# Alternating least squares in nonlinear principal components

Masahiro Kuroda,[1*] Yuichi Mori,[2] Iizuka Masaya[3] and Michio Sakakihara[4]

Principal components analysis (PCA) is probably the most popular descriptive multivariate method for analyzing quantitative data with ratio and interval scale measures. When applying PCA to nominal and ordinal data, the data are processed by a method such as optimal scaling, which nonlinearly transforms nominal and ordinal data into quantitative data. Therefore, PCA with optimal scaling is called nonlinear PCA. Nonlinear PCA reveals nonlinear relationships among variables with different measurement levels and therefore presents a more flexible alternative to ordinary PCA. The alternating least squares algorithm is utilized for nonlinear PCA. The algorithm alternates between optimal scaling for quantifying nominal and ordinal data and ordinary PCA for analyzing optimally scaled data. This article discusses two nonlinear PCA algorithms, namely, PRINCIPALS and PRINCALS. © 2013 Wiley Periodicals, Inc.

## INTRODUCTION

Principal components analysis (PCA) is probably the most popular descriptive multivariate method for analyzing quantitative data with ratio and interval scale measures. However, since PCA assumes that all variables are measured by ratio and interval scales, it is not directly applicable to qualitative data such as nominal and ordinal data. This paper discusses an extension of PCA that handles mixed quantitative and qualitative data. In applying PCA to mixed data, the qualitative data must be quantified. Optimal scaling is a quantification technique that optimally assigns numerical values to qualitative scales within the restrictions of the measurement characteristics of

*Correspondence to: kuroda@soci.ous.ac.jp

[1]Department of Socio-Information, Faculty of Informatics, Okayama University of Science, Okayama, Japan

[2]Department of Socio-Information, Faculty of Informatics, Okayama University of Science, Okayama, Japan

[3]Graduate School of Environmental Science, Okayama University, Okayama, Japan

[4]Department of Information Science, Faculty of Informatics, Okayama University of Science, Okayama, Japan

the qualitative variables.[15] The qualitative data are then nonlinearly transformed into quantitative data. Thus, PCA with optimal scaling is called nonlinear PCA.[5] Nonlinear PCA reveals nonlinear relationships among variables with different measurement levels and therefore presents a more flexible alternative to ordinary PCA.

Nonlinear PCA utilizes the alternating least squares (ALS) algorithm. This algorithm divides the parameters of nonlinear PCA into model parameters and data parameters, and finds their least squares estimates by updating each of them in turn. The algorithm consists of two steps: first, the data parameters are estimated using optimal scaling, and second, the model parameters are computed by PCA of the optimally scaled data obtained in the previous step. We discuss the two ALS algorithms typically employed in nonlinear PCA; PRINCIPALS[16] and PRINCALS.[5]

## QUANTIFICATION OF QUALITATIVE DATA

Let $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_p)$ be an $n \times p$ matrix of observations on $n$ objects and $p$ variables. To quantify

$\mathbf{X}_j$ of the qualitative variable $j$ with $K_j$ categories, the vector is coded by using an $n \times K_j$ indicator matrix

$$\mathbf{G}_j = (g_{jik}) = \begin{pmatrix} g_{j11} & \cdots & g_{j1K_j} \\ \vdots & \vdots & \vdots \\ g_{jn1} & \cdots & g_{jnK_j} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{j1} \cdots \mathbf{g}_{jK_j} \end{pmatrix},$$

where

$$g_{jik} = \begin{cases} 1 & \text{if object } i \text{ belongs to category } k, \\ 0 & \text{if object } i \text{ belongs to some} \\ & \text{other category } k' (\neq k). \end{cases}$$

For example, given

$$\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2 \ \mathbf{X}_3) = \begin{pmatrix} 3 & 1 & 4 \\ 1 & 2 & 3 \\ 2 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 1 & 1 \end{pmatrix},$$

the indicator matrix of $\mathbf{X}$ is

$$\begin{aligned} \mathbf{G} &= (\mathbf{G}_1 \ \mathbf{G}_2 \ \mathbf{G}_3) \\ &= \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Optimal scaling finds $K_j \times 1$ category quantifications $\mathbf{Y}_j$ under the restrictions imposed by the measurement level of variable $j$ and transforms $\mathbf{X}_j$ into an optimally scaled vector $\mathbf{X}_j^* = \mathbf{G}_j \mathbf{Y}_j$. For nominal scale data, the quantification is unrestricted. Ordinal scale data are quantified by monotone regression on account of the monotonicity restriction.

## NONLINEAR PCA

For a data matrix $\mathbf{X}$ of $n$ objects by $p$ quantitative variables, PCA postulates that $\mathbf{X}$ is approximated by the bilinear form

$$\widehat{\mathbf{X}} = \mathbf{Z}\mathbf{A}^\top,$$

where $\mathbf{Z}$ is an $n \times r$ matrix of $n$ component scores on $r$ $(1 \leq r \leq p)$ components and $\mathbf{A}$ is a $p \times r$ matrix of $p$ component loadings on $r$ components. PCA is

formulated in terms of the loss function

$$\begin{aligned} \sigma(\mathbf{Z}, \mathbf{A}) &= \text{tr}\left(\mathbf{X} - \widehat{\mathbf{X}}\right)^\top \left(\mathbf{X} - \widehat{\mathbf{X}}\right) \\ &= \text{tr}\left(\mathbf{X} - \mathbf{Z}\mathbf{A}^\top\right)^\top \left(\mathbf{X} - \mathbf{Z}\mathbf{A}^\top\right). \quad (1) \end{aligned}$$

The minimum of the loss function (Eq. (1)) over $\mathbf{Z}$ and $\mathbf{A}$ is found by singular value decomposition of $\mathbf{X}$ or by eigen-decomposition of $\mathbf{X}^\top\mathbf{X}$.

To handle any combination of quantitative and qualitative data, nonlinear PCA requires the optimal quantification of qualitative data for obtaining $\mathbf{X}^*$ in addition to estimating $\mathbf{Z}$ and $\mathbf{A}$, and thus the loss function has to be minimized over $\mathbf{Z}$, $\mathbf{A}$ and $\mathbf{X}^*$.

In nonlinear PCA, we define two types of loss functions; the first derived from a low-rank approximation of $\mathbf{X}^*$ extended to Eq. (1) and the second obtained by homogeneity analysis with restrictions. We show the loss functions in the following subsections. The ALS algorithm is used to minimize the loss functions.

## Low-Rank Matrix Approximation

In the presence of qualitative variables in $\mathbf{X}$, the loss function (Eq. (1)) is expressed as

$$\begin{aligned} \sigma_L(\mathbf{Z}, \mathbf{A}, \mathbf{X}^*) &= \text{tr}\left(\mathbf{X}^* - \widehat{\mathbf{X}}\right)^\top \left(\mathbf{X}^* - \widehat{\mathbf{X}}\right) \\ &= \text{tr}\left(\mathbf{X}^* - \mathbf{Z}\mathbf{A}^\top\right)^\top \left(\mathbf{X}^* - \mathbf{Z}\mathbf{A}^\top\right) \quad (2) \end{aligned}$$

and is minimized over $\mathbf{Z}$, $\mathbf{A}$, and $\mathbf{X}^*$ under the restrictions

$$\mathbf{X}^{*\top}\mathbf{1}_n = \mathbf{0}_p \quad \text{and} \quad \text{diag}\left[\frac{\mathbf{X}^{*\top}\mathbf{X}^*}{n}\right] = \mathbf{I}_p, \quad (3)$$

where $\mathbf{1}_n$ and $\mathbf{0}_p$ are vectors of ones and zeros of length $n$ and $p$, respectively, and $\mathbf{I}_p$ is the $p \times p$ identity matrix. Since optimal scaling for $\mathbf{X}^*$ can be performed separately and independently for each variable, the loss function (Eq. (2)) can be rewritten as

$$\begin{aligned} \sigma_L(\mathbf{Z}, \mathbf{A}, \mathbf{X}^*) &= \sum_{j=1}^p \left(\mathbf{X}_j^* - \mathbf{Z}\mathbf{A}_j^\top\right)^\top \left(\mathbf{X}_j^* - \mathbf{Z}\mathbf{A}_j^\top\right) \\ &= \sum_{j=1}^p \sigma_{Lj}\left(\mathbf{Z}, \mathbf{A}_j, \mathbf{X}_j^*\right). \quad (4) \end{aligned}$$

Thus, we can minimize $\sigma(\mathbf{Z}, \mathbf{A}, \mathbf{X}^*)$ by independently minimizing each $\sigma_{Lj}\left(\mathbf{Z}, \mathbf{A}_j, \mathbf{X}_j^*\right)$ under the measurement restrictions on variable $j$.

## Homogeneity Analysis

Homogeneity analysis is a method to maximize the homogeneity of several categorical variables and determine quantifications of the categories of each variable such that the homogeneity is maximized.[5] Let $\mathbf{Z}$ be $n \times r$ object scores (component scores) and $\mathbf{W}_j$ be $K_j \times r$ category quantifications of variable $j$ ($j = 1, \ldots, p$). Homogeneity analysis finds the $\mathbf{Z}$ and $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_p)$ that well matches $\mathbf{G}_j \mathbf{W}_j$ to $\mathbf{Z}$ over $j$. The loss function measuring the departure from homogeneity is given by

$$
\sigma_H(\mathbf{Z}, \mathbf{W}) = \sum_{j=1}^{p} \mathrm{tr}\, (\mathbf{Z} - \mathbf{G}_j \mathbf{W}_j)^\top (\mathbf{Z} - \mathbf{G}_j \mathbf{W}_j),
$$

$$
= \sum_{j=1}^{p} \sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j) \tag{5}
$$

and is minimized over $\mathbf{Z}$ and $\mathbf{W}$ under the restrictions

$$
\mathbf{Z}^\top \mathbf{1}_n = \mathbf{0}_r \quad \text{and} \quad \mathbf{Z}^\top \mathbf{Z} = n \mathbf{I}_r. \tag{6}
$$

The minimum of $\sigma_H(\mathbf{Z}, \mathbf{W})$ is obtained by separately minimizing each $\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j)$.

Gifi defines nonlinear PCA as homogeneity analysis imposing *rank-one restrictions*[5]

$$
\mathbf{W}_j = \mathbf{Y}_j \mathbf{A}_j, \tag{7}
$$

where $\mathbf{Y}_j$ is a $K_j \times 1$ vector of category quantifications and $\mathbf{A}_j$ is a $1 \times r$ vector of weights (component loadings). Variables on which rank-one restrictions are imposed are called *single variables* and variables without restrictions are *multiple variables*.

To minimize $\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j)$ under rank-one restrictions (Eq. (7)), we first obtain the least squares estimate $\widetilde{\mathbf{W}}_j$ of $\mathbf{W}_j$. For a fixed $\widetilde{\mathbf{W}}_j$, $\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j)$ can be partitioned as

$$
\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j) = \mathrm{tr}\, (\mathbf{Z} - \mathbf{G}_j \mathbf{W}_j)^\top (\mathbf{Z} - \mathbf{G}_j \mathbf{W}_j),
$$

$$
= \mathrm{tr}\, (\mathbf{Z} - \mathbf{G}_j \widetilde{\mathbf{W}})^\top (\mathbf{Z} - \mathbf{G}_j \widetilde{\mathbf{W}}_j),
$$

$$
+ \mathrm{tr}\, (\mathbf{Y}_j \mathbf{A}_j - \widetilde{\mathbf{W}}_j)^\top (\mathbf{G}_j^\top \mathbf{G}_j)(\mathbf{Y}_j \mathbf{A}_j - \widetilde{\mathbf{W}}_j). \tag{8}
$$

We then minimize the second term on the right hand side of Eq. (8) over $\mathbf{Y}_j$ and $\mathbf{A}_j$ under the restrictions imposed by the measurement level of variable $j$.

Note that, for all variables that are single, minimizing the loss function (Eq. (5)) is equivalent to minimizing the loss function (Eq. (2)) under restrictions (Eq. (3) and Eq. (6)).

## THE ALS ALGORITHM FOR NONLINEAR PCA

The ALS algorithm is a computational algorithm for minimizing a least squares loss function and is applied to solve various matrix optimization problems.[7] When the closed-form solutions of parameter matrices for minimizing the least squares loss function are unavailable, the ALS algorithm finds the least squares estimates of these parameters by updating each of the parameter matrices in turn, keeping the others fixed.

Suppose we wish to minimize a loss function $\sigma(\theta_1, \theta_2, \theta_3)$ over the parameter matrices $\theta_1$, $\theta_2$, and $\theta_3$. Let $\theta^{(t)}$ be the $t$-th estimate of $\theta$. The ALS algorithm updates the estimates of $\theta_1$, $\theta_2$, and $\theta_3$ by solving the least squares problem for each parameter:

$$
\theta_1^{(t+1)} = \arg \min_{\theta_1} \sigma \left( \theta_1, \theta_2^{(t)}, \theta_3^{(t)} \right),
$$

$$
\theta_2^{(t+1)} = \arg \min_{\theta_2} \sigma \left( \theta_1^{(t+1)}, \theta_2, \theta_3^{(t)} \right),
$$

$$
\theta_3^{(t+1)} = \arg \min_{\theta_3} \sigma \left( \theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3 \right).
$$

If each update of the ALS algorithm improves the value of the loss function and if the function is bounded, the function will be locally minimized over the entire set of parameters.[8]

## PRINCIPALS

The ALS algorithm PRINCIPALS, developed by Young et al.,[16] minimizes the loss function (Eq. (4)). The algorithm accepts single nominal, ordinal, and numerical variables, and alternates between two estimation steps. The first step estimates the model parameters $\mathbf{Z}$ and $\mathbf{A}$ for ordinary PCA, and the second estimates the data parameter $\mathbf{X}^*$ for optimally scaled data.

For the initialization of PRINCIPALS, the initial data $\mathbf{X}^{*(0)}$ are determined under the measurement restrictions for each variable and are then standardized to satisfy restriction (Eq. (3)). The observed data $\mathbf{X}$ may be used as $\mathbf{X}^{*(0)}$ after centering and normalizing each column of $\mathbf{X}$ under restriction (Eq. (3)). Given the initial data $\mathbf{X}^{*(0)}$, PRINCIPALS iterates the following two steps:

- *Model estimation step*: Obtain $\mathbf{A}^{(t+1)}$ and $\mathbf{Z}^{(t+1)}$ from

$$
\mathbf{A}^{(t+1)} = \mathbf{X}^{*(t)\top} \mathbf{Z}^{(t)} \left( \mathbf{Z}^{(t)\top} \mathbf{Z}^{(t)} \right)^+,
$$

$$
\mathbf{Z}^{(t+1)} = \mathbf{X}^{*(t)} \mathbf{A}^{(t+1)} \left( \mathbf{A}^{(t+1)\top} \mathbf{A}^{(t+1)} \right)^+,
$$

where the superscript $+$ indicates the Moore-Penrose inverse of a matrix.

- *Optimal scaling step*: Obtain $\mathbf{X}^{*\,(t+1)}$ by separately estimating $\mathbf{X}_j^*$ for each $j$. Compute $\widehat{\mathbf{X}}_j^{(t+1)} = \mathbf{Z}^{(t+1)}\mathbf{A}_j^{(t+1)\top}$. Compute $\mathbf{Y}_j^{(t+1)}$ as

$$\mathbf{Y}_j^{(t+1)} = \left(\mathbf{G}_j^\top \mathbf{G}_j\right)^{-1}\mathbf{G}_j^\top\widehat{\mathbf{X}}_j^{(t+1)}.$$

Recompute $\mathbf{Y}_j^{(t+1)}$ by monotone and linear regression on ordinal and numerical variables, respectively. Compute $\mathbf{X}_j^{*(t+1)} = \mathbf{G}_j\mathbf{Y}_j^{(t+1)}$ and scale $\mathbf{X}_j^{*(t+1)}$ by centering and normalizing.

In the *Model estimation step*, $\mathbf{A}^{(t+1)}$ and $\mathbf{Z}^{(t+1)}$ are found by solving the singular value decomposition of $\mathbf{X}^{*\,(t)}$ or the eigen-decomposition of $\mathbf{X}^{*\,(t)\top}\mathbf{X}^{*\,(t)}/n$.[16]

**Singular value decomposition of $\mathbf{X}^{*(t)}$:** From the Eckart-Young decomposition theorem,[1] we obtain

$$\mathbf{Z}^{(t+1)} = \mathbf{U}_r, \quad \mathbf{A}^{(t+1)} = \mathbf{V}_r\mathbf{D}_r^{1/2},$$

where $\mathbf{U}_r$ is an $n \times r$ matrix of the $r$ normalized eigenvectors of $\mathbf{X}^{*\,(t)\top}\mathbf{X}^{*\,(t)}$ corresponding to the $r$ dominant eigenvectors, $\mathbf{V}_r$ is a $p \times r$ matrix of the $r$ normalized eigenvectors of $\mathbf{X}^{*\,(t)}\mathbf{X}^{*\,(t)\top}$ corresponding to the $r$ dominant eigenvectors, and $\mathbf{D}_r$ is an $r \times r$ diagonal matrix of eigenvalues of $\mathbf{X}^{*\,(t)\top}\mathbf{X}^{*\,(t)}$.

**Eigen-decomposition of $\mathbf{X}^{*\,(t)\top}\mathbf{X}^{*\,(t)}/n$:** By solving

$$\left[\frac{\mathbf{X}^{*(t)\top}\mathbf{X}^{*(t)}}{n}\right]\mathbf{A} = \mathbf{A}\mathbf{D}_r$$

subject to $\mathbf{A}^\top\mathbf{A} = \mathbf{I}_r$, we obtain $\mathbf{A}^{(t+1)}$ and compute $\mathbf{Z}^{(t+1)} = \mathbf{X}^{*\,(t)}\mathbf{A}^{(t+1)}$.

## PRINCALS

The ALS algorithm PRINCALS, developed by Gifi,[5] can handle multiple nominal variables in addition to the single nominal, ordinal, and numerical variables accepted in PRINCIPALS. We denote the set of multiple variables by $\mathcal{J}_M$ and the set of single variables having single nominal and ordinal scales and numerical measurements by $\mathcal{J}_S$. From Eqs. (5) and (8), the loss function to be minimized by PRINCALS is given by

$$\sigma_H(\mathbf{Z}, \mathbf{W}) = \sum_{j\in\mathcal{J}_M}\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j) + \sum_{j\in\mathcal{J}_S}\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j).$$

For the initialization of PRINCALS, we determine the initial values of $\mathbf{Z}$ and $\mathbf{W}$. The matrix $\mathbf{Z}^{(0)}$ is initialized with random numbers under restriction (Eq. (6)), and $\mathbf{W}_j^{(0)}$ is obtained as $\mathbf{W}_j^{*(0)} = \mathbf{G}_j\left(\mathbf{G}_j^\top\mathbf{G}_j\right)^{-1}\mathbf{G}_j^\top\mathbf{Z}^{(0)}$. For each variable $j \in \mathcal{J}_S$, $\mathbf{Y}_j^{*(0)}$ is defined as the first $K_j$

successive integers under the normalization restriction. The vector $\mathbf{A}_j$ is initialized as $\mathbf{A}_j^{(0)} = \mathbf{Z}^{(0)\top}\mathbf{G}_j\mathbf{Y}_j^{*(0)}$, and rescaled to unit length. Given these initial values, PRINCALS iterates the following steps[10]:

- *Estimation of category quantifications*: Compute $\mathbf{W}_j^{(t+1)}$ for $j = 1, \ldots, p$ as

$$\mathbf{W}_j^{(t+1)} = (\mathbf{G}_j^\top\mathbf{G}_j)^{-1}\mathbf{G}_j^\top\mathbf{Z}^{(t)}.$$

- For the multiple variables in $\mathcal{J}_M$, set $\mathbf{W}_j^{(t+1)}$ to the estimate of multiple category quantifications.

- For the single variables in $\mathcal{J}_S$, compute $\mathbf{A}_j^{(t+1)}$ as

$$\mathbf{A}_j^{(t+1)\top} = \mathbf{W}_j^{(t+1)\top}\left(\mathbf{G}_j^\top\mathbf{G}_j\right)\mathbf{Y}_j^{(t)}/\mathbf{Y}_j^{(t)\top}\left(\mathbf{G}_j^\top\mathbf{G}_j\right)\mathbf{Y}_j^{(t)}$$

and $\mathbf{Y}_j^{(t+1)}$ as

$$\mathbf{Y}_j^{(t+1)} = \mathbf{W}_j^{(t+1)}\mathbf{A}_j^{(t+1)\top}/\mathbf{A}_j^{(t+1)}\mathbf{A}_j^{(t+1)\top}.$$

Recompute $\mathbf{Y}_j^{(t+1)}$ by monotone and linear regression on ordinal and numerical variables, respectively. Update $\mathbf{W}_j^{(t+1)} = \mathbf{Y}_j^{(t+1)}\mathbf{A}_j^{(t+1)}$ for ordinal and numerical variables.

- *Update of object scores*: Compute $\mathbf{Z}^{(t+1)}$ by

$$\mathbf{Z}^{(t+1)} = \frac{1}{p}\sum_{j=1}^{p}\mathbf{G}_j\mathbf{W}_j^{(t+1)}. \qquad (9)$$

Column-wise center and orthonormalize $\mathbf{Z}^{(t+1)}$.

## SOFTWARE

The development of the ALS algorithm with optimal scaling for nonlinear PCA was initiated by Young et al.[16] Their study resulted in PRINCIPALS. SAS provided the procedure PRINQUAL based on PRINCIPALS.[13] The result from PRINQUAL is displayed in the output data set and PRINQUAL produces an iteration history table.[14] A biplot in PCA can be also available for graphical representation of the results.

De Leeuw and his coworkers followed up Young et al.'s study of the ALS algorithm with optimal scaling and developed PRINCALS. This algorithm is implemented in CATPCA in SPSS.[11] CATPCA can display biplots of object scores and quantification scores of variables in a joint space. These plots reveal the relationships among objects and variables.

**TABLE 1** | Guttman-Bell data: Variables and Objects

| Description of the Variables | | | | | |
|---|---|---|---|---|---|
| Variable 1 = *Intensity of interaction* | | 1 = slight; 2 = low; 3 = moderate; 4 = high | | | |
| Variable 2 = *Frequency of interaction* | | 1 = slight; 2 = nonrecurring; 3 = infrequent; 4 = frequent | | | |
| Variable 3 = *Feeling of belong* | | 1 = none; 2 = slight; 3 = variable; 4 = high | | | |
| Variable 4 = *Physical proximity* | | 1 = distant; 2 = close | | | |
| Variable 5 = *Formality of relationship* | | 1 = no relationship; 2 = formal; 3 = informal | | | |

| Classification of the Object | Variable 1 | Variable 2 | Variable 3 | Variable 4 | Variable 5 |
|---|---|---|---|---|---|
| *Crowd* | 1 | 1 | 1 | 2 | 2 |
| *Modern community, Neighborhood* | 2 | 3 | 2 | 2 | 2 |
| *Public* | 1 | 2 | 2 | 1 | 1 |
| *Primary Group* | 4 | 4 | 4 | 2 | 3 |
| *Mob* | 4 | 1 | 4 | 2 | 3 |
| *Secondary Group* | 3 | 3 | 3 | 1 | 2 |
| *Audience* | 2 | 1 | 2 | 2 | 2 |

**TABLE 2** | Category Quantifications

| $W_1$ | | $W_2$ | | $W_3$ | | $W_4$ | | $W_5$ | |
|---|---|---|---|---|---|---|---|---|---|
| −0.12 | 0.03 | 0.05 | 0.17 | −0.12 | 0.03 | −0.16 | −0.18 | −0.13 | 0.02 |
| −0.09 | 0.02 | −0.04 | −0.13 | −0.10 | 0.02 | 0.06 | 0.07 | −0.10 | 0.02 |
| −0.09 | 0.02 | −0.04 | −0.13 | −0.10 | 0.02 | | | 0.26 | −0.05 |
| .26 | −0.06 | −0.04 | −0.13 | 0.26 | −0.06 | | | | |

**TABLE 3** | Single Quantifications

| $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ |
|---|---|---|---|---|
| −0.27 | −0.44 | −0.27 | −0.60 | −0.30 |
| −0.21 | 0.33 | −0.23 | 0.24 | −0.22 |
| −0.21 | 0.33 | −0.23 | | 0.60 |
| .60 | 0.33 | 0.60 | | |

**TABLE 4** | Component Loadings

| $A_1^\top$ | $A_2^\top$ | $A_3^\top$ | $A_4^\top$ | $A_5^\top$ |
|---|---|---|---|---|
| 0.44 | −0.12 | 0.43 | 0.26 | 0.44 |
| −0.10 | −0.40 | −0.11 | 0.29 | −0.08 |

**TABLE 5** | Object Scores

| Classification of the Object | $Z_1$ | $Z_2$ |
|---|---|---|
| *Crowd* | −0.07 | 0.23 |
| *Modern Community, Neighborhood* | −0.08 | 0.00 |
| *Public* | −0.17 | −0.17 |
| *Primary Group* | 0.25 | −0.17 |
| *Mob* | 0.27 | 0.06 |
| *Secondary Group* | −0.15 | −0.18 |
| *Audience* | −0.05 | 0.23 |

The ALS algorithm HOMALS was designed for homogeneity analysis and is provided in the R package homals.[4] HOMALS corresponds to PRINCALS for all variables that are multiple nominal. The homals package allows for the restrictions imposed by the measurement level of the variables and is applicable to nonlinear PCA. It also offers various plotting options for representing the results.

# EXAMPLES

We illustrate nonlinear PCA of two data. The homals package is used to estimate the parameters and plot the results.

## Example 1: Guttman-Bell Data

The Guttman-Bell data[6] shown in Table 1 deal with attitudes of social groups consisting of seven objects and five variables with a total of 17 categories. We assume that variables 1–3 are single ordinal, and
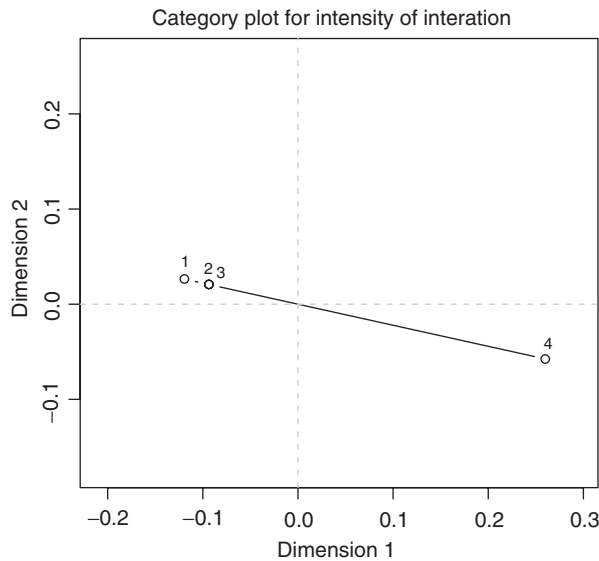
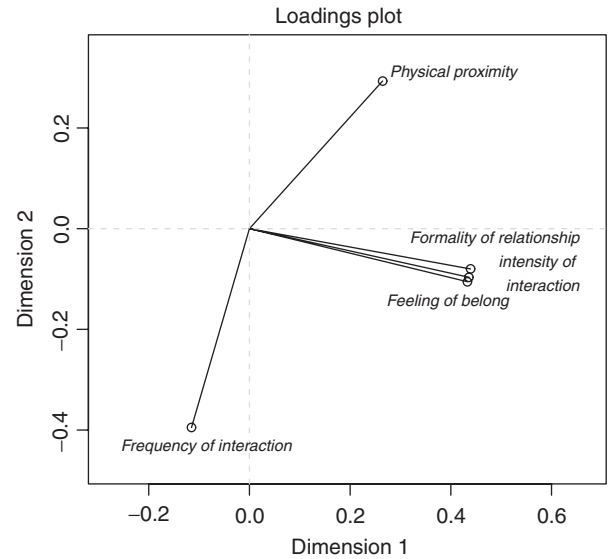**FIGURE 1** | Plot of category quantifications: *Intensity of interaction.*

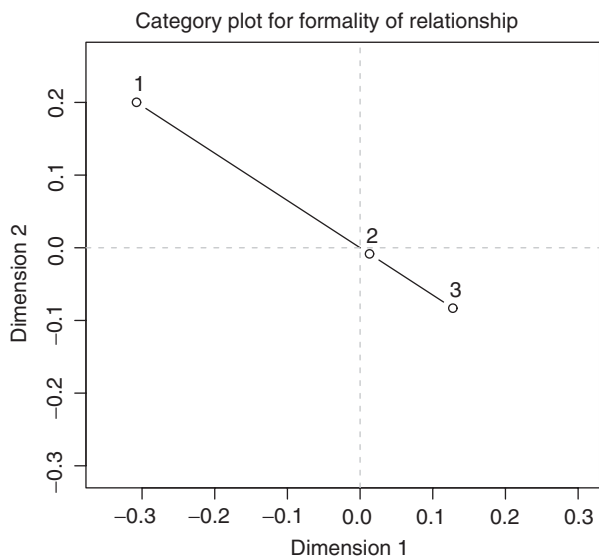

**FIGURE 3** | Plot of component loadings.



**FIGURE 2** | Plot of category quantifications: *Formality of relationship.*
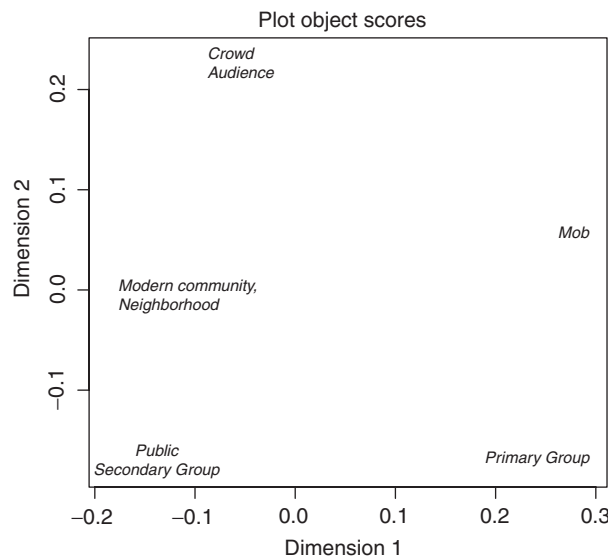


**FIGURE 4** | Plot of object scores.

variables 4 and 5 are single nominal. Thus, the first three variables impose rank-one and monotonicity restrictions, and the last two variables impose rank-one restrictions.

Tables 2–5 report the parameter estimates for $r = 2$. The results are plotted in Figures 1–5.

Figures 1 and 2 plot the category quantifications of *Intensity of interaction* and *Formality of relationship*. In our assumption, *Intensity of interaction* and *Formality of relationship* are ordinal and nominal variables, respectively. These points are given in the first and fifth columns of Table 2. The rank-one restrictions ensure that the category

quantifications of both variables lie on a straight line through the origin. As shown in the plot and table, categories 2 and 3 in *Intensity of interaction* are assigned the same value. The same results appear for *Frequency of interaction* and *Feeling of belong*. The cause of these results is the monotonicity restrictions on the variables.

Figure 3 plots the component loadings in Table 4. This figure shows vectors going in three different directions from the origin. The first group is *Physical proximity*; the second is *Formality of relationship*, *Intensity of interaction*, and *Feeling of belong*; and the third is *Frequency of interaction*.
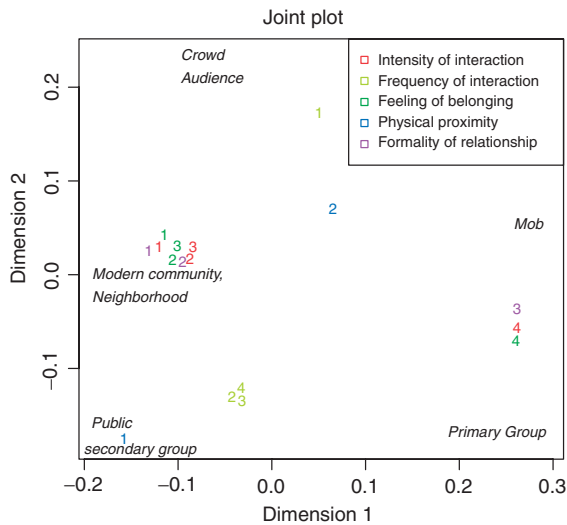
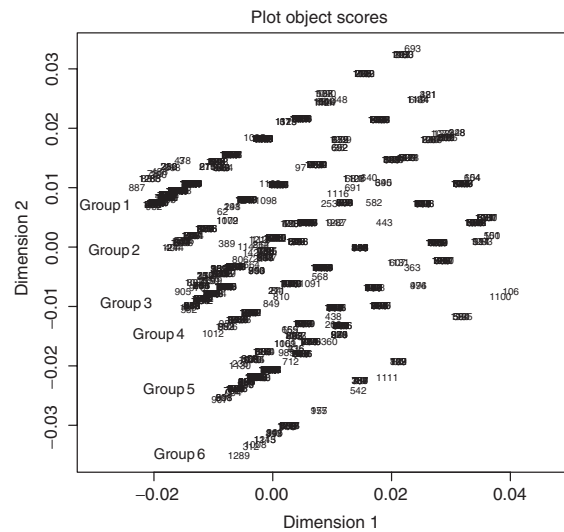**FIGURE 5** | Joint plot of object scores and category quantifications.



**FIGURE 6** | Plot of object scores.

**TABLE 6** | GALO Data: Variables

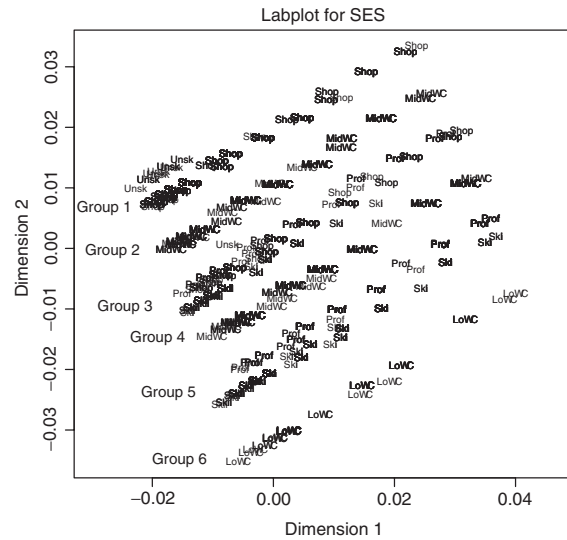| Description of the variables | |
|---|---|
| Variable 1 = Gender: | M = Male, F = Female |
| Variable 2 = IQ: | 1–9 |
| Variable 3 = Advice: | Agr = Agricultural; Ext = Extended primary education; |
| | Gen = General; Grls = Secondary school for girls; |
| | Man = Manual, including housekeeping; |
| | None = No further education; Uni = Pre-University |
| Variable 4 = SES: | LoWC = Lower white collar; MidWC = Middle white collar; |
| | Prof = Professional, Managers; Shop = Shopkeepers; |
| | Skil = Schooled labor; Unsk = Unskilled labor |
| Variable 5 = School: | 1–37 |



**FIGURE 7** | Plot of object scores labeled by SES.

## Example 2: GALO Data

The GALO data[12] shown in Table 6 were collected from 1290 school children in the sixth grade of elementary schools in the city of Groningen (the Netherlands). The variables are Gender, IQ, Advice, SES, and School. IQ (original range 60–144) is divided into nine ordered categories and the schools are enumerated from 1 to 37. In the analysis, we use four variables (variables 1 to 4) and assume that variables 1, 3, and 4 are nominal, and variable 2 is ordinal. Rank-one restrictions are imposed on variables 1, 3, and 4, and rank-one and monotone restrictions are imposed on variable 2.

In Figure 6, which plots the object scores, we identify six object groups. Figure 7 plots the object

Figure 4 plots the object scores **Z** in Table 5. The two object groups, *Crowd* and *Audience*, and *Public* and *Secondary Group*, have similar response patterns, whereas the other objects *Modern Community*, *Neighborhood*, *Primary Group*, and *Mob* have unique response patterns.

Figure 5 is the joint plot of **Z** and **W**. The plot shows that *Modern Community*, *Neighborhood* is associated with categories 1, 2, and 3 of *Intensity of interaction* and *Feeling of belong* and with categories 1 and 2 of *Formality of relationship*. We find that *Public* and *Secondary Group* are also associated with category 1 of *Physical Proximity*.

scores labeled by SES. This figure indicates that each group is characterized by SES: *Shop* and *Unsk* for Group 1; *MidWC* for Groups 2 and 4; *Prof*, *Shop*, and *Skil* for Group 3; *Prof* and *Skil* for Group 5; *LoWC* for Group 6. We also observe that, for each group, the object scores with the same categories in SES lie along a straight line.

## Conclusion

Nonlinear PCA is applicable to complicated multivariate data structures consisting of nominal, ordinal, and numerical variables. The use of optimal scaling for quantifying the qualitative data makes PCA applicable to combined quantitative and qualitative data. Because mixed data are often collected in statistical surveys, the ability of nonlinear PCA to handle such data confers a distinct advantage and broadens the research possibilities in many scientific disciplines. Nonlinear PCA finds the component scores and loadings of ordinary PCA and optimally scaled data, and moreover finds category quantifications of variables. The joint plot facilitates interpretation of the relationships among objects and variables. An available alternative to nonlinear PCA is multiple correspondence analysis,[2,3] which is equivalent to homogeneity analysis if all the row margins of the indicator matrix are equal.

The ALS algorithm is one of the ways to solve the matrix optimization problem. For nonlinear PCA, the algorithm alternates between the least squares fitting of PCA to optimally scaled data and the estimation of optimally scaled data. The popular ALS algorithms for nonlinear PCA, namely, PRINCIPALS and PRINCALS, are deduced by solving the minimization problem of a loss function. PRINCIPALS estimates the component scores and loadings for PCA and optimally scaled data. PRINCALS finds category quantifications of variables in addition to the parameters of PRINCIPALS, and also accepts multiple variables. When all variables are single, both algorithms yield the same results. For all variables treated as numerical, their results are also the same as those from ordinary PCA. The ALS algorithm is the least squares estimation method and its speed of convergence may be very slow in the application of nonlinear PCA to large mixed data. Improving the convergence of the ALS algorithm is an important topic.[9]

## REFERENCES

1. Eckart C, Young G. The approximation of one matrix by another of lower rank. *Psychometrika* 1936, 1:211–218.

2. De Leeuw J, Wang DY, Michailidis G. Correspondence analysis techniquesIn:. In: *Multivariate Analysis, Design of Experiments, and Survey Sampling*. New York: Marcel Dekker; 1999, 523–545.

3. De Leeuw J, Michailidis G. Graph layout techniques and multidimensional data analysisIn:. In: *Game Theory, Optimal Stopping, Probability and Statistics* IMS Lecture Notes Monogr. Ser., 35, Inst. Math. Statist.. Beachwood, OH: Institute of Mathematical Statistics; 2000, 219–248.

4. De Leeuw J, Mair P. A general framework for multivariate analysis with optimal scaling: the R package homals. *J Stat Softw* 2009, 31:1–21.

5. Gifi A. *Nonlinear Multivariate Analysis*. Chichester, England: John Wiley & Sons; 1990.

6. Guttman L. The quantification of a class of attributes: a theory and method of scale constructionIn:. In: Horst P, ed. *The Prediction of Personal Adjustment*. New York: Social Science Research Council; 1941, 319–348.

7. Kiers HAL. Setting up alternating least squares and iterative majorization algorithm for solving various matrix optimization problems. *Comput Stat Data An* 2002, 41:157–170.

8. Krijnen WP. Convergence of the sequence of parameters generated by alternating least squares algorithms. *Comput Stat Data An* 2006, 51:481–489.

9. Kuroda M, Mori Y, Iizuka M, Sakakihara M. Acceleration of the alternating least squares algorithm for principal components analysis. *Comput Stat Data An* 2011, 55:143–153.

10. Michailidis G, De Leeuw J. The Gifi system of descriptive multivariate analysis. *Stat Sci* 1998, 13:307–336.

11. Meulman JJ, Heiser WJ. *IBM SPSS Categories 20*. Chicago: SPSS, Inc.; 2012.

12. Peschar JL. *School, Milieu, Beroep*. Groningen, The Netherlands: Tjeek Willink; 1975.

13. SAS Institute, Inc. Algorithms for the PRINQUAL and TRANSREG procedures. In: *SAS Technical Report R-108*. Cary, NC: SAS Institute, Inc.; 1990.

14. SAS Institute, Inc. The PRINQUAL procedure. In: *SAS/STATR 12.1 User's Guide*. Cary, NC: SAS Institute, Inc.; 2012.

15. Young FW. Quantitative analysis of qualitative data. *Psychometrika* 1981, 46:357–388.

16. Young FW, Takane Y, De Leeuw J. The principal components of mixed measurement level multivariate data: an alternating least squares method with optimal scaling features. *Psychometrika* 1978, 43:279–281.