
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Sign eigenanalysis and its applications to optimization problems and robust statistics

PLEASE CITE THE PUBLISHED VERSION

<http://www.sciencedirect.com/science/article/pii/S0167947304002324>

PUBLISHER

© Elsevier

VERSION

AM (Accepted Manuscript)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Li, Baibing. 2019. "Sign Eigenanalysis and Its Applications to Optimization Problems and Robust Statistics".
figshare. <https://hdl.handle.net/2134/9223>.

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Sign eigenanalysis and its applications to optimization problems and robust statistics

Baibing Li *

School of Mathematics and Statistics
University of Newcastle
Newcastle upon Tyne, NE1 7RU, UK

Abstract

Sign eigenvectors for a real square matrix, \mathbf{A} , are defined to be sign vectors for which all of its elements either retain the same signs or become to their opposite signs after the linear transformation \mathbf{A} , where a sign vector is a vector with the elements equal to either 1 or -1 . Existence of sign eigenvectors for symmetric positive semi-definite matrices is investigated. It is shown that the sign eigenanalysis is closely related to some certain optimization problems and can be applied to develop robust statistical inference procedures in the L_1 norm. A numerical example is given to illustrate the applications to robust multivariate statistical analysis.

Keywords: Eigenproblem; L_1 norm; Optimization; Robust statistical inference

*Tel. +44 191 2227235; Fax: +44 191 2228020. E-mail address: baibing.li@ncl.ac.uk

1. Introduction

For an $n \times n$ matrix, \mathbf{A} , an eigenvector associated with an eigenvalue, λ , is defined to be an n -vector, \mathbf{y} , satisfying

$$\mathbf{A}\mathbf{y} = \lambda\mathbf{y}.$$

Eigenanalysis is closely related to the optimization of quadratic forms in the L_2 norm. For instance, suppose that a matrix \mathbf{A} is symmetric positive semi-definite ($\mathbf{A} \geq 0$), then it is well known that an eigenvector \mathbf{y} associated with the largest eigenvalue is a solution to the following optimization problem (Rao, 1973):

$$\max_{\mathbf{y}} \mathbf{y}^T \mathbf{A} \mathbf{y}, \quad \text{subject to } \|\mathbf{y}\|_2 = 1,$$

where $\|\mathbf{y}\|_2 = (\sum_{i=1}^n y_i^2)^{1/2}$ is the L_2 norm of a vector $\mathbf{y} = [y_1, \dots, y_n]^T$.

The above optimization problem is widely used in multivariate statistical inference such as principal component analysis; see for instance, Rao (1973).

In this paper, we define a sign eigenvector for a real square matrix, \mathbf{A} , to be a sign vector for which all of its elements either retain the same signs or become to their opposite signs after the linear transformation, \mathbf{A} . We investigate existence of sign eigenvectors for symmetric positive semi-definite matrices and relate the sign eigenanalysis to some certain optimization problems which are useful to develop robust statistical inference procedures in the L_1 norm. For instance, similar to the above results in the L_2 norm, we will show that for a matrix $\mathbf{A} \geq 0$, a sign eigenvector associated with the largest sign eigenvalue is a solution to the following optimization problem:

$$\max_{\mathbf{z}} \mathbf{z}^T \mathbf{A} \mathbf{z}, \quad \text{where } \mathbf{z} \text{ is a sign vector with the elements equal to either } 1 \text{ or } -1.$$

2. Definition and main results

Consider an $n \times n$ real matrix, \mathbf{A} . Define a space of sign-vectors to be

$$\mathbf{Z} = \left\{ \mathbf{z} \mid \mathbf{z} = [z_1, \dots, z_n]^T, \ z_i \text{ is } +1 \text{ or } -1, i = 1, \dots, n \right\},$$

and a sign function, $S(x)$, to be

$$S(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}.$$

When \mathbf{x} is a vector, $S(\mathbf{x})$ is a vector the same size as \mathbf{x} containing the signs of the elements of \mathbf{x} .

Definition. For an $n \times n$ real matrix, \mathbf{A} , define a sign eigenvector of \mathbf{A} associated with a sign eigenvalue λ to be a sign vector, $\mathbf{z} \in \mathbf{Z}$, satisfying $S(\mathbf{A}\mathbf{z}) = S(\lambda)\mathbf{z}$, where the corresponding sign eigenvalue is defined to be $\lambda = \mathbf{z}^T \mathbf{A} \mathbf{z} / (\mathbf{z}^T \mathbf{z}) = \mathbf{z}^T \mathbf{A} \mathbf{z} / n$.

According to this definition, a sign eigenvector of the matrix \mathbf{A} is a sign vector for which all of its elements either retain the same signs or become to their opposite signs after the linear transformation, \mathbf{A} . In addition, for a matrix $\mathbf{A} \geq 0$, a sign eigenvector may be simply defined to be a sign vector $\mathbf{z} \in \mathbf{Z}$ satisfying $\mathbf{z} = S(\mathbf{A}\mathbf{z})$ since its associated sign eigenvalue λ is non-negative and thus $S(\lambda) = 1$.

Example. For a 2×2 matrix $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$, the sign vector $\mathbf{z} = [1, 1]^T$ is a sign eigenvector of \mathbf{A} associated with the sign eigenvalue of 3.5 since $\mathbf{Az} = [4, 3]^T$ and $\mathbf{z} = S(\mathbf{Az})$.

The following theorem shows that, like their counterparts of ordinary eigenvectors in the L_2 norm, there is a relationship between sign eigenvectors and extrema of quadratic forms.

Theorem 1. For any $n \times n$ matrix $\mathbf{A} \geq 0$, there exists an optimal solution, \mathbf{z}^* , to following optimization problem

$$\max_{\mathbf{z} \in \mathbf{Z}} \mathbf{z}^T \mathbf{A} \mathbf{z}, \quad (1)$$

which is a sign eigenvector of the matrix \mathbf{A} corresponding to the largest sign eigenvalue.

The theorem given below guarantees existence of sign eigenvectors for symmetric positive semi-definite matrices.

Theorem 2. For any $n \times n$ matrix $\mathbf{A} \geq 0$, there exists at least one sign eigenvector of \mathbf{A} , $\mathbf{z} \in \mathbf{Z}$, satisfying $\mathbf{z} = S(\mathbf{Az})$.

Proofs of theorems 1 and 2 will be given in next section.

The following counter-example shows that if a matrix is not symmetric positive semi-definite, then it may not have a sign eigenvector. There is no necessary and sufficient condition for the existence of a sign eigenvector for an $n \times n$ real matrix.

Example. For $n=2$, the set \mathbf{Z} consists of four elements, $\pm[1, 1]^T$ and $\pm[1, -1]^T$. It is easy to verify that none of them is a sign eigenvector of matrix $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & -2 \end{bmatrix}$, and thus the matrix \mathbf{A} does not have a sign eigenvector.

3. Applications to optimization problems and robust statistics

In this section, we consider some applications of sign eigenvectors to optimization problems and robust statistical inference in the L_1 norm which are summarized in theorems 3 and 4. Proofs of theorems 1 and 2 are then given. At the end of this section we discuss the issue of algorithms.

For an $n \times n$ symmetric positive semi-definite matrix, \mathbf{A} , denote a square root of \mathbf{A} as \mathbf{B} satisfying $\mathbf{A} = \mathbf{B}\mathbf{B}^T$, where $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]^T$ is an $n \times m$ matrix and \mathbf{b}_i ($i=1, \dots, n$) are m -vectors.

Consider the following optimization problem:

$$\max_{\mathbf{x}, \mathbf{z}} \mathbf{z}^T \mathbf{B} \mathbf{x}, \quad \text{subject to } \|\mathbf{x}\|_2 = 1 \quad \text{and } \mathbf{z} \in \mathbf{Z}, \quad (2)$$

where \mathbf{x} is an m -vector.

We first consider the case where the matrix \mathbf{B} has some zero row-vectors, i.e. $\mathbf{b}_j^T = \mathbf{0}$ for some j . From the optimization point of view, those zero row-vectors are not of interest since they neither contribute to the objective function in problem (2) nor to the objective functions in problems (4) and (5) discussed later. Therefore, they may simply be removed from the analysis. Those elements of the optimal solution \mathbf{z}^* to the problem (2) or (5), which correspond to the zero row-vectors, may be either 1 or -1 since they

do not change values of the associated objective function. We thus eliminate the case where the matrix \mathbf{B} has zero row-vectors in the sequel of this section.

Lemma 1. If $\mathbf{b}_i \neq 0$ ($i=1, \dots, n$), then any optimal solution to the problem (2), $(\mathbf{x}^*, \mathbf{z}^*)$, satisfies $\mathbf{b}_i^T \mathbf{x}^* \neq 0$ for $i=1, \dots, n$.

Lemma 2. If $\mathbf{b}_i \neq 0$ ($i=1, \dots, n$), then any optimal solution to the problem (2), $(\mathbf{x}^*, \mathbf{z}^*)$, satisfies

$$\mathbf{x}^* = \mathbf{B}^T \mathbf{z}^* / \|\mathbf{B}^T \mathbf{z}^*\|_2, \quad (3a)$$

$$\text{and } \mathbf{z}^* = S(\mathbf{B}\mathbf{x}^*). \quad (3b)$$

The proofs for both lemmas 1 and 2 are given in the Appendix.

Note that lemma 2 above gives a necessary condition of an optimal solution to (2). Substituting (3a) into (3b) we have

$$\mathbf{z}^* = S(\mathbf{B}\mathbf{B}^T \mathbf{z}^*),$$

hence, a necessary condition of an optimal solution $(\mathbf{x}^*, \mathbf{z}^*)$ to the problem (2) is that \mathbf{z}^* is a sign eigenvector of $\mathbf{A} = \mathbf{B}\mathbf{B}^T$.

The following theorem relates sign eigenvectors to an optimization problem in the L_1 norm.

Theorem 3. Suppose that $\mathbf{b}_i \neq 0$ for $i=1, \dots, n$. Then the vector \mathbf{x}^* is an optimal solution to the following problem:

$$\max_{\mathbf{x}} \sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}|, \quad \text{subject to } \|\mathbf{x}\|_2 = 1, \quad (4)$$

if and only if $(\mathbf{x}^*, \mathbf{z}^*)$ is an optimal solution to the problem (2), where $\mathbf{z}^* = S(\mathbf{B}\mathbf{x}^*)$.

Proof. Suppose that $(\mathbf{x}^*, \mathbf{z}^*)$ is an optimal solution to the problem (2). Then we have

$$\sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}^*| \leq \max_{\|\mathbf{x}\|_2=1} \sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}| = \max_{\|\mathbf{x}\|_2=1} \sum_{i=1}^n S(\mathbf{b}_i^T \mathbf{x}) \mathbf{b}_i^T \mathbf{x} \leq \max_{\substack{\|\mathbf{x}\|_2=1 \\ \mathbf{z} \in \mathcal{Z}}} \mathbf{z}^T \mathbf{B} \mathbf{x}.$$

From lemma 2, we obtain $\max_{\substack{\|\mathbf{x}\|_2=1 \\ \mathbf{z} \in \mathcal{Z}}} \mathbf{z}^T \mathbf{B} \mathbf{x} = \sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}^*|$. Hence, \mathbf{x}^* is an optimal solution to

(4).

Next, suppose that \mathbf{x}^* is an optimal solution to (4) but $(\mathbf{x}^*, \mathbf{z}^*)$ is not an optimal solution to the problem (2). Denote $(\mathbf{x}_0, \mathbf{z}_0)$ as an optimal solution to (2), thus $\mathbf{z}_0^T \mathbf{B} \mathbf{x}_0 > \mathbf{z}^{*T} \mathbf{B} \mathbf{x}^*$. From lemma 2 we have $\mathbf{z}_0^T \mathbf{B} \mathbf{x}_0 = [S(\mathbf{B}\mathbf{x}_0)]^T \mathbf{B} \mathbf{x}_0 = \sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}_0|$. In

addition, since $\mathbf{z}^* = S(\mathbf{B}\mathbf{x}^*)$, we have $\mathbf{z}^{*T} \mathbf{B} \mathbf{x}^* = \sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}^*|$. This leads to

$$\sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}_0| > \sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}^*| \text{ and contradicts the assumption that } \mathbf{x}^* \text{ is an optimal solution to}$$

(4). \square

The theorem 4 given below relates the optimization problem (1) to problem (2).

Theorem 4. Let $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]^T$ be an $n \times m$ matrix, where $\mathbf{b}_i \neq 0$ ($i=1, \dots, n$) are m -vectors. Then the vector \mathbf{z}^* is an optimal solution to the following problem

$$\max_{\mathbf{z} \in \mathbf{Z}} \mathbf{z}^T \mathbf{B} \mathbf{B}^T \mathbf{z}, \quad (5)$$

if and only if $(\mathbf{x}^*, \mathbf{z}^*)$ is an optimal solution to the problem (2), where $\mathbf{x}^* = \mathbf{B}^T \mathbf{z}^* / \|\mathbf{B}^T \mathbf{z}^*\|_2$.

Proof. Suppose $(\mathbf{x}^*, \mathbf{z}^*)$ is an optimal solution to the problem (2) but \mathbf{z}^* is not an optimal solution to the problem (5). Denote \mathbf{z}_0 as an optimal solution to (5), and let $\mathbf{x}_0 = \mathbf{B}^T \mathbf{z}_0 / \|\mathbf{B}^T \mathbf{z}_0\|_2$. Then noting that $(\mathbf{x}^*, \mathbf{z}^*)$ is an optimal solution to the problem (2) we have

$$\mathbf{z}_0^T \mathbf{B} \mathbf{x}_0 < \mathbf{z}^{*T} \mathbf{B} \mathbf{x}^*. \quad (6)$$

Inserting $\mathbf{x}^* = \mathbf{B}^T \mathbf{z}^* / \|\mathbf{B}^T \mathbf{z}^*\|_2$ and $\mathbf{x}_0 = \mathbf{B}^T \mathbf{z}_0 / \|\mathbf{B}^T \mathbf{z}_0\|_2$ into (6) results in $\mathbf{z}_0^T \mathbf{B} \mathbf{B}^T \mathbf{z}_0 < \mathbf{z}^{*T} \mathbf{B} \mathbf{B}^T \mathbf{z}^*$ which contradicts the assumption that \mathbf{z}_0 is an optimal solution to (5). This completes the proof of sufficiency. The necessity can be proved similarly. \square

Next, we consider proofs of theorems 1 and 2. If a square root of the matrix \mathbf{A} , $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]^T$, satisfies $\mathbf{b}_i \neq \mathbf{0}$ for all $i=1, \dots, n$, then theorem 1 is immediate from theorem 4 and lemma 2. Note that if a matrix \mathbf{A} does not have zero columns (nor zero rows since \mathbf{A} is symmetric), then the condition of $\mathbf{b}_i \neq \mathbf{0}$ for all $i=1, \dots, n$ satisfies. On the other hand, if there exist some rows of \mathbf{B} satisfying $\mathbf{b}_j^T = \mathbf{0}$ then theorem 1 still holds if the corresponding elements of an optimal solution to problem (1) are simply taken as 1. In addition, theorem 2 follows immediately by noting that there always exists an optimal solution to the optimization problem (5).

Finally, we consider the issue of algorithms. From theorems 2-4, to solve the problem (2) or (4), we can first solve the problem (5) by enumeration and then calculate an optimal solution to (2) or (4) through (3a). This algorithm was first suggested by Choulakian (2001). Theorems 2-4 demonstrate that the algorithm proposed by Choulakian (2001) is correct although Li et al. (2002) showed that the Choulakian's proof itself was questionable.

In practice, the sizes of real problems may be quite large, thus this enumeration algorithm could be very expensive in terms of computational costs since it involves $O(2^n)$ choices. In this case, we may apply the following alternating algorithm which was proposed by O'Leary and Peleg (1983) to solve a similar problem.

Specifically, let \mathbf{A} be an $n \times n$ symmetric positive semi-definite matrix and let $\mathbf{z}_0 \in \mathbf{Z}$ be an initial guess of the alternating algorithm. Calculate \mathbf{z}_1 to be a vector that solves $\max_{\mathbf{z} \in \mathbf{Z}} \mathbf{z}_0^T \mathbf{A} \mathbf{z}$. Let \mathbf{z}_2 be a vector that solves $\max_{\mathbf{z} \in \mathbf{Z}} \mathbf{z}_1^T \mathbf{A} \mathbf{z}$, etc. In general, define \mathbf{z}_k to be a vector that solves $\max_{\mathbf{z} \in \mathbf{Z}} \mathbf{z}_{k-1}^T \mathbf{A} \mathbf{z}$.

It is clear that solving these sub-problems, $\max_{\mathbf{z} \in \mathbf{Z}} \mathbf{z}_{k-1}^T \mathbf{A} \mathbf{z}$ ($k=1, 2, \dots$), is straightforward, yielding a solution, $\mathbf{z}_k = S(\mathbf{A} \mathbf{z}_{k-1})$. A sign eigenvector \mathbf{z}^* is obtained after \mathbf{z}_k converges, i.e. $\mathbf{z}_k = \mathbf{z}_{k-1}$. Note that this heuristic algorithm cannot guarantee a convergence to

those sign eigenvectors which are associated with the largest sign eigenvalue of \mathbf{A} . See an example below.

Example. Consider a 2×2 matrix $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} > 0$ and set the initial guess as $\mathbf{z}_0 = [1, 1]^T$. According to O’Leary and Peleg algorithm, it converges instantly. The corresponding sign eigenvalue is $\lambda = \mathbf{z}_0^T \mathbf{A} \mathbf{z}_0 / (\mathbf{z}_0^T \mathbf{z}_0) = 1$. It is easy to verify, however, a sign eigenvector associated with the largest sign eigenvalue, 3, is $[1, -1]^T$.

To some extents, the above algorithm is akin to the power method in the L_2 norm which is used to calculate an (ordinary) eigenvector. Specifically, starting from an initial guess \mathbf{u}_0 , a series of vectors $\{\mathbf{u}_k\}$ in the power method is defined to be $\mathbf{v}_k = \mathbf{A} \mathbf{u}_{k-1}$ and $\mathbf{u}_k = \mathbf{v}_k / \|\mathbf{v}_k\|_2$. An eigenvector associated with the largest eigenvalue of a matrix \mathbf{A} is thus calculated after convergence; see for example, Golub and Van Loan (1996), pp406.

The optimization problem (4) may be applied to multivariate statistical analysis to construct robust statistical inference procedures in the L_1 norm. For instance, Galpin and Hawkins (1987) developed a robust principal component analysis procedure in the L_1 norm which was based on the optimization problem (4).

4. A Numerical example

A numerical example is given in this section to illustrate sign eigenanalysis and its application to robust principal component analysis (PCA) in the L_1 norm.

Galpin and Hawkins (1987) considered the following robust PCA formulation. The first loading vector in PCA, \mathbf{x}_1 , is defined to be a solution to the following problem:

$$\max_{\mathbf{x}} \sum_{i=1}^n |\mathbf{b}_i^T \mathbf{x}|, \quad \text{subject to } \|\mathbf{x}\|_2 = 1,$$

where $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]^T$ is a centred observation matrix. The second loading vector, \mathbf{x}_2 , is a solution to the above problem constrained by the condition of orthogonality between the loading vectors \mathbf{x}_1 and \mathbf{x}_2 . The remaining loading vectors, $\mathbf{x}_3, \mathbf{x}_4, \dots$, are defined similarly.

Example. Consider a 13×2 data matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2]^T$ shown in Table 1.

Table 1. An artificial data set

\mathbf{y}_1	0.281	0.284	0.262	0.276	0.308	0.302	0.252
\mathbf{y}_2	-0.020	0.074	0.072	0.043	0.056	0.092	0.014
\mathbf{y}_1	0.241	0.323	0.311	0.324	0.355	0.290	
\mathbf{y}_2	0.050	0.096	0.095	0.045	0.085	0.077	

It should be noted that the first data point, (0.281, -0.020), is an outlier. In this problem the median is used as an estimate of location. Now let $\mathbf{B} = [\mathbf{Y} - \mathbf{1}\mathbf{M}^T]$, and

$\mathbf{A}=\mathbf{B}\mathbf{B}^T$, where \mathbf{M} is a column-vector comprising the median of \mathbf{y}_1 and \mathbf{y}_2 and $\mathbf{1}$ is a 13-dimensional vector of ones.

According to Galpin and Hawkins (1987), to derive the major axis, \mathbf{x}_1 , of PCA in the L_1 norm, we have to solve the problem (4) or, equivalently, to solve the sign eigenvector associated with the largest sign eigenvalue of \mathbf{A} . Keeping the first element of a sign eigenvector being positive, the matrix \mathbf{A} has the following two distinct sign eigenvectors:

$$\mathbf{z}_1=[1, 1, 1, 1, -1, -1, 1, 1, -1, -1, -1, -1, -1]^T,$$

$$\mathbf{z}_2=[1, 1, 1, 1, 1, -1, 1, 1, -1, -1, 1, -1, -1]^T,$$

where \mathbf{z}_1 is a sign eigenvector associated the largest sign eigenvalue, 0.0127. From (3a) the corresponding major axis of PCA is $\mathbf{x}_1 = \mathbf{B}^T \mathbf{z}_1 / \|\mathbf{B}^T \mathbf{z}_1\|_2 = [0.8050, 0.5933]^T$, which is the optimal solution to the problem (4). The first component is therefore $\mathbf{B}\mathbf{x}_1$. The minor axis, $\mathbf{x}_2=[0.5933, -0.8050]^T$, is taken as a vector orthogonal to the major axis, \mathbf{x}_1 . Figure 1 shows the scatter plot (upper left) and the resultant axes using the PCA in the L_1 norm (lower left). For comparison, Figure 1 also displays the axes using the L_2 PCA without and with the outlier, (0.281, -0.020) (upper and lower right respectively).

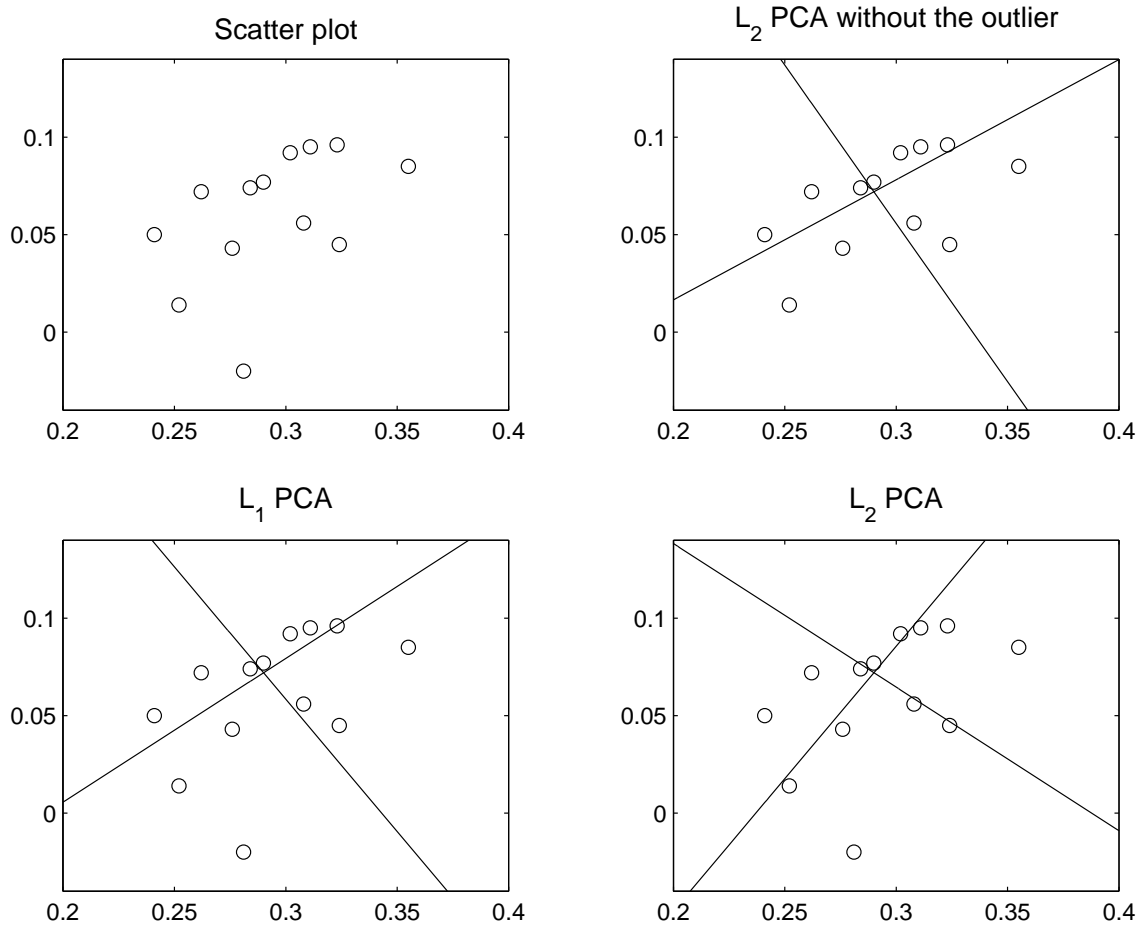


Figure 1. Comparison of several approaches to principal component analysis

It can be seen from Figure 1 (lower left) that the major axis derived by L_1 PCA is drawn slightly towards the outlier when compared with the axes obtained by removing

the outlier (upper right). In contrast, if the outlier is not removed, the major axis in L_2 PCA, Figure 1 lower right, is drawn significantly towards the outlier, indicating that L_2 PCA has greater sensitivity to outliers.

Acknowledgement

The author wishes to thank the referees for their constructive comments on an earlier version of the manuscript.

Appendix. Proofs of Lemmas.

Proof of Lemma 1.

If lemma 1 does not hold, then there exists an index, i , say $i=1$, such that $\mathbf{b}_1^T \mathbf{x}^* = 0$, and the corresponding element of the optimal solution $\mathbf{z}^* = [z_1^*, \dots, z_n^*]^T$, z_1^* , can be either 1 or -1 . Let $\mathbf{t}_1 = [1, z_2^*, \dots, z_n^*]^T$ and $\mathbf{t}_2 = [-1, z_2^*, \dots, z_n^*]^T$. Then $\mathbf{t}_1^T \mathbf{B} \mathbf{b}_1$ and $\mathbf{t}_2^T \mathbf{B} \mathbf{b}_1$ cannot be zero simultaneously (otherwise, from $\mathbf{t}_2^T \mathbf{B} \mathbf{b}_1 = 0$ we have $\sum_{j=2}^n z_j \mathbf{b}_j^T \mathbf{b}_1 = \mathbf{b}_1^T \mathbf{b}_1 > 0$, and from $\mathbf{t}_1^T \mathbf{B} \mathbf{b}_1 = 0$ we have $\sum_{j=2}^n z_j \mathbf{b}_j^T \mathbf{b}_1 = -\mathbf{b}_1^T \mathbf{b}_1 < 0$). Without loss of generality we suppose that $\mathbf{t}_1^T \mathbf{B} \mathbf{b}_1 \neq 0$. Define $\mathbf{v} = \mathbf{b}_1$ if $\mathbf{t}_1^T \mathbf{B} \mathbf{b}_1 > 0$, otherwise $\mathbf{v} = -\mathbf{b}_1$, such that $\mathbf{t}_1^T \mathbf{B} \mathbf{v} > 0$. Consider $\mathbf{x}(\varepsilon) = \mathbf{x}^* + \varepsilon \mathbf{v}$. The objective function evaluated at $(\mathbf{x}(\varepsilon), \mathbf{t}_1)$ is:

$$J(\varepsilon) = \frac{\mathbf{t}_1^T \mathbf{B} \mathbf{x}(\varepsilon)}{\|\mathbf{x}(\varepsilon)\|_2} = \frac{\mathbf{t}_1^T \mathbf{B} \mathbf{x}^* + \varepsilon \mathbf{t}_1^T \mathbf{B} \mathbf{v}}{(\mathbf{x}^{*T} \mathbf{x}^* + \varepsilon^2 \mathbf{v}^T \mathbf{v})^{1/2}}.$$

It is easy to verify that $dJ(\varepsilon)/d\varepsilon = \left\{ (\mathbf{t}_1^T \mathbf{B} \mathbf{v}) \|\mathbf{x}^*\|_2^2 - \varepsilon (\mathbf{t}_1^T \mathbf{B} \mathbf{x}^*) \|\mathbf{v}\|_2^2 \right\} \|\mathbf{x}(\varepsilon)\|_2^{-3}$. Since $\mathbf{t}_1^T \mathbf{B} \mathbf{x}^* = \mathbf{z}^{*T} \mathbf{B} \mathbf{x}^* > 0$ and $\|\mathbf{x}^*\|_2 \neq 0$, there exist some small $\varepsilon > 0$ such that $dJ(\varepsilon)/d\varepsilon > 0$. This contradicts the assumption that $(\mathbf{x}^*, \mathbf{z}^*)$ is an optimal solution to the problem (2). \square

Proof of Lemma 2.

Equation (3a) is immediate by differentiating the Lagrange function of problem (2), $L = \mathbf{z}^T \mathbf{B} \mathbf{x} - \mu(\mathbf{x}^T \mathbf{x} - 1)$, with respect to \mathbf{x} , and setting the derivative to zero, where μ is a Lagrange multiplier.

Next, to prove (3b), we note that $\mathbf{z}^{*T} \mathbf{B} \mathbf{x}^* = \sum_{i=1}^n z_i^* \mathbf{b}_i^T \mathbf{x}^*$. From Lemma 1 we have $\mathbf{b}_i^T \mathbf{x}^* \neq 0$, thus, to attain the maximum, z_i^* must have the same signs as $\mathbf{b}_i^T \mathbf{x}^*$ ($i=1, \dots, n$), i.e. $\mathbf{z}^* = S(\mathbf{B} \mathbf{x}^*)$. \square

References

- Choulakian, V., 2001. Robust Q-mode principal component analysis in L_1 . *Comput. Statist. Data Anal.* 37, 135-150.
- Galpin, J.S. and Hawkins, D. M., 1987. Methods of L_1 estimation of a covariance matrix. *Comput. Statist. Data Anal.* 5, 305-319.
- Golub, G. H. and Van Loan, C. F., 1996. *Matrix computations*. 3rd Ed. The Johns Hopkins University Press, London.
- Li, B., Martin, E.B., and Morris, A.J., 2002. On principal component analysis in L_1 . *Comput. Statist. Data Anal.* 40, 471-474.
- O'Leary, D. P. and Peleg, S., 1983. Digital image compression by outer product expansion. *IEEE Transactions on Communications* COM-31, 441-444.
- Rao, C.R., 1973. *Linear statistical inference and its applications*. Wiley, New York.