

## A Justification of Rotation in Principal Component Analysis: Projective viewpoint of PCA

主成分分析 (PCA) における回転の正当化：射影の観点から見たPCA

OKAMOTO Yasuharu  
岡 本 安 晴

# A Justification of Rotation in Principal Component Analysis: Projective viewpoint of PCA<sup>1)</sup>

主成分分析 (PCA) における回転の正当化：射影の観点から見た PCA

OKAMOTO, Yasuharu  
岡本安晴

[Abstract]

---

Rotation in principal component analysis (PCA) is justified from projective point of view. PCA is interpreted as orthogonal projection of data, and components are constructed as coordinates of the projected data with respect to some basis. From the viewpoint of projection, any bases for the space of the projected data are equally valid, but a particular one may have substantive meaning in the research area. Transition from the current basis to another one induces change in coordinates, which is represented by transformation matrix, orthogonal or oblique.

Optimal orthogonal projection is obtained from the singular value decomposition (SVD) of the data matrix and results in the solution numerically equivalent to usual PCA. But, from the viewpoint of projection, transformation of components is interpreted as transformation of coordinates induced by change of bases, which can be understood geometrically with clear intuition. The principal components in usual PCA are coordinates of the projected data for the basis composed of right singular vectors of the data matrix.

---

## Introduction

When principal components of data matrix  $\mathbf{Z}$  are extracted successively with a certain optimal criterion, which constrains components to be orthogonal with each other, directions of the components are uniquely determined, except for reflection. Because of this uniqueness of the components with respect to the optimality condition, rotation in principal component analysis (PCA) is inhibited (e.g. Timm, 1975, p. 540). Against this theoretical criticism of the

rotation, users of PCA may feel that rotation of principal components is useful from substantive point of view, i.e. rotated components have substantive meanings. Justification of the rotation is discussed in this paper.

### A Justification

Let  $Z$  be a standardized data matrix of order  $n \times k$  with elements  $z_{ij}$ , which represents the value of the  $j$ th variable for the  $i$ th case. PCA can be done using the singular value decomposition (SVD) of  $Z/n^{1/2}$

$$\frac{1}{n^{1/2}}Z = U\Lambda V',$$

where  $U$  and  $V$  are column orthonormal matrices of order  $n \times s$  and  $k \times s$ , respectively.  $s$  is the rank of  $Z$  and  $\Lambda$  is a diagonal matrix with nonnegative elements. We can put

$$U = [U_1 U_2], \quad V = [V_1 V_2]$$

and

$$\Lambda = \begin{bmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \Lambda_2 \end{bmatrix}, \quad \Lambda_1 = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_r \end{bmatrix}, \quad \Lambda_2 = \begin{bmatrix} \lambda_{r+1} & & 0 \\ & \ddots & \\ 0 & & \lambda_s \end{bmatrix}, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s > 0$$

where  $U_1$  and  $U_2$  are of order  $n \times r$  and  $n \times (s-r)$ , and  $V_1$  and  $V_2$  of order  $k \times r$  and  $k \times (s-r)$ . For PCA with  $r$  principal components, we derive the following decomposition

$$\begin{aligned} Z &= Z(V_1 V_1') + Z[\mathbf{I} - (V_1 V_1')] \\ &= ZP + Z(\mathbf{I} - P) \end{aligned} \tag{1}$$

where  $P = V_1 V_1'$ .  $P$  is a orthogonal projection, i.e. satisfies the following two conditions

$$P^2 = P, \quad \text{and} \quad P' = P$$

Two parts of the sum are orthogonal with each other,

$$(ZP)[Z(\mathbf{I} - P)]' = Z(P - P^2)Z' = 0$$

$r$  principal components  $C = [c_1 \dots c_r]$  are given by the following equation (e. g. Yanai et al., 1990, p. 230),

$$C = ZV_1\Lambda_1^{-1} = n^{1/2}U_1$$

Hence, we have

$$\mathbf{ZP} = \mathbf{C}(\Lambda_1 \mathbf{V}_1') = [\mathbf{c}_1 \cdots \mathbf{c}_r](\Lambda_1 \mathbf{V}_1') \quad (2)$$

Eq. (2) shows that the first term of the decomposition (1) is a space spanned by the  $r$  principal components. With any nonsingular matrix  $\mathbf{T}$  of order  $r \times r$ , we can transform eq. (2) as follows,

$$\mathbf{ZP} = (\mathbf{CT})\{\mathbf{T}^{-1}(\Lambda_1 \mathbf{V}_1')\} \quad (3)$$

Eq. (3) means that the space spanned by the initial set of principal components can be also spanned by  $\mathbf{CT}$ , which are components derived by rotation  $\mathbf{T}$  for basis  $\mathbf{T}^{-1}(\Lambda_1 \mathbf{V}_1')$ . When we decompose  $\mathbf{Z}$  with the initial  $r$  principal components as eq. (1), the term  $\mathbf{ZP}$ , which is spanned by the initial principal components does not change under rotation  $\mathbf{T}$  as is shown by eq. (3). This invariance of the space  $\mathbf{ZP}$  justifies rotation of principal components.

### More Technical Details

Ten Berge and Kiers (1996)<sup>2)</sup> classifies PCAs into three types, Hotelling PCA, Pearson PCA and Rao PCA. In Hotelling PCA, components  $\mathbf{ZW}$  is extracted so that

$$tr\left\{\frac{1}{n}(\mathbf{ZW})'(\mathbf{ZW})\right\} \quad (4)$$

is maximized under the constraint

$$\mathbf{W}'\mathbf{W} = \mathbf{I}, \quad (5)$$

where  $\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_r]$  is a matrix of weights for  $r$  components. For any orthogonal rotation  $\mathbf{T}$ , we have

$$tr\left\{\frac{1}{n}(\mathbf{ZWT})'(\mathbf{ZWT})\right\} = tr\left\{\frac{1}{n}(\mathbf{ZW})'(\mathbf{ZW})\right\}$$

and

$$(\mathbf{WT})'(\mathbf{WT}) = \mathbf{T}'\mathbf{W}'\mathbf{WT} = \mathbf{T}'\mathbf{T} = \mathbf{I}$$

That is, if  $\mathbf{ZW}$  maximizes eq. (4) under the constraint (5),  $\mathbf{ZWT}$  also do the same. Hence, in Hotelling PCA, orthogonal rotation is allowed.

In Pearson PCA or Rao PCA, linear combinations  $\mathbf{ZWP}'$  of components  $\mathbf{ZW}$  are used to approximate the data  $\mathbf{Z}$  or the correlation matrix  $\mathbf{Z}'\mathbf{Z}/n$ , where  $\mathbf{P}$  is a pattern matrix.

Because

$$\mathbf{ZWP}' = (\mathbf{ZWT})(\mathbf{P}(\mathbf{T}')^{-1})'$$

for any nonsingular matrix  $\mathbf{T}$ , any rotation, which may be oblique, can be performed in Pearson PCA or in Rao PCA.

According to ten Berge and Kiers, among the three types of PCA, only Hotelling PCA does not allow principal components to be rotated obliquely. But projective interpretation of Hotelling PCA, which is presented in this paper, accepts oblique rotation.

### Projective Interpretation of Hotelling PCA

In Hotelling PCA,  $r$  principal components are derived one by one using weight  $\mathbf{w}_i$ , where  $\mathbf{w}_i'\mathbf{w}_i = 1$  and  $\mathbf{w}_i'\mathbf{w}_j = 0$ , for  $i \neq j$ . Each component  $\mathbf{Zw}_i$  is determined successively to maximize variance  $\frac{1}{n}(\mathbf{Zw}_i)'(\mathbf{Zw}_i)$ . Ten Berge et al. explain that Rao (1973) rederived equation (6) as a criterion for simultaneous derivation of  $r$  principal components in Hotelling PCA.

$$\text{tr}\left\{\frac{1}{n}(\mathbf{ZW})'(\mathbf{ZW})\right\} = \text{tr}(\mathbf{W}'\mathbf{R}\mathbf{W}), \quad (6)$$

where

$$\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_r], \quad \mathbf{W}'\mathbf{W} = \mathbf{I}_r \quad (7)$$

and

$$\mathbf{R} = \frac{1}{n}\mathbf{Z}'\mathbf{Z} \text{ is the correlation matrix of } \mathbf{Z}.$$

I propose that  $\mathbf{ZW}$  should be considered to be an orthogonal projection of  $\mathbf{Z}$  rather than a matrix of mere linear combinations of the columns of  $\mathbf{Z}$ . When  $\mathbf{Zw}_i$  is interpreted as a orthogonal projection of  $\mathbf{Z}$  onto a one-dimensional space  $\mathbf{w}_i$ , a natural extension of  $\mathbf{Zw}_i$  to  $r$ -dimensional space is an orthogonal projection  $\mathbf{ZP}$  of  $\mathbf{Z}$  by an orthogonal projection matrix  $\mathbf{P}$  of rank  $r$ . In this extension by an orthogonal projection, the criterion is given by the following equation:

$$\text{tr}\left\{\frac{1}{n}(\mathbf{ZP})'(\mathbf{ZP})\right\} = \text{tr}(\mathbf{P}'\mathbf{R}\mathbf{P}), \quad (8)$$

where

$$\mathbf{P}^2 = \mathbf{P}, \quad \mathbf{P}' = \mathbf{P} \quad \text{and} \quad \text{rank}(\mathbf{P}) = r. \quad (9)$$

(6) and (8) are of the same form except matrices  $\mathbf{W}$  and  $\mathbf{P}$ . But, conceptualizations are quite different. In (6),  $\mathbf{Z}\mathbf{W}$  is a matrix of linear combinations of  $\mathbf{Z}$  and the constraint  $\mathbf{W}'\mathbf{W} = \mathbf{I}_r$  is imposed to prevent the components from coinciding with each other (ten Berge et al. 1996, Pp. 336-337). As for (8), the constraints (9) are used to represent that  $\mathbf{P}$  is an orthogonal projection matrix of rank  $r$ . (8) represents r-dimensional dispersion of the projected data points around the origin in the k-dimensional space. The orthogonal projection matrix  $\mathbf{P}$  which maximizes (8) is derived in the following section.

### Optimal Orthogonal Projection

To seek the optimal orthogonal projection, let consider the singular value decomposition (SVD) of  $\frac{1}{n^{1/2}}\mathbf{Z}$ . Put the SVD as follows:

$$\frac{1}{n^{1/2}}\mathbf{Z} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}' = [\mathbf{U}_1 \mathbf{U}_2] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_s \end{bmatrix} \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix}.$$

Then we have

$$\mathbf{R} = \frac{1}{n}\mathbf{Z}'\mathbf{Z} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' = \lambda_1^2 \mathbf{v}_1 \mathbf{v}'_1 + \dots + \lambda_s^2 \mathbf{v}_s \mathbf{v}'_s$$

Hence, from (8), we have

$$\begin{aligned} \text{tr} \left\{ \frac{1}{n} (\mathbf{Z}\mathbf{P})' (\mathbf{Z}\mathbf{P}) \right\} &= \text{tr}(\mathbf{P}'\mathbf{R}\mathbf{P}) \\ &= \text{tr} \left\{ \mathbf{P}' (\lambda_1^2 \mathbf{v}_1 \mathbf{v}'_1 + \dots + \lambda_s^2 \mathbf{v}_s \mathbf{v}'_s) \mathbf{P} \right\} \\ &= \lambda_1^2 \text{tr} \left\{ (\mathbf{v}'_1 \mathbf{P}) (\mathbf{v}'_1 \mathbf{P})' \right\} + \dots + \lambda_s^2 \text{tr} \left\{ (\mathbf{v}'_s \mathbf{P}) (\mathbf{v}'_s \mathbf{P})' \right\} \\ &= \lambda_1^2 \|\mathbf{v}'_1 \mathbf{P}\|^2 + \dots + \lambda_s^2 \|\mathbf{v}'_s \mathbf{P}\|^2 \\ &\leq \lambda_1^2 + \dots + \lambda_s^2, \end{aligned}$$

since, in general,

$$\|\mathbf{v}'_i \mathbf{P}\|^2 \leq \|\mathbf{v}_i\|^2 = 1 \quad \text{for an orthogonal projection matrix } \mathbf{P}.$$

Because

$$\text{rank}(\mathbf{P}) = r,$$

we can get as the maximum value of (8)

$$\lambda_1^2 + \dots + \lambda_r^2$$

when

$$\mathbf{v}'_1 = \mathbf{v}'_1 \mathbf{P}, \dots, \mathbf{v}'_r = \mathbf{v}'_r \mathbf{P}.$$

That is, the optimal orthogonal projection matrix  $\mathbf{P}$  is given by

$$\mathbf{P} = [\mathbf{v}_1 \dots \mathbf{v}_r] \begin{bmatrix} \mathbf{v}'_1 \\ \vdots \\ \mathbf{v}'_r \end{bmatrix} = \mathbf{V}_1 \mathbf{V}'_1$$

For more details, see the appendix.

### Principal Components

A point in projection  $\mathbf{ZP}$  can be represented by coordinates corresponding to any basis of  $\mathbf{ZP}$ . One of possible bases is  $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ . When we adopt this orthonormal basis  $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ , coordinates of the points in  $\mathbf{ZP}$  are given as follows:

$$\begin{aligned} \mathbf{ZP}[\mathbf{v}_1 \dots \mathbf{v}_r] &= n^{1/2} [\mathbf{U}_1 \mathbf{U}_2] \mathbf{\Lambda} \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix} \mathbf{v}_1 \mathbf{v}'_1 \mathbf{v}_1 \\ &= n^{1/2} [\mathbf{U}_1 \mathbf{U}_2] \mathbf{\Lambda} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \\ &= n^{1/2} \mathbf{U}_1 \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{bmatrix} \\ &= n^{1/2} [\lambda_1 \mathbf{u}_1 \dots \lambda_r \mathbf{u}_r] \end{aligned}$$

Standardizing the above coordinates, we get the so-called principal components  $\mathbf{C}$ , where

$$\mathbf{C} = \mathbf{ZP}[\mathbf{v}_1 \dots \mathbf{v}_r] \mathbf{\Lambda}^{-1} = n^{1/2} \mathbf{U}_1 \quad (10)$$

Now, put

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_1 & 0 \\ 0 & \mathbf{\Lambda}_2 \end{bmatrix}, \quad \mathbf{\Lambda}_1 = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_2 = \begin{bmatrix} \lambda_{r+1} & & 0 \\ & \ddots & \\ 0 & & \lambda_s \end{bmatrix}$$

Then rewriting  $\mathbf{ZP}$  as follows

$$\begin{aligned} \mathbf{ZP} &= n^{1/2} [\mathbf{U}_1 \mathbf{U}_2] \begin{bmatrix} \mathbf{\Lambda}_1 & 0 \\ 0 & \mathbf{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix} \mathbf{V}_1 \mathbf{V}'_1 \\ &= n^{1/2} [\mathbf{U}_1 \mathbf{U}_2] \begin{bmatrix} \mathbf{\Lambda}_1 & 0 \\ 0 & \mathbf{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_r \\ \mathbf{0} \end{bmatrix} \mathbf{V}'_1 \\ &= n^{1/2} [\mathbf{U}_1 \mathbf{U}_2] \begin{bmatrix} \mathbf{\Lambda}_1 \mathbf{V}'_1 \\ \mathbf{0} \end{bmatrix} \\ &= n^{1/2} \mathbf{U}_1 \mathbf{\Lambda}_1 \mathbf{V}'_1, \end{aligned}$$

We get

$$\mathbf{ZP} = n^{1/2} \mathbf{U}_1 (\mathbf{\Lambda}_1 \mathbf{V}'_1) \quad (11)$$

Comparing (11) with (10), we have

$$\mathbf{ZP} = \mathbf{C} (\mathbf{\Lambda}_1 \mathbf{V}'_1).$$

Hence, denoting the pattern matrix for the principal components  $\mathbf{C}$  by  $\mathbf{P}_{at}$ , we have

$$\mathbf{P}_{at} = (\mathbf{\Lambda}_1 \mathbf{V}'_1)' = \mathbf{V}_1 \mathbf{\Lambda}_1$$

and

$$\begin{aligned} \mathbf{Z} &= \mathbf{ZP} + \mathbf{Z}(\mathbf{I} - \mathbf{P}) \\ &= \mathbf{C} \mathbf{P}'_{at} + \mathbf{Z}(\mathbf{I} - \mathbf{P}) \end{aligned}$$

Because  $\mathbf{P}$  is an orthogonal projection matrix, we have

$$(\mathbf{ZP})\{\mathbf{Z}(\mathbf{I} - \mathbf{P})\}' = (\mathbf{C} \mathbf{P}'_{at})\{\mathbf{Z}(\mathbf{I} - \mathbf{P})\}' = \mathbf{0}$$

That is, columns of  $\mathbf{ZP}$  and  $\mathbf{Z}(\mathbf{I} - \mathbf{P})$  are orthogonal (uncorrelated). From this orthogonal relation, we have following Pythagorean equation

$$tr(\mathbf{Z}'\mathbf{Z}) = tr\left\{\mathbf{ZP}'(\mathbf{ZP})\right\} + tr\left\{\mathbf{Z}(\mathbf{I} - \mathbf{P})'(\mathbf{Z}(\mathbf{I} - \mathbf{P}))\right\}$$

The above equation shows maximization of  $tr\left\{\mathbf{ZP}'(\mathbf{ZP})\right\}$  (Hotelling PCA) is equivalent to minimization of  $tr\left\{\mathbf{Z}(\mathbf{I} - \mathbf{P})'(\mathbf{Z}(\mathbf{I} - \mathbf{P}))\right\}$  (Pearson PCA)(Kiers, personal communication, 2006, Aug.).

The structure matrix  $\mathbf{S}$  is given as follows



$$\begin{aligned}
 \mathbf{S} &= \frac{1}{n} \mathbf{Z}' \mathbf{C} \\
 &= \frac{1}{n^{1/2}} [\mathbf{V}_1 \mathbf{V}_2] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{U}'_1 \\ \mathbf{U}'_2 \end{bmatrix} (n^{1/2} \mathbf{U}_1) \\
 &= \mathbf{V}_1 \Lambda_1 \\
 &= \mathbf{P}'_{at}
 \end{aligned}$$

### Rotations

Any basis  $\{\mathbf{b}_1, \dots, \mathbf{b}_r\}$  of  $\mathbf{ZP}$  can be represented by transformation of the basis  $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ . Hence, for some nonsingular matrix  $\mathbf{T}$ , we have

$$\begin{bmatrix} \mathbf{b}'_1 \\ \vdots \\ \mathbf{b}'_r \end{bmatrix} = \mathbf{T} \begin{bmatrix} \mathbf{v}'_1 \\ \vdots \\ \mathbf{v}'_r \end{bmatrix} = \mathbf{T} \mathbf{V}'_1$$

From (11), we have

$$\begin{aligned}
 \mathbf{ZP} &= n^{1/2} \mathbf{U}_1 (\Lambda_1 \mathbf{V}'_1) \\
 &= \mathbf{C} \begin{pmatrix} \Lambda_1 \mathbf{T}^{-1} \begin{bmatrix} \mathbf{b}'_1 \\ \vdots \\ \mathbf{b}'_r \end{bmatrix} \end{pmatrix} \\
 &= (\mathbf{C} \Lambda_1 \mathbf{T}^{-1}) \begin{bmatrix} \mathbf{b}'_1 \\ \vdots \\ \mathbf{b}'_r \end{bmatrix}
 \end{aligned}$$

Hence, coordinates  $\mathbf{C}_T$  of  $\mathbf{ZP}$  for a rotated basis  $\mathbf{T} \mathbf{V}'_1$  by rotation  $\mathbf{T}$ , which may be oblique rather than orthogonal, are given by

$$\mathbf{C}_T = \mathbf{C} \Lambda_1 \mathbf{T}^{-1}$$

To seek the pattern matrix  $\mathbf{P}_T$  for the rotated components  $\mathbf{C}_T$ , consider the following equations:

$$\begin{aligned}
 \mathbf{ZP} &= \mathbf{C} \mathbf{P}'_{at} \\
 &= \mathbf{C} \Lambda_1 \mathbf{T}^{-1} \mathbf{T} \Lambda_1^{-1} \mathbf{P}'_{at}
 \end{aligned}$$

$$= \mathbf{C}_T (\mathbf{T} \boldsymbol{\Lambda}_1^{-1} \mathbf{P}'_{at})$$

Hence, we have

$$\mathbf{P}_T = (\mathbf{T} \boldsymbol{\Lambda}_1^{-1} \mathbf{P}'_{at})' = \mathbf{P}_{at} \boldsymbol{\Lambda}_1^{-1} \mathbf{T}'$$

The structure matrix  $\mathbf{S}_T$  for rotation  $\mathbf{T}$  is derived as follows:

$$\begin{aligned} \mathbf{S}_T &= \frac{1}{n} \mathbf{Z}' \mathbf{C}_T \\ &= \frac{1}{n} \mathbf{Z}' \mathbf{C} \boldsymbol{\Lambda}_1 \mathbf{T}^{-1} \\ &= \mathbf{S} \boldsymbol{\Lambda}_1 \mathbf{T}^{-1} \\ &= \mathbf{P}_{at} \boldsymbol{\Lambda}_1^{-1} \mathbf{T}' (\mathbf{T}')^{-1} \boldsymbol{\Lambda}_1 \boldsymbol{\Lambda}_1 \mathbf{T}^{-1} \\ &= \mathbf{P}_T (\boldsymbol{\Lambda}_1 \mathbf{T}^{-1})' (\boldsymbol{\Lambda}_1 \mathbf{T}^{-1}) \end{aligned}$$

### Summary

Principal component analysis (PCA) is interpreted as an orthogonal projection of data  $\mathbf{Z}$ . Rotation in PCA is considered in this projected space and interpreted as transformation of coordinates of projection  $\mathbf{ZP}$ , which accompanies with change of a basis of  $\mathbf{ZP}$ . The so-called principal components are coordinates of  $\mathbf{ZP}$  for the special basis, which consists of the right singular vectors  $\mathbf{V}_1$  of  $\mathbf{Z}$ . Any other bases than  $\mathbf{V}_1$  are also valid for  $\mathbf{ZP}$ . Transformations between bases are not constrained to be orthogonal. They can be oblique from the projective point of view presented in this paper.

### Appendix

Ten Berge (personal communication, June 21, 2006) pointed out that the maximum value of (8) can be given by generalized Kristof theorem (ten Berge, 2005, pp. 28-29). First, transform (8) as follows,

$$tr(\mathbf{P}' \mathbf{R} \mathbf{P}) = tr(\mathbf{P}' \mathbf{V} \boldsymbol{\Lambda}^2 \mathbf{V}' \mathbf{P}) = tr(\mathbf{V}' \mathbf{P} \mathbf{P}' \mathbf{V} \boldsymbol{\Lambda}^2)$$

$\mathbf{V}$  and  $\mathbf{P}$  are suborthonormal (s.o.), so  $\mathbf{V}' \mathbf{P} \mathbf{P}' \mathbf{V}$  is also s.o. (property 2 in ten Berge,

2005, p. 28). Rank of  $\mathbf{V}'\mathbf{P}\mathbf{P}'\mathbf{V}$  satisfies the following relation,

$$\text{rank}(\mathbf{V}'\mathbf{P}\mathbf{P}'\mathbf{V}) \leq \text{rank}(\mathbf{P}) = r$$

Hence, applying generalized Kristof theorem to  $\text{tr}(\mathbf{V}'\mathbf{P}\mathbf{P}'\mathbf{V}\Lambda^2)$ , we have

$$\text{tr}\left\{\frac{1}{n}(\mathbf{ZP})'(\mathbf{ZP})\right\} = \text{tr}(\mathbf{V}'\mathbf{P}\mathbf{P}'\mathbf{V}\Lambda^2) \leq \lambda_1^2 + \dots + \lambda_r^2$$

In the following, I present a proof of the above relation, which is based on the fact that  $\mathbf{P}$  is an orthogonal projection. This proof would give the reader more concrete image.

Since  $\mathbf{P}$  is an orthogonal projection matrix and  $\text{rank}(\mathbf{P}) = r$ , we can put

$$\mathbf{P} = [\mathbf{p}_1 \cdots \mathbf{p}_r] \begin{bmatrix} \mathbf{p}'_1 \\ \vdots \\ \mathbf{p}'_r \end{bmatrix}$$

for some orthonormal column vectors  $\mathbf{p}_i$ s (Golub and van Loan, 1996). Then we have

$$\begin{aligned} \mathbf{v}'_i\mathbf{P} &= \mathbf{v}'_i\mathbf{p}_1\mathbf{p}'_1 + \dots + \mathbf{v}'_i\mathbf{p}_r\mathbf{p}'_r \\ \|\mathbf{v}'_i\mathbf{P}\|^2 &= (\mathbf{v}'_i\mathbf{p}_1)^2 + \dots + (\mathbf{v}'_i\mathbf{p}_r)^2 \\ \|\mathbf{v}'_1\mathbf{P}\|^2 + \dots + \|\mathbf{v}'_s\mathbf{P}\|^2 &= (\mathbf{v}'_1\mathbf{p}_1)^2 + \dots + (\mathbf{v}'_1\mathbf{p}_r)^2 + \\ &\quad \dots \\ &\quad + (\mathbf{v}'_s\mathbf{p}_1)^2 + \dots + (\mathbf{v}'_s\mathbf{p}_r)^2 \end{aligned}$$

Noting that

$$(\mathbf{v}'_i\mathbf{p}_j)^2 + \dots + (\mathbf{v}'_s\mathbf{p}_j)^2 = (\mathbf{p}'_j\mathbf{v}_i)^2 + \dots + (\mathbf{p}'_j\mathbf{v}_s)^2 = \|\mathbf{p}'_j\mathbf{P}_v\|^2 \leq 1,$$

where

$$\mathbf{P}_v = \mathbf{V}\mathbf{V}' \text{ is an orthogonal projection matrix,}$$

we have

$$\|\mathbf{v}'_1\mathbf{P}\|^2 + \dots + \|\mathbf{v}'_s\mathbf{P}\|^2 = \|\mathbf{p}'_1\mathbf{P}_v\|^2 + \dots + \|\mathbf{p}'_r\mathbf{P}_v\|^2 \leq r$$

To maximize

$$\text{tr}\left\{\frac{1}{n}(\mathbf{ZP})'(\mathbf{ZP})\right\} = \lambda_1^2 \|\mathbf{v}'_1\mathbf{P}\|^2 + \dots + \lambda_s^2 \|\mathbf{v}'_s\mathbf{P}\|^2$$

under the constraint

$$\|\mathbf{v}'_j\mathbf{P}\|^2 \leq 1 \quad \text{and} \quad \|\mathbf{v}'_1\mathbf{P}\|^2 + \dots + \|\mathbf{v}'_s\mathbf{P}\|^2 \leq r,$$

consider the following problem:

Maximize the function

$$f(a_1, \dots, a_s) = \lambda_1^2 a_1 + \dots + \lambda_s^2 a_s, \quad \lambda_1^2 \geq \dots \geq \lambda_s^2 > 0$$

under the constraints

$$0 \leq a_i \leq 1, \quad a_1 + \dots + a_s \leq r$$

First, we can assume

$$a_1 + \dots + a_s = r,$$

which is shown as follows:

if  $a_1 + \dots + a_s < r$ , then we can find an  $a_i$  such that  $a_i < 1$  and put

$$a_{i1} = a_i + \varepsilon, \quad \text{where } \varepsilon = \min\{1 - a_i, r - (a_1 + \dots + a_s)\}.$$

Replacing  $a_i$  by  $a_{i1}$ , we have

$$f(a_1, \dots, a_i, \dots, a_s) < f(a_1, \dots, a_{i1}, \dots, a_s).$$

From the above discussion, we can assume

$$a_1 + \dots + a_s = r.$$

Moreover, we can assume

$$a_1 = \dots = a_r = 1, \quad a_{r+1} = \dots = a_s = 0,$$

for which  $f(a_1, \dots, a_s)$  takes its maximum value under the constraints. This can be seen from the following discussions.

If  $a_1 < 1$ , then we can find  $\lambda_i$  such that  $a_i > 0$ ,  $i > 1$  and put

$$a_{11} = a_1 + \varepsilon, \quad a_{i1} = a_i - \varepsilon, \quad \text{where } \varepsilon = \min\{1 - a_1, a_i\}.$$

Replacing  $a_1$  and  $a_i$  by  $a_{11}$  and  $a_{i1}$  respectively, we have

$$f(a_1, \dots, a_i, \dots, a_s) \leq f(a_{11}, \dots, a_{i1}, \dots, a_s)$$

Hence, we can assume

$$a_1 = 1.$$

If  $a_1 = \dots = a_{j-1} = 1$  and  $a_j < 1$ ,  $1 < j \leq r$ , then we can find  $a_i > 0$ ,  $j < i$  and put

$$a_{j1} = a_j + \varepsilon, \quad a_{i1} = a_i - \varepsilon, \quad \text{where } \varepsilon = \min\{1 - a_j, a_i\}.$$

Replacing  $a_j$  and  $a_i$  by  $a_{j1}$  and  $a_{i1}$  respectively, we have

$$f(a_1, \dots, a_j, \dots, a_i, \dots, a_s) \leq f(a_1, \dots, a_{j_1}, \dots, a_{i_1}, \dots, a_s)$$

Hence, we can assume

$$a_j = 1, \quad 1 \leq j \leq r$$

From the above discussions, we can choose  $a_i$ 's, which maximizes  $f(a_1, \dots, a_s)$  under the constraint

$$0 \leq a_i \leq 1, \quad a_1 + \dots + a_s \leq r,$$

such that

$$a_1 = \dots = a_r = 1, \quad a_{r+1} = \dots = a_s = 0.$$

Now, putting

$$\mathbf{P} = \mathbf{V}_1 \mathbf{V}_1',$$

we have

$$a_1 = \|\mathbf{v}'_1 \mathbf{P}\|^2 = 1, \quad \dots, \quad a_r = \|\mathbf{v}'_r \mathbf{P}\|^2 = 1, \\ a_{r+1} = \|\mathbf{v}'_{r+1} \mathbf{P}\|^2 = 0, \quad \dots, \quad a_s = \|\mathbf{v}'_s \mathbf{P}\|^2 = 0$$

## References

- Golub, G. H. and van Loan, C. F. (1996). *Matrix Computations*, 3<sup>rd</sup> ed. Baltimore and London: The Johns Hopkins University Press.
- Okamoto, Y. (2005). On the special type of principal component analysis: It should belong to factor analysis. *Japan Women's University Journal, Faculty of Integrated Arts and Social Sciences*, 16, 43-50.
- Rao, C. R. (1973). *Linear statistical inference and its applications*, 2<sup>nd</sup> ed. New York: John Wiley & Sons.
- Ten Berge, J. M. F. (2005). *Least Squares optimization in multivariate analysis*. (Pdf version of the monograph published by DSWO Press, Leiden, 1993).
- Ten Berge, J. M. F. and Kiers, H. A. L. (1996). Optimality criteria for principal component analysis and generalizations. *British Journal of Mathematical and Statistical Psychology*, 49, 335-345.
- Timm, N. H. (1975). *Multivariate Analysis with Applications in Education and Psychology*.

A Justification of Rotation in Principal Component Analysis: Projective viewpoint of PCA

Wadsworth Publishing Company.

Yanai, H., Shigemasu, K., Mayekawa, S. and Ichikawa, M. (1990). *Factor Analysis: Theory and methods*. Asakura-Shoten (in Japanese).

- 
- 1) This work is based on my two manuscript, "A Justification of Rotation in Principal Component Analysis" (2006, Apr., distributed as a pdf file attached to an e-mail) and "On Rotations in Hotelling PCA: Reformulation of Hotelling PCA by projection" (2006, Jun. distributed as a pdf file attached to an e-mail).
  - 2) K. Adachi referred to ten Berge and Kiers (1996) in his comment on my manuscript (Apr. 2006).