Web-based Supplementary Materials for

"Biclustering via Sparse Singular Value Decomposition"

by Mihee Lee, Haipeng Shen, Jianhua Z. Huang and J. S. Marron

This document contains the following supplementary materials for Lee et al. (2009):

- Additional simulation studies

- Analysis results for the food nutritional data

- Additional analysis of the lung cancer data using RoBiC and Plaid

Web Appendix A: Additional Simulation Studies

Case 1: Uniform Nonzero Entries

In this section, we report one additional simulation study where the nonzero entries of the true signal matrix are all the same. In addition to comparing SSVD with SVD, Plaid (Lazzeroni and Owen, 2002) and RoBiC (Asgarian and Greiner, 2008), we also investigate how the adaptive lasso weight parameters γ_1 and γ_2 affect the performance of SSVD.

For this simulation, we consider as the truth a 100×50 matrix $\mathbf{X}^*$ whose nonzero entries are the same. This is the simplest example because every nonzero element of $\mathbf{X}^*$ is the same, and hence is equally likely to be chosen by a sparse procedure.

Let $\boldsymbol{u}$ be a unit vector of length 100 with $u_i = 1/\sqrt{50}$ for $i = 1, \ldots, 50$, and $u_i = 0$ otherwise. In addition, let $\boldsymbol{v}$ be a unit vector of length 50 with $v_j = 1/5$ for $j = 1, \ldots, 25$, and $v_j = 0$ otherwise. Then $\mathbf{X}^* = s\boldsymbol{u}\boldsymbol{v}^T$ is a rank-1 matrix with uniform nonzero entries, i.e.

$$\mathbf{X}^* = s \left(\frac{1}{\sqrt{50}} \right) \left(\frac{1}{\sqrt{25}} \right) \begin{bmatrix} \mathbf{J} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix},$$

where $\mathbf{J}$ is the 50×25 matrix with entries 1, and $\mathbf{O}$ is the 50×25 zero matrix. For this simulation study, we set $s = 30$; hence the nonzero entries of $\mathbf{X}^*$ are 0.8485.

1

We generate the data matrix $\mathbf{X}$ as the sum of the true signal $\mathbf{X}^*$ and the noise $\boldsymbol{\epsilon} = [\epsilon_{i,j}]$, where $\{\epsilon_{i,j} : i = 1, \ldots, 100, \ j = 1, \ldots, 50\}$ is a random sample from the standard normal distribution. The nonzero entries of $\mathbf{X}^*$ (0.8485) are rather small compared to the noise. More specifically, the signal-to-noise ratio is only 0.135. The simulation is repeated 100 times.

As discussed in Section 3, we use BIC to choose the degree of sparsity in each updating step of the SSVD algorithm. For simplicity purposes, we set $\gamma_1 = \gamma_2$ (denoted by γ) in deciding the adaptive weight vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$. As for Plaid, we use the most flexible model described in Lazzeroni and Owen (2002), which performs the best among all plaid models.

Note that the goal of the analysis is to detect the sparse structure of the true signal matrix $\mathbf{X}^*$. We compare the performance among the following methods: SVD, Plaid, RoBiC, and SSVD with $\gamma = 0, \ 0.5, \ 2$. For notational convenience, we use the notation $Adt(\gamma)$ for the SSVD procedure based on the adaptive lasso penalty with the weight parameter γ. Here, we want to comment that the entries of $\mathbf{X}$ can be understood either as a multiplicative model, which is assumed by SVD, SSVD and RoBiC, or as an additive model, which is assumed by Plaid.

Table 1 reports the analysis results, in terms of the average number of zero elements in the estimated 100 singular vectors in both directions (Column 1), the average number (and proportion) of correctly identified zeros (Column 2), the average number (and proportion) of correctly identified nonzeros (Column 3), and the misclassification rate (Column 4). For example, considering the $\boldsymbol{u}$ vector, $\widehat{\boldsymbol{u}}$ obtained by Adt(2) contains 50.47 zero elements on average while the truth contains 50 zeros; in addition, on average, 49 (98%) of the zero entries are correctly estimated as zero, and 48.53 (97.06%) of the nonzero elements are correctly classified as nonzeros; as a result, 2.47% of the 100 entries are misclassified.

The first three rows of Table 1 compare the results of the SSVD with different adaptive weight parameter γ. Generally speaking, as γ increases, SSVD performs better in detecting zero elements, but slightly worse in finding nonzero entries. Adt(2) performs the best in terms of the misclassification rate. Base on this, we only report the results from Adt(2)

Table 1: Comparison of the performance among SVD, Plaid, RoBiC and SSVD

		Avg. # of zeros (True)	Avg. # of correctly identified zeros	Avg. # of correctly identified nonzeros	Misclassification rate
LASSO	u	41.12 (50)	40.95 (81.90%)	49.83 (99.66%)	9.22%
(Adt(0))	v	20.69 (25)	20.69 (82.76%)	25.00 (100.0%)	8.62%
Adt(0.5)	u	44.61 (50)	44.33 (88.66%)	49.72 (99.44%)	5.95%
	v	23.21 (25)	23.21 (92.84%)	25.00 (100.0%)	3.58%
Adt(2)	u	50.47 (50)	49.00 (98.00%)	48.53 (97.06%)	2.47%
	v	24.69 (25)	24.68 (98.72%)	24.99 (99.96%)	0.66%
RoBiC	u	48.37 (50)	47.06 (94.12%)	48.69 (97.38%)	4.25%
	v	24.87 (25)	24.84 (99.36%)	24.97 (99.88%)	0.38%
Plaid	u	45.00 (50)	44.70 (89.40%)	49.70 (99.40%)	5.60%
	v	23.20 (25)	23.20 (92.80%)	25.00 (100.0%)	3.60%
SVD	u	0.00 (50)	0.00 (0.00%)	50.00 (100.0%)	50.0%
	v	0.00 (25)	0.00 (0.00%)	25.00 (100.0%)	50.0%

for SSVD in the numerical studies below in this online supplement, as well as those in the main paper.

We now compare Adt(2) with the three other methods: the standard SVD, Plaid and RoBiC. As expected, SVD completely fails at detecting the sparse structure of the true signal. This confirms the necessity of incorporating sparsity in the model estimation. The improvement of Adt(2) over the standard SVD is clear in all categories. RoBiC performs slightly worse than Adt(2). The performance of Plaid is similar to that of Adt(0.5), which is worse than Adt(2) and RoBiC. We note that RoBiC yields an *outlying* estimate in one simulation.

Case 2: Rank-2 Approximation

For this simulation study, we consider a case where the rank of the true signal $\mathbf{X}^*$ is 2. Suppose $\mathbf{X}^*$ has the following form

$$\mathbf{X}^* = s_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + s_1 \boldsymbol{u}_2 \boldsymbol{v}_2^T,$$

where $s_1 = 1000$ and $s_2 = 100$, and

$$
\begin{aligned}
\widetilde{\boldsymbol{u}}_1 &= [r(20,2),\, r(10,4),\, r(3,8),\, r(1,16),\, r(0,70)]^T, \\
\widetilde{\boldsymbol{v}}_1 &= [r(1,20),\, r(0,30)]^T, \\
\widetilde{\boldsymbol{u}}_2 &= [r(0,6), 5, -5,\, r(0,6),\, r(10,4),\, r(-10,4),\, r(0,8),\, r(30,6),\, r(0,14)]^T, \\
\widetilde{\boldsymbol{v}}_2 &= [r(0,10),\, r(1,5),\, r(-1,5),\, r(0,30)], \\
\boldsymbol{u}_1 &= \widetilde{\boldsymbol{u}}_1/\|\widetilde{\boldsymbol{u}}_1\|, \quad \boldsymbol{v}_1 = \widetilde{\boldsymbol{v}}_1/\|\widetilde{\boldsymbol{v}}_1\|, \quad \boldsymbol{u}_2 = \widetilde{\boldsymbol{u}}_2/\|\widetilde{\boldsymbol{u}}_2\| \quad \text{and} \quad \boldsymbol{v}_2 = \widetilde{\boldsymbol{v}}_2/\|\widetilde{\boldsymbol{v}}_2\|.
\end{aligned}
$$

One can easily check that $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are orthogonal, so are $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Thus $\mathbf{X}^*$ is a 100×50 matrix of rank 2. As in the earlier simulation, the noise matrix $\boldsymbol{\epsilon}$ is generated from the standard normal distribution, and we then simulate the observation matrix $\mathbf{X}$ as the sum of $\mathbf{X}^*$ and $\boldsymbol{\epsilon}$. The simulation is again repeated 100 times.

Our SSVD procedure extracts the first sparse singular triplet $\{\widehat{s}_1, \widehat{\boldsymbol{u}}_1, \widehat{\boldsymbol{v}}_1\}$ from the raw data $\mathbf{X}$. Then, it is applied to the residual matrix $\mathbf{X} - \widehat{s}_1 \widehat{\boldsymbol{u}}_1 \widehat{\boldsymbol{v}}_1^T$ to extract the second triplet $\{\widehat{s}_2, \widehat{\boldsymbol{u}}_2, \widehat{\boldsymbol{v}}_2\}$. The SSVD estimate of $\mathbf{X}^*$ then has the form of $\widehat{s}_1 \widehat{\boldsymbol{u}}_1 \widehat{\boldsymbol{v}}_1^T + \widehat{s}_2 \widehat{\boldsymbol{u}}_2 \widehat{\boldsymbol{v}}_2^T$.

Table 2 compares the estimation results among the standard SVD, RoBiC and Adt(2). Plaid is not considered here for two reasons: first, the result of Plaid tends to be similar with (or worse than) that of RoBiC; second, the existing packages for Plaid either lack an automatic procedure to output the estimation results or only output the locations of the detected biclusters without estimating the Plaid model.

Again, the strawman SVD fails completely. One can easily see that Adt(2) performs well in both layers: it detects almost perfectly the zero/nonzero entries. On the other hand, RoBiC has trouble detecting the nonzero elements of $\boldsymbol{u}_1$ in particular, which is caused by the large variation in the entries of $\boldsymbol{u}_1$.

Table 2: Case 2: Comparison of performance among SVD, RoBiC and Adt(2)

		Avg. # of zeros (True)	Avg. # of correctly identified zeros	Avg. # of correctly identified nonzeros	Misclassification rate
Adt(2)	u_1	69.99 (70)	69.99 (99.99%)	30.00 (100.0%)	0.01%
	v_1	30.00 (30)	30.00 (100.0%)	20.00 (100.0%)	0.00%
	u_2	83.82 (84)	83.82 (99.79%)	16.00 (100.0%)	0.18%
	v_2	39.93 (40)	39.93 (99.83%)	10.00 (100.0%)	0.14%
RoBiC	u_1	92.00 (70)	70.00 (100.0%)	8.00 (26.67%)	22.00%
	v_1	30.00 (30)	30.00 (100.0%)	20.00 (100.0%)	0.32%
	u_2	84.84 (84)	83.99 (99.99%)	15.15 (94.69%)	0.86%
	v_2	40.00 (40)	40.00 (100.0%)	10.00 (100.0%)	0.00%
SVD	u_1	0.00 (70)	0.00 (0.00%)	30.00 (100.0%)	70.00%
	v_1	0.00 (30)	0.00 (0.00%)	20.00 (100.0%)	60.00%
	u_2	0.00 (84)	0.00 (0.00%)	16.00 (100.0%)	84.00%
	v_2	0.00 (40)	0.00 (0.00%)	10.00 (100.0%)	80.00%

Finally, we compare the estimated underlying signal $\mathbf{X}^*$ obtained by RoBiC and Adt(2). Figure 1 plots the heatmaps of the true signal matrix $\mathbf{X}^*$ and the average estimated signal given by RoBiC and Adt(2). For better visualization, only the first 50 rows and 25 columns are plotted. The same color scale is used for the three panels. Once again, it is clear that Adt(2) recovers the true sparse signal much better than RoBiC.
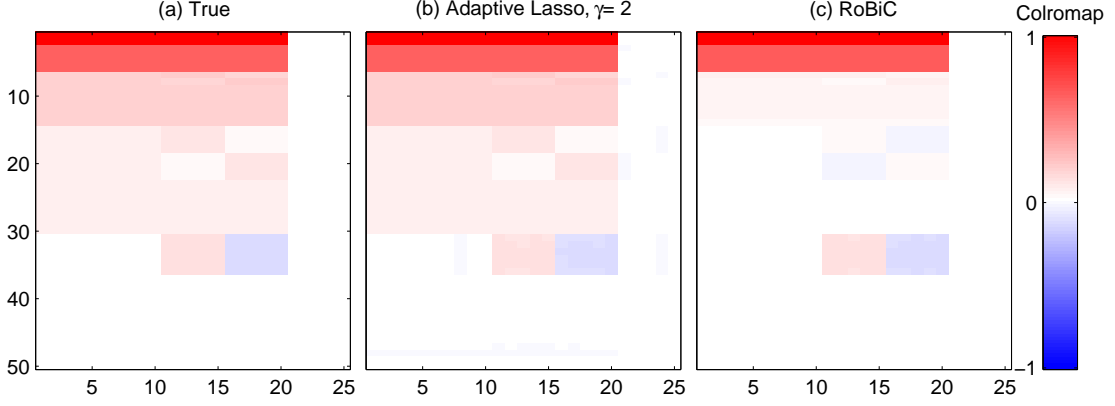
Figure 1: Case 2: Comparison of the heatmaps between RoBiC and Adt(2). Panel (a): the true $\mathbf{X}^*$, Panel (b): the average Adt(2) estimator of $\mathbf{X}^*$, Panel (c): the average RoBiC estimator of $\mathbf{X}^*$.

Web Appendix B: Food Nutritional Data

In this section we apply SSVD to the food nutritional data analyzed in Lazzeroni and Owen (2002) using Plaid. This data set is available at *http://www.ntwrks.com/~mikev/chart1.html*.

The data contain 6 nutritional measures for 961 foods: grams of fat, calories of food energy, grams of carbohydrate, grams of protein, milligrams of cholesterol, and grams of saturated fat. Since the foods are measured in different serving sizes, all the nutrition values are first standardized by dividing by the food weight. Similar to Lazzeroni and Owen (2002), the six variables are then centered and scaled. The final data matrix $\mathbf{X}$ is then a 961×6 matrix, where its entry $X_{i,j}$ is the standardized amount of the jth nutritional fact in the ith food. Note that each column of $\mathbf{X}$ is centered and scaled.

When applying the SSVD, we use the adaptive lasso penalty on both $\boldsymbol{u}$ and $\boldsymbol{v}$ with a common weight parameter $\gamma = 2$. We extract pairs of sparse singular vectors sequentially; since the (sparse) singular value s_k is almost zero after $k = 3$, as shown in Panel (a) of Figure 2, we decide to use only the first three layers. The BIC is used to select the sparsities for each pair of sparse vectors. We compare the SSVD results with those reported in Lazzeroni and Owen (2002).

One thing we want to mention here is that our iterative algorithm converges quite fast.
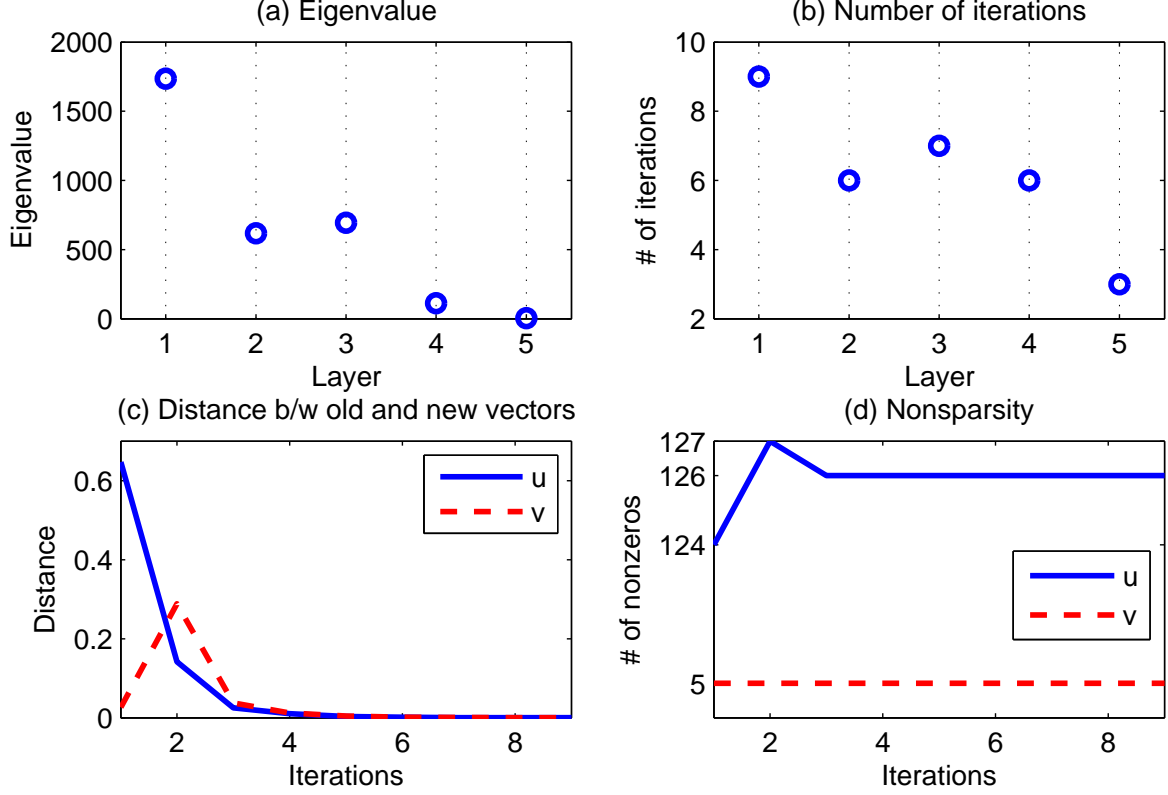
6

Figure 2: Food Data: Convergence Plots. Panel (a) displays the (sparse) eigenvalues according to the number of layers. It is clearly seen that the eigenvalue is almost zero after the third layer, which suggests to use the first three layers. Panel (b) shows how many iterations are required to extract each layer. For the first layer, Panel (c) displays the distance between u_{new} and u_{old} (similarly between v_{new} and v_{old}) as a function of number of iterations; Panel (d) plots the chosen sparsities at each iteration.

Panel (b) of Figure 2 shows that the algorithm always converges within 10 steps when extracting the first five layers. The two lower panels show more details of the convergence when extracting the first layer: Panel (c) displays the distance between the singular vector (u or v) and its update as a function of number of iterations, which becomes very small after about 4 iterations; Panel (d) shows that the resulting non-sparsity becomes stable after only 3 iterations.

Figure 3 plots the ordered leading entries of $\widehat{u}_1$ and $\widehat{u}_2$ as well as the entries of $\widehat{v}_1$

and $\widehat{\boldsymbol{v}}_2$. The first SSVD layer consists of 126 foods and 5 nutritional facts: Fat (0.6284), Saturated fat (0.5649), Food energy (0.5052), Carbohydrate (-0.1406) and Cholesterol (0.1052), where the numbers in the parentheses are the entries of $\widehat{\boldsymbol{v}}_1$. This implies that most of the variation in the first layer can be explained by Fat, Saturated Fat and Food energy. In addition, the fact that these variables are clustered together reveals a strong relationship between these nutritional facts. The second SSVD layer involves only eight foods and two nutritional variables: Cholesterol (0.9964) and Protein (0.0843). The third layer consists of 3 nutritional facts, Protein (0.9459), Cholesterol (0.2280) and Carbohydrate (-0.2308), with 111 foods; compared to the second layer which is mainly explained by cholesterol, most of variation in this layer can be explained by protein level.
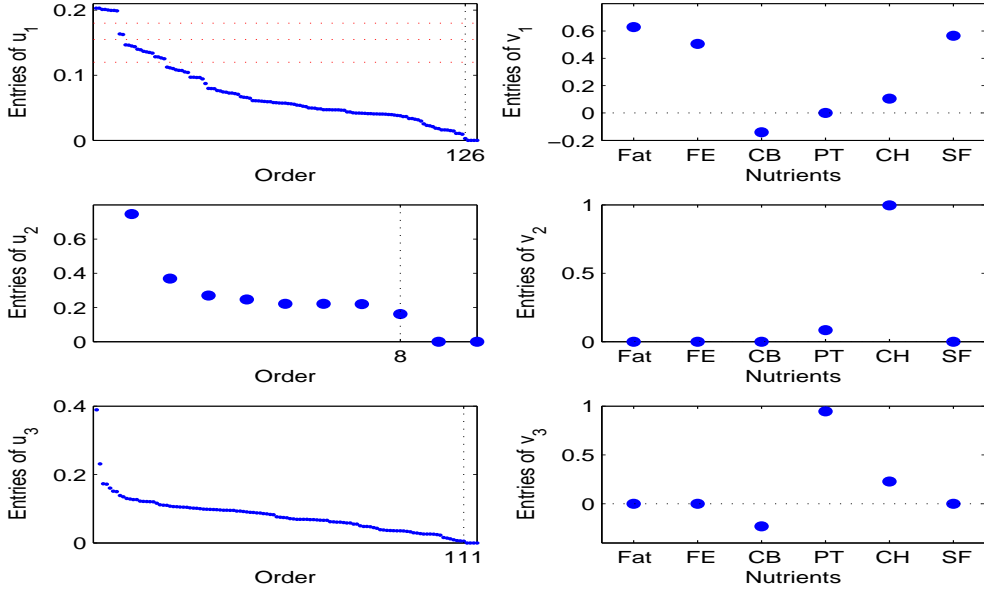

Figure 3: Food Data: The leading entries of $\widehat{\boldsymbol{u}}_k$ and the entries of $\widehat{\boldsymbol{v}}_k$, $k = 1, 2, 3$. The entries of $\widehat{\boldsymbol{u}}_k$ are ordered decreasingly. The variables for $\widehat{\boldsymbol{v}}_k$ are, from left to right, "Fat", "Food Energy", "Carbohydrate", "Protein", "Cholesterol" and "Saturated Fat".

Below we want to further understand the types of foods contained in each layer. The first layer contains many foods, and is harder to interpret. We thus further divide the foods into several groups according to the ordered entries in $\widehat{\boldsymbol{u}}_1$, as indicated by the dashed horizontal lines in the upper left panel of Figure 3 and tabulated below in Table 3. The four

groups are, in order, for solid oils, fats, liquid oils, and margarine and nuts. These foods contain more fat, saturated fat, and food energy per weight unit. For example, solid fats are well known to contain lots of saturated fat. The last part of the first layer (not shown here) contains the cheese-cream group and the meat-sausage group.

The interpretation of the second layer is much simpler than the first layer. As shown in the right-lower panel of Figure 3, Cholesterol explains most of the variation in this layer. This is reflected in the foods within this layer (Table 3): eggs and livers, which are well known to be high in cholesterol (and in protein). Similarly, the foods in the third layer can be interpreted as high-protein, high-cholesterol but low-carbohydrate foods.

For the same data set, as reported in Lazzeroni and Owen (2002), Plaid extracts ten layers, and the last five layers are dropped since they are small. The nutritional facts in the first SSVD layer are also detected by Plaid as the significant variables in its first layer. The top 20 foods (Table 3) are the same as those detected by Plaid, see Table 2 in Lazzeroni and Owen (2002). The second layer of SSVD is related to the fifth layer of Plaid which selects Cholesterol as the only significant variable and contains 59 foods with the leading eight the same as those detected by SSVD (Table 3). In addition, the third layer of SSVD corresponds to the second layer of Plaid.

Table 3: Food Data: Leading food groups in the first three layers

Foods		$\hat{u}_{1,i}$	Foods		$\hat{u}_{1,i}$
Layer 1, Group 1					
BUTTER, SALTED	1/2 CUP	0.2029	BUTTER, UNSALTED	1/2 CUP	0.2029
BUTTER, SALTED	1 TBSP	0.2010	BUTTER, UNSALTED	1 TBSP	0.2010
LARD	1 CUP	0.2002	BUTTER, SALTED	1 PAT	0.1995
BUTTER, UNSALTED	1 PAT	0.1995	LARD	1 TBSP	0.1989
Layer 1, Group 2					
FATS, CKING/VEGETBL SHORT.	1 TBSP	0.1635	FATS, CKING/VEGETBL SHORT.	1 CUP	0.1625
Layer 1, Group 3					
SOYBEAN-COT.SEED OIL, HYDRGN	1 TBSP	0.1467	SOYBEAN-COT.SEED OIL, HYDRGN	1 CUP	0.1462
PEANUT OIL	1 TBSP	0.1450	PEANUT OIL	1 CUP	0.1438
SOYBEAN OIL, HYDRGN	1 TBSP	0.1400	SOYBEAN OIL, HYDRGN	1 CUP	0.1391
OLIVE OIL	1 TBSP	0.1367	OLIVE OIL	1 CUP	0.1359
CORN OIL	1 TBSP	0.1350	CORN OIL	1 CUP	0.1339
SUNFLOWER OIL	1 CUP	0.1283	SUNFLOWER OIL	1 TBSP	0.1282
SAFFLOWER OIL	1 TBSP	0.1265	SAFFLOWER OIL	1 CUP	0.1253
Layer 1, Group 4 (Top 14)					
IMITATION CREAMERS, POWDERED	1 TSP	0.1099	MARGARINE, REG.,HARD,80% FAT	1/2 CUP	0.1124
MARGARINE, REG.,HARD,80% FAT	1 PAT	0.1112	MARGARINE, REG.,HARD,80% FAT	1 TBSP	0.1104
IMITATION CREAMERS, POWDERED	1 TSP	0.1083	MARGARINE, REG.,SOFT,80% FAT	8 OZ	0.1075
CHOCOLATE, BITTER	1 OZ	0.1073	MARGARINE, REG.,SOFT,80% FAT	1 TBSP	0.1051
MAYONNAISE, REG.	1 TBSP	0.1042	MACADAMIA, OILRSTD,SALTED	1 OZ	0.0969
MACADAMIA, OILRSTD,UNSALT	1 OZ	0.0969	MACADAMIA, OILRSTD,SALTED	1 CUP	0.0965
MACADAMIA, OILRSTD,UNSALT	1 CUP	0.0965	BRAZIL NUTS	1 OZ	0.0943
Layer 2					
EGGS, RAW, YOLK	1 YOLK	0.7464	CHICKEN LIVER, COOKED	1 LIVER	0.3692
BEEF LIVER, FRIED	3 OZ	0.2699	EGGS, COOKED, FRIED	1 EGG	0.2465
EGGS, COOKED, HARD-COOKED	2 EGG	0.2217	EGGS, RAW, WHOLE	1 EGG	0.2217
EGGS, COOKED, POACHED	1 EGG	0.2202	EGGS, COOKED, SCRAMBLED/OMELET	1 EGG	0.1619
Layer 3 (Top 16)					
GELATIN, DRY	1 ENVELP	0.3893	SEAWEED, SPIRULINA, DRIED	1 OZ	0.2313
PARMESAN CHEESE, GRATED	1 OZ	0.1731	PARMESAN CHEESE, GRATED	1 CUP	0.1716
PARMESAN CHEESE, GRATED	1 TBSP	0.1604	LAMB,CHOPS,ARM,BRAISED,LEAN	1.7 OZ	0.1514
YEAST, BAKERS, DRY, ACTIVE	1 PKG	0.1499	TUNA, CANND, DRND,WATR, WHITE	3 OZ	0.1384
PORK SHOULDER, BRAISD, LEAN	2.4 OZ	0.1347	LAMB,CHOPS,ARM,BRAISED,LEAN+FT	2.2 OZ	0.1299
BEEF, DRIED, CHIPPED	2.5 OZ	0.1282	BEEF, CKD,BTTM ROUND,LEAN ONLY	2.8 OZ	0.1267
PORK CHOP, LOIN, BROIL, LEAN	2.5 OZ	0.1265	BEEF HEART, BRAISED	3 OZ	0.1223
CHICKEN, FRIED, FLOUR, BREAST	3.5 OZ	0.1216	YEAST, BREWERS, DRY	1 TBSP	0.1210

Web Appendix C: Lung Cancer Data

In this section, we apply RoBiC and Plaid to analyze the lung cancer gene expression data, and compare the analysis results with those obtained by SSVD, reported in Section 2 of Lee et al. (2009). To make the comparison meaningful, all three methods are applied to extract the first three bicluster layers.

The three RoBiC layers consist of 872, 898 and 1,509 genes respectively, and 2,518 unique genes in total. As for Plaid, the three layers include 7,754, 7,862 and 6,433 genes and 12,125 unique genes in total, much more than those detected by either RoBiC or SSVD which selected 4,545 unique genes.

The analysis results are presented in Figures 4 to 9 below. The corresponding plots based on the SSVD analysis can be found in Figures 1 to 3 of Lee et al. (2009). In particular,

- Figures 4 and 5 sequentially show the image plots of the three RoBiC and Plaid layers, respectively. The plot of SSVD is Figure 1 in the paper.

  For better visualization of the gene grouping, the columns of each layer are rearranged based on an ascending ordering of the gene effects. For RoBiC, such effects are reflected by the entries of the right sparse singular vectors $\widehat{\boldsymbol{v}}_k$; for Plaid, they are measured by the estimated column effects $\widehat{\beta}_k$, see the Plaid model equation on Page 13 of the paper.

  For RoBiC (Figure 4), 10,000 zeroed-out genes are excluded when plotting, and the boundaries of the white areas are indicated. Similarly for Plaid (Figure 5), 2,500 zeroed-out genes are not plotted.

- Figures 6 and 7 show the scatter plots among the subject effects in the first three biclusters, for RoBiC and Plaid respectively. The plot of SSVD is Figure 2 in the paper. Different cancer types are indicated with different symbols and colors.

  The subject effects are measured by the entries of the sparse left singular vectors $\widehat{\boldsymbol{u}}_k$ for RoBiC, or the estimated row effects $\widehat{\alpha}_k$ for Plaid, as defined in the Plaid model equation on Page 13 of the paper.

- Similar to Figure 3 in the paper for SSVD, Figure 8 and 9 plot the raw data matrix and its RoBiC or Plaid low-rank matrix approximations obtained by adding up the first k bicluster layers, for $k = 1, 2, 3$.

  The dashed vertical lines separate out the various gene groups obtained by the first three biclusters. To reveal the details better, only the first 5,000 genes are plotted in Figure 8 for RoBiC. In Figure 9 for Plaid, all the 12,625 genes are plotted, because the three layers contain 12,125 genes in total.

Several observations can be made when comparing these figures with Figures 1 to 3 in the paper, which are summarized below. The comparison suggests that SSVD gives better biclustering results than either RoBiC or Plaid, especially in terms of subject grouping and low-rank approximation of the original matrix.

- Based on Figures 4 and 5,

  - RoBiC tends to only pick up the stronger contrasts identified by SSVD. For example, the first RoBiC bicluster picks up the contrast between Carcinoid and Normal patients, but not the contrast between Colon and SmallCell subjects; the second bicluster only includes part of the Colon group. This is consistent with what is observed through the simulation studies: RoBiC tends to miss the smaller effects when the signal has many levels, which is caused by the hinge model. In addition, the third layer only includes one Normal subject and one SmallCell patient.

  - as for Plaid, the first layer includes all the 56 subjects, the second layer contains Normal and SmallCell subjects, and the third one covers the Carcinoid group and part of the Colon group.

- Based on Figures 6 and 7,

  - RoBiC leads to a different clustering of the subjects based on the first three left singular vectors. Note that each panel has a cluster at the origin that contains some subjects from the Colon, Normal and SmallCell groups.
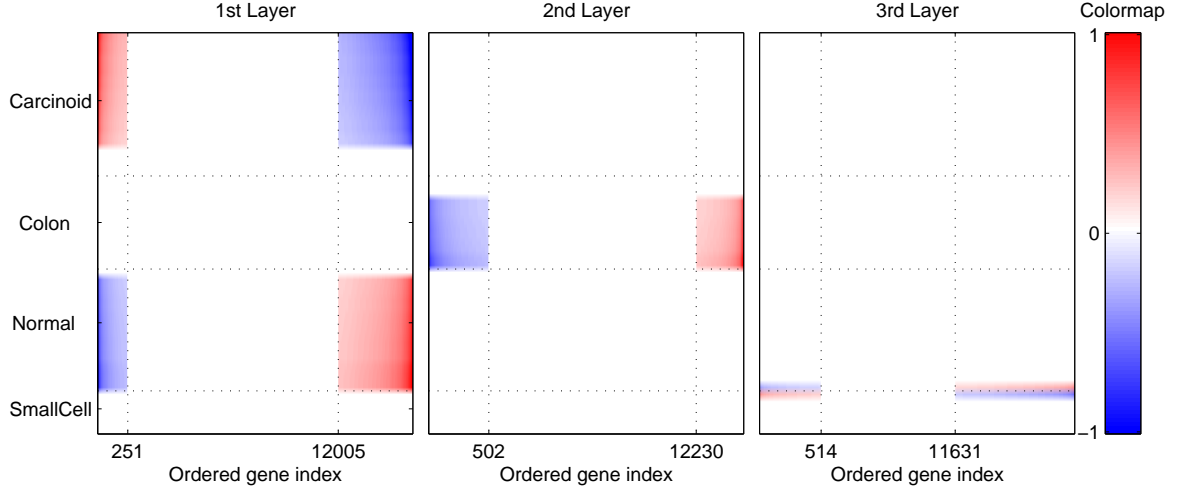
12

Figure 4: Lung Cancer Data: Image plots of the first three RoBiC layers $\widehat{s}_k\widehat{\boldsymbol{u}}_k\widehat{\boldsymbol{v}}_k^T$ ($k = 1, 2, 3$). In each panel, the genes are rearranged according to an increasing order of the entries of $\widehat{\boldsymbol{v}}_k$, and subjects are also rearranged according to the values of $\widehat{\boldsymbol{u}}_k$ within each subject group. (In each panel, 10,000 genes in the middle white area are excluded when plotting.)

- the clustering result from Plaid is much worse than either SSVD or RoBiC: the four cancer groups are never clearly separated.

• Based on Figures 8 and 9,

- the low-rank approximation performance of RoBiC is worse than SSVD, while Plaid gives the worst approximation. The appearance of the rank-3 Plaid approximation is mainly due to the dominant first layer.

References

Asgarian, N. and Greiner, R. (2008). Using rank-1 biclusters to classify microarray data. *Technical Report, University of Alberta, Canada* .

Lazzeroni, L. and Owen, A. (2002). Plaid models for gene expression data. *Statistica Sinica* **12,** 61–86.
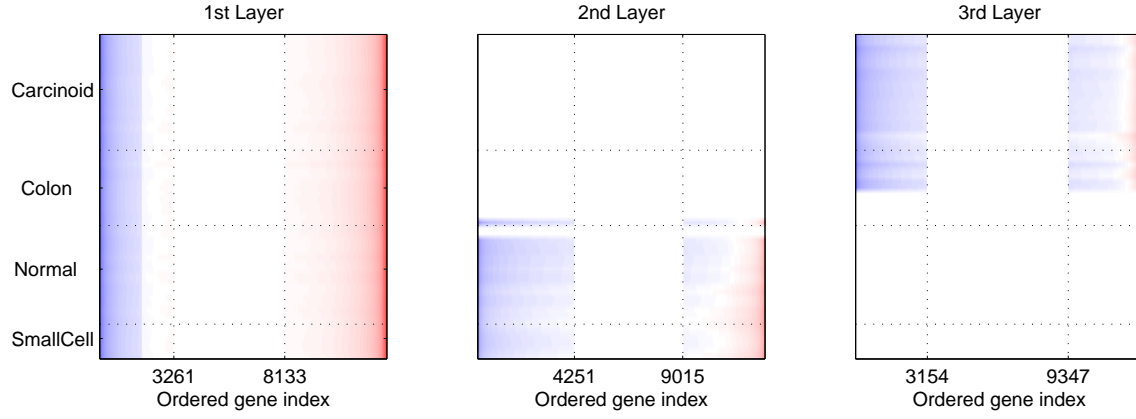
Figure 5: Lung Cancer Data: Image plots of the first three Plaid layers $(\hat{\mu}_k + \hat{\alpha}_k + \hat{\beta}_k)\hat{\rho}_k\hat{\kappa}_k$ $(k = 1, 2, 3)$. In each panel, the genes are rearranged according to an increasing order of the entries of $\hat{\beta}_k\hat{\kappa}_k$, and subjects are also rearranged according to the values of $\hat{\alpha}_k\hat{\rho}_k$ within each subject group. (In each panel, 2,500 genes in the middle white area are excluded when plotting.)

Lee, M., Shen, H., Huang, J. Z., and Marron, J. S. (2009). Biclustering via sparse singular value decomposition. *Biometrics, submitted* .
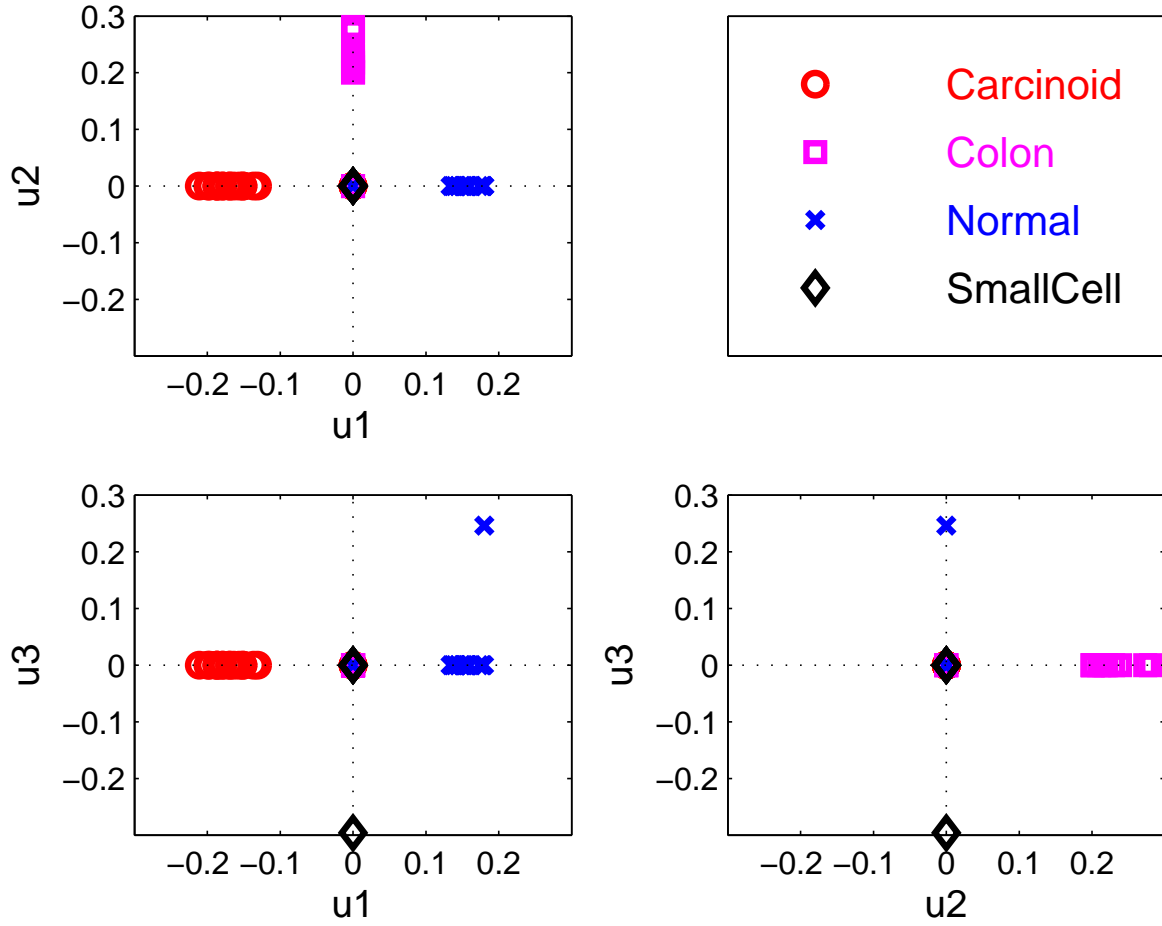
Figure 6: Lung Cancer Data: Scatter plots of the entries of the first three left sparse singular vectors $\widehat{\boldsymbol{u}}_k$ ($k = 1, 2, 3$), obtained by RoBiC.
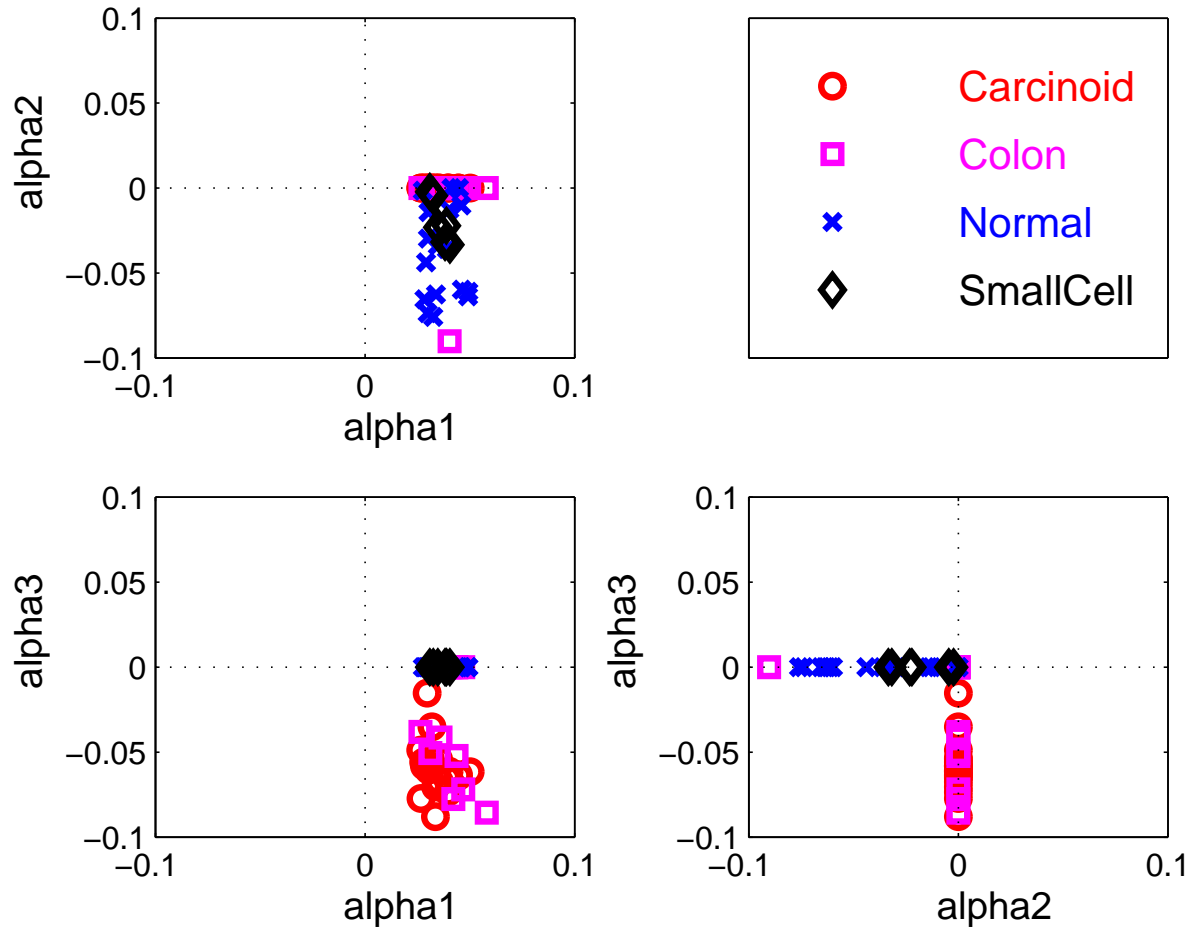
Figure 7: Lung Cancer Data: Scatter plots of the entries of the first three row-effect vectors $\widehat{\alpha}_k$ $(k = 1, 2, 3)$, obtained by Plaid.
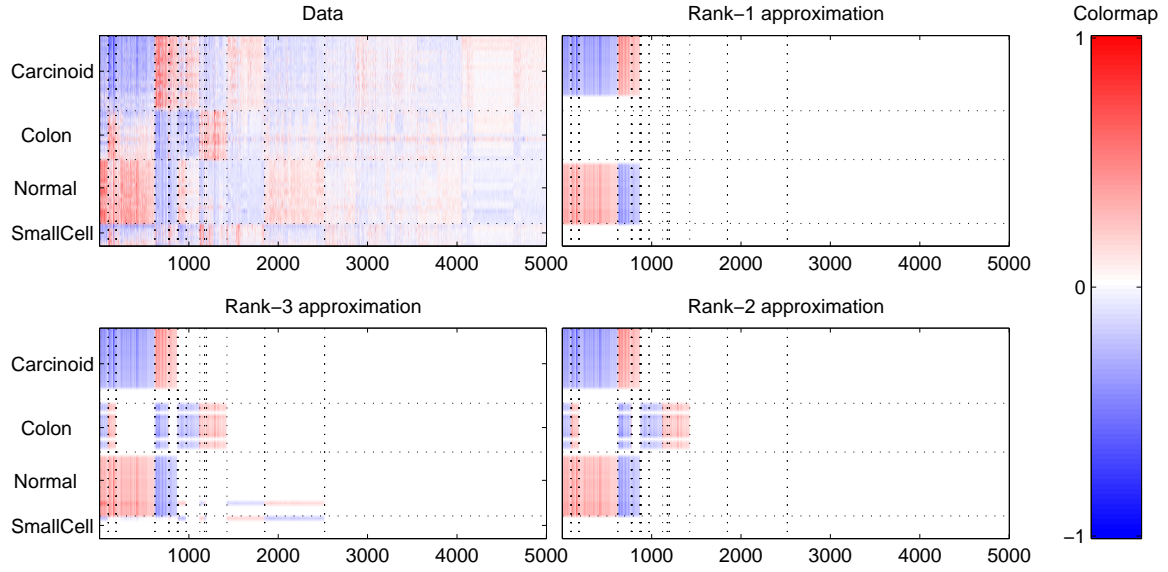
Figure 8: Lung Cancer Data: Comparison of the raw data matrix and the best RoBiC rank-k approximations ($k = 1, 2, 3$). (Only the first 5,000 genes are plotted to better reveal the details.)
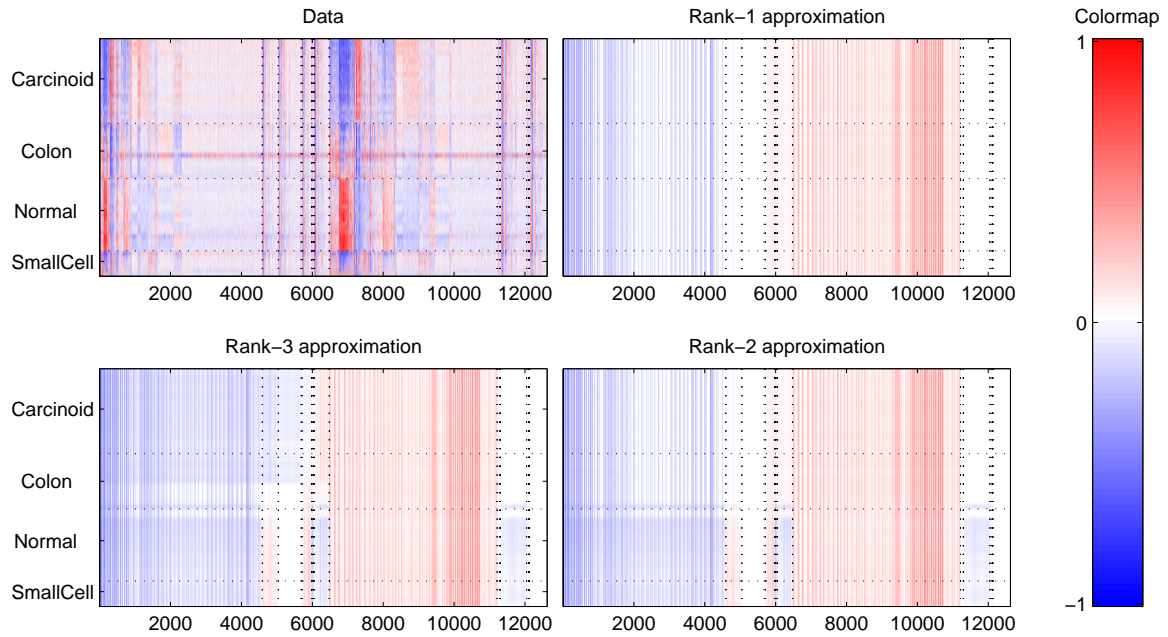

Figure 9: Lung Cancer Data: Comparison of the raw data matrix and the best Plaid rank-k approximations ($k = 1, 2, 3$).

17