Journal of the Korean
Statistical Society
Vol. 25, No. 1, 1996

# Resistant Singular Value Decomposition and Its Statistical Applications

**Yong-Seok Choi[1] and Myung-Hoe Huh[2]**

**Abstract**

The singular value decomposition is one of the most useful methods in the area of matrix computation. It gives dimension reduction which is the centeral idea in many multivariate analyses. But this method is not resistant, i.e., it is very sensitive to small changes in the input data. In this article, we derive the resistant version of singular value decomposition for principal component analysis. And we give its statistical applications to biplot which is similar to principal component analysis in aspects of the dimension reduction of an $n \times p$ data matrix. Therefore, we derive the resistant principal component analysis and biplot based on the resistant singular value decomposition. They provide graphical multivariate data analyses relatively little influenced by outlying observations.

**Key Words** : Eigensystem; Biplot; Principal component analysis; Resistant version; Singular value decomposition.

[1]Department of Statistics, Research Institute of Information and Communication, Pusan National University, Pusan, 609-735, Korea.
[2]Department of Statistics, Korea University, Seoul, 136-701, Korea.

## 1. INTRODUCTION

The singular value decomposition is one of the most useful methods in the area of matrix computation. The singular value decomposition of an $n \times p$ data matrix $\mathbf{X}$ easily gives dimensional reduction which is a central idea in principal component analysis. Traditionally, the eigensystem of $\mathbf{X}'\mathbf{X}$ is used for dimensional reduction in principal component analysis. As a practical matter, however, there are reasons for preferring the use of the singular value decomposition (Belsley, et al. (1980, p. 99)). In particular, we note that in operating directly on the $n \times p$ matrix $\mathbf{X}$, the singular value decomposition avoids the additional computational burden of forming $\mathbf{X}'\mathbf{X}$ and it gives useful means giving graphical relations of rows and columns of $\mathbf{X}$ in lower dimensional space. In other words, the dimensional reduction of the matrix $\mathbf{X}$ can be easily established by the singular value decomposition.

Some multivariate analyses (principal component biplot, correspondence analysis and principal factor anaysis) are similar to principal component analysis in aspects of the dimensional reduction of an $n \times p$ data matrix $\mathbf{X}$. In particular, principal component biplot is a technique not only for standard $n \times p$ data matrix but also for any two-way array of data. Gabriel (1971) showed that the computational techniques and geometry of principal component biplot were equivalent to those of principal component analysis in spirit. In fact, all techniques which are similar to principal component analysis adopt the singular value decomposition for main algebraic tool. However, we note that the singular value decomposition of the data matrix is not resistant in the sense that it is very sensitive to small changes in the data matrix.

The aims of this article are:

1) to develop the resistant version of singular value decomposition, and

2) to give its applications to principal component analysis and biplot.

The main approach is similar to that of robust regression. The development of robust regression becomes a popular topic for statistical research (Cook and Weisberg (1982, pp. 200-204); Li (1985)).

In Section 2, we derive the algorithm for a resistant version of principal component analysis. This is based on the resistant eigensystem calculation using iterative procedure. In Section 3, we derive the resistant singular value decomposition, which is similar to construction of the singular value decomposition from eigensystem and principal components. In Section 4, we develop

the resistant principal component biplot using the resistant singular value decomposition of Section 3. In Section 5, numerical illustrations are given. In Section 6, concluding remarks are given. Finally, Appendix provides the proofs of optimization problem and goodness of resistant approximation of Section 2.

## 2. RESISTANT PRINCIPAL COMPONENT ANALYSIS

Consider an $n \times p$ data matrix $\mathbf{X} = (x_{ij})$ , $i = 1, ..., n; j = 1, ..., p$. Subtracting out the mean of each variables such as $\bar{x}_{.j} = \sum_{i=1}^{n} x_{ij}/\ n$, we obtain a new $n \times p$ data matrix $\widetilde{\mathbf{X}} = (x_{ij} - \bar{x}_{.j}), i = 1, ..., n; j = 1, ..., p$, which is called variables-centered.

As noted by some authors (Lebart, et al. (1984, Chapter 1), Jolliffe (1986, Chapter 3) and Seber (1984, Chapter 5)), traditionally, principal component analysis is based on the eigensystem of cross product matrix $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$, or is achieved by the singular value decomposition of a data matrix $\widetilde{\mathbf{X}}$ itself. Here, we note that the data matrix $\widetilde{\mathbf{X}}$ is centered by subtracting the mean of each variables from original data.

By the way, it is well known that the sample mean as location estimator is not resistant. So if there exist outliers in data, principal component analysis which depends on both eigensystem of $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$ and singular value decomposition of $\widetilde{\mathbf{X}}$, do not give the desirable results. We note that if $\mathbf{S}$ is a $p$-variate sample variance-covariance matrix corresponding to $\widetilde{\mathbf{X}}$ of rank $r$, then $\mathbf{S} = \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/n$. That is to say, the sample mean which is not resistant influences the sample variance-covariance matrix. And then the sample variance-covariance is, unfortunately, very sensitive to outliers. So it is necessary to obtain the robust sample variance-covariance matrix without the influence of outliers in data. Some authors (Gnanadesikan and Kettenring (1972), Campbell (1980), Huber (1981, pp. 199-242), Matthews (1984), Devlin, et al. (1981) and Rivest and Plante (1988)) give a variety of approaches for obtaining robust covariance and correlation matrices. Specially, Devlin, et al. (1981) compare the performances of several robust procedures for obtaining a correlation matrix. They recommended the use of robust location estimators instead of the sample mean of each variables with very high-dimensional data. More recently, Rivest and Plante (1988) give robust principal component based on the Marona (1976)'s multivariate generation of M-estimators.

Now first, instead of mean of variables in a variable-centered data matrix $\widetilde{\mathbf{X}}$, define a resistant version of mean of the $j^{th}$ variables as $\mathbf{T}_j(\mathbf{X})$, $j = 1, ..., p$. By subtracting out it from a data matrix $\mathbf{X}$, we obtain a new data matrix $\widetilde{\mathbf{X}}^* = (x_{ij}^*) = (x_{ij} - \mathbf{T}_j(\mathbf{X}))$, $i = 1, ..., n$; $j = 1, ...p$. Let $\widetilde{\mathbf{X}}^* = (\tilde{\mathbf{x}}_1^*, ..., \tilde{\mathbf{x}}_n^*)'$ be the row representation where $\tilde{\mathbf{x}}_i^* = (\tilde{x}_{i1}^*, ..., \tilde{x}_{ip}^*)'$, $i = 1, ..., n$, can be viewed as $n$ points in a $p$-dimensional space $\mathcal{R}^p$. Consider the $\mathcal{S} = span(\mathbf{v}_1, ..., \mathbf{v}_s)$ as a subspace of dimension $s (1 \leq s \leq p)$ of $\mathcal{R}^p$ and then this is spanned by orthonormal $p \times s$ vector $\mathbf{V} = (\mathbf{v}_1, ..., \mathbf{v}_s)$.

Letting $\hat{\mathbf{x}}_i^*$ in $\mathcal{S}$ be the nearest point of an arbitrary point $\tilde{\mathbf{x}}_i^*$ in $\mathcal{R}^p$, we have

$$\tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* = (\mathbf{I}_p - \mathbf{V}\mathbf{V}')\tilde{\mathbf{x}}_i^*.$$

So the squared distance $d_i^{*2}$ of $\tilde{\mathbf{x}}_i^*$ from $\hat{\mathbf{x}}_i^*$ in $\mathcal{S}$ is given by

$$d_i^{*2} = \| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|^2 = \tilde{\mathbf{x}}_i^{*'}(\mathbf{I}_p - \mathbf{V}\mathbf{V}')\tilde{\mathbf{x}}_i^*.$$

Let $d_i^* = \| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \| = [\tilde{\mathbf{x}}_i^{*'}(\mathbf{I}_p - \mathbf{V}\mathbf{V}')\tilde{\mathbf{x}}_i^*]^{1/2}$, and consider the minimization of

$$D_s^* = \sum_{i=1}^{n} \rho(d_i^*) = \sum_{i=1}^{n} \rho(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|), \tag{2.1}$$

where

$$\rho(t) = \begin{cases} t^2/2, & \text{for } | t | \leq c, \\ c | t | - c^2/2, & \text{for } | t | > c. \end{cases}$$

Here $\rho(\cdot)$ is Huber's type and also we can use other types of $\rho(\cdot)$, say that of Andrews, as given in Li (1985, Table 8.2, p. 293). We will show that the minimization of (2.1) yields a resistant version of principal component analysis and singular value decomposition of $\widetilde{\mathbf{X}}^*$. From this result, we have the resistant singular value decomposition.

Consider the problem of minimization of (2.1) subject to $\mathbf{v}_j'\mathbf{v}_k = 0 \, (j \neq k)$ and $\|\mathbf{v}_j\| = 1$, $j, k = 1, ..., s$. By the analogous Lagrangian method, we will obtain the $\mathbf{v}_k$, $k = 1, ..., s$, sequentially, for this optimization problem. For the details of procedures solving this problem, see Section 1 of Appendix. Therefore, the optimal $s$-dimensional subspace is found by

$$(\widetilde{\mathbf{X}}^{*'}\mathbf{D}_{\mathbf{w}}\widetilde{\mathbf{X}}^*)\mathbf{v}_k = \lambda_{kk}^*\mathbf{v}_k, \; k = 1, ..., s. \tag{2.2}$$

Here $\mathbf{v}_k, k = 1, ..., s$ are eigenvectors corresponding to the first $s$ largest eigenvalues $\lambda_k^{*2} = \lambda_{kk}^*, k = 1, ..., s$ of $\widetilde{\mathbf{X}}^{*'}\mathbf{D}_{\mathbf{w}}\widetilde{\mathbf{X}}^*$, $\mathbf{D}_{\mathbf{w}} = diag(w_1, .., w_n)$ with $w_i =$

$\psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)/ \| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|$, $i = 1, ..., n$, and $\psi(\cdot)$ is the derivative of $\rho(\cdot)$. We call (2.2) a resistant eigensystem of $\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*$.

We note that in the above procedure, calculation of the resistant eigensystem can be done by using iterative procedure with traditional eigensystem. And the algorithm for a resistant version of principal component analysis proceeds as follows:

**Step 1:** Take as a tentative vector $\mathbf{v}_k$ the $k^{th}$ eigenvector from the eigensystem of $\widetilde{\mathbf{X}}^{*'}\widetilde{\mathbf{X}}^*$ where $\widetilde{\mathbf{X}}^*$ is made by subtracting out the resistant version of mean for the $j^{th}$ variables of original data matrix $\mathbf{X}$.

**Step 2:** Determine the $w_i = \psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)/\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|$, $i = 1, \ldots, n$ and then calculate a resistant estimator of sample mean as

$$\mathbf{T}_j(\mathbf{X}) = \sum_{i=1}^{n} w_i x_{ij} / \sum_{i=1}^{n} w_i, \; j = 1, ..., p,$$

and make a new $\widetilde{\mathbf{X}}^*$ by subtracting out $\mathbf{T}_j(\mathbf{X}), j = 1, ..., p$, from the original $\mathbf{X}$.

**Step 3:** Determine the eigenvalues and eigenvectors from the resistant eigensystem of $\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*$ using the $w_i, i = 1, ..., n$, from Step 2. We call these resistant eigenvalues and eigenvectors.

**Step 4:** Repeat Steps 2 to 3 until, on each successive procedure, the absolute difference between the tentative and updated eigenvectors becomes less than or equal to $\varepsilon$ which is sufficiently small.

Also as discussed in Lebart et al. (1984, Chapter 1), Jolliffe (1986, Chapter 3), and Murtagh and Heck (1987), since the principal components in the subspace of $\mathcal{R}^p$ are generated by $\mathbf{v}_k, k = 1, ..., s$, the coordinates of the points are the components of $\widetilde{\mathbf{X}}^*\mathbf{v}_k$. We note that the $\mathbf{D_w}$ norm of the vector $\widetilde{\mathbf{X}}^*\mathbf{v}_k$ is $\lambda_k^*$ because $\|\widetilde{\mathbf{X}}^*\mathbf{v}_k\|_{\mathbf{D_w}}^2 = \mathbf{v}_k'(\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*)\mathbf{v}_k = \lambda_k^{*2}$.

Also in dual space $\mathcal{R}^n$, we obtain the normalized coordinates vector $\mathbf{u}_k$ such that $\mathbf{u}_i'\mathbf{D_w}\mathbf{u}_k = 0\,(j \neq k)$ and $\|\mathbf{u}_j\|_{\mathbf{D_w}} = 1$, $j, k = 1, ..., s$, corresponding to the eigenvalue $\lambda_k^{*2} \neq 0$ is given by

$$\mathbf{u}_k = \widetilde{\mathbf{X}}^*\mathbf{v}_k / \|\widetilde{\mathbf{X}}^*\mathbf{v}_k\|_{\mathbf{D_w}} = (1/\lambda_k^*)\widetilde{\mathbf{X}}^*\mathbf{v}_k. \tag{2.3}$$

**Remark:** In deriving (2.2), we assumed that the scale parameter $\sigma$ (a measure of spread) was fixed. Without loss of generality, we assumed $\sigma = 1$. In practice, however, $\sigma$ must be estimated. In particular, Li (1985) and Hoaglin, et al. (1976, Chapter 12) discussed this problem in robust regression and in robust estimation of location respectively. In our case, before each iterative step in the algorithm for resistant principal component anlysis, we note that $\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|^2/\sigma^2 (i = 1, ..., n)$ where $\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|$ is the residual vector which is orthognal to the $s$-dimensional subspace, is an approximate $\chi^2$ distribution with $p - s$ degrees of freedom. So we use the median scale estimator

$$\hat{\sigma} = (med_i(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|^2)/\chi^2_{.50(p-s)})^{1/2}$$

where $\chi^2_{.50(p-s)}$ is 50 percentile point of $\chi^2$ distribution with $p - s$ degrees of freedom. Therefore in computing (2.2) actually, we used $\mathbf{D_w} = diag(w_1, ..., w_n)$ where $w_i = \psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|/\hat{\sigma})/(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \| /\hat{\sigma})$. Note that an important reason for the widespread use of the median scale is its excellent resistance: the median scale has to be a reasonably robust estimator of scale (Hoaglin, et al. (1976, pp. 365-414)).

So far, for the resistant principal component analysis it has been seen that the eigenvectors corresponding with the $s (1 \leq s \leq p)$ largest eigenvalues yield the optimal $s$-dimensional subspace of $\mathcal{R}^n$ and its normalize coordinates.

Now we provide a measure of the quality of $s$-dimensional approximation in resistant principal component analysis. We call this a goodness of resistant approximation. As well known in traditional goodness of approximation in principal component analysis, this can be calculated by

$$\rho_s^* = \sum_{k=1}^{s} \lambda_k^{*2} / \sum_{k=1}^{r} \lambda_k^{*2}. \tag{2.4}$$

For the proof of (2.4), see Section 2 of Appendix.

## 3. RESISTANT SINGULAR VALUE DECOMPOSITION

Generally, it is well known that the singular value decomposition easily gives dimensional reduction in principal component analysis. Although singular value decomposition is derived from the eigensystem and principal components, this is an alternative view of principal component analysis in

the useful aspects as discussed in Section 1 (Lebart, et al. (1984, Chapter 1); Jolliffe (1986, Chapter 3); Seber (1984, Chapter 5)). Similary the resistant eigensystem (2.2) and principal components in (2.3) give the resistant singular value decomposition.

**Theorem.** If $\widetilde{\mathbf{X}}^*$ is the $n \times p$ data matrix defined in Section 2, $\widetilde{\mathbf{X}}^*$ can be written as

$$\widetilde{\mathbf{X}}^* = \mathbf{U}\mathbf{D}_{\lambda^*}\mathbf{V}',$$

where $\mathbf{D}_\mathbf{w} = diag(w_1, .., w_n)$ with $w_i = \psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)/\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|, i = 1, ..., n$, $\mathbf{U}$ is an $n \times p$ matrix such that $\mathbf{U}'\mathbf{D}_\mathbf{w}\mathbf{U} = \mathbf{I}_p$, $\mathbf{V}$ is a $p \times p$ matrix of eigenvectors of $\widetilde{\mathbf{X}}^{*'}\mathbf{D}_\mathbf{w}\widetilde{\mathbf{X}}^*$ such that $\mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}_p$, and $\mathbf{D}_{\lambda^*} = diag(\lambda_1^*, ..., \lambda_p^*)$ with $\lambda_k^{*^2}$ is the $k^{th}$ eigenvalue of $\widetilde{\mathbf{X}}^{*'}\mathbf{D}_\mathbf{w}\widetilde{\mathbf{X}}^*$.

**Proof.** From the normalized coordinates vector $\mathbf{u}_k$ in (2.3) corresponding to the eigenvalue $\lambda_k^{*^2} \neq 0$, we obtain $\widetilde{\mathbf{X}}^*\mathbf{v}_k = \lambda_k^*\mathbf{u}_k$. Postmultiplying by $\mathbf{v}_k$ and summing over $k$ gives

$$\widetilde{\mathbf{X}}^* \sum_{k=1}^{p} \mathbf{v}_k\mathbf{v}_k' = \sum_{k=1}^{p} \lambda_k^*\mathbf{u}_k\mathbf{v}_k',$$

Hence, this reduces to

$$\widetilde{\mathbf{X}}^* = \mathbf{U}\mathbf{D}_{\lambda^*}\mathbf{V}',$$

where $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}_{\lambda^*}$ are defined in the theorem. The proof is complete.

We note that the resistant singular value decomposition in Theorem is a form of generalized singular value decomposition. Therefore the resistant principal component analysis can be considered as a special case of the generalized principal component analysis (Greenacre (1984, Appendix A) and Jolliffe (1986, pp. 223-226)).

## 4. RESISTANT PRINCIPAL COMPONENT BIPLOT

Biplot is a graphical dispaly of the rows and columns of an $n \times p$ matrix $\widetilde{\mathbf{X}}$, allowing the visual appraisal of large data matrices. It is also closely related to principal component analysis. This fact is well discussed in some works: Gabriel (1971), Jolliffe (1986, pp. 72-85), and Choi (1991). In particular,

Gabriel (1971) and Choi (1991) provide a principal component biplot having several interesting statistical properties in the interpretation of rows and columns. Its main algebraic tool is the singular value decomposition. But, it is known that this approach is not resistant, as pointed out by Bradu and Gabriel (1978).

Therefore, the purpose of this section is to obtain the resistant version of principal component biplot based on the resistant singular value decomposition in Theorem of Section 3. We call this a resistant principal component biplot by the analogy with resistant principal component analysis.

For deriving a resistant principal component biplot, the algorithm is similar to the principal compomnent biplot well discussed in Gabriel (1971) and Choi (1991).

First, consider the resistant singular value decomposition of $n \times p$ data matrix $\widetilde{\mathbf{X}}^*$ of rank $r$ centered at a robust location estimate. It is natural that the resistant singular value decomposition of $\widetilde{\mathbf{X}}^*$ can be written by

$$\begin{aligned} \widetilde{\mathbf{X}}^* &= \sum_{k=1}^{r} \mathbf{u}_k \lambda_k^* \mathbf{v}_k', \\ &= \sum_{k=1}^{r} (\mathbf{u}_k \lambda_k^{*m})(\lambda_k^{*1-m} \mathbf{v}_k'), \\ &= (\mathbf{U}\mathbf{D}_{\lambda^{*m}})(\mathbf{V}\mathbf{D}_{\lambda^{*1-m}})', \end{aligned} \tag{4.1}$$

where $0 \leq m \leq 1$, $\mathbf{U}, \mathbf{V}$ and $\mathbf{D}_{\lambda^*}$ are well defined in Theorem of Section 3, and $\mathbf{D}_{\lambda^{*m}} = diag(\lambda_1^{*m}, ..., \lambda_r^{*m})$ and $\mathbf{D}_{\lambda^{*1-m}} = diag(\lambda_1^{*1-m}, ..., \lambda_r^{*1-m})$.

Second, consider the $n \times p$ weighted data matrix $\mathbf{D_w}^{1/2}\widetilde{\mathbf{X}}^*$ where $\mathbf{D_w}$ and $\widetilde{\mathbf{X}}^*$ are pointed out in previous section. Naturally, it leads to a weighted variance-covariance matrix

$$\mathbf{S}^* = \widetilde{\mathbf{X}}^{*\prime}\mathbf{D_w}\widetilde{\mathbf{X}}^*/n^*, \tag{4.2}$$

where $n^* = \sum_{i=1}^{n} w_i = \mathbf{1}_n'\mathbf{D_w}\mathbf{1}_n$. Note that this is another form of robust covariance estimators in many works discussed in Section 2. Since the resistant eigensystem of $\widetilde{\mathbf{X}}^{*\prime}\mathbf{D_w}\widetilde{\mathbf{X}}^*$ as defined in Section 2 can be written by

$$\widetilde{\mathbf{X}}^{*\prime}\mathbf{D_w}\widetilde{\mathbf{X}}^* = \mathbf{V}\mathbf{D}_{\lambda^{*2}}\mathbf{V}' = \sum_{k=1}^{r} \lambda_k^{*2} \mathbf{v}_k \mathbf{v}_k', \tag{4.3}$$

where $\mathbf{V} = (\mathbf{v}_1, ..., \mathbf{v}_r)$ is the $p \times r$ orthonormal matrix such that $\mathbf{v}_j'\mathbf{v}_k = 0 (j \neq k)$ and $\|\mathbf{v}_j\| = 1$, $j, k = 1, ..., r$, (4.2) becomes

$$n^*\mathbf{S}^*\mathbf{v}_k = \lambda_k^{*2}\mathbf{v}_k, \; k = 1, ..., r.$$

So we have the form of the resistant eigensystem for resistant principal component analysis as discussed in Section 2. Now if we set $m = 0$ in (4.1), we have a resistant factorization of $\widetilde{\mathbf{X}}^*$ as

$$\widetilde{\mathbf{X}}^* = \sum_{k=1}^{r} \mathbf{u}_k(\lambda_k^* \mathbf{v}_k') = \mathbf{A}^* \mathbf{B}^{*'}, \tag{4.4}$$

where $\mathbf{A}^* = (\mathbf{u}_1, ..., \mathbf{u}_r)$ and $\mathbf{B}^* = (\mathbf{v}_1\lambda_1^*, ..., \mathbf{v}_r\lambda_r^*)$. In strict analogy with Gabriel (1971), the following lemma gives the properties of an optimal $s$-dimensional resistant principal component biplot.

**Lemma.** Consider the resistant rank $s$ approximation $\widetilde{\mathbf{X}}_{(s)}^*$ to $\widetilde{\mathbf{X}}^*$ of (4.4). For the purpose of biplotting in the $s$-dimensional resistant principal component biplot, choose the resistant factors such as

$$\mathbf{A}_{(s)}^* = n^{*1/2}(\mathbf{u}_1, ..., \mathbf{u}_s) \text{ and } \mathbf{B}_{(s)}^* = n^{*-1/2}(\mathbf{v}_1\lambda_1^*, ..., \mathbf{v}_s\lambda_s^*). \tag{4.5}$$

Then $\widetilde{\mathbf{X}}_{(s)}^*$ can be factorized as

$$\widetilde{\mathbf{X}}_{(s)}^* = \mathbf{A}_{(s)}^* \mathbf{B}_{(s)}^{*'}.$$

And we have the relationships such as

$$\begin{aligned} \widetilde{\mathbf{X}}^* &\simeq \mathbf{A}_{(s)}^* \mathbf{B}_{(s)}^{*'}, \\ \widetilde{\mathbf{X}}^* \mathbf{S}^{*-1} \widetilde{\mathbf{X}}^{*'} &\simeq \mathbf{A}_{(s)}^* \mathbf{A}_{(s)}^{*'}, \\ \mathbf{S}^* &\simeq \mathbf{B}_{(s)}^* \mathbf{B}_{(s)}^{*'}, \end{aligned}$$

where $\simeq$ denotes "is approximated by means of a least squares fit of rank $s$".

From a viewpoint of the rank $s$ resistant approximation as defined in Section 2 of Appendix, it is natural that the quality of the rank $s$ resistant approximation $\widetilde{\mathbf{X}}_{(s)}^*$ is evaluated with the criterion

$$\rho_s^* = \sum_{k=1}^{s} \lambda_k^{*2} / \sum_{k=1}^{r} \lambda_k^{*2}.$$

## 5. NUMERICAL ILLUSTRATIONS

**Example 1.** To illustrate our resistant version of principal component analysis, we use the census-tract data (Johnson and Wichern, 1992, Table 8.2, p. 392). The data provides fourteen tract informations on five socioeconomic variables for the Madison, Wisconsin area.

The optimal 2-dimensional display of principal component scores is given in Fig. 1 with the goodness of approximation 0.9323, i.e., 93.23% of the total variation is explained by the first two principal axes. In Fig. 1, the capital letters and numbers are the row (forteen tracts) and the column (five socioeconomic variables) markers respectively. We note that the row markers B, C, G, J, K, L and M have similar patterns of behavior characterized by low values on four variables 1 (total populaton), 3 (total emloyment) and 4 (health services employment), whereas row marker F has a higher score for these variables. In fact, since the variables 1, 3 and 4 have the same characteristic, natually, their angles in Fig. 1 must be small. Moreover, since the variables 2 (median school years) and 5 (median value home) have the same characteristic, their angle must be small. But these interpretations in Fig. 1 are not clear.

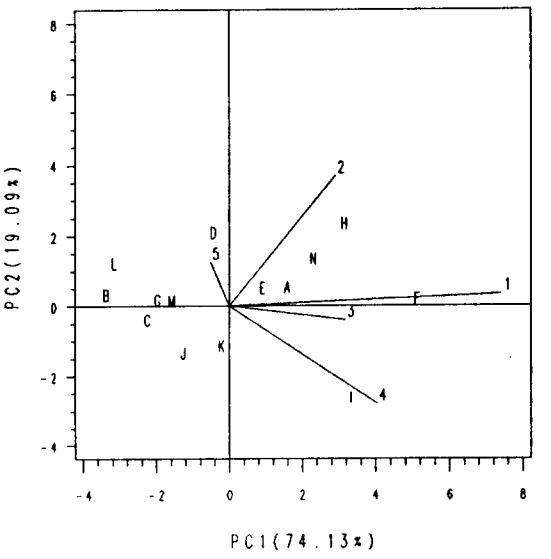


**Fig. 1.** Optimal 2-dimensional display by principal component analysis of the census-tract data

Now consider resistant principal component analysis with Andrews' $\psi(\cdot)$ function given by

$$\psi(t) = \begin{cases} c\sin(t/c), & \text{for } 0 \le t < c\pi, \\ 0, & \text{for } t \ge c\pi, \end{cases}$$

with $c = 0.89$. As in Remark of Section 2, we use 0.29 for the median scale estimate. Actually, $(c\pi)^2$ is 95 percentile point of $\chi^2$ distribution with 3 degrees of freedom.

The final weights used in computing resistant eigensystem (2.2) are in the diagonal matrix

$$\mathbf{D_w} = \text{diag}(0.00,\ 0.00,\ 0.90,\ 0.43,\ 0.76,\ 0.81,\ 1.00,\ 0.00,\ 0.91,\ 0.88,\ 0.77,\ 0.00,\ 0.00,\ 0.00).$$

We note that in $\mathbf{D_w}$, the elements having 0.00 (zero) are notable observations The optimal 2-dimensional display of resistant principal component analysis is shown in Fig. 2 with the goodness of resistant approximation 99.08%. By reducing the influence of the notable tracts, Fig. 2 gives somewhat lucid interpretations of principal component analysis.

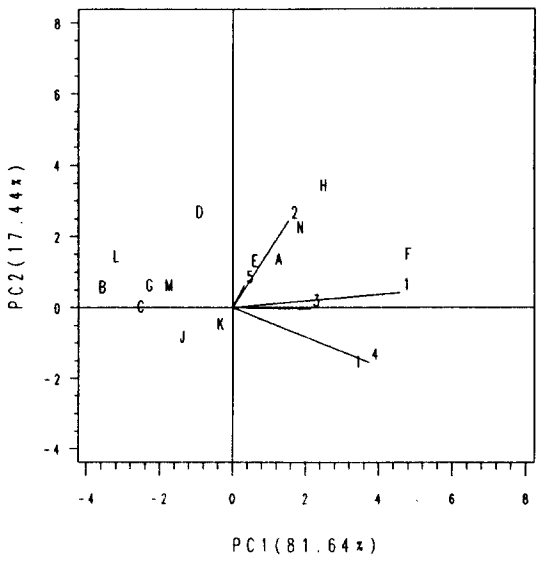


**Fig. 2.** Optimal 2-dimensional display by resistant principal component analysis of the census-tract data

Another example using the protein consumption data (Greenacre (1984, Table 9.9, p. 278)) is illustrated and discussed by Choi (1991).

**Example 2.** The data (Gabriel and Zamir (1979)) is the numbers of science Ph. D. degrees awarded in the United States during the years 1960-1961 and 1970-1975.

The first two largest eigenvalues and their proportions of total variation are 52.92 (96.46%) and 1.45 (2.61%). Thus the principal component biplot is shown in Fig. 3 with the goodness of approximation 99.10%. In Fig. 3, note that the row (twelve sciences) and column (years 1960-1961 and 1970-1975) markers are denoted by the capital letters and numbers respectively.

We note that D (Chemistry), E (Earth Sciences) and G (Agricultural Sciences) have smaller increases in column markers 1 (1960) and 2 (1961), but as to the 3-8(1970-1975) changes, these appear less pronounced. A (Engineering) and F (Biological Sciences) are increasing in 1970-1975. So we know that the column markers 1-2 (the 1960's) and 3-8 (the 1970's) give different patterns for the row markers respectively. But these interpretations are not clear for the columns markers 3-8 (the 1970's).

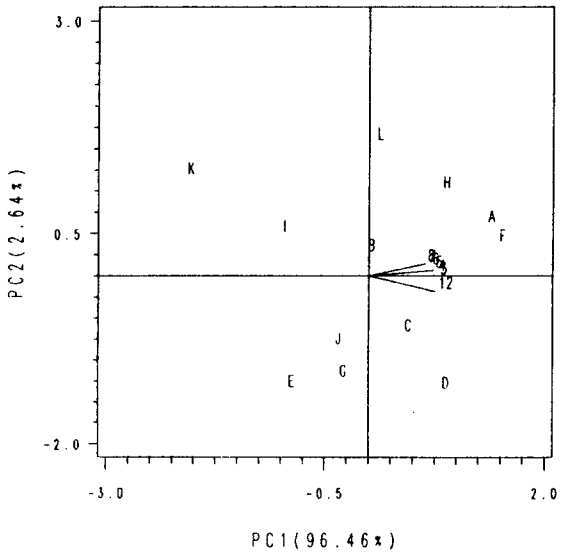


**Fig. 3.** Principal component biplot of the science doctorate award data

Now consider resistant principal component biplot. In order to obtain the resistant factors (4.5), we use the Andrews' $\psi(\cdot)$ function as defined in Example 1 where $(c\pi)^2$ is 95 percentile point of $\chi^2$ distribution with 6 degrees of freedom and hence $c = 1.13$. Of course, we use 0.05 for the median scale estimate. Then we have the $12 \times 12$ diagonal matrix

$$\mathbf{D_w} = diag(0.00,\ 0.00,\ 0.00,\ 0.41,\ 0.88,\ 1.00,\ 0.72,\ 0.00,\ 0.50,\ 0.51,\ 0.00,\ 0.98).$$

In this case, we have the first two largest eigenvalues and their proportions of total variation are 16.45 (95.91%) and 0.67 (3.93%).

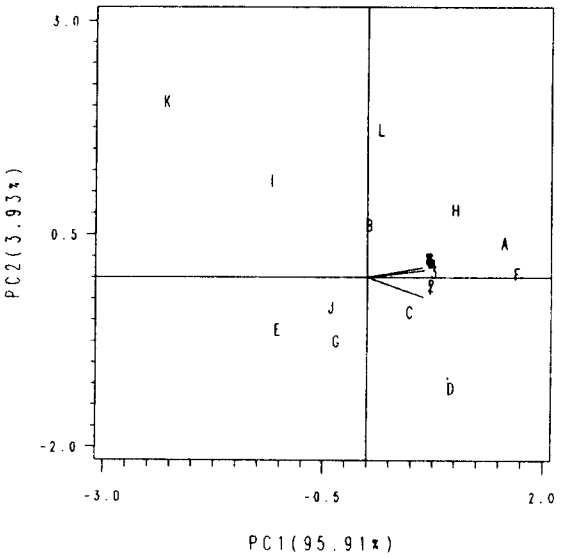


**Fig. 4.** Resistant principal component biplot of the science doctorate award data

The resistant principal component biplot with the resistant goodness of approximation 99.84% is given in Fig. 4. It gives the optimal 2-dimensioanal display of rows and columns of the data with greater dispersion by reducing the influence of notable rows A (Engineering), B (Mathematics), C (Physics and Astrometry), H (Psychology) and K (Anthropology) with weights 0.00.

For another example using the ability test data from du Toit, et al. (1986, Table 1.1, p. 6), see Choi (1991).

## 6. CONCLUDING REMARKS

In this paper, we limit ourselves to reducing the influence of outliers in the singular value decomposition and obtaining the resistant version of it. And we give its applications to principal component analysis and biplot. We note that in the iterative procedure deriving the resistant version of it, the iteration is continued until a reasonable degree of convergence is reached. But the degree of convergence is not known exactly. And in choosing resistant estimators of location and scale, we need to study problems of resistance as breakdown point and influence functions of them.

## APPENDIX : Proofs

### 1. Proof of the optimization problem in Section 2.

Consider the problem in minimization of $D_s^* = \sum_i \rho(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)$ subject to $\mathbf{v}_j'\mathbf{v}_k = 0\,(j \neq k)$ and $\|\mathbf{v}_j\| = 1$, $j, k = 1, ..., s$. By the traditional Lagrangian method, we will obtain the $\mathbf{v}_k$, $k = 1, ..., s$, sequentially, for this optimization problem.

First, define the Lagragian expression in order to find a best vector $\mathbf{v}$ in $\mathcal{S}$ for one-dimensional subspace:

$$\mathcal{L} = \sum_i \rho(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|) + (\lambda^*/2)(\mathbf{v}'\mathbf{v} - 1)$$

where $\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \| = [\tilde{\mathbf{x}}_i^{*'}\tilde{\mathbf{x}}_i^* - (\mathbf{v}'\tilde{\mathbf{x}}_i^*)^2]^{1/2}$ and $\lambda^*/2$ is a Lagrange multiplier. Differentiating the objective function $\rho(\cdot)$ and constraint terms in Lagrangian expressions with respect to $\mathbf{v}$, we obtain

$$-\sum_i [\psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)/\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|]\tilde{\mathbf{x}}_i^*(\tilde{\mathbf{x}}_i^{*'}\mathbf{v}) + \lambda^*\mathbf{v} = 0,$$

where $\psi(\cdot)$ is the derivative of $\rho(\cdot)$.

Setting $w_i = \psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)/\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|$, we have

$$\sum_i (w_i \tilde{\mathbf{x}}_i^* \tilde{\mathbf{x}}_i^{*'})\mathbf{v} = \lambda^*\mathbf{v},$$

which leads to

$$(\widetilde{\mathbf{X}}^{*'}\mathbf{D}_\mathbf{w}\widetilde{\mathbf{X}}^*)\mathbf{v} = \lambda^*\mathbf{v}.$$

Therefore, $\mathbf{v}$ is the eigenvector corresponding to the eigenvalue $\lambda^*$ of $\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*$ where $\mathbf{D_w} = diag(w_1, ..., w_n)$. Hence the optimal value $\mathbf{v}_1$ is the eigenvector associated with the largest eigenvalue $\lambda_1^{*^2} = \lambda^*$ of the matrix $\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*$.

Next, let us find the second vector $\mathbf{v}_2$ in $\mathcal{S}$ for two-dimensional subspace, that is the best approximation for the set of points obviously contains the subspace defined by $\mathbf{v}_1$. This problem is to find $\mathbf{v}_2$ minimizing $\sum_i \rho(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)$, subject to $\mathbf{v}_1'\mathbf{v}_2 = 0$ and $\mathbf{v}_2'\mathbf{v}_2 = 1$ where $\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \| = [\tilde{\mathbf{x}}_i^{*'}\tilde{\mathbf{x}}_i^* - (\mathbf{v}_1'\tilde{\mathbf{x}}_i^*)^2 - (\mathbf{v}_2'\tilde{\mathbf{x}}_i^2)]^{1/2}$. So we have Lagrangian equation:

$$\sum_i \rho(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|) + (\lambda_{22}^*/2)(\mathbf{v}_2'\mathbf{v}_2 - 1) + \mu_2(\mathbf{v}_2'\mathbf{v}_1)$$

where $\lambda_{22}^*/2$ and $\mu_2$ are Lagrange multipliers. Differentiating it with respect to $\mathbf{v}_2$, we have

$$-\sum_i [\psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)/\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|]\tilde{\mathbf{x}}_i^*(\tilde{\mathbf{x}}_i^{*'}\mathbf{v}_2) + \lambda_{22}^*\mathbf{v}_2 + \mu_2\mathbf{v}_1 = 0.$$

Let $w_i = \psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)/\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|$ and then premultiply both sides by - $\mathbf{v}_1'$, we have

$$\mathbf{v}_1'(\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*)\mathbf{v}_2 = \lambda_{22}^*\mathbf{v}_1'\mathbf{v}_2 + \mu_2\mathbf{v}_1'\mathbf{v}_1.$$

Since $\mathbf{v}_1'(\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*)\mathbf{v}_2 = \lambda_1^{*^2}\mathbf{v}_1'\mathbf{v}_2 = 0$, $\mu_2$ must equal to 0. So we have

$$(\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*)\mathbf{v}_2 = \lambda_{22}^*\mathbf{v}_2.$$

By the similar fashion, the optimal $s$-dimensional subspace is obtained by solving

$$(\widetilde{\mathbf{X}}^{*'}\mathbf{D_w}\widetilde{\mathbf{X}}^*)\mathbf{v}_k = \lambda_{kk}^*\mathbf{v}_k,\ k = 1, ..., s.$$

The proof is complete.

## 2. Proof of the goodness of resistant approximation in Section 2.

The following discussion gives a lower rank resistant approximation. Moreover, this gives the goodness of resistant approximation based on the rank $s$ resistant approximation (Greenacre (1984, Appendix A) and Jolliffe (1986, pp. 223-226)). This is similar to the generalized lower rank least squares approximation.

Let $\widetilde{\mathbf{X}}^*$ be the matrix of rank $r$ and $\mathbf{D_w}$ be the diagonal matrix with diagonal element $w_i, i = 1, ..., n$, defined in (2.2) of Section 2. Let $\mathbf{M}^*$ be an $n \times p$ matrix of rank $s (s < r)$. The resistant matrix approximation

$$\mathbf{M}^* = \widetilde{\mathbf{X}}^*_{(s)} = \sum_{k=1}^{s} \lambda_k^* \mathbf{u}_k \mathbf{v}_k'$$

minimizes

$$\| \widetilde{\mathbf{X}}^* - \mathbf{M}^* \|^2_{\mathbf{D_w}} = \sum_{i=1}^{n} w_i (\tilde{\mathbf{x}}_i^* - \mathbf{m}_i^*)'(\tilde{\mathbf{x}}_i^* - \mathbf{m}_i^*)$$

among all matrices $\mathbf{M}^*$ of rank less than or equal to $s$, where $\tilde{\mathbf{x}}_i^{*'}$ and $\mathbf{m}_i^{*'}$ are the rows of $\widetilde{\mathbf{X}}^*$ and $\mathbf{M}^*$ respectively. The minimum value is $\sum_{k=s+1}^{r} \lambda_k^{*2}$.

Since $w_i = \psi(\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|)/\| \tilde{\mathbf{x}}_i^* - \hat{\mathbf{x}}_i^* \|$, which is the $i^{th}$ diagonal element of $\mathbf{D_w}$, gives the resistant procedures, we may call $\mathbf{M}^* = \widetilde{\mathbf{X}}^*_{(s)}$ the rank $s$ resistant approximation of $\widetilde{\mathbf{X}}^*$. And as well known in traditional goodness of approximation, the goodness of resistant approximation can be calculated by

$$\begin{aligned} \rho_s^* &= 1 - \| \widetilde{\mathbf{X}}^* - \widetilde{\mathbf{X}}^*_{(s)} \|^2_{\mathbf{D_w}} / \| \widetilde{\mathbf{X}}^* \|^2_{\mathbf{D_w}}, \\ &= 1 - \sum_{k=s+1}^{r} \lambda_k^{*2} / \sum_{k=1}^{r} \lambda_k^{*2} = \sum_{k=1}^{s} \lambda_k^{*2} / \sum_{k=1}^{r} \lambda_k^{*2}. \end{aligned}$$

## REFERENCES

( 1) Belsley, D., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* Wiley, New York.

( 2) Bradu, D. and Gabriel, K. R. (1978). The biplot as a diagnostic tool for models of two-way tables. *Technometrics*, **20**, 47-68.

( 3) Campbell, N. A. (1980). Robust procedures in multivariate analysis I: robust covariance estimation. *Applied Statistics*, **29**, 231-237.

( 4) Choi, Y. S. (1991). Resistant Principal Component Analysis, Biplot And Correspondence Analysis. *Unpublished Ph.D. dissertation, University of Korea.*

( 5) Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman and Hall, New York.

( 6) Devlin, S. J., Gnanadesikan, R. and Kettenring, J. R. (1981). Robust estimation and outlier detection with correlation coefficients. *Biometrika*, **62**, 531-545.

( 7) du Toit, S. H. C., Steyn, A. G. W. and Stumpf, R. H. (1986). *Graphical Exploratory Data Analysis.* Springer-Verlag, New York.

( 8) Gabriel, K. R. (1971). The biplot graphics display of matrices with applications to principal component analysis. *Biometrika*, **58**, 453-467.

( 9) Gabriel, K. R. and Zamirs. (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, **21**, 489-498.

(10) Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, **28**, 81-124.

(11) Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis.* Academic Press, New York.

(12) Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (1976). *Understanding Robust and Exploratory Data Analysis.* Wiley, New York.

(13) Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.

(14) Jolliffe, I. T. (1986). *Principal Component Analysis.* Springer-Verlag, New York.

(15) Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis.* Prentice-Hall, New Jersey.

(16) Lebart, L., Morineau, A. and Warwick, K. (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices.* Wiley, New York.

(17) Li, G. (1985). Robust Regression. *Exploring Data Tables, Trends and Shapes*, edited by Hoaglin, D. C., Mosteller, F., and Tukey, J. W., 281-343, Wiley, New York.

(18) Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, **4**, 51-67.

(19) Matthews, J. N. S. (1984). Robust methods in the assessment of multivariate normality. *Applied Statistics*, **33**, 272-277.

(20) Murtagh, F. and Heck, A. (1987). *Multivariate Data Analysis.* D. Reidel Publishing Company, Dordrecht, Holland.

(21) Rivest, L. P. and Plante, N. (1988). N'analyse en composantes principales robuste. *Rev Statistique Appliquee*, **36**, 55-66.

(22) Seber, G. A. (1984). *Multivariate Observations.* Wiley, New York.