

On the Special Type of Principal Component Analysis¹⁾: It should belong to factor analysis

ある種の主成分分析について：因子分析と考えられる場合

OKAMOTO, Yasuharu

岡本安晴

[Abstract]

There are two types of criteria used in principal components analysis (PCA), (1) criteria of one by one extraction of components and (2) the join loss criterion. There are also two types of variables, observed and unobserved. In this paper, the two types of criteria are connected respectively with the two types of variables, and distinctive features of the join loss criterion are discussed. Both types of criteria are proposed to extract the major part of the information in the data. But, one of distinctive features of criterion (2) against criterion (1) is that criterion (2) is set for latent variables, while criterion (1) is set for linear combinations of the observed variables. It is also pointed out that the solution of criterion (2) can be rotated, although rotation of the solution of criterion (1) is inhibited. Rotation is a common procedure in factor analysis. In this paper, based on these features, criterion (2) is proposed to be considered as that for factor analysis. Positive evaluation of correlated errors, which result from criterion (2), is also discussed.

1. Introduction

In some cases, terms such as composite, latent and synthetic variables are used synonymously (e.g. Thomson, 2004, p. 9). But, this synonymousness may obscure conceptual distinction between observed and unobserved variables. In this paper, I emphasize this conceptual distinction, and then propose a new interpretation about a certain type of principal component analysis (PCA), which should be distinguished from the other type of PCA with respect to observed and unobserved distinction and classified as factor analysis (FA).

Latent variables, which are not directly observable and should be inferred, play important role in science, because science is about isolating and understanding latent constructs (Thompson, 2004, p. 159). Synonymous terms with latent variables are unobserved or unmeasured variables and factors (Bollen, 1989, p. 11). Latent variables or factors manifest their effects through observed variables, and their relations are represented, e.g., by measurement equations in structural equation models (SEM). Factors in FA are hypothetical components

(Mulaik, 1972, p. 96) and unobservable ones (Morrison, 1976, p. 303). Factors in FA are inferred using equations, which are special cases of SEM.

Regarding the distinction of observed and unobserved quantities, methods of PCA can be classified into two groups. In the first group, principal components are constructed as weighted combinations of observed variables. But the second type of PCA sets components as unobservable variables, which should be inferred and is characterized as minimizing eq. (1) to be presented and explained in the next section. In the next section, the above distinction as to the types of PCA will be discussed.

2. New Interpretation

The first type of PCA uses one by one extraction procedures. In one by one extraction, for data $\mathbf{X} = (\mathbf{x}_1 \cdots \mathbf{x}_p)$, where \mathbf{X} is $N \times p$ matrix and (i, j) element x_{ij} of \mathbf{X} represents the observed value of variable j for subject i , we construct a linear combination of the observed variables $\mathbf{c} = w_1 \mathbf{x}_1 + \cdots + w_p \mathbf{x}_p$, which is called a component. Each component is extracted successively for some specified criterion, e. g., maximization of variance (Morrison, 1976), minimization of difference (Lebart, Marineau, & Warwick, 1984), maximization of sum of covariances (Yanai, Shigemasu, Maekawa, & Ichikawa, 1990), and so on. Successive components are extracted under the orthogonality condition. Hence, in general, directions of the components are uniquely determined except for reflection, so rotation of the extracted components is inhibited.

The second type of PCA uses the criterion called the join loss criterion (Gifi, 1990, p. 152),

$$G_j(\mathbf{X}, \mathbf{F}, \mathbf{A}) = tr[(\mathbf{X} - \mathbf{FA})(\mathbf{X} - \mathbf{FA})'] \quad (1)$$

and extracts the components simultaneously. In Eq. (1), \mathbf{X} is a data matrix, $\mathbf{F} = (\mathbf{f}_1 \cdots \mathbf{f}_q)$ a matrix of components, which should be called more appropriately factors according to the reason discussed in this paper, and \mathbf{A} a matrix of weights⁽ⁱⁱ⁾. When \mathbf{F} and \mathbf{A} minimize $G_j(\mathbf{X}, \mathbf{F}, \mathbf{A})$, for any nonsingular matrix \mathbf{T} , \mathbf{FT}^{-1} and \mathbf{AT}' also minimize it, because

$$\mathbf{FA}' = (\mathbf{FT}^{-1})(\mathbf{AT}')'$$

This means that rotation is allowed in contrast to the first type of criterion, i.e. one by one extraction.

To show that Eq. (1) can be considered to be a model of FA, put

$$\mathbf{X} = \mathbf{FA}' + \mathbf{E}, \quad (2)$$

The join loss criterion, minimization of $G_j(\mathbf{X}, \mathbf{F}, \mathbf{A})$, means minimization of squared sum of errors in Eq. (2), the elements of \mathbf{E} . That is, a solution of the join loss criterion is a least squares estimate for the model Eq. (2), which is the same form of models of factor analysis.

For example, McLachlan and Krishnan (1997) explain the model of factor analysis in the same form of Eq. (2) except for the constant term, although their criterion is that of maximum likelihood and EM algorithm is used. The join loss criterion and the factor analysis by EM algorithm differ only in methods of parameter estimation, and are, as models for representation of data, essentially identical in form and concept. This fact indicates factor analytic aspect of the join loss criterion.

Here, I emphasize conceptual difference between factors $\mathbf{F} = (\mathbf{f}_1 \cdots \mathbf{f}_q)$ and components $\mathbf{C} = (\mathbf{c}_1 \cdots \mathbf{c}_q)$, which result respectively from the join loss criterion and from one by one extraction. In one by one extraction, components $\mathbf{c}_s; s=1, \dots, q$ are constructed as linear sum of the observed variables and optimum linear sums are searched successively. That is,

$$\mathbf{c}_s = w_1 \mathbf{x}_1 + \cdots + w_p \mathbf{x}_p \in S(\mathbf{x}_1 \cdots \mathbf{x}_p),$$

where $S(\mathbf{x}_1 \cdots \mathbf{x}_p)$ is the linear space spanned by $\mathbf{x}_1, \dots, \mathbf{x}_p$.

On the other hand, when the join loss criterion Eq. (1) is set, the solution \mathbf{F} in Eq. (1) is more appropriate to be considered as a matrix of latent variables, because in setting Eq. (1), \mathbf{F} is not assumed to be linear combination of the observed variables. That is, latent variables \mathbf{f}_i in eq. (1) are not constrained to belong to $S(\mathbf{x}_1 \cdots \mathbf{x}_p)$. Conceptually, it is allowed that

$$\mathbf{f}_i \notin S(\mathbf{x}_1 \cdots \mathbf{x}_p).$$

Only estimation of the values of these latent variables is given as linear sum of the observed variables. It is reasonable that the estimator is given as a function of the observed values. The form of the join loss criterion sets the solution conceptually as latent variables, and the actual solution (numerical estimate of the values of the latent variables) is given as linear sum of the observed variables. The common form of mathematical expressions of the solutions for the one by one extraction and for the join loss criterion should not obscure the conceptual distinction between variables constructed as linear combination of the observed variables and latent variables (factors), although estimation of the latter is given by the linear combination. Latent variables, which are to be estimated, and estimators, which estimate latent variables, are conceptually different.

Although some authors treat PCA as one of methods of FA (e.g. Thompson, 2004), others consider that PCA is essentially different from FA (e.g. Yanai et al., 1990, p.7). When we consider that PCA is a different method from FA, we should talk about only the first type of PCA, in which one by one extraction methods are used and components are constructed as linear weighted compositions of observed variables. The join loss criterion, which is proposed for PCA, should be considered to belong to FA. In the following, aspects of the join loss criterion considered as that of factor analysis are discussed.

3. What does the New Interpretation Produce

In this section, two distinctive aspects of the join loss criterion, which are desirable in some sense as a model of factor analysis (FA), are discussed. One of them is concerned with Heywood cases and the other with correlation between residual errors.

3.1. No Heywood Cases

In some cases of factor analysis, it occurs that communality becomes larger than 1.0, which is well known as a Heywood case. In case of the join loss criterion, communality never becomes larger than 1.0. This fact is obvious, because solutions of the join loss criterion are estimates of latent variables which are numerically identical to those of principal component analysis with one by one extraction methods. In appendix A, this fact is formally proven. Not being anxious about Heywood cases, we can always analyze data by factor analysis with the join loss criterion.

3.2. Correlation between Residual Errors

When we adopt the join loss criterion, residual errors are correlated. In standard models of factor analysis, the residual errors are assumed to be uncorrelated (e.g. Harman, 1976). But the assumption of uncorrelated residual errors seems to be technical. If we allow residual errors to be correlated, the number of free parameters of the model of FA becomes larger than the degree of freedom of the data, i.e. the correlation matrix in this case. By the constraint that the residual errors are independent, we can reduce the number of free parameters so that it does not exceed the degree of freedom of the data and the model becomes identifiable except rotation.

Interpretation of specific factors by Harman (1976, p. 19) suggests possibility of correlation between residual errors. Haebara (2002a) decomposes the residual error term of item X into two portions, e_x and e'_x ; e_x represents the internal state of the subject on the day when the item is answered, and varies between days. e'_x varies within the day and represents instantaneous fluctuation of the subject. Haebara proposes that portions e_x of items should be considered to be correlated between items X s. Kano (2002a, b) discusses utilities of nonzero covariances of residual errors in structural equation modeling. Considering these discussions, it seems natural to allow residual errors in factor analysis to be correlated. Mathematically, error terms of E can be considered more appropriately as residuals and the residuals may include minor factors, each of which influences few variables and is neglected and not represented as a common factor in the analysis. Exact expression of covariances between the residual errors for the join loss criterion is given in appendix B.

4. Summary

Principal component analysis (PCA) and factor analysis (FA) are distinguished as conceptually different models by some authors (e.g. Yanai et al., 1990). PCA with one by one extraction method constructs the optimal linear combinations of the observed variables, and FA seeks latent variables called factors. The join loss criterion is proposed as a criterion of PCA (Gifi, 1990), which extracts maximum information of the data. In this paper, some aspects of the join loss criterion, which are contrasted to ones of traditional one-by-one extraction criteria, are pointed out and new interpretation that the join loss criterion should be considered to be that of FA is proposed. The contrasted aspects are (1) latent character of solutions of the join loss criterion, and (2) justification of rotation of factors obtained for the join loss criterion. In eq. (1) (the join loss criterion), components f_i should be considered to be latent variables and are not constrained to be in the subspace spanned by the data matrix X . In one by one extraction methods, components are constructed as composite variables of observed ones, which are constrained within the subspace spanned by X . As to rotation, in case of one by one extraction of components of PCA, those components are sought among linear combination of the observed variables under orthogonality constraint, and rotation of the obtained components are inhibited. But, eq. (1) allows rotation of factors F .

Positive aspects of the join loss criterion, when it is considered as that of FA, are pointed out. As shown in appendix A, the join loss criterion never brings about Heywood cases, by which standard methods of FA sometimes fail. The other aspect is correlation between residual errors (appendix B), which is considered to be natural as a model of human performance.

References

- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons, Inc.
- Gifi, A. (1990). *Nonlinear multivariate analysis*. John Wiley & Sons, Inc. (Reprinted with corrections 1991)
- Haebara, T. (2002). On the relevance of goodness of model fit to the purpose of research – Should we keep the number of indicator variables small? - *The Japanese Journal of Behaviormetrics*, 29, 160-166 (in Japanese).
- Harman, H. H. (1976). *Modern factor analysis (Third edition revised)*. The University of Chicago Press.
- Kano, Y. (2002a). Does structural equation modeling outperform traditional factor analysis, analysis of variance and path analysis? *The Japanese Journal of Behaviormetrics*, 29, 138-159 (in Japanese).
- Kano, Y. (2002b). Rejoinder: Use of error covariances and the role of specific factors. *The*

Japanese Journal of Behaviormetrics, 29, 182-197 (in Japanese).

Lebart, L., Morineau, A. and Warwick, K. M. (1984). *Multivariate descriptive statistical analysis:*

Correspondence analysis and related techniques for large matrices. John Wiley & Sons.

McLachlan, G. J. and Krishnan, T. (1997). *The EM algorithm and extensions.* John Wiley & Sons, Inc.

Morrison, D. F. (1976). *Multivariate statistical methods*, 2nd ed. McGraw-Hill Book Company.

Mulaik, S. A. (1972). *The foundations of factor analysis.* McGraw-Hill Book Company.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* American Psychological Association.

Yanai, H., Shigemasu, K., Maekawa, S., and Ichikawa, M. (1990). *Factor analysis: Theory and procedure.* Asakura-Shoten (in Japanese).

Appendix A: No Heywood Cases

Let h_j^2 be communality of variable j , i.e.

$$\begin{bmatrix} h_1^2 & & 0 \\ & \ddots & \\ 0 & & h_p^2 \end{bmatrix} = \text{diag}\left\{\frac{1}{N}(\mathbf{FA}')'(\mathbf{FA}')\right\} = \text{diag}(\mathbf{AA}'),$$

where orthonormal condition $\frac{1}{N}\mathbf{F}'\mathbf{F} = \mathbf{I}$ is used.

In the above equation, we can put

$$\mathbf{A}' = \frac{1}{\sqrt{N}}\mathbf{\Lambda}\mathbf{Q}', \quad \mathbf{F} = \sqrt{N}\mathbf{P}$$

where N is the number of rows of \mathbf{X} , $\mathbf{\Lambda}$ the appropriate diagonal matrix of singular values of \mathbf{X} , \mathbf{Q} the corresponding matrix of singular vectors (see Eqs. (A.4) and (A.5)).

Hence, we have

$$\mathbf{AA}' = \left(\frac{1}{\sqrt{N}}\mathbf{\Lambda}\mathbf{Q}'\right)' \left(\frac{1}{\sqrt{N}}\mathbf{\Lambda}\mathbf{Q}'\right) = \frac{1}{N}\mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}' \quad (\text{A.1})$$

Put

$$\mathbf{Q} = [\mathbf{q}_1 \quad \cdots \quad \mathbf{q}_q], \quad \mathbf{q}_k = \begin{bmatrix} q_{1k} \\ \vdots \\ q_{pk} \end{bmatrix},$$

then we have

$$\begin{aligned} \frac{1}{N} \mathbf{Q} \Lambda^2 \mathbf{Q}' &= \frac{1}{N} (\lambda_1^2 \mathbf{q}_1 \mathbf{q}_1' + \dots + \lambda_q^2 \mathbf{q}_q \mathbf{q}_q') \\ &= \frac{1}{N} \begin{bmatrix} \sum_{k=1}^q \lambda_k^2 q_{1k}^2 & & * \\ & \ddots & \\ * & & \sum_{k=1}^q \lambda_k^2 q_{pk}^2 \end{bmatrix}, \end{aligned} \quad (\text{A.2})$$

By Eqs. (A.1) and (A.2), we have

$$h_i^2 = \frac{1}{N} \sum_{k=1}^q \lambda_k^2 q_{ik}^2 \quad (\text{A.3})$$

Now, put the singular decomposition of \mathbf{X} as follows;

$$\mathbf{X} = \mathbf{P}_0 \Lambda_0 \mathbf{Q}_0' \quad (\text{A.4})$$

where

$$\begin{aligned} \mathbf{P}_0 &= [\mathbf{P} \ \mathbf{P}_E], \quad \Lambda_0 = \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda_E \end{bmatrix}, \quad \mathbf{Q}_0 = [\mathbf{Q} \ \mathbf{Q}_E] \quad (\text{A.5}) \\ \mathbf{P} &= [\mathbf{p}_1 \ \dots \ \mathbf{p}_q], \quad \mathbf{P}_E = [\mathbf{p}_{q+1} \ \dots \ \mathbf{p}_p] \\ \Lambda &= \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_q \end{bmatrix}, \quad \Lambda_E = \begin{bmatrix} \lambda_{q+1} & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix} \\ \mathbf{Q} &= [\mathbf{q}_1 \ \dots \ \mathbf{q}_q], \quad \mathbf{Q}_E = [\mathbf{q}_{q+1} \ \dots \ \mathbf{q}_p] \end{aligned}$$

Hence we have

$$\frac{1}{N} \mathbf{X}' \mathbf{X} = \frac{1}{N} \begin{bmatrix} \sum_{k=1}^q \lambda_k^2 q_{1k}^2 & & * \\ & \ddots & \\ * & & \sum_{k=1}^q \lambda_k^2 q_{pk}^2 \end{bmatrix} \quad (\text{A.6})$$

By Eqs. (A.3) and (A.6), we have

$$h_i^2 = \frac{1}{N} \sum_{k=1}^q \lambda_k^2 q_{ik}^2 \leq \frac{1}{N} \sum_{k=1}^q \lambda_k^2 q_{jk}^2 = \frac{1}{N} \sum_{i=1}^N x_{ij}^2 = 1,$$

where \mathbf{X} is assumed to be a matrix of standardized scores.

Appendix B: Covariances between Residual Errors

By Eq. (A.4), we have

$$\mathbf{E} = \mathbf{X} - \mathbf{FA}$$

$$\begin{aligned}
 &= [\mathbf{P} \quad \mathbf{P}_E] \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda_E \end{bmatrix} \begin{bmatrix} \mathbf{Q}' \\ \mathbf{Q}'_E \end{bmatrix} - \mathbf{P} \Lambda \mathbf{Q}' \\
 &= \mathbf{P}_E \Lambda_E \mathbf{Q}'_E
 \end{aligned}$$

Hence we have

$$\begin{aligned}
 \frac{1}{N} \mathbf{E}' \mathbf{E} &= \frac{1}{N} \mathbf{Q}_E \Lambda_E \mathbf{P}'_E \mathbf{P}_E \Lambda_E \mathbf{Q}'_E \\
 &= \frac{1}{N} \mathbf{Q}_E \Lambda_E^2 \mathbf{Q}'_E \tag{B.1}
 \end{aligned}$$

In general, matrix Eq. (B.1) is not diagonal, which means that the residual errors are not independent.

-
- i) This work is based on the paper presented at the 31st annual meeting of the Behaviormetric Society of Japan in 2003.
 - ii) Gifi (1990) presents the join loss criterion in the following format: $\sigma_f(\mathbf{X}, \mathbf{A}) \equiv SSQ(\mathbf{H} - \mathbf{X}\mathbf{A}')$, where \mathbf{X} is a matrix of components, \mathbf{A} a matrix of weights, and \mathbf{H} a data matrix. The notation in Eq. (1) follows that of FA.