

Chapter 2

Nonlinear Principal Component Analysis

Abstract Principal components analysis (PCA) is a commonly used descriptive multivariate method for handling quantitative data and can be extended to deal with mixed measurement level data. For the extended PCA with such a mixture of quantitative and qualitative data, we require the quantification of qualitative data in order to obtain optimal scaling data. PCA with optimal scaling is referred to as nonlinear PCA, (Gifi, *Nonlinear Multivariate Analysis*. Wiley, Chichester, 1990). Nonlinear PCA including optimal scaling alternates between estimating the parameters of PCA and quantifying qualitative data. The alternating least squares (ALS) algorithm is used as the algorithm for nonlinear PCA and can find least squares solutions by minimizing two types of loss functions: a low-rank approximation and homogeneity analysis with restrictions. PRINCIPALS of Young et al. (Principal components of mixed measurement level multivariate data: an alternating least squares method with optimal scaling features 43:279–281, 1978) and PRINCALS of Gifi (*Nonlinear Multivariate Analysis*. Wiley, Chichester, 1990) are used for the computation.

Keywords Optimal scaling · Quantification · Alternating least squares algorithm · Low-rank approximation · Homogeneity analysis

2.1 Principal Component Analysis

Let $\mathbf{Y} = (\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_p)$ be a data matrix of n objects by p numerical variables and each column of $\mathbf{Y}$ be standardized, i.e., $\mathbf{y}_i^\top \mathbf{1}_n = 0$ and $\mathbf{y}_i^\top \mathbf{y}_i / n = 1$ for $i = 1, \dots, p$, where $\mathbf{1}_n$ is an $n \times 1$ vector of ones.

Principal component analysis (PCA) linearly transforms $\mathbf{Y}$ of p variables into a substantially smaller set of uncorrelated variables that contains much of the information of the original data set. Then PCA simplifies the description of $\mathbf{Y}$ and reveals the structure of $\mathbf{Y}$ and the variables.

PCA postulates that $\mathbf{Y}$ is approximated by the bilinear form

$$\hat{\mathbf{Y}} = \mathbf{Z}\mathbf{A}^\top, \quad (2.1)$$

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 $p \times r$ weight matrix that gives the coefficients of the linear combinations. PCA is formulated in terms of the loss function

$$\sigma(\mathbf{Z}, \mathbf{A}) = \text{tr}(\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}) = \text{tr}(\mathbf{Y} - \mathbf{Z}\mathbf{A}^\top)^\top (\mathbf{Y} - \mathbf{Z}\mathbf{A}^\top). \quad (2.2)$$

The minimum of the loss function (2.1) over $\mathbf{Z}$ and $\mathbf{A}$ is found by the eigen-decomposition of $\mathbf{Y}^\top \mathbf{Y}/n$ or the singular value decomposition of $\mathbf{Y}$.

2.1.1 Eigen-Decomposition of $\mathbf{Y}^\top \mathbf{Y}/n$

Let $\mathbf{S} = \mathbf{Y}^\top \mathbf{Y}/n$ be a $p \times p$ symmetric matrix. Then we have the following relation between the eigenvalues and eigenvectors of $\mathbf{S}$:

$$\mathbf{S}\mathbf{a}_i = \lambda_i \mathbf{a}_i, \quad \mathbf{a}_i^\top \mathbf{a}_i = 1 \quad \text{and} \quad \mathbf{a}_i^\top \mathbf{a}_j = 0 \quad (i \neq j) \quad (2.3)$$

for $i, j = 1, 2, \dots, p$. We denote the $p \times p$ matrix having p eigenvectors as columns by $\mathbf{A}$ and the $p \times p$ matrix having p eigenvalues as its diagonal elements by $\mathbf{D}_p$:

$$\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_p) \quad \text{and} \quad \mathbf{D}_p = \text{diag}(\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_p),$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. The relation between the eigenvalues and eigenvectors given by Eq. (2.3) can be expressed by

$$\mathbf{S}\mathbf{A} = \mathbf{A}\mathbf{D}_p, \quad \text{and} \quad \mathbf{A}^\top \mathbf{A} = \mathbf{I}_p,$$

where $\mathbf{I}_p$ is a $p \times p$ identity matrix. We obtain $\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_r)$ by solving

$$\mathbf{S}\mathbf{A} = \mathbf{A}\mathbf{D}_r$$

subject to $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_r$, and then compute $\mathbf{Z} = \mathbf{Y}\mathbf{A}$. Note that

$$\mathbf{Z}^\top \mathbf{Z} = \mathbf{A}^\top \mathbf{Y}^\top \mathbf{Y} \mathbf{A} = n\mathbf{I}_p.$$

2.1.2 Singular Value Decomposition of $\mathbf{Y}$

Let $\mathbf{Y}$ have rank l ($l \leq p$). From the Eckart-Young decomposition theorem (Eckart and Young 1936), $\mathbf{Y}$ has the following matrix decomposition

$$\mathbf{Y} = \mathbf{U}\mathbf{D}^{1/2}\mathbf{V}^\top, \quad (2.4)$$

where $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ have the following properties:

- $\mathbf{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_l)$ is an $n \times l$ matrix of left singular vectors satisfying $\mathbf{u}_i^\top \mathbf{u}_i = 1$ and $\mathbf{u}_i^\top \mathbf{u}_j = 0$, and $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_l$.
- $\mathbf{V} = (\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_l)$ is a $p \times l$ matrix of right singular vectors satisfying $\mathbf{v}_i^\top \mathbf{v}_i = 1$ and $\mathbf{v}_i^\top \mathbf{v}_j = 0$, and $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_l$.
- $\mathbf{D}$ is a $l \times l$ diagonal matrix of eigenvalues of $\mathbf{Y}^\top \mathbf{Y}$ or $\mathbf{Y}\mathbf{Y}^\top$.

We perform spectral decomposition of $\mathbf{Y}^\top \mathbf{Y}$:

$$\mathbf{Y}^\top \mathbf{Y} = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^\top + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^\top + \dots + \lambda_l \mathbf{v}_l \mathbf{v}_l^\top, \quad (2.5)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l \geq 0$ are eigenvalues of $\mathbf{Y}^\top \mathbf{Y}$ in descending order, and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l$ are the corresponding normalized eigenvectors of length one. The matrices $\mathbf{V}$ and $\mathbf{D}^{1/2}$ based on the decomposition (2.5) are defined as

$$\mathbf{V} = (\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_l), \quad \text{and} \quad \mathbf{D}^{1/2} = \text{diag}(\sqrt{\lambda_1} \ \sqrt{\lambda_2} \ \dots \ \sqrt{\lambda_l}).$$

From Eq. (2.4), we have

$$\mathbf{Z} = \mathbf{Z}\mathbf{A}^\top \mathbf{A} = \mathbf{Y}\mathbf{A} = \mathbf{U}\mathbf{D}^{1/2}.$$

Then the matrix $\mathbf{U}$ under restrictions $\mathbf{u}_i^\top \mathbf{u}_i = 1$ and $\mathbf{u}_i^\top \mathbf{u}_j = 0$ is given by

$$\mathbf{U} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} \mathbf{Y}\mathbf{v}_1 & \frac{1}{\sqrt{\lambda_2}} \mathbf{Y}\mathbf{v}_2 & \dots & \frac{1}{\sqrt{\lambda_l}} \mathbf{Y}\mathbf{v}_l \end{pmatrix}.$$

2.2 Quantification of Qualitative Data

Optimal scaling is a quantification technique that optimally assigns numerical values to qualitative scales within the restrictions of the measurement characteristics of the qualitative variables (Young 1981).

Let $\mathbf{y}_j$ of $\mathbf{Y}$ be a qualitative vector with K_j categories. To quantify $\mathbf{y}_j$, the vector is coded by using an $n \times K_j$ indicator matrix

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

where

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

For example, given

$$\mathbf{Y} = (\mathbf{y}_1 \ \mathbf{y}_2 \ \mathbf{y}_3) = \begin{pmatrix} \text{Blue} & \text{Yes} & 4 \\ \text{Red} & \text{No} & 3 \\ \text{Green} & \text{Yes} & 1 \\ \text{Green} & \text{No} & 2 \\ \text{Blue} & \text{Yes} & 1 \end{pmatrix},$$

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

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

Thus we have

$$\mathbf{y}_1 = \mathbf{G}_1 \begin{pmatrix} \text{Red} \\ \text{Green} \\ \text{Blue} \end{pmatrix}, \quad \mathbf{y}_2 = \mathbf{G}_2 \begin{pmatrix} \text{Yes} \\ \text{No} \end{pmatrix}, \quad \mathbf{y}_3 = \mathbf{G}_3 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Optimal scaling finds $K_j \times 1$ category quantifications $\mathbf{q}_j$ under the restrictions imposed by the measurement level of variable j and transforms $\mathbf{y}_j$ into an optimally scaled vector $\mathbf{y}_j^* = \mathbf{G}_j \mathbf{q}_j$. There are different ways for quantifying observed data of nominal, ordinal and numerical variables:

- Nominal scale data: The quantification is unrestricted. Objects i and $h (\neq i)$ in the same category for variable j obtain the same quantification. Thus, if $y_{ji} = y_{jh}$ then $y_{ji}^* = y_{jh}^*$.
- Ordinal scale data: The quantification is restricted to the order of categories. If observed categories y_{ji} and y_{jh} for objects i and h in variable j have order $y_{ji} > y_{jh}$ then quantified categories have order $y_{ji}^* \geq y_{jh}^*$.
- Numerical data: The observed vector $\mathbf{y}_j$ for variable j replaces $\mathbf{y}_j^*$ by standardizing with zero mean and unit variance.

2.3 Nonlinear PCA

PCA assumes that data are quantitative and thus it is not directly applicable to qualitative data such as nominal and ordinal data. When PCA handles mixed quantitative and qualitative data, the qualitative data must be quantified. In nonlinear PCA, the qualitative data of nominal and ordinal variables are nonlinearly transformed into

quantitative data. Thus, PCA with optimal scaling is called nonlinear PCA (Gifi 1990).

Nonlinear PCA reveals nonlinear relationships among variables with different measurement levels and therefore presents a more flexible alternative to ordinary PCA. Nonlinear PCA can find solutions by minimizing two types of loss functions; a low-rank approximation of $\mathbf{Y}^*$ extended to Eq. (2.2) and homogeneity analysis with restrictions. We show the loss functions and provide the ALS algorithm used for minimizing these loss functions.

2.3.1 Low-Rank Matrix Approximation

In the presence of qualitative variables in $\mathbf{Y}$, the loss function (2.2) is expressed as

$$\sigma_L(\mathbf{Z}, \mathbf{A}, \mathbf{Y}^*) = \text{tr}(\mathbf{Y}^* - \hat{\mathbf{Y}})^\top (\mathbf{Y}^* - \hat{\mathbf{Y}}) = \text{tr}(\mathbf{Y}^* - \mathbf{Z}\mathbf{A}^\top)^\top (\mathbf{Y}^* - \mathbf{Z}\mathbf{A}^\top) \quad (2.6)$$

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

$$\mathbf{Y}^{*\top} \mathbf{1}_n = \mathbf{0}_p \quad \text{and} \quad \text{diag} \left[\frac{\mathbf{Y}^{*\top} \mathbf{Y}^*}{n} \right] = \mathbf{I}_p, \quad (2.7)$$

where $\mathbf{1}_n$ and $\mathbf{0}_p$ are vectors of ones and zeros of length n and p , respectively. Optimal scaling for $\mathbf{Y}^*$ can be performed separately and independently for each variable, and then the loss function (2.6) can be rewritten as

$$\sigma_L(\mathbf{Z}, \mathbf{A}, \mathbf{Y}^*) = \sum_{j=1}^p (\mathbf{y}_j^* - \mathbf{Z}\mathbf{a}_j^\top)^\top (\mathbf{y}_j^* - \mathbf{Z}\mathbf{a}_j^\top) = \sum_{j=1}^p \sigma_{Lj}(\mathbf{Z}, \mathbf{a}_j, \mathbf{y}_j^*). \quad (2.8)$$

when minimizing independently each $\sigma_{Lj}(\mathbf{Z}, \mathbf{a}_j, \mathbf{y}_j^*)$ under the measurement restrictions on variable j , we can minimize $\sigma(\mathbf{Z}, \mathbf{A}, \mathbf{Y}^*)$.

2.3.2 Homogeneity Analysis

Homogeneity analysis maximizes the homogeneity of several categorical variables and quantifies the categories of each variable such that the homogeneity is maximized (Gifi 1990). 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 ($j = 1, \dots, p$). The loss function measuring the departure from homogeneity is given by

$$\begin{aligned}
\sigma_H(\mathbf{Z}, \mathbf{W}) &= \sum_{j=1}^p \text{tr}(\mathbf{Z} - \mathbf{G}_j \mathbf{W}_j)^\top (\mathbf{Z} - \mathbf{G}_j \mathbf{W}_j) \\
&= \sum_{j=1}^p \sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j)
\end{aligned} \tag{2.9}$$

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

$$\mathbf{Z}^\top \mathbf{1}_n = \mathbf{0}_r \quad \text{and} \quad \mathbf{Z}^\top \mathbf{Z} = n\mathbf{I}_r. \tag{2.10}$$

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

Gifi (1990) defines nonlinear PCA as homogeneity analysis imposing a *rank-one restriction* whose form is

$$\mathbf{W}_j = \mathbf{q}_j \mathbf{a}_j, \tag{2.11}$$

where $\mathbf{q}_j$ is a $K_j \times 1$ vector of category quantifications and $\mathbf{a}_j$ is a $1 \times r$ vector of weights (component loadings). Nominal variables on which restriction (2.11) is imposed are called *single nominal variables* and variables without restrictions are *multiple nominal variables*.

To minimize $\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j)$ under restriction (2.11), we first obtain the least squares estimate $\tilde{\mathbf{W}}_j$ of $\mathbf{W}_j$. For a fixed $\tilde{\mathbf{W}}_j$, $\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j)$ can be partitioned as

$$\begin{aligned}
\sigma_{Hj}(\mathbf{Z}, \mathbf{W}_j) &= \text{tr}(\mathbf{Z} - \mathbf{G}_j \mathbf{W}_j)^\top (\mathbf{Z} - \mathbf{G}_j \mathbf{W}_j) \\
&= \text{tr}(\mathbf{Z} - \mathbf{G}_j \tilde{\mathbf{W}}_j)^\top (\mathbf{Z} - \mathbf{G}_j \tilde{\mathbf{W}}_j) \\
&\quad + \text{tr}(\mathbf{q}_j \mathbf{a}_j - \tilde{\mathbf{W}}_j)^\top (\mathbf{G}_j^\top \mathbf{G}_j)(\mathbf{q}_j \mathbf{a}_j - \tilde{\mathbf{W}}_j).
\end{aligned} \tag{2.12}$$

We then minimize the second term on the right hand side of Eq.(2.12) over $\mathbf{q}_j$ and $\mathbf{a}_j$ under the restrictions imposed by the measurement level of variable j .

Each column vector of $\mathbf{Y}^*$ under restriction (2.11) is computed by

$$\mathbf{y}_j^* = \mathbf{G}_j \mathbf{q}_j.$$

Then Eq.(2.9) under restriction (2.10) is expressed as

$$\begin{aligned}
\sigma_H(\mathbf{Z}, \mathbf{W}) &= \sum_{i=1}^p \text{tr}(\mathbf{Z} - \mathbf{G}_i \mathbf{W}_i)^\top (\mathbf{Z} - \mathbf{G}_i \mathbf{W}_i) \\
&= np - 2 \sum_{i=1}^p \text{tr}(\mathbf{a}_i^\top \mathbf{y}_i^{*\top} \mathbf{Z}_i) + \sum_{i=1}^p \text{tr}(\mathbf{a}_i^\top \mathbf{y}_i^{*\top} \mathbf{y}_i^* \mathbf{a}_i) \\
&= np - 2\text{tr}(\mathbf{A}^\top \mathbf{Y}^{*\top} \mathbf{Z}) + \text{tr}(\mathbf{A}^\top \mathbf{A}).
\end{aligned}$$

when expanding Eq. (2.6) under restriction (2.7), we also obtain

$$\begin{aligned}\sigma_L(\mathbf{Z}, \mathbf{A}, \mathbf{Y}^*) &= \text{tr}(\mathbf{Y}^* - \mathbf{Z}\mathbf{A}^\top)^\top (\mathbf{Y}^* - \mathbf{Z}\mathbf{A}^\top)^\top \\ &= np - 2\text{tr}(\mathbf{A}^\top \mathbf{Y}^{*\top} \mathbf{Z}) + \text{tr}(\mathbf{A}\mathbf{A}^\top).\end{aligned}$$

Thus, minimizing the loss function (2.9) is equivalent to minimizing the loss function (2.6) under restrictions (2.7) and (2.10).

2.3.3 Alternating Least Squares Algorithm for Nonlinear PCA

The minimization of loss functions (2.6) and (2.9) has to take place with respect to both parameters of $\mathbf{Y}^*$ and $(\mathbf{Z}, \mathbf{A})$ and both of $\mathbf{Z}$ and $\mathbf{W}$, although we can not find simultaneously the solutions of these parameters. The alternating least squares (ALS) algorithm is utilized to solve such minimization problem.

We describe the general procedure of the ALS algorithm. Let $\sigma(\theta_1, \theta_2)$ be a loss function and (θ_1, θ_2) be the parameter matrices of the function. We denote the t -th estimate of θ as $\theta^{(t)}$. To minimize $\sigma(\theta_1, \theta_2)$ over θ_1 and θ_2 , the ALS algorithm updates the estimates of θ_1 and θ_2 by solving the least squares problem for each parameter:

$$\begin{aligned}\theta_1^{(t+1)} &= \arg \min_{\theta_1} \sigma(\theta_1, \theta_2^{(t)}), \\ \theta_2^{(t+1)} &= \arg \min_{\theta_2} \sigma(\theta_1^{(t+1)}, \theta_2).\end{aligned}$$

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 (Krijnen 2006).

We show the two ALS algorithms typically employed in nonlinear PCA; PRINCIPALS (Young et al. 1978) and PRINCALS (Gifi 1990).

2.3.3.1 PRINCIPALS

PRINCIPALS developed by Young et al. (1978) is the ALS algorithm that minimizes the loss function (2.8). PRINCIPALS 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 obtains the estimate of the data parameter $\mathbf{Y}^*$ for optimally scaled data.

For the initialization of PRINCIPALS, the initial data $\mathbf{Y}^{*(0)}$ are determined under the measurement restrictions for each variable and are then standardized to satisfy restriction (2.7). The observed data $\mathbf{Y}$ may be used as $\mathbf{Y}^{*(0)}$ after standardizing each

column of $\mathbf{Y}$ under restriction (2.7). Given the initial data $\mathbf{Y}^{*(0)}$, PRINCIPALS iterates the following two steps:

- *Model estimation step*: By solving the eigen-decomposition of $\mathbf{Y}^{*(t)\top} \mathbf{Y}^{*(t)} / n$ or the singular value decomposition of $\mathbf{Y}^{*(t)}$, obtain $\mathbf{A}^{(t+1)}$ and compute $\mathbf{Z}^{(t+1)} = \mathbf{Y}^{*(t)} \mathbf{A}^{(t+1)}$. Update $\hat{\mathbf{Y}}^{(t+1)} = \mathbf{Z}^{(t+1)} \mathbf{A}^{(t+1)\top}$.
- *Optimal scaling step*: Obtain $\mathbf{Y}^{*(t+1)}$ by separately estimating $\mathbf{y}_j^*$ for each variable j . Compute $\mathbf{q}_j^{(t+1)}$ for nominal variables as

$$\mathbf{q}_j^{(t+1)} = (\mathbf{G}_j^\top \mathbf{G}_j)^{-1} \mathbf{G}_j^\top \hat{\mathbf{y}}_j^{(t+1)}.$$

Re-compute $\mathbf{q}_j^{(t+1)}$ for ordinal variables using the monotone regression (Kruskal 1964). For nominal and ordinal variables, update $\mathbf{y}_j^{*(t+1)} = \mathbf{G}_j \mathbf{q}_j^{(t+1)}$ and stan-

Table 2.1 Sleeping bag data from Prediger (1997)

	Temperature	Weight	Price	Material	Quality rate
One kilo bag	7	940	149	Liteloft	3
Sund	3	1880	139	Hollow ber	1
Kompakt basic	0	1280	249	MTI Loft	3
Finmark tour	0	1750	179	Hollow ber	1
Interlight Lyx	0	1900	239	Thermolite	1
Kompakt	−3	1490	299	MTI Loft	2
Touch the cloud	−3	1550	299	Liteloft	2
Cat's meow	−7	1450	339	Polarguard	3
Igloo super	−7	2060	279	Terraloft	1
Donna	−7	1850	349	MTI Loft	2
Tyin	−15	2100	399	Ultraloft	2
Travellers dream	3	970	379	Goose-downs	3
Yeti light	3	800	349	Goose-downs	3
Climber	−3	1690	329	Duck-downs	2
Viking	−3	1200	369	Goose-downs	3
Eiger	−3	1500	419	Goose-downs	2
Climber light	−7	1380	349	Goose-downs	3
Cobra	−7	1460	449	Duck-downs	3
Cobra comfort	−10	1820	549	Duck-downs	2
Fox re	−10	1390	669	Goose-downs	3
Mont Blanc	−15	1800	549	Goose-downs	3

standardize $\mathbf{y}_j^{*(t+1)}$. For numerical variables, standardize observed vector $\mathbf{y}_j$ and set $\mathbf{y}_j^{*(t+1)} = \mathbf{y}_j$.

2.3.3.2 PRINCALS

PRINCALS is the ALS algorithm developed by Gifi (1990) and can handle multiple nominal variables in addition to the single nominal, ordinal and numerical variables. We denote the set of multiple variables by $\mathbf{J}_M$ and the set of single variables having single nominal and ordinal scales and numerical measurements by $\mathbf{J}_S$. From Eqs. (2.9) and (2.12), the loss function to be minimized by PRINCALS is given by

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

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 (2.10), and $\mathbf{W}_j^{(0)}$ is obtained as $\mathbf{W}_j^{(0)} = (\mathbf{G}_j^\top \mathbf{G}_j)^{-1} \mathbf{G}_j^\top \mathbf{Z}^{(0)}$. For each variable $j \in \mathbf{J}_S$, $\mathbf{q}_j^{(0)}$ is defined as the first K_j successive integers under the normalization restriction. The vector $\mathbf{a}_j$ is initialized as $\mathbf{a}_j^{(0)} = \mathbf{Z}^{(0)\top} \mathbf{G}_j \mathbf{q}_j^{(0)}$, and rescaled to unit length. Given these initial values, PRINCALS iterates the following steps (Michailidis and de Leeuw 1998):

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

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

Table 2.2 Quantification of Material and Quality rate

Material	
Duck-downs	−0.168
Goose-downs	−0.147
Hollow ber	0.473
Liteloft	0.046
MTI loft	0.014
Polarguard	−0.005
Terraloft	0.313
Thermolite	0.390
Ultraloft	−0.250
Quality rate	
1	−0.450
2	0.106
3	0.106

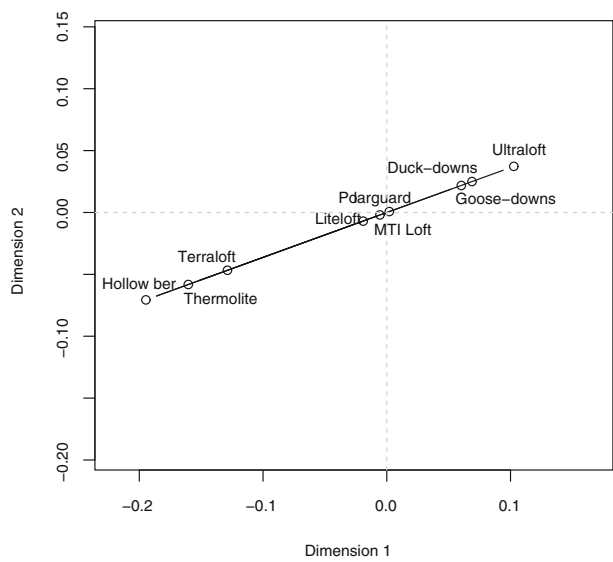


Fig. 2.1 Category plot for material

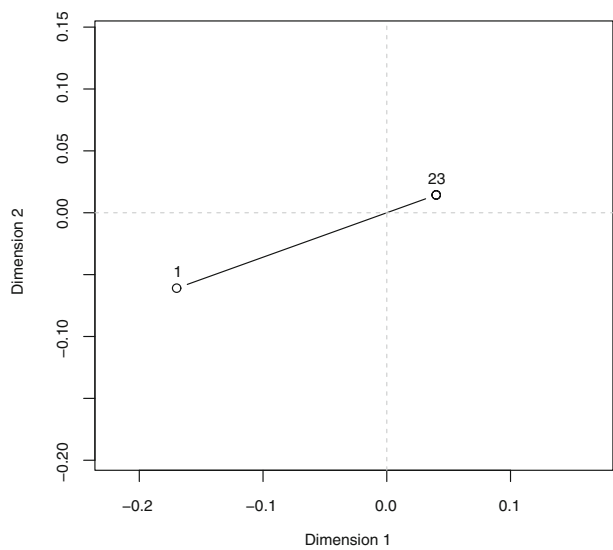


Fig. 2.2 Category plot for Quality rate

Table 2.3 Optimal scaled sleeping bag data

	Temperature	Weight	Price	Material	Quality rate
One kilo bag	0.425	−0.324	−0.342	0.046	0.106
Sund	0.290	0.252	−0.388	0.473	−0.450
Kompakt basic	0.155	−0.216	−0.202	0.014	0.106
Finmark tour	0.155	0.108	−0.295	0.473	−0.450
Interlight Lyx	0.155	0.288	−0.248	0.390	−0.450
Kompakt	0.019	−0.036	−0.109	0.014	0.106
Touch the cloud	0.019	0.036	−0.109	0.046	0.106
Cat's meow	−0.116	−0.108	−0.016	−0.005	0.106
Igloo super	−0.116	0.324	−0.155	0.313	−0.450
Donna	−0.116	0.216	0.031	0.014	0.106
Tyin	−0.387	0.360	0.171	−0.250	0.106
Travellers dream	0.290	−0.288	0.124	−0.147	0.106
Yeti light	0.290	−0.360	0.031	−0.147	0.106
Climber	0.019	0.072	−0.062	−0.168	0.106
Viking	0.019	−0.252	0.078	−0.147	0.106
Eiger	0.019	0.000	0.217	−0.147	0.106
Climber light	−0.116	−0.180	0.031	−0.147	0.106
Cobra	−0.116	−0.072	0.264	−0.168	0.106
Cobra comfort	−0.251	0.180	0.311	−0.168	0.106
Fox re	−0.251	−0.144	0.357	−0.147	0.106
Mont Blanc	−0.387	0.144	0.311	−0.147	0.106

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

$$\mathbf{a}_j^{(t+1)\top} = \mathbf{W}_j^{(t+1)\top} (\mathbf{G}_j^\top \mathbf{G}_j) \mathbf{q}_j^{(t)} / \mathbf{q}_j^{(t)\top} (\mathbf{G}_j^\top \mathbf{G}_j) \mathbf{q}_j^{(t)}$$

and compute $\mathbf{q}_j^{(t+1)}$ for nominal variables by

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

Re-compute $\mathbf{q}_j^{(t+1)}$ for ordinal variables using the monotone regression in a similar manner as for PRINCIPALS. For numerical variables, standardize observed

Table 2.4 Component scores

	$\mathbf{Z}_1$	$\mathbf{Z}_2$
One kilo bag	−0.104	0.185
Sund	−0.208	−0.048
Kompakt basic	−0.010	−0.084
Finmark tour	−0.172	−0.120
Interlight Lyx	−0.167	−0.040
Kompakt	0.019	−0.065
Touch the cloud	0.010	−0.044
Cat's meow	0.024	−0.059
Igloo super	−0.132	−0.036
Donna	0.006	0.056
Tyin	0.078	0.105
Travellers dream	−0.006	0.224
Yeti light	−0.014	0.194
Climber	0.042	0.003
Viking	0.078	−0.109
Eiger	0.080	−0.013
Climber light	0.053	−0.061
Cobra	0.079	−0.008
Cobra comfort	0.111	0.010
Fox re	0.136	−0.107
Mont Blanc	0.097	0.0182

Table 2.5 Factor loadings

	Temperature	Weight	Price	Material	Quality rate
$\mathbf{Z}_1$	−0.330	−0.204	0.396	−0.411	0.377
$\mathbf{Z}_2$	0.179	0.348	0.045	−0.149	0.136

vector $\mathbf{y}_j$ and compute $\mathbf{q}_j^{(t+1)} = (\mathbf{G}_j^\top \mathbf{G}_j)^{-1} \mathbf{G}_j^\top \mathbf{y}_j$. Update $\mathbf{W}_j^{(t+1)} = \mathbf{q}_j^{(t+1)} \mathbf{a}_j^{(t+1)}$ for ordinal and numerical variables.

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

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

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

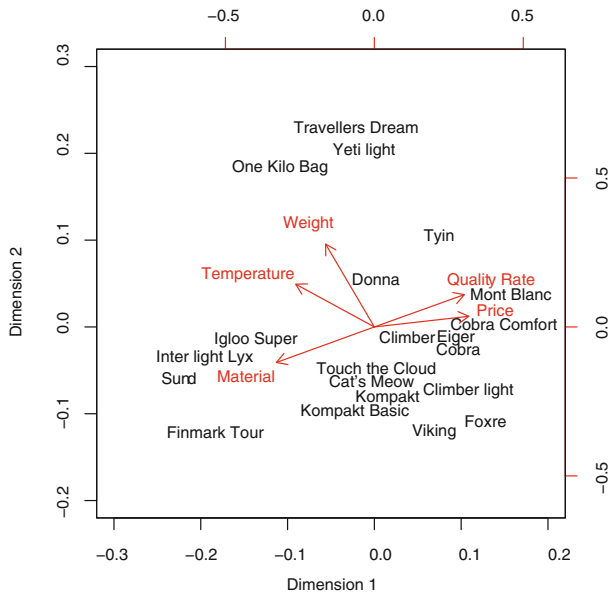


Fig. 2.3 Biplot of the first two principal components

2.4 Example: Sleeping Bags

We illustrate nonlinear PCA using sleeping bag data from Prediger (1997) given in Table 2.1. The data were collected on 21 sleeping bags with Temperature, Weight, Price, Material and Quality Rate. Quality Rate is scaled from 1 to 3 such that the higher value is the better one. The first three variables are numerical, Material is nominal and Quality Rate is ordinal. The computation for quantifying qualitative data and PCA is performed by the R package `homals` of De Leeuw and Mair (2009) that provides the ALS algorithm for homogeneity analysis. When imposing the rank-one restrictions on Material and Quality Rate, `homals` is the same as PRINCALS. We set $r = 2$ and obtain the following results.

Table 2.2 reports the quantified values of Material and Quality Rate. Then Material are quantified without order restriction due to the nominal variable, while the quantification of Quality Rate is restricted to the order of categories. Figures 2.1 and 2.2 are the plots of the category quantifications of Material and Quality Rate. These figures graphically show the order restrictions for these variables. Table 2.3 shows optimal scaled sleeping bag data. The component scores and factor loadings are given in Tables 2.4 and 2.5, respectively. Figure 2.3 is the biplot of the first two principal components. We can interpret the data using ordinary PCA.

References

- De Leeuw, J., Mair, P.: A general framework for multivariate analysis with optimal scaling: the R package *homals*. *J. Stat. Softw.* **31**, 1–21 (2009)
- Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–218 (1936)
- Gifi, A.: *Nonlinear Multivariate Analysis*. Wiley, Chichester (1990)
- Krijnen, W.P.: Convergence of the sequence of parameters generated by alternating least squares algorithms. *Comput. Stat. Data Anal.* **51**, 481–489 (2006)
- Kruskal, J.B.: Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**, 115–129 (1964)
- Michailidis, G., de Leeuw, J.: The Gifi system of descriptive multivariate analysis. *Stat. Sci.* **13**, 307–336 (1998)
- Prediger, S.: Symbolic objects in formal concept analysis. In: Mineau, G., Fall, A (eds.) *Proceedings of the 2nd International Symposium on Knowledge, Retrieval, Use, and Storage for Efficiency* (1997)
- Young, F.W.: Quantitative analysis of qualitative data. *Psychometrika* **46**, 357–388 (1981)
- Young, F.W., Takane, Y., de Leeuw, J.: Principal components of mixed measurement level multivariate data: an alternating least squares method with optimal scaling features. *Psychometrika* **43**, 279–281 (1978)

Nonlinear Principal Component Analysis and Its
Applications

Mori, Y.; Kuroda, M.; Makino, N.

2016, X, 80 p. 17 illus., 8 illus. in color., Softcover

ISBN: 978-981-10-0157-4