

Robust Singular Value Decomposition Based on Weighted Least Absolute Deviation Regression

Kang-Mo Jung^{1,a}

^aDepartment of Informatics & Statistics, Kunsan National University and
Department of Statistics, University of California at Davis

Abstract

The singular value decomposition of a rectangular matrix is a basic tool to understand the structure of the data and particularly the relationship between row and column factors. However, conventional singular value decomposition used the least squares method and is not robust to outliers. We propose a simple robust singular value decomposition algorithm based on the weighted least absolute deviation which is not sensitive to leverage points. Its implementation is easy and the computation time is reasonably low. Numerical results give the data structure and the outlying information.

Keywords: Outliers, robust statistics, singular value decomposition, weighted least absolute deviation

1. Introduction

The singular value decomposition(SVD) of a rectangular matrix is a valuable tool in unveiling its structure and has found wide-ranging applications such as data analysis, signal processing, pattern recognition, image processing and compression. In statistical analysis, SVD is fundamental for using the biplot (Bradu and Gabriel, 1978), correspondence analysis (Greenacre, 1984) and principal component analysis. Recently Liu *et al.* (2003) used SVD for analyzing microarray data. It is known that gene expression data are well suited to analysis using SVD.

The conventional approach to the SVD does not permit any missing elements, for example microgene data is usually not complete. Gabriel and Zamir (1979) proposed an alternative method to address this problem, an alternating least squares(ALS) algorithm. However, they used a least squares algorithm which is known to be highly sensitive to outliers. In the extreme case, an individual cell, if sufficiently outlying, can draw even the leading principal component toward itself (Hawkins *et al.*, 2001). It is therefore desirable to construct a robust SVD.

To achieve robustness, Hawkins *et al.* (2001) considered a robust fit in alternating regression using least absolute deviation(ALAD) regression. Liu *et al.* (2003) used the least trimmed squares(LTS) method (Rousseeuw, 1984) in alternating fitting. Croux *et al.* (2003) implemented the weighted LAD estimator in factor analysis model. Chen *et al.* (2008) compared the above robust estimators in the sense of computational issues. They concluded that the LTS estimator is subject to dead cycles and it is computationally expensive, even if the LTS estimator is robust against high leverage outliers

This work was financially supported by the Kunsan National University's Long-term Overseas Research Program for Faculty Member in the year 2009.

¹ Professor, Department of Informatics & Statistics, Kunsan National University, Kunsan 573-701, Korea.
E-mail: kmjung@kunsan.ac.kr

with its fast algorithm available in R and SAS. The LAD estimator always converges, but it is sensitive to leverage points (Croux *et al.*, 2003). We propose a robust SVD based on a weighted LAD estimator (AWLAD).

In Section 2 we review SVD and discuss the deficiencies of the conventional SVD such as non-robustness and intractability of missing data. In Section 3 we describe robust alternating regression estimators to resolve the above deficiencies that are followed by our proposed algorithm of AWLAD with the consideration of computational time and robustness in Section 4. Some numerical results of the proposed algorithm are given in Section 5.

2. Conventional Singular Value Decomposition

Let \mathbf{X} denote an $n \times p$ matrix of real-valued data and rank r , where without loss of generality $n \leq p$, and therefore $r \leq p$. And let x_{ij} denote the element of \mathbf{X} . The equation for SVD of the matrix \mathbf{X} is

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T,$$

where \mathbf{U} is the $n \times p$ column-orthogonal matrix of left eigenvectors, \mathbf{V} the $p \times p$ orthogonal matrix of right eigenvectors, and \mathbf{D} the $p \times p$ diagonal matrix with elements d_i s which are called singular values. Let \mathbf{u}_i and \mathbf{v}_i for the i th left and right eigenvectors, respectively. Then the matrix \mathbf{X} can be rewritten as

$$\mathbf{X} = \sum_{i=1}^r d_i \mathbf{u}_i \mathbf{v}_i^T. \quad (2.1)$$

When we sum up to small k instead of r , the term $\mathbf{X}^{(k)} = \sum_{i=1}^k d_i \mathbf{u}_i \mathbf{v}_i^T$ is the closest rank- k matrix to \mathbf{X} . The term *closest* means that $\mathbf{X}^{(k)}$ minimizes the sum of the squares of the difference of the elements of \mathbf{X} and $\mathbf{X}^{(k)}$.

The SVD is often found through a principal component analysis on either $\mathbf{X}^T \mathbf{X}$ or $\mathbf{X} \mathbf{X}^T$. The eigenvalues and eigenvectors of $\mathbf{X}^T \mathbf{X}$ becomes the squares of eigenvalues and the right eigenvectors \mathbf{V} of the matrix \mathbf{X} , respectively, because we have

$$\mathbf{X}^T \mathbf{X} = \mathbf{V} \mathbf{D} \mathbf{U}^T \mathbf{U} \mathbf{D} \mathbf{V}^T = \mathbf{V} \mathbf{D}^2 \mathbf{V}^T.$$

And then $\mathbf{U} = \mathbf{X} \mathbf{V} \mathbf{D}^{-1}$. The conventional SVD has two serious deficiencies. First, it requires the complete data matrix \mathbf{X} . Even if only one element was missed, then the computation is implausible. The missing data occurs often in experimental data such as microarray data. Second, the conventional SVD uses a least squares method and it is very sensitive to outliers or susceptible observations. Such outliers should be accommodated when dealing with microarray data, where a sprinkling of entries are found to be very large or small (Liu *et al.*, 2003).

3. Robust Alternating Regressions

Gabriel and Zamir (1979) introduced so called alternating regression (or criss-cross regression) with least square fits to resolve the missing information deficiency. The procedure seeks the leading eigen triple $d_1, \mathbf{u}_1, \mathbf{v}_1$ of the least squares approximation to \mathbf{X} . The ALS algorithm iteratively for $d_i, \mathbf{u}_i, \mathbf{v}_i$ for $i = 2, \dots, p$. Denote by $\theta = (d_1, \dots, d_p, \mathbf{u}_1^T, \dots, \mathbf{u}_p^T, \mathbf{v}_1^T, \dots, \mathbf{v}_p^T)$ the vector of all singular values and eigenvectors and let

$$\hat{x}_{ij}(\theta) = \sum_{l=1}^p d_l u_{il} v_{jl}$$

be the fitted value of x_{ij} by the singular value decomposition. Given \mathbf{u}_1 we can get the regression coefficient $d_1 \mathbf{v}_1$ using a least squares method by minimizing for $j = 1, \dots, p$,

$$\sum_{i=1}^n (x_{ij} - \hat{x}_{ij}(\theta))^2. \quad (3.1)$$

By scaling the vector to unit length we have \mathbf{v}_1 and the scale factor d_1 . And then given \mathbf{v}_1 we can get \mathbf{u}_1 by a similar procedure. The eigen triple $d_1, \mathbf{u}_1, \mathbf{v}_1$ gives the first term of the conventional SVD.

Gabriel and Zamir (1979) showed that the procedure converges to the true leading eigenvalue and eigenvector triple, because each step of the ALS decreases the residual sum of squares from that of the previous step. After the first leading eigen triple has been found, the data matrix \mathbf{X} is replaced by $\mathbf{X} - d_1 \mathbf{u}_1 \mathbf{v}_1^T$ and the procedure is conducted to obtain the second leading eigen triple $d_2, \mathbf{u}_2, \mathbf{v}_2$.

The ALS approach is stable and is a remedy for missing data. However it is well known that the least squares estimator is non-robust against outliers. In the extreme situation of a microarray context, a single cell may draw even the leading principal component toward itself. It is therefore desirable to develop a robust SVD. As a robust method instead of a least squares criterion Hawkins *et al.* (2001) considered the LAD estimator by minimizing the absolute residual sum instead of (3.1). For example, given \mathbf{u}_1 we use the criterion for $j = 1, \dots, p$,

$$\sum_{i=1}^n |x_{ij} - \hat{x}_{ij}(\theta)|. \quad (3.2)$$

The LAD estimator always converges; however, it is sensitive to leverage points (Croux *et al.*, 2003). Liu *et al.* (2003) used the LTS estimator in alternating fitting. Chen *et al.* (2008) compared robust algorithms in computational speed and concluded that the LTS estimator is subject to dead cycles that do not improve the procedure. Therefore, the LTS in alternating regression has computational burden of time, even if the LTS estimator is robust against high leverage outliers.

Croux *et al.* (2003) proposed the weighted LAD estimator with a focus on a factor analysis model. They used multiple regressions instead of simple regression iteratively. However, the procedure does not guarantee the leading singular values and it cannot be adopted to the SVD. Gabriel and Zamir (1979) described that the multiple regressions does not prove the convergence to the minimum of the criterion in ALS. Thus, we propose a robust SVD based on an iteratively weighted LAD estimator.

4. Weighted LAD Algorithm

Given \mathbf{u}_1 the criterion (3.2) is replaced by the criterion of minimizing the weighted LAD norm

$$\sum_{i=1}^n w_i |x_{ij} - \hat{x}_{ij}(\theta)|, \quad (4.1)$$

where w_i is defined by

$$w_i(\theta) = \min \left(1, \frac{z_{0.05}^2}{\text{RD}_i^2} \right), \quad i = 1, \dots, n \quad (4.2)$$

and $z_{0.05}$ is the upper 5% critical value of the standard normal distribution and

$$\text{RD}_i^2 = \frac{(u_{i1} - T(U))^2}{C(U)}$$

are robust distances (Rousseeuw and van Zomeren, 1990) computed from the collection $U = \{u_{i1} \mid 1 \leq i \leq n\}$. The robust location and scale estimators $T(U)$ and $C(U)$ are taken as the median and median absolute deviation, which is a univariate version of the minimum volume estimator (Croux *et al.*, 2003). Giloni *et al.* (2006) proposed another method to determine weights for each observation. Analogously given \mathbf{v}_1 the weight l_j is defined on \mathbf{v}_1 .

We use the criterion (4.1) instead of (3.2). It makes the estimator robust to leverage points and faster than the LTS estimator. The SVD based on the weighted LAD are outlined below.

Step 1. Set the starting point \mathbf{u}_1 as the vector of median absolute values of each row.

Step 2. For $i = 1, \dots, n$, compute the weight w_i defined in (4.2), which decreases the weight of outlying i th row in the set of estimated vectors U . Then fit the weighted LAD regression coefficient by computing

$$c_j = \operatorname{argmin} \sum_{i=1}^n w_i |x_{ij} - c_j u_{i1}|$$

for $j = 1, \dots, p$.

Step 3. Calculate the right eigenvector $\mathbf{v}_1 = \mathbf{c}/\|\mathbf{c}\|$, where $\|\mathbf{c}\|$ is the Euclidean norm and $d_1 = \|\mathbf{c}\|$.

Step 4. Compute the weight l_j for $j = 1, \dots, p$ and fit the weighted LAD regression coefficient by computing

$$g_i = \operatorname{argmin} \sum_{j=1}^p l_j |x_{ij} - g_i v_{j1}|$$

for $i = 1, \dots, n$.

Step 5. Calculate the left eigenvector $\mathbf{u}_1 = \mathbf{g}/\|\mathbf{g}\|$.

Step 6. Iterate Steps 2–6 until the convergence is satisfied.

The procedure continues until the maximum absolute value among the previous estimators and the current estimators is less than 10^{-5} .

Now the first eigen triple is obtained. For the second eigen triple the above steps are performed with $\mathbf{X} = \mathbf{X} - d_1 \mathbf{u}_1 \mathbf{v}_1^T$. This iteration will be succeeded for $1, \dots, p$.

The singular values of the conventional SVD can be usually arranged in decreasing order and the corresponding eigenvectors can also be done. However, this is not necessarily the case with robust SVD methods (Hawkins *et al.*, 2001). Note that eigenvectors by the proposed algorithm may not be orthogonal. Orthogonality requires only least squares criterion.

5. Numerical Examples

We compare the results of ALS, ALAD and AWLAD for simulation data and a real data by constructing a biplot.

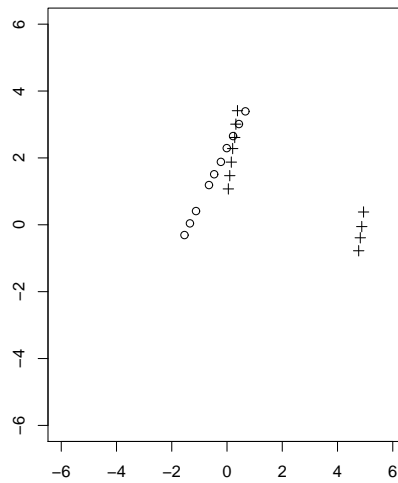


Figure 1: The biplot based on the SVD by alternating least squares with outliers

5.1. Simulation

The simulation data is generated by the 11 by 11 additive data matrix $x_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$, where $\mu = 1$; $\alpha = -5, -4, \dots, 4, 5$; $\beta = -5, -4, \dots, 4, 5$ and the error e_{ij} comes from $N(0, 0.125)$. This data scheme is used in Hawkins *et al.* (2001). The data matrix is of approximate rank two. An additive model for the data matrix \mathbf{X} can be confirmed by the row and column markers that are both straight lines and lie at right angles. We can get the biplots for the ALS, ALAD and AWLAD that are similar to the plot in which there are two perpendicular lines to each other. This correctly diagnoses an additive model (Bradru and Gabriel, 1978).

Now after adding four outliers (adding 15 to four random cells) in the data we conducted three SVD methods. We will show that the ALS is very sensitive to these four outliers, but AWLAD is not. The lines from the ALS are not straight lines, but the plots from the other robust methods are almost the same as before. It indicates that the ALS is not robust to a contaminated data, but the ALAD and AWLAD are robust.

We contaminate the data by adding four outliers (adding 50 to four consecutive cells). The biplots by those methods are depicted in Figures 1 to 3. As shown in Figures 1 and 2, the ALS and the ALAD do not represent a straight line. In addition, we cannot infer an additive model. However, in Figure 3 the AWLAD presents straight lines which are perpendicular to each other and gives the outlying information on one row of data matrix. Figure 2 may be resulted from the fact that the ALAD estimator is sensitive to leverage points. Also, the right angles between two straight lines in Figure 3 indicates an additive model. In Figure 3 the row marker of cell 4 is not invisible, because the value is above the limit of the vertical axis. It implies that the 4th row vector is outlying and are leverage points.

5.2. Real data

The data on specific volume of rubber is analyzed by Bradru and Gabriel (1978) and Hawkins *et al.* (2001). The data are from a three factor layout: treatment of rubber (2 levels), temperature (4 quantitative levels), pressure (6 quantitative levels). The data is incomplete in that no observations on one cell. See Bradru and Gabriel (1978) for details.

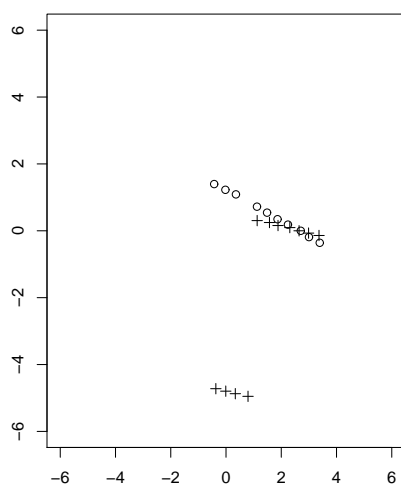


Figure 2: The biplot based on the SVD by alternating least absolute deviation with outliers

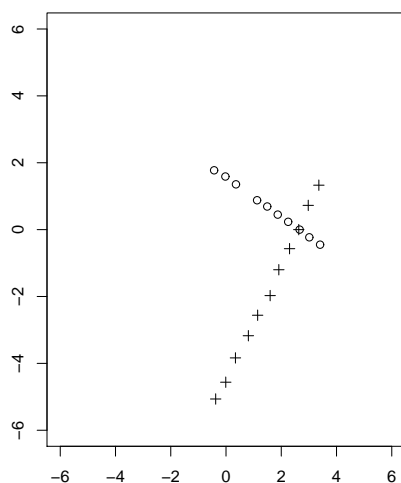


Figure 3: The biplot based on the SVD by alternating weighted least absolute deviation with outliers

A three factor layout can be a two way model by combining two factors, temperature and treatment as rows. We performed the ALS, ALAD and AWLAD estimators, because alternating regressions can treat missing data. Three methods estimated the missing value at 176.75, 175.13, 175.13, respectively. The three methods had a similar biplot that all the column markers (pressure levels) constitute a straight line and all the row markers (temperature and treatment) roughly constitute a straight line that is perpendicular to the line of column markers (Hawkins *et al.*, 2001).

As in the previous subsection, we contaminate the data by adding 500 to a specific level (400 level of pressure and peroxide cured level of treatment). Then Figure 4 shows the biplot based on the ALAD. It does not give an additive structure of data and is not satisfactory. The results by the ALS is similar to that by the ALAD. But the biplot by our proposed algorithm is similar to that by the rubber data with no outliers. It is shown in Figure 5 that gives the information on outlying cells. Figure 5 also implies that the 2nd column is a candidate of outliers. It coincides exactly with the structure of

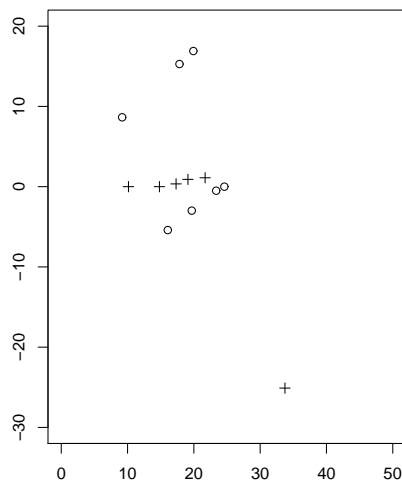


Figure 4: The biplot based on the SVD by alternating least absolute deviation for the rubber data

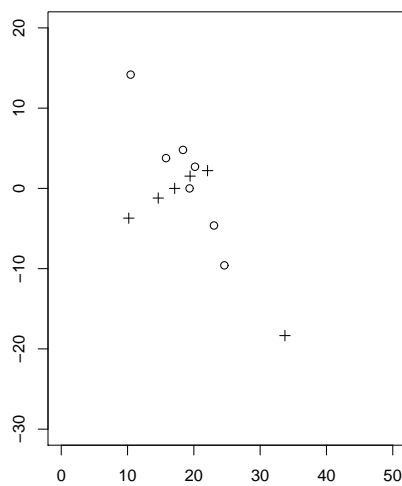


Figure 5: The biplot based on the SVD by alternating weighted least absolute deviation for the rubber data

the contaminated data.

The AWLAD algorithm is more robust than the ALAD algorithm, when in particular the data has leverage points. Furthermore, through numerical examples we can see that the biplot based on the AWLAD algorithm gives the outlying information. The AWLAD is a simple algorithm to implement.

References

- Bradu, D. and Gabriel, K. R. (1978). The biplot as a diagnostic tool for models of two-way tables, *Technometrics*, **20**, 47–68.
- Chen, C., He, X. and Wei, Y. (2008). Lower rank approximation of matrices based on fast and robust alternating regression, *Journal of Computational and Graphical Statistics*, **17**, 186–200.

- Croux, C., Filzmoser, P., Pison, G. and Rousseeuw, P. J. (2003). Fitting multiplicative models by robust alternating regressions, *Statistics and Computing*, **13**, 23–36.
- Gabriel, K. R. and Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights, *Technometrics*, **21**, 489–498.
- Gilnoi, A., Simonoff, J. S. and Sengupta, B. (2006). Robust weighted LAD regression, *Computational Statistics and Data Analysis*, **50**, 3124–3140.
- Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*, Academic Press, London.
- Hawkins, D. M., Liu, L. and Young, S. S. (2001). Robust singular value decomposition, *National Institute of Statistical Sciences Technical Report 122*.
- Liu, L., Hawkins, D. W., Ghosh, S. and Young, S. S. (2003). Robust singular value decomposition analysis of microarray data, *Proceedings of the National Academy of Sciences of the USA*, **100**, 13167–13172.
- Rousseeuw, P. J. (1984). Least median of squares regression, *Journal of the American Statistical Association*, **79**, 871–880.
- Rousseeuw, P. J. and van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points, *Journal of the American Statistical Association*, **85**, 633–639.

Received August 2010; Accepted October 2010