Singular Value Decomposition and Its Visualization

Author(s): Lingsong Zhang, J. S. Marron, Haipeng Shen and Zhengyuan Zhu

Source: *Journal of Computational and Graphical Statistics*, Dec., 2007, Vol. 16, No. 4 (Dec., 2007), pp. 833-854

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of America

Stable URL: https://www.jstor.org/stable/27594278

Singular Value Decomposition and Its Visualization

Lingsong ZHANG, J. S. MARRON,
Haipeng SHEN, and Zhengyuan ZHU

Singular value decomposition (SVD) is a useful tool in functional data analysis (FDA). Compared to principal component analysis (PCA), SVD is more fundamental, because SVD simultaneously provides the PCAs in both row and column spaces. We compare SVD and PCA from the FDA view point, and extend the usual SVD to variations by considering different centerings. A generalized scree plot is proposed to select an appropriate centering in practice. Several useful matrix views of the SVD components are introduced to explore different features in data, including SVD surface plots, image plots, curve movies, and rotation movies. These methods visualize both column and row information of a two-way matrix simultaneously, relate the matrix to relevant curves, show local variations, and highlight interactions between columns and rows. Several toy examples are designed to compare the different variations of SVD, and real data examples are used to illustrate the usefulness of the visualization methods.

Key Words: Exploratory data analysis; Functional data analysis; Principal component analysis.

1. INTRODUCTION

Functional data analysis (FDA) is the study of curves (and more complex objects) as data (Ramsay and Silverman 2002, 2005). Methods related to principal component analysis (PCA) have provided many insights. Compared to the PCA method, singular value decomposition (SVD) can be thought of as more fundamental, because SVD provides not only a direct approach to calculate the principal components (PCs), but also derives the PCAs in the row and column spaces simultaneously. In this article, we view a set of curves as a two-way data matrix, explore the connections and differences between SVD and PCA from a FDA view point, and propose several visualization methods for the SVD components.

Lingsong Zhang is Research Fellow, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115 (E-mail: *zhang@hsph.harvard.edu*). J. S. Marron (E-mail: *marron@email.unc.edu*) is Amos Hawley Distinguished Professor, Haipeng Shen (*haipeng@email.unc.edu*) is Assistant Professor, and Zhengyuan Zhu (*zhuz@email.unc.edu*) is Assistant Professor, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599-3260.

Journal of Computational and Graphical Statistics, Volume 16, Number 4, Pages 833–854
DOI: 10.1198/106186007X256080

Let X be a data matrix. In the statistical literature, the rows of X are often viewed as observations for an experiment, and the columns of X are thought of as the covariates. SVD provides a useful factorization of the data matrix X, while PCA provides a nearly parallel factoring, via eigen-analysis of the sample covariance matrix, that is, $X^T X$, when X is column centered at 0. The eigenvalues for $X^T X$ are then the squares of the singular values for X, and the eigenvectors for $X^T X$ are the singular rows for X. Here, we extend the usual (column centered) PCA method into a general SVD framework, and consider four types of SVDs based on different centerings. Several criteria are discussed for model selection, that is, selecting the appropriate type of SVD, including approximation performance, complexity, and interpretability. We introduce a generalized scree plot, which provides a visual aid for model selection in terms of complexity and approximation performance, and presents a simple way to understand the tradeoff between these two criteria.

Visualization methods can be very helpful in finding underlying features of a dataset. In the context of PCA or SVD, common visualization methods include the biplot (Gabriel 1971), scatterplots between singular columns or singular rows (see Jolliffe 2002, sect. 5.1), etc. The biplot shows the relations between the rows and columns, and the scatterplot can be used to cluster them. However, for FDA datasets, these plots fail to show the functional curves.

In the FDA field, it is also common to plot singular columns or singular rows as curves. Marron et al. (2004) provided a visualization method for functional data (using functional PCA), which shows the functional objects (curves), projections on the PCs and the residuals. When considering a time series of curves, different colors are used to show the time ordering. These methods can also be applied in the SVD framework, but to understand all the structure in the data, they need to be applied twice, once for the rows and once for the columns. In addition, when the time series structure is complicated, color coding curves might not be enough to reveal the time effect, because of overplotting problems.

In this article, we propose several *matrix views* of the SVD components, which may reveal new underlying features of the dataset, which are hard to find by other methods. The major visualization tool is a set of surface plots. To enhance the interpretability of the surface plot, we introduce the SVD rotation movies, the SVD curve movies, and the SVD image plots. These visualization methods reveal different underlying information about the SVD components. In fact, the visualization method in Marron et al. (2004) (e.g., the first column in Figure 3 of their article) can be viewed as one particular snapshot of the SVD rotation movie. These visualization methods are motivated by an Internet traffic dataset, which is discussed in Section 2. Another real application is reported in Section 6 as well. A Spanish mortality dataset is analyzed to illustrate that the image plot highlights the cohort effect (i.e., the interaction between age groups and years).

The following parts are organized as follows. The motivating example in network traffic analysis is in Section 2. Section 3 gives a brief introduction of SVD, compares it with PCA and discusses the four types of centerings. Section 4 describes the generation of the plots and the movies in detail. Section 5 shows several toy examples to illustrate the four types of centerings. More applications are reported in Section 6.

2. MOTIVATING EXAMPLE

Internet traffic, measured over time at a single location, forms a very noisy time series. Figure 1(a) shows an example of network traffic data collected at the main Internet link of the UNC campus network, as packet counts per half hour over a period of seven weeks, which cover part of two sessions of UNC summer school in 2003. The approximately 49 tall thin spikes represent peak daytime network usage, which suggests that there is a strong daily effect. The tallest spikes are grouped into clusters of five corresponding to weekdays, with in-between gaps corresponding to weekends.

Because of the expected similarity of the daily shapes, and as a device for studying potential contrasts between these shapes (e.g., differences between weekdays and weekends), we analyze the time series in an unusual way. We rearrange the data as a 49×48 matrix, so that each row represents a day, and each column represents one half-hour interval within a day. This treatment is similar to the singular-spectrum analysis method (Golyandina et al. 2001). Figure 1(b) is a mesh plot showing the structure of the data matrix. This shows that the data matrix is noisy, but we can still see a weekly pattern and also clear daily shapes from it.

One way to analyze the dataset is to treat the daily shapes, the *rows* of the data matrix, as functions (curves), and to use some functional data analysis methods to understand the underlying features. PCA has proven to be very useful for this purpose. But observe that the *columns* of the matrix are also curves (i.e., time series) of interest as well. In particular, for each fixed half-hour interval, these are the counts over days. A natural eigen-analysis for simultaneous PCA of rows and columns is SVD of the matrix. This decomposition is motivated by the analysis of call center datasets in Shen and Huang (2005). Similar to the PCA method, SVD provides a useful first tool for exploratory data analysis. The SVD method decomposes the data matrix into a sum of several matrices (which are the SVD components), where each matrix has rank one. Each component provides insights into features of the data.

To find the underlying characteristics using SVD, we provide a set of surface plots of the SVD components, which help to examine the data matrix in both directions (i.e., both rows as data, and columns as data). Long (1983) used a similar method to illustrate matrix approximation of SVD from a mathematical education view point. Interpretation of the surface plots is aided by two types of movies. One movie shows the plots from different angles, and the other movie highlights the row curves and column curves. These movies can be used to demonstrate the time-varying features of the components, and highlight some special features or outliers. These visualization methods can be used alone or with other visualization methods, to find interesting structure in the data.

For the network traffic data, the SVD surface plot is in Figure 2. The top left panel is the mesh plot for the original data matrix (the same as Figure 1(b)). The top middle subplot is the first SVD component, which turns out to be essentially a smoothed version of the original data matrix, showing a clear weekly pattern. This first SVD component also indicates that weekdays and weekends might not share the same daily shape (or magnitude of daily shapes). The top right panel is the second component, which has a clear shape for

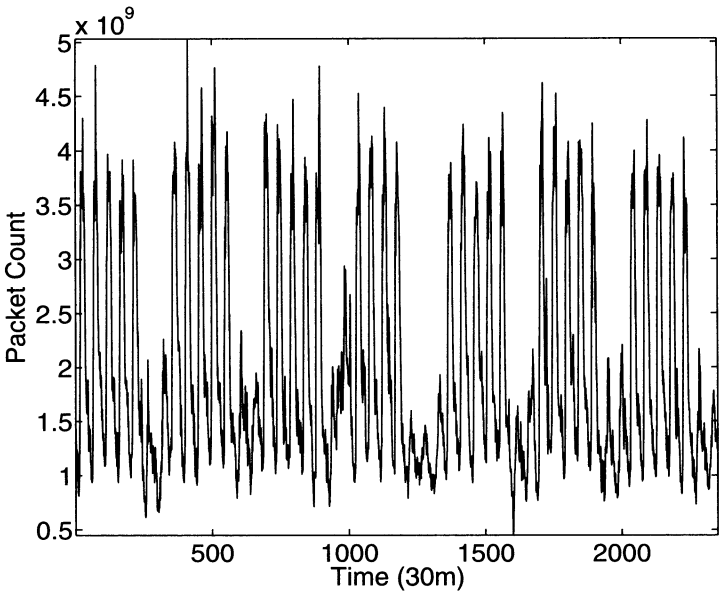

Figure 1. (a) Original time series plot of packet counts per half-an-hour over 49 days. (b) Mesh plot for the original network traffic data.

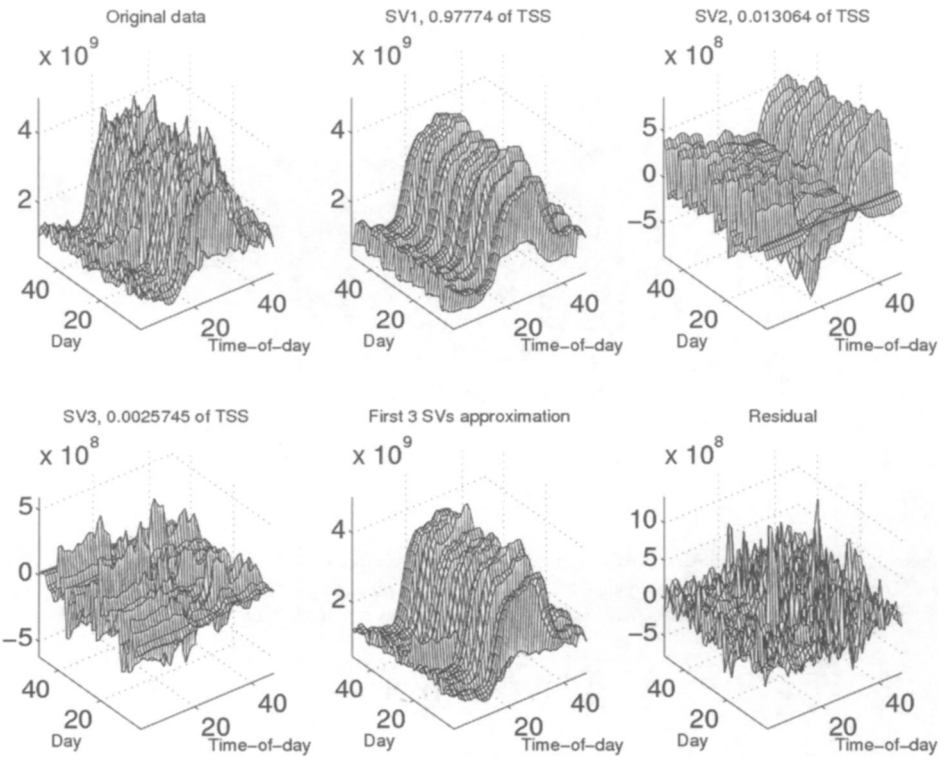

Figure 2. SVD surface plots for the network traffic data. SV1–SV3 are a decomposition of the data matrix. These are combined to give the model in the bottom middle panel, with the corresponding residual shown in the bottom right panel.

weekends, and is relatively flat for weekdays. This indicates the existence of a weekday–weekend effect, and suggests analyzing weekend and weekday data separately as a good option. The bottom left is the third component, which shows that there are very large bumps in some days. Those bumps might indicate that the corresponding dates have some special features, as discussed in the following.

The SVD movies for these data highlight those features. Figure 3 shows one carefully selected snapshot of the SVD curve movie for the third component. It highlights Sunday, June 29, as a possible outlier, by showing a large bump in the movie. This day had a network workload that was between the weekday and weekend data. A check of the university calendar reveals that this was the first Sunday of the second summer session, when a large number of students returned from home after the summer session break, which made the network usage different from other weekends. All movies for the network traffic dataset can be easily viewed at *http://www.unc.edu/~lszhang/research/network/SVDmovie* (Zhang 2006). More analysis of this dataset is presented in Section 6.1.

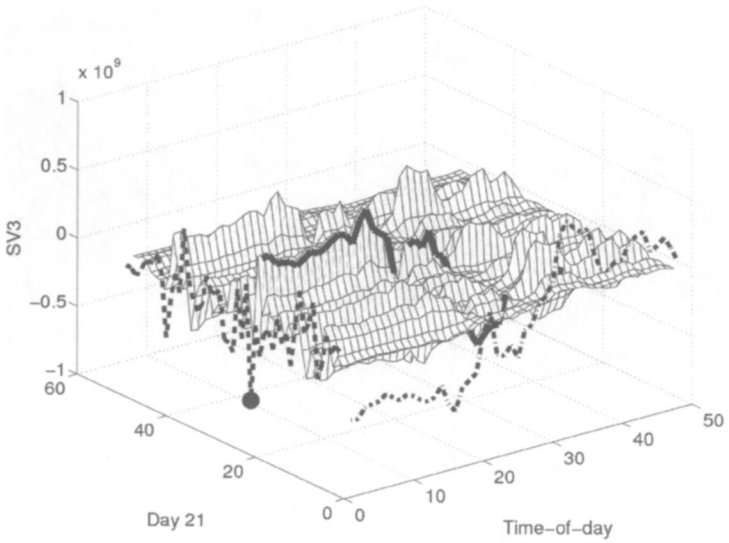

Figure 3. One carefully chosen snapshot of the SVD curve movie of the third SVD component for the network data. The movie highlights some outlying days by showing high spikes on the surface. For instance, this snapshot shows the 21st row, which suggests that June 29 is a special day, as explained in the text.

3. SVD AND PCA

In this section, we give a brief introduction of SVD (Section 3.1), its relation with PCA (Section 3.2), and a visual aid for selecting different centerings (Section 3.3).

3.1 SVD AND ITS PROPERTIES

Let $X = (x_{ij})_{m\times n}$ with $\text{rank}(X) = r$. Let $\{\mathbf{r}_i : i = 1, \ldots, m\}$, $\{\mathbf{c}_j : j = 1, \ldots, n\}$ be the row and column vectors of the matrix X, respectively. The SVD of X is defined as

$$X = USV^T = s_1\mathbf{u}_1\mathbf{v}_1^T + s_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + s_r\mathbf{u}_r\mathbf{v}_r^T,$$

where $U = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r)$, $V = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r)$, and $S = \text{diag}\{s_1, s_2, \ldots, s_r\}$ with $s_1 \geq s_2 \geq \cdots \geq s_r > 0$. The singular columns $\{\mathbf{u}_i\}$ form an orthonormal basis for the column space spanned by $\{\mathbf{c}_j\}$, and the singular rows $\{\mathbf{v}_j\}$ form an orthonormal basis for the row space spanned by $\{\mathbf{r}_i\}$. The vectors $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are called *singular columns* and *singular rows*, respectively (Gabriel and Odoroff 1984); the scalars $\{s_i\}$ are called *singular values*; and the matrices $\{s_i\mathbf{u}_i\mathbf{v}_i^T\}(i = 1, \ldots, r)$ are referred to as *SVD components*.

The SVD factorization has an important approximation property. Let A be a rank k $(k \leq r)$ (approximation) matrix, and define $R = X - A = (r_{ij})_{m\times n}$ as its residual matrix. We then define the residual sum of squares (RSS) of the matrix A as the sum of squares (SS) of the elements in R, that is, $\text{RSS}(A) = \text{SS}(R) = \sum_{i=1}^m \sum_{j=1}^n r_{ij}^2$. Householder and Young (1938) showed that

$$\underset{A:\text{rank}(A)=k}{\arg\min}\ \text{RSS}(A) = A_k = \sum_{l=1}^{k} s_l\mathbf{u}_l\mathbf{v}_l^T,$$

and the corresponding residual matrix R is $R = X - A_k = \sum_{l=k+1}^{r} s_l \mathbf{u}_l \mathbf{v}_l^T$. In other words, the SVD provides the best rank k approximation of the data matrix X.

3.2 FOUR TYPES OF CENTERINGS FOR SVD

As noted earlier, SVD and PCA are closely related. SVD as defined above provides a decomposition of X. PCA is very similar with the only difference being column mean centering. Our matrix view raises a natural question of: "why not do row mean centering?" We will study and compare four possible types of centerings and correspondingly four types of SVD: *no centering* (Simple SVD or SSVD), *column centering* (Column SVD or CSVD), *row centering* (Row SVD or RSVD), and centering in both row and column directions, which is referred to as *double centering* (Double SVD or DSVD).

Another natural choice of centering is removing the overall mean and then applying SVD. When all the elements in X are relatively larger numbers, removing the overall mean will decrease the magnitudes of these entries, and thus improve the numerical stability of the SVD. However, Gabriel (1978) mentioned that, for the model with an overall constant plus multiplicative terms, the least squares estimation of the components is not equivalent to fitting the overall mean and then applying the SVD for the residual part. With a functional dataset, the overall mean does not provide informative curve information of the dataset. There are cases where removing the overall mean will cause the data to lose some good features, for example, the orthogonality of the curves (see Zhang 2006). Note that the column/row/double centerings automatically remove the overall mean. Thus, we will not discuss the case of just removing the overall mean. However, our programs do provide the option of applying SVD after just removing the overall mean.

Let $\overline{x}$ be the sample overall mean of all the elements in X, $\overline{x}_c$ be the sample column mean vector with each elements being the mean of each column, and $\overline{x}_r$ be the sample row mean vector with each element being the mean of the corresponding row. We define the sample column mean matrix as $\mathrm{CM} = 1_{m\times 1}\overline{x}_c^T$, sample row mean matrix as $\mathrm{RM} = \overline{x}_r 1_{1\times n}$, and the sample double mean matrix as $\mathrm{DM} = \overline{x}_r 1_{1\times n} + 1_{m\times 1}\overline{x}_c^T - \overline{x}1_{m\times n}$.

We can use the same formula $X = A + R$ for all these four types of SVD, where A is the approximation matrix, and R is the residual matrix. For no centering, $A = A^{(s)}$ is the sum of the first several SVD components of X, which is the best approximation matrix at the corresponding rank. For column centering, $A = \mathrm{CM} + A^{(c)}$ where $A^{(c)}$ is the sum of the first several SVD components of $X - \mathrm{CM}$. Similarly, for row centering and double centering, we have $A = \mathrm{RM} + A^{(r)}$ and $A = \mathrm{DM} + A^{(d)}$, where $A^{(r)}$ is the sum of the first several SVD components of $X - \mathrm{RM}$ and $A^{(d)}$ is the sum of the first several SVD components of $X - \mathrm{DM}$. Note that A here is not the same for the four centerings, while we just use the same notation for convenience. Note that all these approximation matrices are least squares estimations of some models; see Gabriel (1978) for details. Note that CM and RM have rank 1, and DM is at most rank 2.

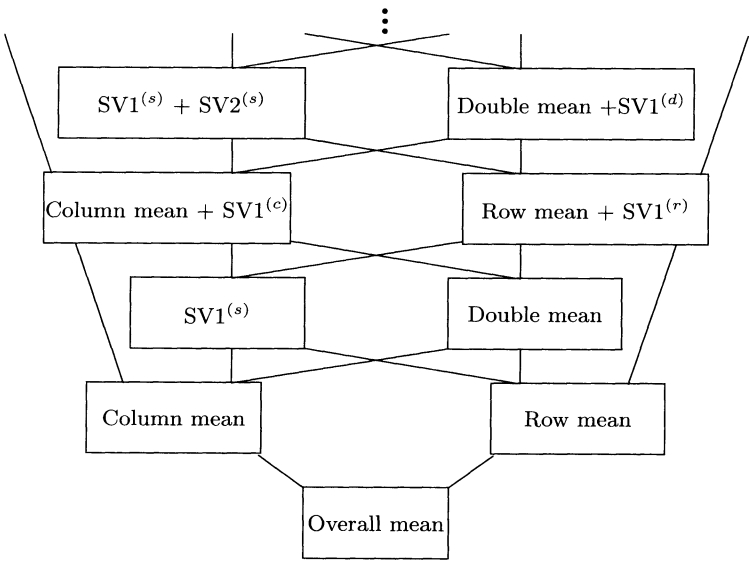

Figure 4. Relationship between approximations using the four types of SVD. A lower level is always worse than an upper level one in approximation, while it also provides a simpler model than the upper level.

3.2.1 Approximation Diagram

Figure 4 shows how the four types of centerings relate to each other in terms of approximation. The boxes on the same horizontal level have noncomparable RSS (i.e., either can give a better approximation). Lower level boxes always have larger RSS than upper level boxes. The overall mean provides the worst approximation (i.e., the largest RSS) among all these mean matrices. The line segments show direct comparisons given in Equations (3.1)–(3.4). For example, the column mean provides a larger RSS than both the double mean and the first SSVD component, and so does the row mean (as shown in Equations (3.1)–(3.2) with $\text{rank}(A^{(s)}) = 1$). The first SSVD component has larger RSS than either RSVD or CSVD with mean matrix plus one SVD component, and so does the double mean.

This diagram also shows the level of complexity, in the sense that models on the same level share similar complexity of the model, and the lower level models are simpler (i.e., they have fewer components) than the upper level models. For example, as shown in Figure 4, a model using either the column mean or the row mean is simpler than using the double mean or the first SSVD component. In fact, the double mean can be viewed as the sum of the row mean and the column mean (i.e., an additive model), while the first SSVD component can be viewed as the product of the row and the column mean (i.e., a multiplicative model). This is why we treat them as having the same level of complexity.

3.2.2 Theoretical Comparisons

The following are theoretical comparisons between the four types of SVD, in terms of approximation performance. Some of these results are reported in a different form in

Golyandina et al. (2001).

1. If $\text{rank}\left(A^{(c)}\right) = \text{rank}\left(A^{(r)}\right) = \text{rank}\left(A^{(d)}\right)$, we have

$$\begin{aligned} \text{RSS}\left(\text{CM} + A^{(c)}\right) &\geq \text{RSS}\left(\text{DM} + A^{(d)}\right), \text{ and} \\ \text{RSS}\left(\text{RM} + A^{(r)}\right) &\geq \text{RSS}\left(\text{DM} + A^{(d)}\right). \end{aligned} \tag{3.1}$$

   These inequalities show a case that double centering gives a better approximation than either column centering or row centering. Note that the rank of the approximation matrix of double centering ($\text{DM} + A^{(d)}$) is usually greater than the other two.

2. If $\text{rank}\left(A^{(c)}\right) = \text{rank}\left(A^{(r)}\right) = \text{rank}\left(A^{(s)}\right) - 1$, we have

$$\text{RSS}\left(\text{CM} + A^{(c)}\right) \geq \text{RSS}\left(A^{(s)}\right), \text{ and } \text{RSS}\left(\text{RM} + A^{(r)}\right) \geq \text{RSS}\left(A^{(s)}\right). \tag{3.2}$$

   Similarly, the above two inequalities show that the SSVD approximation is better than either column or row centering, if the approximation matrices have the same rank.

3. If $\text{rank}(A^{(c)}) = \text{rank}(A^{(r)}) = \text{rank}(A^{(s)})$, we have

$$\text{RSS}\left(\text{CM} + A^{(c)}\right) \leq \text{RSS}\left(A^{(s)}\right), \text{ and } \text{RSS}\left(\text{RM} + A^{(r)}\right) \leq \text{RSS}\left(A^{(s)}\right). \tag{3.3}$$

   This suggests that, with the same number of SVD components, the RSVD and CSVD models provide better approximation than the SSVD model. In this case, the approximation matrix of RSVD and CSVD have a rank that is one larger than that of SSVD.

4. If $\text{rank}(A^{(c)}) = \text{rank}(A^{(r)}) = \text{rank}(A^{(d)}) + 1$, we have

$$\begin{aligned} \text{RSS}\left(\text{CM} + A^{(c)}\right) &\leq \text{RSS}\left(\text{DM} + A^{(d)}\right), \text{ and} \\ \text{RSS}\left(\text{RM} + A^{(r)}\right) &\leq \text{RSS}\left(\text{DM} + A^{(d)}\right). \end{aligned} \tag{3.4}$$

   This shows that if approximation matrices have the same rank, the CSVD and RSVD models have better approximation performance than DSVD.

5. In terms of the RSS, there is no clear relationship between column centering and row centering (either could be better), nor between double centering and no centering.

When a dataset is being explored, sometimes the context may suggest the most appropriate centering. Otherwise, we suggest trying all four centerings and deciding which one is preferable. The following criteria should be considered: the model should have a small RSS, few components, and be easy to interpret. In some situations, the aim of the problem and the constraints of the related context should also be considered, as shown in Section 5.

3.3 MODEL SELECTION AND GENERALIZED SCREE PLOT

In this subsection, a generalized scree plot is proposed as a visual aid to determine the appropriate centering and the number of components. In the context of PCA, the scree plot (Cattell 1966; Jolliffe 2002, sect. 6.1.3) is widely recommended to attempt to determine an "appropriate" number of PC components. Note that in some contexts, the log-scale scree plot might convey an entirely different impression.

Here we define the scree plot for the four types of centerings in a novel way, since the mean matrices and their degree of approximation need to be incorporated into the plot. Let A_k be the approximation matrix of the four types of centerings with rank k ($k+1$ for DSVD), that is, the sum of the first k SSVD components; or column (/row/double) mean matrix plus the first $k-1$ CSVD (/RSVD/DSVD) components. R_k is the residual matrix corresponding to A_k. We denote $l_k = \mathrm{SS}(R_k)/\mathrm{TSS}$, where the total sum of squares (TSS) is $\mathrm{TSS} = \mathrm{SS}(X)$ as the *residual proportion* of the TSS. The plot of (k, l_k) is called the residual proportion plot. It is obvious that l_k is nonincreasing, and we can use similar rules to decide the appropriate number of components.

As noticed in Section 3.2, if the approximation matrices of CSVD/RSVD/SSVD have the same rank k, and DSVD has rank $k+1$, we know SSVD is the best rank k approximation, while the CSVD/RSVD has larger RSS than the first k SSVD components, but has smaller RSS than the first $k-1$ SSVD components. These comparisons also hold in terms of model complexity, where smaller can be replaced by simpler. If we plot (k, l_k) for the SSVD in the integer grid of k, the residual proportion of the RSVD or the CSVD should be in between that of the SSVD with rank k and $k-1$. Thus, we use the half-grid $k-1/2$ here to show the approximation performance of the RSVD/CSVD (i.e., plot $(k-1/2, l_k)$ for RSVD/CSVD). The DSVD with one larger rank has noncomparable RSS with SSVD, so we plot it at the same level as SSVD.

The resulting plot described above is defined as our *generalized scree plot*, where simpler models are always to the left. The above special treatment (i.e., plotting $(k-1/2, l_k)$) makes the points in the plot correspond to (the models of) the rectangles in the approximation diagram of Figure 4. And the levels from the bottom to the top in Figure 4 correspond to the grids on the horizontal axis from the left to the right. Thus, the generalized scree plot simultaneously visualizes the approximation performance and model complexity. In order to choose the appropriate centering and number of components, one possible way to decide the number of components uses the usual interpretation of scree plot. After this, one can select which is the leftmost. Zhang (2006) provides a MATLAB function, `gscreeplot.m`, to generate the generalized scree plot, which allows various options, including log-scale scree plot. Note that the scree plot assumes a strong signal, with large variation relative to the noise component. If this assumption is violated, other considerations should be used to select the optimal model.

Figure 5 shows the generalized scree plot for the network dataset in Section 2. From the plot, we find that all four models use two components for the major modes of variation, and they have similar approximation performance. In terms of model complexity, we might use either RSVD or CSVD as the final model to find underlying features of the network

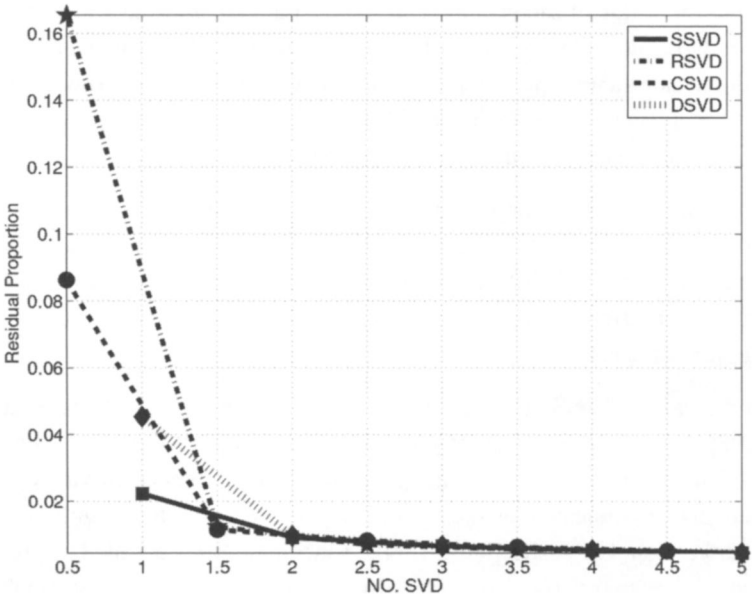

Figure 5. Generalized scree plot for the network traffic data. It shows that all four types of centerings use two components to explain the major modes of variation. Thus, the RSVD/CSVD model with two components is the best in terms of model complexity and approximation performance. By looking at all the surface plots, we choose SSVD with three components as the model to analyze the network data, because it provides the best interpretability in this context.

dataset. By looking at the surface plots [the SSVD surface plot is in Figure 2, and the other three can be viewed at Zhang (2006)] of all four types of centerings, we find the SSVD model with three components provides the best interpretability among the four centerings, which is why this one was chosen in Section 2. This is an example of overruling the usual interpretation of the scree plot. However, viewing the generalized scree plot can be a good starting point in terms of model selection. See Section 5 for more discussion.

4. GENERATION OF THE PLOTS AND MOVIES

Details of the generation of the surface plots, image plots, and two SVD movies (SVD rotation movie and SVD curve movie) are provided in the following.

Surface plots: The surface plots consist of $k + 3$ ($k < \text{rank}(X)$) subplots. The first one is the mesh plot for the original data matrix. The next k subplots are the mesh plots for the first k SVD components ($s_i \mathbf{u}_i \mathbf{v}_i^T$). The $(k + 2)$nd subplot is the mesh plot for the k-component approximation matrix, that is, the summation matrix of the first k components. And the last one is the mesh plot for the residual. The surface plots can be generated using the MATLAB function `svd3dplot.m`, which can be obtained from Zhang (2006).

Image plots: The image plots provide a special view angle of the SVD components, by showing the image view of the original data matrix, the first k SVD components, the reconstruction of the k components, and the residual after approximation. In our design,

we let the minimal value of each component share a light color (white), and the maximum value share a dark color (black). The SVD images show a good view angle to highlight the local variations, data subgroups and interactions between columns and rows. The program `svd3dplot.m` in Zhang (2006) with option (`'iimage', 1`) generates the image plots. See the demographic data (Section 6.2) as an example.

SVD rotation movie: Viewing the surface plots from different angles helps to find data features. Another program `svdviewbycomp.m` in Zhang (2006) generates movies for different SVD components with different angles of view. It rotates the surface plot of an SVD component, so that different features are more clear from different view angles. The SVD rotation movie for the network data is explained in Section 6.1.

SVD curve movie: The SVD curve movie illustrates how the SVD components relate to classical FDA. The curve movie of the ith SVD component ($s_i \mathbf{u}_i \mathbf{v}_i^T$) is based on the mesh plot of the component. Within each mesh plot, two reference curves with different line types are used to indicate how the surface is generated from the singular row and singular column vectors. The dashed curve (in the row direction) is a scaled singular column ($c_{\mathbf{u}_i} \mathbf{u}_i$), where $c_{\mathbf{u}_i} = \max_{kl}((\mathbf{u}_i \mathbf{v}_i^T)_{kl}) / \max_k(|\mathbf{u}_{ik}|)$; and the dash-dot curve (in the column direction) is a scaled singular row ($c_{\mathbf{v}_i} \mathbf{v}_i$), where $c_{\mathbf{v}_i} = \max_{kl}((\mathbf{u}_i \mathbf{v}_i^T)_{kl}) / \max_k(|\mathbf{v}_{ik}|)$. The scaled constants are chosen so that the reference curves share a comparable vertical range to the surface. A solid line varies on the surface, from the first row to the last and then back to the first row, while a big dot moves along the singular column (the dashed curve) with respect to the solid line. Then the solid line varies from the first column to the last and then back to the first column, while the corresponding dot varies along the singular row (the dash-dot curve). The motion shows how the curves change in both directions, and highlights features of interest, such as outliers. The function, `svd3dplot.m` generates the SVD curve movie, with appropriate options. As an example, Figure 3 in Section 2 has provided one snapshot of the SVD curve movie for the third SSVD component of the network dataset.

Note that, for large datasets, it is helpful to restrict the range for rows or columns to get a zoomed SVD movie, which demonstrates local features. Another function, `svd3dzoomplot.m` in Zhang (2006) can generate the zoomed version of the SVD movies and plots. See Zhang et al. (2006) for a chemometrics dataset, which is used to illustrate the utility of zoomed surface plot and zoomed curve movies. It is not discussed here to save space.

5. FOUR TYPES OF SVD AND TOY EXAMPLES

In this section, we use simulated examples to illustrate model selection among the four types of SVD. These examples make it clear that sometimes we do not have a "best" choice. Also for real applications, it is not enough to use only the generalized scree plot to select appropriate models. However, the generalized scree plot is still useful to give an initial impression of which model might be a better candidate. It is also useful when the user does not have time for deep exploration of the datasets. If time permits, we strongly

recommend the application of the four types of centerings simultaneously, and the use of some visualization methods, including the matrix views or other information, to select the most interpretable model.

We designed several simulated datasets to illustrate the above idea of model selection. They also show that each of the four types of centerings can be the most appropriate model. In this section, we show two interesting datasets. These simulated datasets are designed as 49×48 matrices, the same as the network traffic dataset we discussed in Section 2. In this setting, each row can be viewed as one daily usage profile, and each column can be treated as a cross-day times series of one specific time in a day. In addition, these toy examples are designed to have clear weekly patterns. A large number of plots and other simulated examples, similar to those actually shown in this section, are available at *http://www.unc.edu/~lszhang/research/network/SVDmovie* (Zhang 2006) to save space.

5.1 EXAMPLE 1

The first example is used to illustrate a situation where CSVD gives the best approximation performance. This example is designed to be the sum of a column vector (i.e., $\mu_c(j)$ in (5.1)), a multiplicative component (i.e., $f_2(i)g_2(j)$ in (5.1)), and some noise. The model can be written as

$$h_1(i, j) = \mu_c(j) + f_1(i)g_1(j) + \varepsilon(i, j), \tag{5.1}$$

where $\mu_c(j) = \sin(j\pi/24)$; $g_1(j) = -\cos(j\pi/24)$; $f_1(i) = 1$ when $\text{mod}(i, 7) \neq 0$ and 6, or $f_1(i) = 2$ when otherwise; and $\varepsilon(i, j) \overset{\text{iid}}{\sim} N(0, .04)$. Note that we will use the same notation ε for different realizations of all the simulated examples. In terms of network usages, the weekdays and weekends do not have the same usage magnitudes (due to the multiplicative component), nor the same usage patterns (because of the column vector component).

The generalized scree plot (available at Zhang 2006) suggests that there are two components for the CSVD/DSVD/SSVD models, while the RSVD model uses three components. The RSVD is the worst model among the four, because the row mean matrix explains a very low proportion of the TSS. These suggest that the CSVD is the best model among the four types of centerings, in terms of complexity and approximation performance. We will use the surface plots to compare the interpretabilities of these four centerings.

By examining the surface plots of all four centerings, we find the RSVD and SSVD provide similar decompositions with the exception that the RSVD has an additional row mean matrix. This similarity is due to the fact that the row mean matrix explains a very low proportion of the TSS, as discussed earlier. The surface plots of the CSVD (the first row in Figure 6) show the common usage pattern (i.e., the column mean) in the top middle panel. The next CSVD component (the top right panel) shows the contrast between the weekdays and weekends. This also shows that after removing a common daily usage profile, the contrast curves between the weekdays and weekends have the same shape, but have different contrast magnitudes. It is hard to answer the question of whether the contrast between them is different in daily shapes or in different usage magnitudes.

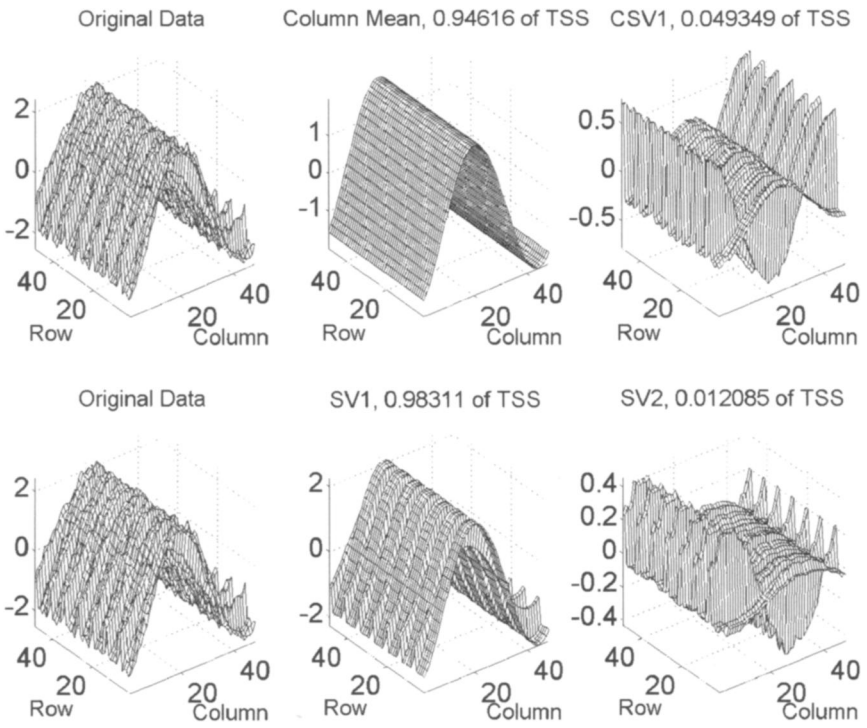

Figure 6. The first row shows the surface plots of the CSVD for model (5.1), and the second row provides the surface plots of the SSVD. For this example, the SSVD and the CSVD are candidates of good approximation performance. The CSVD is better than the SSVD, because it contains a simpler model. However, the SSVD shows that weekdays and weekends have different daily shapes and different magnitudes of daily usages.

Meanwhile, the surface plots for the SSVD (the second row in Figure 6) use two components for the major variation of the data matrix. If the daily usages share the same usage pattern but with different magnitudes, one SSVD component should be enough for the major modes of variation. Thus, the two SSVD components for this example suggest that weekdays and weekends do have different usage patterns and different magnitudes. This suggests that the SSVD model for this example has better interpretability than the CSVD model.

5.2 EXAMPLE 2

This example is used to illustrate a case where the CSVD is the best model when model complexity and approximation performance are considered. On the other hand, the worst model in terms of these two criteria, the RSVD, provides the best model in terms of interpretability. The model of this example contains an overall mean level shift (μ) and two multiplicative components. The mathematical description of this model is

$$h_2(i, j) = \mu + f_2(i)g_2(j) + f_3(i)g_3(j) + \varepsilon(i, j), \tag{5.2}$$

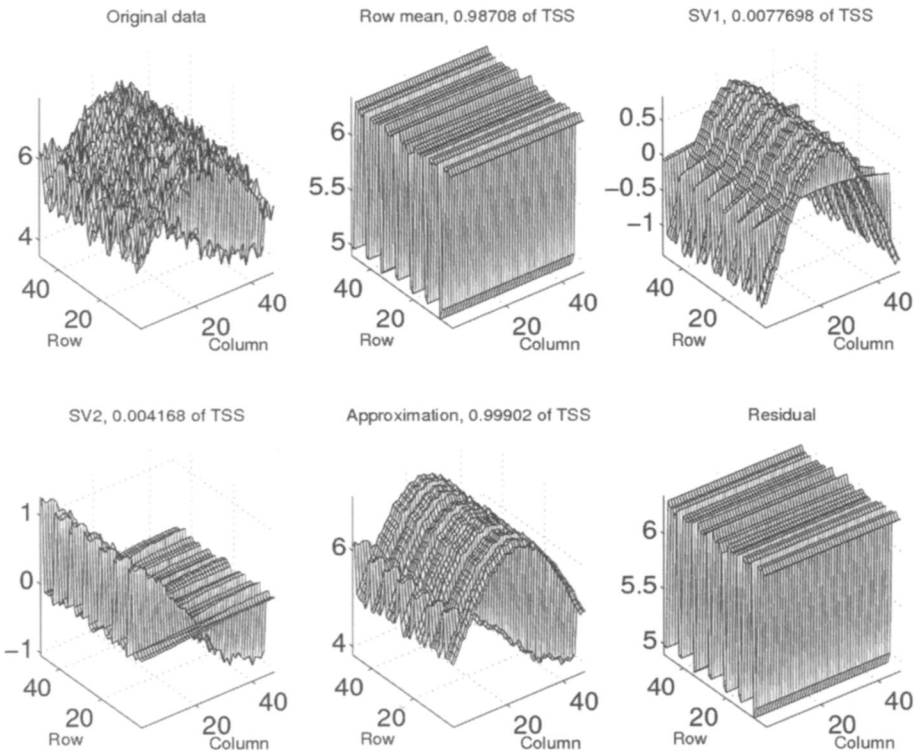

Figure 7. The surface plot of the RSVD for the second toy examples, from Equation (5.2). The RSVD is the worst model in terms of approximation performance. But it correctly picks up the weekly pattern (row mean matrix), the shapes of the weekdays (SV1) and the shapes of the weekends (SV2).

where $f_2(i) = 2$, $f_3(i) = 0$ when $\text{mod}(i, 7) \neq 0$ and 6, and $f_2(i) = 0$, $f_3(i) = 1$ when otherwise; $g_2(j) = \sin(j\pi/48)$; and $g_3(j) = \cos(j\pi/48)$. Here the two multiplicative components are orthogonal to each other in both column and row spaces. We set $\mu = 5$, such that all elements in h_2 are positive.

The generalized scree plot (available at Zhang 2006) shows that the RSVD uses three components for the major modes of variation, while other SVDs use two components, which suggest that the RSVD is the "worst" model among the four types of centerings. The CSVD model is the best in terms of complexity and approximation performance. However, after comparing all the surface plots (of the four types of decomposition), we find that the RSVD, the "worst" model discussed earlier, provides the clearest decomposition using three components. The surface plots for the RSVD is in Figure 7. The row mean matrix (the top middle panel) grasps the weekly pattern. The first RSVD component (the top right panel) shows sine curves for the weekdays, and is nearly flat in the weekends. And the second RSVD component (the bottom left panel) picks the cosine curves in the weekends, and stays close to zero for the weekdays. This gives a perfect separation of the curves in model (5.2). For this toy example, we might choose CSVD as the final model, because it is the best model in terms of approximation performance and complexity. On the other hand,

RSVD can also be selected as the best choice, because it provides the best separation of the curves, thus it has the best interpretability.

6. REAL APPLICATIONS

In this section, we further illustrate the utility of our visualization methods. Section 6.1 continues the analysis of the network data discussed in Section 2. The image plot is illustrated in Section 6.2 to highlight the interactions of age groups and years in a Spanish mortality dataset.

6.1 FURTHER ANALYSIS OF INTERNET TRAFFIC DATA

Here we further analyze the network traffic dataset in Section 2 by using the SVD rotation movie. The surface plot gives an insightful visualization of the SVD components in Section 2. Viewing the surface plot from different angles will help to highlight different interesting features. Here we recommend viewing the full SVD rotation movie for the SVD components. The left panel in Figure 8 shows a carefully chosen snapshot of the rotation movie for the first SSVD component. A careful examination shows that besides the typical weekly pattern, the fourth weekend seems to be a long weekend, and the third weekend is kind of short. In fact, Friday July 4 makes the fourth weekend special, and Sunday June 29 is in the third weekend, which has been discussed in Section 2.

The right panel in Figure 8 is a carefully selected snapshot of the rotation movie for the second SVD component. We find that the 19th row of the second SVD is unusual because although it is a weekday, it looks more like a weekend, but with a smaller bump. The 19th row was Friday June 27, the last day for late registration of the UNC summer school second session. Checking the original dataset, we found that the network usage on that morning oscillates a lot, which might be explained by bursts of usage after the end of each class time, followed by rapid departure of the students.

The above, along with the discussion in Section 2, shows that the SVD rotation movie and SVD curve movie are useful for outlier identification. Note that the above outlying days can also be identified by other visualization methods. For example, scatterplots between singular rows can be very helpful in finding outlying days. Figure 9 shows the scatterplot between $\mathbf{u}_1$ and $\mathbf{u}_2$. From the scatterplot we find two groups of points. Investigation suggests that they correspond to the weekdays and weekends, as indicated using dots (for weekdays) and plus (for weekends) respectively. Friday, July 4, was a holiday, and it falls into the weekend group, which is not surprising because on that day most students and school employees were away. Sunday, June 29, is between those two groups, for reasons discussed above. Friday, July 27, might also be found through this scatterplot, with an unusually low $\mathbf{u}_2$ value. But the confidence here might not be high, because this point is not that far away from the weekday data points. A scatterplot of $\mathbf{u}_1$ and $\mathbf{u}_3$ can also show some outlying days, which is skipped because no new outliers were found.

For this dataset, although the scatterplots among singular columns are useful to find clusters of days and some outlying days, a limitation is that they cannot provide a func-

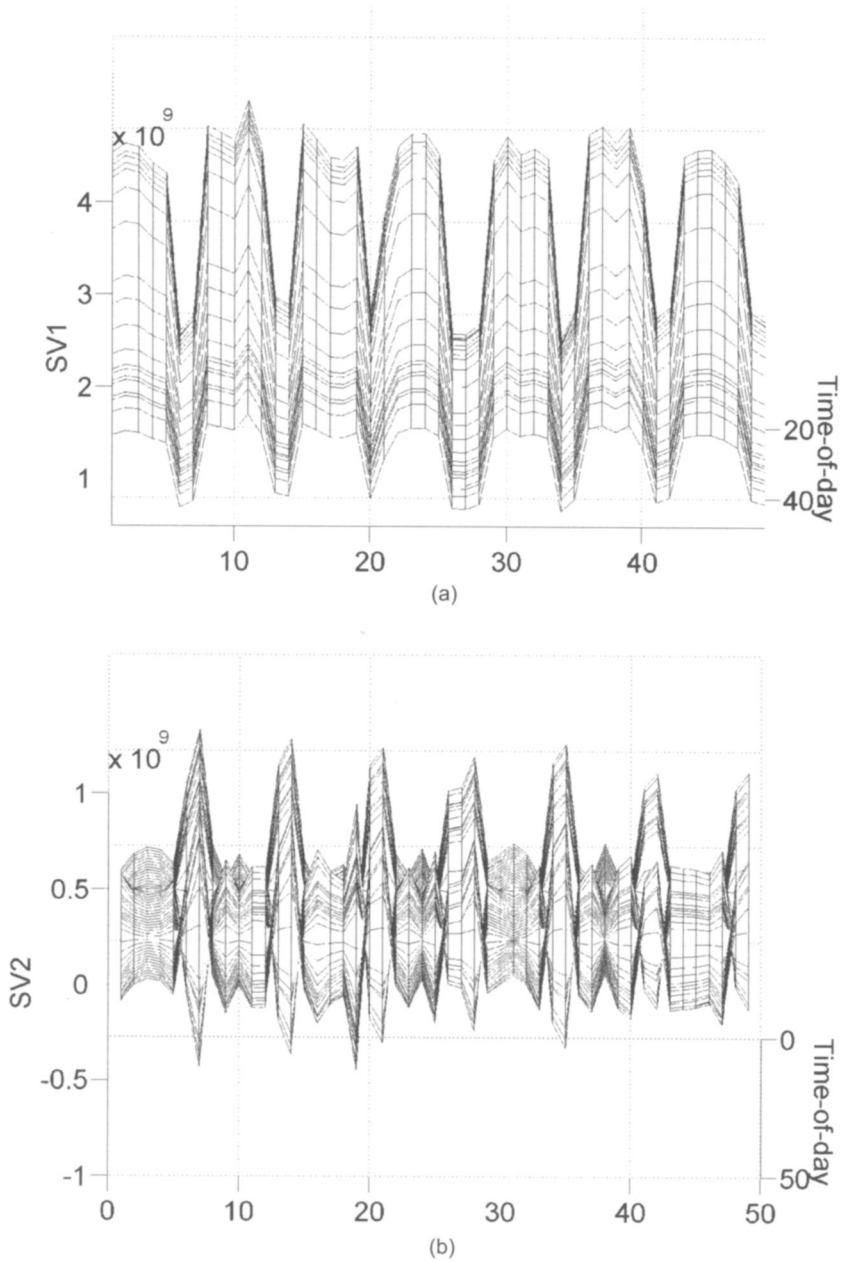

Figure 8. Snapshots of the SSVD rotation movie with carefully chosen angles of view. (a) The first SSVD component of the network traffic dataset, showing the 4th weekend is a long weekend. (In fact, it contains the July 4th holiday, a Friday that year.) (b) the second SSVD component of the network traffic dataset, showing the 19th row is a special weekday, which is similar to the weekends but with a smaller bump.

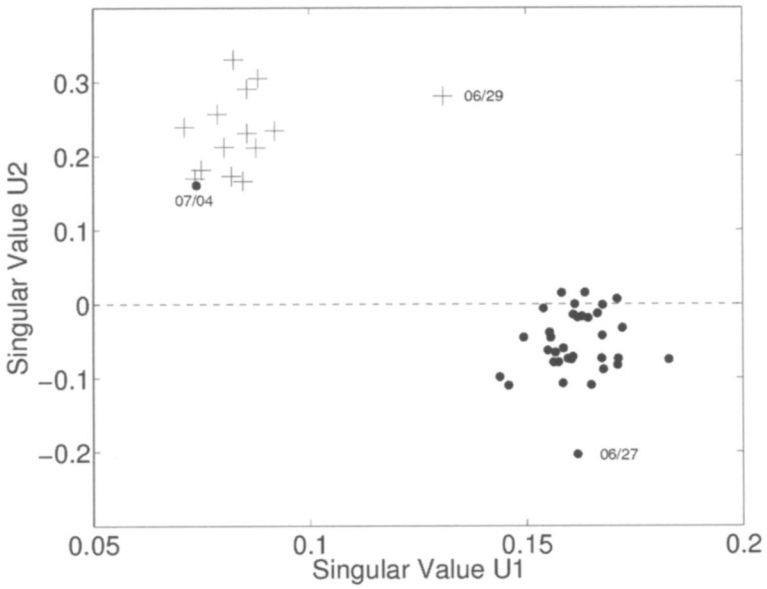

Figure 9. Scatterplots between singular columns $\mathbf{u}_1$ vs. $\mathbf{u}_2$ for the Internet data.

tional view of the daily shapes and cannot directly show what drives those days to be outliers. When the user employs SVD or PCA methods for a functional dataset, we strongly recommend viewing surface plots and the two types of SVD movies during the exploration.

6.2 DEMOGRAPHIC DATA

Understanding mortality is an important research problem in demography. The following analysis is trying to understand the variation of Spanish mortality among different age groups and across the years, following a first functional data analysis by Dr. Andrés M. Alonso Fernández. The Spanish mortality dataset (after a logarithm transformation) used here was collected such that each row represents an age group from 0 to 110, and each column represents a year between 1908 and 2002 (HMD 2005). We can view each column vector as a mortality curve of different age groups; or each row vector as a time series of mortality for a given age group. Figure 10 is a mesh plot of this mortality data. In the year direction, we find the mortality decreases when the year increases. It seems likely that improvements in medical care and in life quality made this happen. We also notice younger people benefit more than older people, shown as larger decreasing magnitudes in the surface plot. In order to understand the variation from 1908 to 2002 (95 years), and also the variation of mortality among different age groups, a natural choice for the analysis is the SSVD method. The generalized scree plot for this dataset, shows that the SSVD and the DSVD have similar approximation performance and complexity. However, the image plots for them show that the SSVD model has better interpretability, as discussed in the following. Note that the logarithm transformation is monotone, so the variation information will be kept after the transformation.

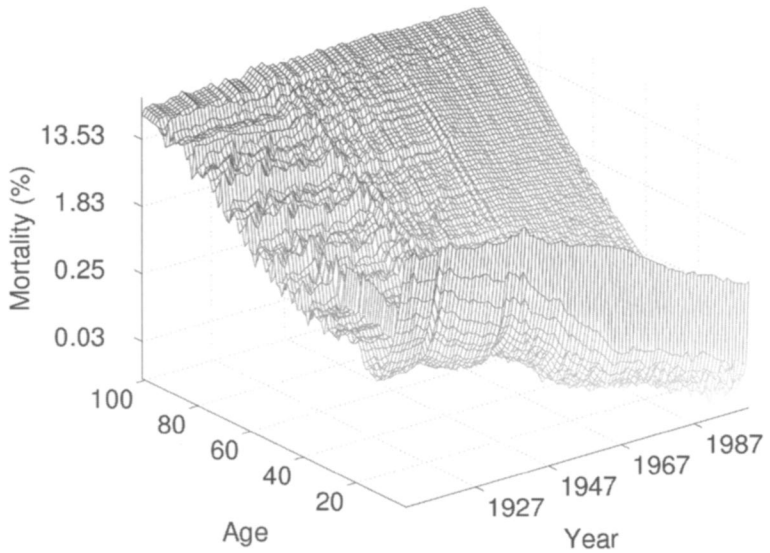

Figure 10. The mesh plot for the Spanish mortality dataset (after the logarithm transformation).

In this section, we use the image plots to reveal the underlying features of this dataset (see the surface plots for this dataset at Zhang 2006). As discussed in Section 4, the image plot is a variation of the surface plot, which is essentially viewing the surface from the top. This visualization helps to reveal the local variations and the interactions between columns and rows. Figure 11 is the image plot for this dataset. As discussed in Section 4, we assign the maximum value of each matrix to a dark color (black), and the minimum of each matrix to a light color (white).

From the image plots of the SSVD, we find that there is a decreasing trend of mortality over time in the first SSVD component (the top middle panel). It also suggests that the decrease among younger people is more significant than for older people. There are two dark regions (which show mortality increases dramatically) in the year direction. Checking the world history, we find these two outlying periods correspond to the 1918 world influenza pandemic, and the 1936–1939 Spanish civil war, which killed millions of people (in an unusual age distribution) in Spain. The second component (the top right panel) highlights an infant effect, because all the other age groups seem to have the same color (across time) in the image, but the infant group has a significantly decreasing trend over time. This shows that the mortality of infants decreased significantly over years (compared to the average).

The third component (the bottom left panel) seems to give an interesting clustering of the data. For example, in the age direction, there are four prominent groups, lower than 3, from 3 to 18, from 18 to 45, and over 45. While in the year direction, there are four major groups, which are from 1908–1935, 1936–1950, 1951–1985, and 1985–2002. Note that the first three years in the year group 1936–1950, shown as relative dark regions in the image, correspond to the Spanish civil war. The year group, from 1985 to 2002, corresponds to the increasing modern traffic fatalities. There is also a small dark region in the year group 1908–1935, which was driven by the 1918 influenza pandemic. In addition, the third com-

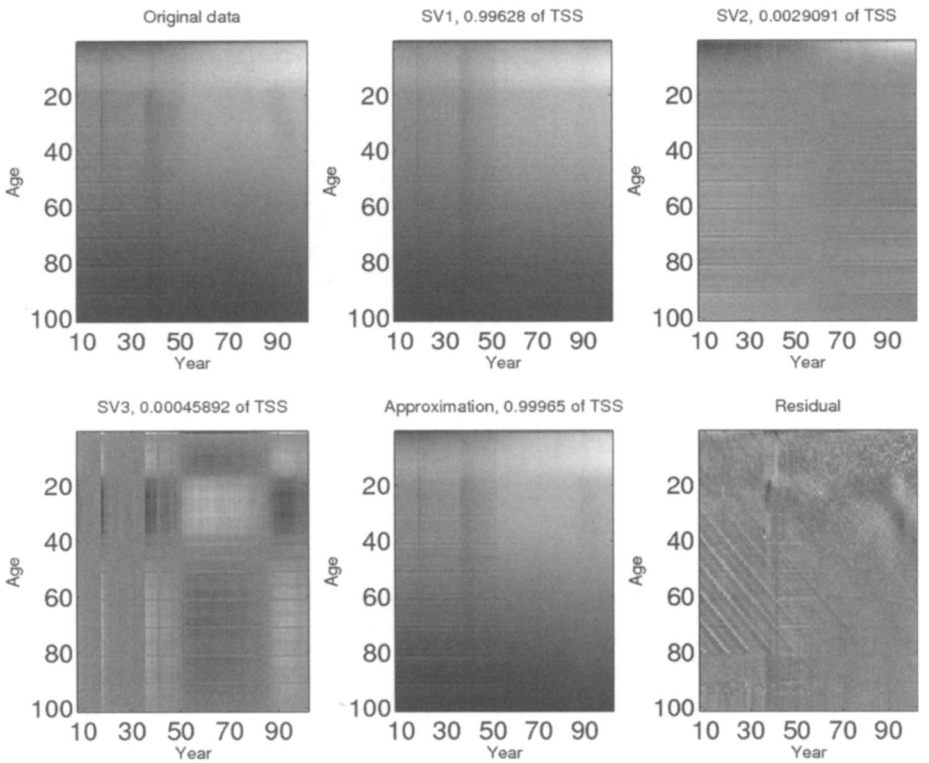

Figure 11. The image plots of the SSVD for the Spanish mortality data (after a logarithm transformation). Light colors correspond to smaller values, and darker colors are larger values. Interesting features in the residual part (the light diagonals in the lower left region) correspond to the cohort effects.

ponent shows several blocks in the matrix, which suggests that the mortalities of distinct age groups were affected differently by the special events. For example, the 1918 influenza pandemic, civil war and the modern traffic fatalities affected younger people more than other age groups, shown as the darker blocks in the picture. Note that these underlying features can also be found through the surface plots (available at Zhang 2006). However, the interpretation is more straightforward when using the image plots.

The surface plot of the residual matrix seems to show only noise, however, the image view of it highlights some interesting features, which might not be seen from other visualization methods. The plot (the bottom right panel in Figure 11) has some light and dark diagonal lines in the left bottom area. These diagonal lines turn out to be the same population group (*cohort*) among the years. The light lines are evenly distributed among the age groups, are 10 years apart, and all end at around 80 years of age. This strange behavior exists for about 40 years, that is, until 1958. A very likely reason for this is a rounding effect in the early years. This is believed to be due to imprecise death record, for example, if somebody died at age 39, the age at death was sometimes recorded as 40. It is clear that the image plot highlights this cohort effect (the interaction between age groups and years), which is harder to find using more conventional PCA visualization methods.

7. DISCUSSION

For noisy data, the SVD components can contain a lot of noise. Both of the referees suggest incorporating some smoothing techniques to reduce the noise in the component matrices. There is a lot of discussion between presmoothing before applying FDA methods or applying smoothing-incorporated FDA methods; for example, see (Ramsay and Silverman 2005, chap. 9). For our functional SVD method, if one prefers, one can presmooth the data matrix using a favorite smoother, and then apply our programs to gain some insights. We are working on incorporating some regularized SVD methods (e.g., Huang et al. 2007) into our programs, which provide a more natural incorporation of smoothing.

A referee observed that our notion of the two-way data matrix requires that the data are observed on common grid points. In cases with irregularly observed data points or missing data, one could preprocess the data using some smoothing, interpolation, or imputation methods to obtain regular observed data matrices, before applying our visualization methods.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. Andrés M. Alonso Fernández of Departamento de Estadística in Universidad Carlos III de Madrid, for providing the Spanish mortality data to us. Thanks also go to all the reviewers for their helpful comments and constructive suggestions. This research was partly supported by U.S. National Science Foundation grant DMS-0308331, DMS-0606577, and DMS-0605434.

[Received June 2005. Revised February 2007.]

REFERENCES

Cattell, R. B. (1966), "The Scree Test for the Number of Factors," *Multivariate Behavioral Research*, 1, 245–276.

Gabriel, K. R. (1971), "The Biplot Graphic Display of Matrices with Application to Principal Component Analysis," *Biometrika*, 58, 453–467.

——— (1978), "Least Squares Approximation of Matrices by Additive and Multiplicative Models," *Journal of Royal Statistical Society*, Series B, 40, 186–196.

Gabriel, K. R., and Odoroff, C. L. (1984), "Resistant Lower Rank Approximation of Matrices," in *Data analysis and Informatics, III*, Amsterdam: North Holland, pp. 23–30.

Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001), *Analysis of Time Series Structure*, New York: Chapman and Hall.

HMD (2005), "Human Mortality Database," University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available online at *http://www.mortality.org* or *http://www.humanmortality.de*.

Householder, A. S., and Young, G. (1938), "Matrix Approximations and Latent Roots," *American Mathematical Monthly*, 45, 165–171.

Huang, J. Z., Shen, H., and Buja, A. (2007), "Principal Component Analysis of Two-Way Functional Data Using Two-Way Regularized Singular Value Decompositions," technical report.

Jolliffe, I. T. (2002), *Principal Component Analysis*, New York: Springer-Verlag.

Long, C. (1983), "Visualization of Matrix Singular Value Decomposition," *Mathematics Magazine*, 56, 161–167.

Marron, J. S., Wendelberger, J. R., and Kober, E. M. (2004), *Time Series Functional Data Analysis*, Los Alamos National Lab, No. LA-UR-04-3911.

Ramsay, J. O., and Silverman, B. W. (2002), *Applied Functional Data Analysis, Methods and Case Studies*, New York: Springer-Verlag.

——— (2005), *Functional Data Analysis* (2nd ed.), New York: Springer-Verlag.

Shen, H., and Huang, J. Z. (2005), "Analysis of Call Center Arrival Data using Singular Value Decomposition," *Applied Stochastic Models in Business and Industry*, 21, 251–263.

Zhang, L. (2006), "Svd Movies and Plots for Singular Value Decomposition and its Visualization," Available online at *http://www.unc.edu/~lszhang/research/network/SVDmovie*.

Zhang, L., Marron, J. S., Shen, H., and Zhu, Z. (2006), "Singular Value Decomposition and Its Visualization," technical report, available at *http://www.unc.edu/~lszhang/research/SVDvisual.pdf*.