. To answer this need, we propose a simple remedy that helps RESIDUE(II) to resolve its dependency on scaling variances. We suggest to take a specific data transformation through which the hidden scaling factors are implicitly removed. We analyze the effect of various data transformation on the two residues [Cho et al., 2004], RESIDUE(I) and RESIDUE(II), for row/column dimension of data matrix having global/local scaling and global/shifting factors.

Both analysis and experimental results reveal that column standard deviation normalization and column Z-score transformation are effective data transformations for both RESIDUE(I) and RESIDUE(II). Especially through MSSRCC with RESIDUE(II) and the two data transformations, we are able to discover coherent patterns with both scaling and shifting factors. The transformed matrix contains the constant global scaling factor 1 and local shifting factors and gives the perfect residue score, i.e., zero RESIDUE(II).

Additionally, we formally characterize effect of data transformations on some other existing data models in the literature as well as six Euclidean co-clustering schemes in Bregman co-clustering algorithms [Banerjee et al., 2007]. Note that RESIDUE(I) and RESIDUE(II) are two special

cases (scheme 2 and scheme 6, respectively) of six Euclidean co-clustering schemes in Bregman co-clustering algorithms.

Supplementary Information

Software and the supplementary information are available at <http://www.cs.utexas.edu/users/dml/Software/cocluster.html>.

Acknowledgements

This research was supported by NSF grant CCF-0431257, NSF Career Award ACI-0093404, and NSF-ITR award IIS-0325116.

References

- J. S. Aguilar-Ruiz. Shifting and scaling patterns from gene expression data. *Bioinformatics*, 21(20):3840–3845, 2005.
- U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. In *Proceedings of the National Academy of Science*, volume 96, pages 6745–6750, 1999.
- S. A. Armstrong, J. E. Staunton, L. B. Silverman, R. Pieters, M. L. den Boer, M. D. Minden, S. E. Sallan, E. S. Lander, T. R. Golub, and S. J. Korsmeyer. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nature Genetics*, 30:41–47, 2002.
- A. Banerjee, I. S. Dhillon, J. Ghosh, S. Merugu, and D. Modha. A generalized maximum entropy approach to bregman co-clustering and matrix approximation. *Journal of Machine Learning Research*, 8:1919–1986, 2007.
- A. Ben-Dor, B. Chor, R. M. Karp, and Z. Yakhini. Discovering local structure in gene expression data: The order-preserving submatrix problem. *Journal of Computational Biology*, 10(3-4):373–384, 2003.
- T. H. Bø and I. Jonassen. New feature subset selection procedures for classification of expression profiles. *Genome Biology*, 3(4), 2002.
- R. Bro and A. K. Smilde. Centering and scaling in component analysis. *Journal of Chemometrics*, 17:16–33, 2003.
- Y. Cheng and G. M. Church. Biclustering of expression data. In *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology (ISMB'00)*, volume 8, pages 93–103, 2000.
- H. Cho, I. S. Dhillon, Y. Guan, and S. Sra. Minimum sum squared residue based coclustering of gene expression data. In *Proceedings of the 4th SIAM International Conference on Data Mining (SDM'04)*, pages 114–125, 2004.
- M. Dettling and P. Bühlmann. Supervised clustering of genes. *Genome Biology*, 3(12), 2002.
- S. Dudoit and J. Fridlyand. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology*, 3(7):research0036.1–0036.21, 2002.
- T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537, 1999.
- J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3/4):325–338, 1966.
- R. A. Harshman and M. E. Lundy. *Research methods for multimode data analysis*, chapter 6. Data preprocessing and the extended PARAFAC model, pages 216–284. New York: Praeger, 1984.
- J. A. Hartigan. Direct clustering of a data matrix. *Journal of the American Statistical Association*, 67(337):123–129, 1972.

- Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein. Spectral biclustering of microarray data: Coclustering of genes and conditions. *Genome Research*, 13(4):703–716, 2003.
- B. R. Kowalski and C. F. Bender. Pattern recognition: A powerful approach to interpreting chemical data. *Journal of the American Chemical Society*, 94(16):5632–5639, 1972.
- P. J. Lewi. Spectral mapping, a technique for classifying biological activity profiles of chemical compounds. *Arzneimittel Forschung*, 26(7):1295–1300, 1976.
- P. J. Lewi. Spectral map analysis: factorial analysis of contrast, especially from log ratios. *Chemometrics and Intelligent Laboratory Systems*, 5(2):105–116, 1989.
- S. C. Madeira and A. L. Oliveira. Biclustering algorithms for biological data analysis: a survey. *IEEE Transactions on Computational Biology and Bioinformatics*, 1(1):24–45, 2004.
- F. C. Sánchez, P. J. Lewi, and D. L. Massart. Effect of different preprocessing methods for principal component analysis applied to the composition of mixtures: detection of impurities in HPLC-DAD. *Chemometrics and Intelligent Laboratory Systems*, 25(2):157–177, 1994.
- M. B. Seasholtz and B. R. Kowalski. The effect of mean centering on prediction in multivariate calibration. *Journal of Chemometrics*, 6(2):103–111, 1992.
- A. Smilde, R. Bro, and P. Geladi. *Multi-way Analysis with applications in the chemical sciences*, chapter 9. Preprocessing, pages 221–255. John Wiley & Sons, 2004.
- W. S. Torgerson. Multidimensional scaling: 1. theory and method. *Psychometrika*, 17:401–419, 1952.
- L. Wouters, H. W. Göhlmann, L. Bijmens, S. U. Kass, G. Molenberghs, and P. J. Lewi. Graphical exploration of gene expression data: A comparative study of three multivariate methods. *Biometrics*, 59:1131–1139, 2003.

A Other data models

In this section, we analyze the characteristics of the two residues for some of other data models in the literature and their variants. Unless otherwise specified, the models consider local scaling and/or local shifting factors. For brevity, we provide the case with no data transformation.

A.1 $a_{ij} = \pi + \gamma_i + \beta_j$

$$\begin{aligned} a_{iJ} &= \pi + \gamma_i + \mu_{\beta_J}, \\ a_{Ij} &= \pi + \mu_{\gamma_I} + \beta_j, \\ a_{IJ} &= \pi + \mu_{\gamma_I} + \mu_{\beta_J}. \end{aligned}$$

Therefore, we have RESIDUE(I),

$$h_{ij} = a_{ij} - a_{IJ} = (\gamma_i - \mu_{\gamma_I}) + (\beta_j - \mu_{\beta_J})$$

and RESIDUE(II),

$$h_{ij} = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ} = 0.$$

A.2 $a_{ij} = \pi_i + \gamma_i + \beta_j$

$$\begin{aligned} a_{iJ} &= \pi_i + \gamma_i + \mu_{\beta_J}, \\ a_{Ij} &= \mu_{\pi_I} + \mu_{\gamma_I} + \beta_j, \\ a_{IJ} &= \mu_{\pi_I} + \mu_{\gamma_I} + \mu_{\beta_J}. \end{aligned}$$

Therefore, we have RESIDUE(I),

$$h_{ij} = a_{ij} - a_{IJ} = (\pi_i - \mu_{\pi_I}) + (\gamma_i - \mu_{\gamma_I}) + (\beta_j - \mu_{\beta_J})$$

and RESIDUE(II),

$$h_{ij} = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ} = 0.$$

A.3 $a_{ij} = \pi (\alpha_i + \delta_j) + (\gamma_i + \beta_j)$

$$\begin{aligned} a_{iJ} &= \pi (\alpha_i + \mu_{\delta_J}) + (\gamma_i + \mu_{\beta_J}), \\ a_{Ij} &= \pi (\mu_{\alpha_I} + \delta_j) + (\mu_{\gamma_I} + \beta_j), \\ a_{IJ} &= \pi (\mu_{\alpha_I} + \mu_{\delta_J}) + (\mu_{\gamma_I} + \mu_{\beta_J}). \end{aligned}$$

Therefore, we have RESIDUE(I),

$$h_{ij} = a_{ij} - a_{IJ} = \pi ((\alpha_i - \mu_{\alpha_I}) + (\delta_j - \mu_{\delta_J})) + (\gamma_i - \mu_{\gamma_I}) + (\beta_j - \mu_{\beta_J})$$

and RESIDUE(II),

$$h_{ij} = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ} = 0.$$

$$\mathbf{A.4} \quad a_{ij} = \pi_i (\alpha_i + \delta_j) + (\gamma_i + \beta_j)$$

$$\begin{aligned} a_{iJ} &= \pi_i (\alpha_i + \mu_{\delta_J}) + (\gamma_i + \mu_{\beta_J}), \\ a_{Ij} &= (\mu_{\pi_I \alpha_I} + \mu_{\pi_I} \delta_j) + (\mu_{\gamma_I} + \beta_j), \\ a_{IJ} &= (\mu_{\pi_I \alpha_I} + \mu_{\pi_I} \mu_{\delta_J}) + (\mu_{\gamma_I} + \mu_{\beta_J}), \end{aligned}$$

where $\mu_{\pi_I \alpha_I} = \frac{1}{|I|} \sum_{i \in I} \pi_i \alpha_i$. Therefore, we have RESIDUE(I),

$$h_{ij} = a_{ij} - a_{IJ} = (\pi_i \alpha_i - \mu_{\pi_I \alpha_I}) + (\pi_i \delta_j - \mu_{\pi_I} \mu_{\delta_J}) + (\gamma_i - \mu_{\gamma_I}) + (\beta_j - \mu_{\beta_J})$$

and RESIDUE(II),

$$h_{ij} = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ} = (\pi_i - \mu_{\pi_I}) (\delta_j - \mu_{\delta_J}).$$

B Euclidean co-clustering schemes in Bregman co-clustering algorithms

Here, we summarize the results for six Euclidean co-clustering schemes in Bregman co-clustering algorithms [Banerjee et al., 2007]. Note that we are referred to the results in Tables 5 and 9, when analyzing effect of data transformations on Scheme 2 (i.e., MSSRCC(I)) and Scheme 6 (i.e., MSSRCC(II)).

Table 4: Residue values along with different data transformation.

Type	Transformation			Metric
	Target	Scaling	Shifting	Scheme 1
NT	n/a	G	G	$(\pi_i - \mu_\pi - \mu_{\pi_I}) \alpha - \beta$
	n/a	G	L	$(\pi_i - \mu_\pi - \mu_{\pi_I}) \alpha + (\beta_j - \mu_\beta - \mu_{\beta_J})$
	n/a	L	G	$(\pi_i \alpha_j - \mu_\pi \mu_{\alpha_J} - \mu_{\pi_I} \mu_\alpha) - \beta$
	n/a	L	L	$(\pi_i \alpha_j - \mu_\pi \mu_{\alpha_J} - \mu_{\pi_I} \mu_\alpha) + (\beta_j - \mu_\beta - \mu_{\beta_J})$
DC	n/a	G	G	0
	n/a	G	L	0
	n/a	L	G	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha)$
	n/a	L	L	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha)$
MC	Row	G	G	0
	Row	G	L	$\beta_j - \mu_{\beta_J}$
	Row	L	G	$\pi_i (\alpha_j - \mu_\alpha) + \mu_\pi (\mu_\alpha - \mu_{\alpha_J})$
	Row	L	L	$\pi_i (\alpha_j - \mu_\alpha) + \mu_\pi (\mu_\alpha - \mu_{\alpha_J}) + (\beta_j - \mu_{\beta_J})$
	Col	G	G	$(\pi_i - \mu_{\pi_I}) \alpha$
	Col	G	L	$(\pi_i - \mu_{\pi_I}) \alpha$
	Col	L	G	$(\pi_i - \mu_\pi) \alpha_j + (\mu_\pi - \mu_{\pi_I}) \mu_\alpha$
	Col	L	L	$(\pi_i - \mu_\pi) \alpha_j + (\mu_\pi - \mu_{\pi_I}) \mu_\alpha$
SDN	Row	G	G	n/a
	Row	G	L	$\frac{1}{\sigma_\beta} ((\pi_i - \mu_\pi - \mu_{\pi_I}) \alpha + (\beta_j - \mu_\beta - \mu_{\beta_J}))$
	Row	L	G	$\frac{1}{\sigma_\alpha} \left((\alpha_j - \mu_\alpha - \mu_{\alpha_J}) + \beta \left(\frac{1}{\pi_i} - \frac{1}{m} \sum_{i=1}^m \frac{1}{\pi_i} - \frac{1}{ I } \sum_{i \in I} \frac{1}{\pi_i} \right) \right)$
	Row	L	L	$\frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i \alpha_j + \beta_j)$ $-\frac{1}{m} \sum_{i=1}^m \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i \mu_{\alpha_J} + \mu_{\beta_J})$ $-\frac{1}{ I } \sum_{i \in I} \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i \mu_\alpha + \mu_\beta)$
	Col	G	G	$\frac{1}{\sigma_\pi} \left((\pi_i - \mu_\pi - \mu_{\pi_I}) - \frac{\beta}{\alpha} \right)$
	Col	G	L	$\frac{1}{\sigma_\pi} \left((\pi_i - \mu_\pi - \mu_{\pi_I}) + \frac{1}{\alpha} (\beta_j - \mu_\beta - \mu_{\beta_J}) \right)$
	Col	L	G	$\frac{1}{\sigma_\pi} \left((\pi_i - \mu_\pi - \mu_{\pi_I}) + \beta \left(\frac{1}{\alpha_j} - \frac{1}{n} \sum_{j=1}^n \frac{1}{\alpha_j} - \frac{1}{ J } \sum_{j \in J} \frac{1}{\alpha_j} \right) \right)$
	Col	L	L	$\frac{1}{\sigma_\pi} \left((\pi_i - \mu_\pi - \mu_{\pi_I}) + \left(\frac{\beta_j}{\alpha_j} - \frac{1}{n} \sum_{j=1}^n \frac{\beta_j}{\alpha_j} - \frac{1}{ J } \sum_{j \in J} \frac{\beta_j}{\alpha_j} \right) \right)$
ZT	Row	G	G	0
	Row	G	L	$\frac{1}{\sigma_\beta} (\beta_j - \mu_{\beta_J})$
	Row	L	G	$\frac{1}{\sigma_\alpha} (\alpha_j - \mu_{\alpha_J})$
	Row	L	L	$\frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i (\alpha_j - \mu_\alpha) + (\beta_j - \mu_\beta))$ $+\frac{1}{ I } \sum_{i \in I} \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i (\mu_\alpha - \mu_{\alpha_J}) + (\mu_\beta - \mu_{\beta_J}))$
	Col	G	G	$\frac{1}{\sigma_\pi} (\pi_i - \mu_{\pi_I})$
	Col	G	L	$\frac{1}{\sigma_\pi} (\pi_i - \mu_{\pi_I})$
	Col	L	G	$\frac{1}{\sigma_\pi} (\pi_i - \mu_{\pi_I})$
	Col	L	L	$\frac{1}{\sigma_\pi} (\pi_i - \mu_{\pi_I})$

Abbreviations: NT – No Transformation; DC – Double Centering; MC – Mean Centering; SDN – Standard Deviation Normalization; ZT – Z-score Transformation; Col – Column; G – Global; and L – Local.

Table 5: Residue values along with different data transformation.

Transformation				Metric
Type	Target	Scaling	Shifting	Scheme 2 (i.e., MSSRCC(I))
NT	n/a	G	G	$(\pi_i - \mu_{\pi_I}) \alpha$
	n/a	G	L	$(\pi_i - \mu_{\pi_I}) \alpha + (\beta_j - \mu_{\beta_J})$
	n/a	L	G	$(\pi_i \alpha_j - \mu_{\pi_I} \mu_{\alpha_J})$
	n/a	L	L	$(\pi_i \alpha_j - \mu_{\pi_I} \mu_{\alpha_J}) + (\beta_j - \mu_{\beta_J})$
DC	n/a	G	G	0
	n/a	G	L	0
	n/a	L	G	$(\pi_i - \mu_{\pi_I}) \alpha_j + (\mu_{\pi_I} - \mu_{\pi_I}) \mu_{\alpha_J} - (\pi_i - \mu_{\pi_I}) \mu_{\alpha}$
	n/a	L	L	$(\pi_i - \mu_{\pi_I}) \alpha_j + (\mu_{\pi_I} - \mu_{\pi_I}) \mu_{\alpha_J} - (\pi_i - \mu_{\pi_I}) \mu_{\alpha}$
MC	Row	G	G	0
	Row	G	L	$\beta_j - \mu_{\beta_J}$
	Row	L	G	$\pi_i (\alpha_j - \mu_{\alpha}) + \mu_{\pi_I} (\mu_{\alpha} - \mu_{\alpha_J})$
	Row	L	L	$\pi_i (\alpha_j - \mu_{\alpha}) + \mu_{\pi_I} (\mu_{\alpha} - \mu_{\alpha_J}) + (\beta_j - \mu_{\beta_J})$
	Col	G	G	$(\pi_i - \mu_{\pi_I}) \alpha$
	Col	G	L	$(\pi_i - \mu_{\pi_I}) \alpha$
	Col	L	G	$(\pi_i - \mu_{\pi_I}) \alpha_j + (\mu_{\pi_I} - \mu_{\pi_I}) \mu_{\alpha_J}$
	Col	L	L	$(\pi_i - \mu_{\pi_I}) \alpha_j + (\mu_{\pi_I} - \mu_{\pi_I}) \mu_{\alpha_J}$
SDN	Row	G	G	n/a
	Row	G	L	$\frac{1}{\sigma_{\beta}} ((\pi_i - \mu_{\pi_I}) \alpha + (\beta_j - \mu_{\beta_J}))$
	Row	L	G	$\frac{1}{\sigma_{\alpha}} \left((\alpha_j - \mu_{\alpha_J}) + \beta \left(\frac{1}{\pi_i} - \frac{1}{ I } \sum_{i \in I} \frac{1}{\pi_i} \right) \right)$
	Row	L	L	$\frac{1}{\sigma_{\alpha}} (\pi_i \alpha_j + \beta_j) - \frac{1}{ I } \sum_{i \in I} \frac{1}{\sigma_{\alpha_i}} (\pi_i \mu_{\alpha_J} - \mu_{\beta_J})$
	Col	G	G	$\frac{1}{\sigma_{\pi}} (\pi_i - \mu_{\pi_I})$
	Col	G	L	$\frac{1}{\sigma_{\pi}} \left((\pi_i - \mu_{\pi_I}) + \frac{1}{\alpha} (\beta_j - \mu_{\beta_J}) \right)$
	Col	L	G	$\frac{1}{\sigma_{\pi}} \left((\pi_i - \mu_{\pi_I}) + \beta \left(\frac{1}{\alpha_j} - \frac{1}{ J } \sum_{j \in J} \frac{1}{\alpha_j} \right) \right)$
	Col	L	L	$\frac{1}{\sigma_{\pi}} \left((\pi_i - \mu_{\pi_I}) + \left(\frac{\beta_j}{\alpha_j} - \frac{1}{ J } \sum_{j \in J} \frac{\beta_j}{\alpha_j} \right) \right)$
ZT	Row	G	G	n/a
	Row	G	L	$\frac{1}{\sigma_{\beta}} (\beta_j - \mu_{\beta_J})$
	Row	L	G	$\frac{1}{\sigma_{\alpha}} (\alpha_j - \mu_{\alpha_J})$
	Row	L	L	$\frac{1}{\sigma_{\alpha}} (\pi_i (\alpha_j - \mu_{\alpha}) + (\beta_j - \mu_{\beta}))$ $+ \frac{1}{ I } \sum_{i \in I} \frac{1}{\sigma_{\alpha'_i}} (\pi_i (\mu_{\alpha} - \mu_{\alpha_J}) + (\mu_{\beta} - \mu_{\beta_J}))$
	Col	G	G	$\frac{1}{\sigma_{\pi}} (\pi_i - \mu_{\pi_I})$
	Col	G	L	$\frac{1}{\sigma_{\pi}} (\pi_i - \mu_{\pi_I})$
	Col	L	G	$\frac{1}{\sigma_{\pi}} (\pi_i - \mu_{\pi_I})$
	Col	L	L	$\frac{1}{\sigma_{\pi}} (\pi_i - \mu_{\pi_I})$

Abbreviations: NT – No Transformation; DC – Double Centering; MC – Mean Centering; SDN – Standard Deviation Normalization; ZT – Z-score Transformation; Col – Column; G – Global; and L – Local.

Table 6: Residue values along with different data transformation.

Type	Transformation			Metric
	Target	Scaling	Shifting	Scheme 3
NT	n/a	G	G	$-(\mu_{\pi_I} \alpha + \beta)$
	n/a	G	L	$-\mu_{\pi_I} \alpha + (\beta_j - \mu_\beta - \mu_{\beta_J})$
	n/a	L	G	$(\pi_i \alpha_j - \pi_i \mu_\alpha - \mu_{\pi_I} \mu_{\alpha_J}) - \beta$
	n/a	L	L	$(\pi_i \alpha_j - \pi_i \mu_\alpha - \mu_{\pi_I} \mu_{\alpha_J}) + (\beta_j - \mu_\beta - \mu_{\beta_J})$
DC	n/a	G	G	0
	n/a	G	L	0
	n/a	L	G	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha)$
	n/a	L	L	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha) - (\mu_\pi - \mu_{\pi_I}) (\mu_\alpha - \mu_{\alpha_J})$
MC	Row	G	G	0
	Row	G	L	$\beta_j - \mu_{\beta_J}$
	Row	L	G	$\pi_i (\alpha_j - \mu_\alpha) + \mu_{\pi_I} (\mu_\alpha - \mu_{\alpha_J})$
	Row	L	L	$\pi_i (\alpha_j - \mu_\alpha) + \mu_{\pi_I} (\mu_\alpha - \mu_{\alpha_J}) + (\beta_j - \mu_{\beta_J})$
	Col	G	G	$(\mu_\pi - \mu_{\pi_I}) \alpha$
	Col	G	L	$(\mu_\pi - \mu_{\pi_I}) \alpha$
	Col	L	G	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha) + (\mu_\pi - \mu_{\pi_I}) \mu_{\alpha_J}$
	Col	L	L	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha) + (\mu_\pi - \mu_{\pi_I}) \mu_{\alpha_J}$
SDN	Row	G	G	n/a
	Row	G	L	$\frac{1}{\sigma_\beta} (-\mu_{\pi_I} \alpha + (\beta_j - \mu_\beta - \mu_{\beta_J}))$
	Row	L	G	$\frac{1}{\sigma_\alpha} ((\alpha_j - \mu_\alpha - \mu_{\pi_I} \alpha) - \mu_{\beta_J})$
	Row	L	L	$\frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i (\alpha_j - \mu_\alpha) + (\beta_j - \mu_\beta))$ $-\frac{1}{ I } \sum_{i \in I} \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i \mu_{\alpha_J} - \mu_{\beta_J})$
	Col	G	G	$-\frac{1}{\sigma_\pi} (\mu_{\pi_I} + \frac{\beta}{\alpha})$
	Col	G	L	$\frac{1}{\sigma_\pi} (-\mu_{\pi_I} + \frac{1}{\alpha} (\beta_j - \mu_\beta - \mu_{\beta_J}))$
	Col	L	G	$\frac{1}{\sigma_\pi} \left(-\mu_{\pi_I} + \beta \left(\frac{1}{\alpha_j} - \frac{1}{n} \sum_{i=1}^n \frac{1}{\alpha_j} - \frac{1}{ J } \sum_{j \in J} \frac{1}{\alpha_j} \right) \right)$
	Col	L	L	$\frac{1}{\sigma_\pi} \left(-\mu_{\pi_I} + \left(\frac{\beta_j}{\alpha_j} - \frac{1}{n} \sum_{i=1}^n \frac{\beta_j}{\alpha_j} - \frac{1}{ J } \sum_{j \in J} \frac{\beta_j}{\alpha_j} \right) \right)$
ZT	Row	G	G	0
	Row	G	L	$\frac{1}{\sigma_\beta} (\beta_j - \mu_{\beta_J})$
	Row	L	G	$\frac{1}{\sigma_\alpha} (\alpha_j - \mu_{\alpha_J})$
	Row	L	L	$\frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i (\alpha_j - \mu_\alpha) + (\beta_j - \mu_\beta))$ $-\frac{1}{ I } \sum_{i \in I} \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i (\mu_{\alpha_J} - \mu_\alpha) + (\mu_{\beta_J} - \mu_\beta))$
	Col	G	G	$\frac{1}{\sigma_\pi} (\mu_\pi - \mu_{\pi_I})$
	Col	G	L	$\frac{1}{\sigma_\pi} (\mu_\pi - \mu_{\pi_I})$
	Col	L	G	$\frac{1}{\sigma_\pi} (\mu_\pi - \mu_{\pi_I})$
	Col	L	L	$\frac{1}{\sigma_\pi} (\mu_\pi - \mu_{\pi_I})$

Abbreviations: NT – No Transformation; DC – Double Centering; MC – Mean Centering; SDN – Standard Deviation Normalization; ZT – Z-score Transformation; Col – Column; G – Global; and L – Local.

Table 7: Residue values along with different data transformation.

Type	Transformation			Metric
	Target	Scaling	Shifting	Scheme 4
NT	n/a	G	G	$(\pi_i - \mu_\pi - \mu_{\pi_I}) \alpha - \beta$
	n/a	G	L	$(\pi_i - \mu_\pi - \mu_{\pi_I}) \alpha - \mu_{\beta_J}$
	n/a	L	G	$(\pi_i \alpha_j - \mu_\pi \alpha_j - \mu_{\pi_I} \mu_{\alpha_J}) - \beta$
	n/a	L	L	$(\pi_i \alpha_j - \mu_\pi \alpha_j - \mu_{\pi_I} \mu_{\alpha_J}) - \mu_{\beta_J}$
DC	n/a	G	G	0
	n/a	G	L	0
	n/a	L	G	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha)$
	n/a	L	L	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha) - (\mu_\pi - \mu_{\pi_I}) (\mu_\alpha - \mu_{\alpha_J})$
MC	Row	G	G	0
	Row	G	L	$-(\mu_{\beta_J} - \mu_\beta)$
	Row	L	G	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha) + \mu_{\pi_I} (\mu_\alpha - \mu_{\alpha_J})$
	Row	L	L	$(\pi_i - \mu_\pi) (\alpha_j - \mu_\alpha) + \mu_{\pi_I} (\mu_\alpha - \mu_{\alpha_J}) + (\mu_\beta - \mu_{\beta_J})$
	Col	G	G	$(\pi_i - \mu_{\pi_I}) \alpha$
	Col	G	L	$(\pi_i - \mu_{\pi_I}) \alpha$
	Col	L	G	$(\pi_i - \mu_\pi) \alpha_j + (\mu_\pi - \mu_{\pi_I}) \mu_{\alpha_J}$
	Col	L	L	$(\pi_i - \mu_\pi) \alpha_j + (\mu_\pi - \mu_{\pi_I}) \mu_{\alpha_J}$
SDN	Row	G	G	n/a
	Row	G	L	$\frac{1}{\sigma_\beta} ((\pi_i - \mu_\pi - \mu_{\pi_I}) \alpha - \mu_{\beta_J})$
	Row	L	G	$\frac{1}{\sigma_\alpha} \left(-\mu_{\alpha_J} + \beta \left(\frac{1}{\pi_i} - \frac{1}{m} \sum_{i=1}^m \frac{1}{\pi_i} - \frac{1}{ I } \sum_{i \in I} \frac{1}{\pi_i} \right) \right)$
	Row	L	L	$\frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i \alpha_j + \beta_j)$ $-\frac{1}{m} \sum_{i=1}^m \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i \alpha_j + \beta_j)$ $-\frac{1}{ I } \sum_{i \in I} \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i \mu_{\alpha_J} + \mu_{\beta_J})$
	Col	G	G	$\frac{1}{\sigma_\pi} \left((\pi_i - \mu_\pi - \mu_{\pi_I}) - \frac{\beta}{\alpha} \right)$
	Col	G	L	$\frac{1}{\sigma_\pi} \left((\pi_i - \mu_\pi - \mu_{\pi_I}) - \frac{\mu_{\beta_I}}{\alpha} \right)$
	Col	L	G	$\frac{1}{\sigma_\pi} \left((\pi_i - \mu_\pi - \mu_{\pi_I}) - \beta \frac{1}{ J } \sum_{j \in J} \frac{1}{\alpha_j} \right)$
	Col	L	L	$\frac{1}{\sigma_\pi} \left((\pi_i - \mu_\pi - \mu_{\pi_I}) - \frac{1}{ J } \sum_{j \in J} \frac{\beta_j}{\alpha_j} \right)$
	Row	G	G	0
	Row	G	L	$-\frac{1}{\sigma_\beta} (\mu_{\beta_J} - \mu_\beta)$
ZT	Row	L	G	$-\frac{1}{\sigma_\alpha} (\mu_{\alpha_J} - \mu_\alpha)$
	Row	L	L	$\frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i (\alpha_j - \mu_\alpha) + (\beta_j - \mu_\beta))$ $-\frac{1}{m} \sum_{i=1}^m \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i (\alpha_j - \mu_\alpha) + (\beta_j - \mu_\beta))$ $-\frac{1}{ I } \sum_{i \in I} \frac{1}{\sqrt{\pi_i^2 \sigma_\alpha^2 + \sigma_\beta^2}} (\pi_i (\mu_{\alpha_J} - \mu_\alpha) + (\mu_{\beta_J} - \mu_\beta))$
	Col	G	G	$\frac{1}{\sigma_\pi} (\pi_i - \mu_{\pi_I})$
	Col	G	L	$\frac{1}{\sigma_\pi} (\pi_i - \mu_{\pi_I})$
	Col	L	G	$\frac{1}{\sigma_\pi} (\pi_i - \mu_{\pi_I})$
	Col	L	L	$\frac{1}{\sigma_\pi} (\pi_i - \mu_{\pi_I})$

Abbreviations: NT – No Transformation; DC – Double Centering; MC – Mean Centering; SDN – Standard Deviation Normalization; ZT – Z-score Transformation; Col – Column; G – Global; and L – Local.

Table 8: Sum squared residue (SSR) values along with different data transformation.

Transformation				Metric
Type	Target	Scaling	Shifting	Scheme 5
NT	n/a	G	G	$-(\mu_\pi + \mu_{\pi_I})\alpha - 2\beta$
	n/a	G	L	$-(\pi_i + \mu_\pi)\alpha - (\mu_\beta + \mu_{\beta_J})$
	n/a	L	G	$(\pi_i\alpha_j - \pi_i\mu_\alpha - \mu_\pi\alpha_j - \mu_{\pi_I}\mu_{\alpha_J}) - 2\beta$
	n/a	L	L	$(\pi_i\alpha_j - \pi_i\mu_\alpha - \mu_\pi\alpha_j - \mu_{\pi_I}\mu_{\alpha_J}) - (\mu_\beta + \mu_{\beta_J})$
DC	n/a	G	G	0
	n/a	G	L	0
	n/a	L	G	$(\pi_i - \mu_\pi)(\alpha_j - \mu_\alpha)$
	n/a	L	L	$(\pi_i - \mu_\pi)(\alpha_j - \mu_\alpha) - (\mu_\pi - \mu_{\pi_I})(\mu_\alpha - \mu_{\alpha_J})$
MC	Row	G	G	0
	Row	G	L	$\mu_\beta - \mu_{\beta_J}$
	Row	L	G	$(\pi_i - \mu_\pi)(\alpha_j - \mu_\alpha) + \mu_{\pi_I}(\mu_\alpha - \mu_{\alpha_J})$
	Row	L	L	$(\pi_i - \mu_\pi)(\alpha_j - \mu_\alpha) + \mu_{\pi_I}(\mu_\alpha - \mu_{\alpha_J}) + (\mu_\beta - \mu_{\beta_J})$
	Col	G	G	$\mu_\pi - \mu_{\pi_I}$
	Col	G	L	$\mu_\pi - \mu_{\pi_I}$
	Col	L	G	$(\pi_i - \mu_\pi)(\alpha_j - \mu_\alpha) + (\mu_\pi - \mu_{\pi_I})\mu_{\alpha_J}$
	Col	L	L	$(\pi_i - \mu_\pi)(\alpha_j - \mu_\alpha) + (\mu_\pi - \mu_{\pi_I})\mu_{\alpha_J}$
SDN	Row	G	G	n/a
	Row	G	L	$\frac{1}{\sigma_\beta}((-\mu_\pi - \mu_{\pi_I})\alpha - (\mu_\beta + \mu_{\beta_J}))$
	Row	L	G	$\frac{1}{\sigma_\alpha}\left(-(\mu_\alpha + \mu_{\alpha_J}) - \beta\left(\frac{1}{m}\sum_{i=1}^m\frac{1}{\pi_i} + \frac{1}{ I }\sum_{i\in I}\frac{1}{\pi_i}\right)\right)$
	Row	L	L	$\frac{1}{\sqrt{\pi_i^2\sigma_\alpha^2\sigma_\beta^2}}(\pi_i(\alpha_j - \mu_\alpha) + (\beta_j - \mu_\beta))$
				$-\frac{1}{m}\sum_{i=1}^m\frac{1}{\sqrt{\pi_i^2\sigma_\alpha^2\sigma_\beta^2}}(\pi_i\alpha_j + \beta_j)$
				$-\frac{1}{ I }\sum_{i\in I}\frac{1}{\sqrt{\pi_i^2\sigma_\alpha^2\sigma_\beta^2}}(\pi_i\mu_{\alpha_J} + \mu_{\beta_J})$
	Col	G	G	$\frac{1}{\sigma_\pi}(-(\mu_\pi + \mu_{\pi_I}))$
	Col	G	L	$\frac{1}{\sigma_\pi}(-(\mu_\pi + \mu_{\pi_I}) - \frac{1}{\alpha}(\mu_\beta + \mu_{\beta_J}))$
	Col	L	G	$\frac{1}{\sigma_\pi}\left(-(\mu_\pi + \mu_{\pi_I}) - \beta\left(\frac{1}{n}\sigma_{j=1}^n\frac{1}{\alpha_j} + \frac{1}{ J }\sigma_{j\in J}\frac{1}{\alpha_j}\right)\right)$
	Col	L	L	$\frac{1}{\sigma_\pi}\left(-(\mu_\pi + \mu_{\pi_I}) - \left(\frac{1}{n}\sigma_{j=1}^n\frac{\beta_j}{\alpha_j} + \frac{1}{ J }\sigma_{j\in J}\frac{\beta_j}{\alpha_j}\right)\right)$
ZT	Row	G	G	0
	Row	G	L	$-\frac{1}{\sigma_\beta}(\mu_{\beta_J} - \mu_\beta)$
	Row	L	G	$-\frac{1}{\sigma_\alpha}(\mu_{\alpha_J} - \mu_\alpha)$
	Row	L	L	$\frac{1}{\sqrt{\pi_i^2\sigma_\alpha^2 + \sigma_\beta^2}}(\pi_i(\alpha_j - \mu_\alpha) + (\beta_j - \mu_\beta))$
				$-\frac{1}{m}\sum_{i=1}^m\frac{1}{\sqrt{\pi_i^2\sigma_\alpha^2 + \sigma_\beta^2}}(\pi_i(\alpha_j - \mu_\alpha) + (\beta_j - \mu_\beta))$
				$+\frac{1}{ I }\sum_{i\in I}\frac{1}{\sqrt{\pi_i^2\sigma_\alpha^2 + \sigma_\beta^2}}(\pi_i(\mu_\alpha - \mu_{\alpha_J}) + (\mu_\beta - \mu_{\beta_J}))$
	Col	G	G	$\frac{1}{\sigma_\pi}(\mu_\pi - \mu_{\pi_I})$
	Col	G	L	$\frac{1}{\sigma_\pi}(\mu_\pi - \mu_{\pi_I})$
	Col	L	G	$\frac{1}{\sigma_\pi}(\mu_\pi - \mu_{\pi_I})$
	Col	L	L	$\frac{1}{\sigma_\pi}(\mu_\pi - \mu_{\pi_I})$

Abbreviations: NT – No Transformation; DC – Double Centering; MC – Mean Centering; SDN – Standard Deviation Normalization; ZT – Z-score Transformation; Col – Column; G – Global; and L – Local.

Table 9: Residue values along with different data transformation.

Type	Transformation			Metric
	Target	Scaling	Shifting	Scheme 6 (i.e., MSSRCC(II))
NT	n/a	G	G	0
	n/a	G	L	0
	n/a	L	G	$(\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$
	n/a	L	L	$(\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$
DC	n/a	G	G	0
	n/a	G	L	0
	n/a	L	G	$(\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$
	n/a	L	L	$(\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$
MC	Row	G	G	0
	Row	G	L	0
	Row	L	G	$(\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$
	Row	L	L	$(\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$
	Col	G	G	0
	Col	G	L	0
	Col	L	G	$(\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$
	Col	L	L	$(\pi_i - \mu_{\pi_I})(\alpha_j - \mu_{\alpha_J})$
SDN	Row	G	G	n/a
	Row	G	L	0
	Row	L	G	0
	Row	L	L	$\frac{1}{\sigma_i}(\pi_i(\alpha_j - \mu_{\alpha_J}) + (\beta_j - \mu_{\beta_J}))$ $-\frac{1}{ I } \sum_{i \in I} \frac{1}{\sigma_i}((\alpha_j - \mu_{\alpha_J}) - (\beta_j - \mu_{\beta_J}))$
	Col	G	G	0
	Col	G	L	0
	Col	L	G	0
	Col	L	L	0
ZT	Row	G	G	n/a
	Row	G	L	0
	Row	L	G	0
	Row	L	L	$\frac{1}{\sigma_{A'_i}}(\pi_i(\alpha_j - \mu_{\alpha_J}) + (\beta_j - \mu_{\beta_J}))$ $-\frac{1}{ I } \sum_{i \in I} \frac{1}{\sigma_{A'_i}}(\pi_i(\alpha_j - \mu_{\alpha_J}) + (\beta_j - \mu_{\beta_J}))$
	Col	G	G	0
	Col	G	L	0
	Col	L	G	0
	Col	L	L	0

Abbreviations: NT – No Transformation; DC – Double Centering; MC – Mean Centering; SDN – Standard Deviation Normalization; ZT – Z-score Transformation; Col – Column; G – Global; and L – Local.