Chapter 1

Principal Components Analysis

June 24, 2013

G Tutz - preliminary

The objective of principal components analysis is to transform a set of p variables $x_1, \ldots, x_p$ into r new variables $z_1, \ldots, z_r$ which account for most of the variation of the original variables. The number r of the so-called principal components $z_1, \ldots, z_r$ is assumed to be small relative to p. Classical principal components analysis is based on linear transformations of the original variables.

1.1 Principal Components for Random Variables

Principal components analysis may be formulated for random variables or observations in a rather similar way. We will first consider principal components for random variables

1.1.1 Basic Concept

For the random variables form one assumes that $\boldsymbol{x}^T = (x_1, \ldots, x_p)$ is a vector of random variables with mean $\boldsymbol{\mu}^T = (\mu_1, \ldots, \mu_p)$ and covariance $\text{cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$.

The first principal component is determined by a weight vector $\boldsymbol{\alpha}_1$ chosen such that

$$z_1 = \boldsymbol{\alpha}_1^T \boldsymbol{x} = \alpha_{11} x_1 + \cdots + \alpha_{1p} x_p$$

has maximal variance, that is,

$$\mathrm{var}(z_1) = \boldsymbol{\alpha}_1^T \boldsymbol{\Sigma}\, \boldsymbol{\alpha}_1 \to \max_{\boldsymbol{\alpha}_1}.$$

under the constraint $||\boldsymbol{\alpha}_1|| = 1$. The constraint is necessary since the variance could be increased without limit by increasing the components of $\boldsymbol{\alpha}_1$.

For the second principal component $z_2 = \boldsymbol{\alpha}_2^T \boldsymbol{x}$ one postulates

$$var(z_2) \to \max_{\boldsymbol{\alpha}_2}$$

with the constraints $||\boldsymbol{\alpha}_2|| = 1$ and $\boldsymbol{\alpha}_1^T \boldsymbol{\alpha}_2 = 0$. The latter constraint is equivalent to postulating $\mathrm{cov}(z_1, z_2) = \mathrm{cov}(\boldsymbol{\alpha}_1^T \boldsymbol{x}, \boldsymbol{\alpha}_2^T \boldsymbol{x}) = \boldsymbol{\alpha}_1^T \boldsymbol{\Sigma}\, \boldsymbol{\alpha}_2 = 0$. The further principal components are obtained by looking for weights which maximize the variance under the restriction that the weight is orthogonal to the weights of the previous principal components.

In summary the objective is to find weights $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ and corresponding linear combinations

$$\begin{aligned} z_1 &= \boldsymbol{\alpha}_1^T \boldsymbol{x} = \alpha_{11} x_1 + \cdots + \alpha_{1p} x_p \\ &\vdots \\ z_r &= \boldsymbol{\alpha}_r^T \boldsymbol{x} = \alpha_{r1} x_1 + \cdots + \alpha_{rp} x_p \end{aligned}$$

such that

$$var(z_j) = \boldsymbol{\alpha}_j^T \boldsymbol{\Sigma} \boldsymbol{\alpha}_j \to \max_{\boldsymbol{\alpha}_j}, \qquad j = 1, \ldots, p, \tag{1.1}$$

with the side constraints $||\boldsymbol{\alpha}_j|| = 1, \boldsymbol{\alpha}_j^T \boldsymbol{\alpha}_s = 0, s = 1, \ldots, j-1$. The second side constraint is equivalent to postulating that $\mathrm{cov}(z_j, z_s) = 0, s = 1, \ldots, j-1$.

Principal Components by Maximization of Variance

Find weights $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$ such that for $z_j = \boldsymbol{\alpha}_j^T \boldsymbol{x}$

$$var(z_j) = \boldsymbol{\alpha}_j^T \boldsymbol{\Sigma}\, \boldsymbol{\alpha}_j \to \max_{\boldsymbol{\alpha}_j}$$

with side constraints $||\boldsymbol{\alpha}_j|| = 1, \boldsymbol{\alpha}_j^T \boldsymbol{\alpha}_s = 0, s = 1, \ldots, j-1$.

1.1.2 Obtaining Solutions

For the first principal component, the problem is to maximize $\mathrm{var}(\boldsymbol{\alpha}_1^T \boldsymbol{x}) = \boldsymbol{\alpha}_1^T \boldsymbol{\Sigma}\, \boldsymbol{\alpha}_1$ subject to $\boldsymbol{\alpha}_1^T \boldsymbol{\alpha}_1 = 1$. Using Lagrange multiplier λ one considers

$$\varphi_1(\boldsymbol{\alpha}_1) = \boldsymbol{\alpha}_1^T \boldsymbol{\Sigma}\, \boldsymbol{\alpha}_1 - \lambda(\boldsymbol{\alpha}_1^T \boldsymbol{\alpha}_1)$$

The derivatives of $\varphi(\boldsymbol{\alpha}_1)$

$$\begin{aligned} \frac{\partial \varphi_1}{\partial \boldsymbol{\alpha}} &= 2\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 - 2\lambda\boldsymbol{\alpha}_1 \\ \frac{\partial \varphi_1}{\partial \lambda} &= \boldsymbol{\alpha}_1^T \boldsymbol{\alpha}_1 \end{aligned}$$

yield the equations

$$\boldsymbol{\Sigma}\,\boldsymbol{\alpha}_1 = \lambda\boldsymbol{\alpha}_1, \qquad \boldsymbol{\alpha}_1^T\boldsymbol{\alpha}_1 = 1$$

which represent an eigenvalue problem. Thus the eigenvector $\boldsymbol{\alpha}_1$ which corresponds to the largest eigenvalue λ_1 is a solution of the maximization problem.

Consider now maximization of $\mathrm{var}(\boldsymbol{\alpha_2}^T\boldsymbol{x}) = \boldsymbol{\alpha}_2^T\boldsymbol{\Sigma}\,\boldsymbol{\alpha}_2$ subject to constraints $||\boldsymbol{\alpha}_1|| = 1$, $\boldsymbol{\alpha}_2^T\boldsymbol{\alpha}_1 = 0$. one considers the function with Lagrange multipliers for the constraints

$$\varphi_2(\boldsymbol{\alpha}_2) = \boldsymbol{\alpha}_2^T\boldsymbol{\Sigma}\,\boldsymbol{\alpha}_2 + \lambda(\boldsymbol{\alpha}_1^T\boldsymbol{\alpha}_2 - 1) + \gamma(\boldsymbol{\alpha}_1^T\boldsymbol{\alpha}_2)$$

The derivatives $\partial\varphi_2/\partial\boldsymbol{\alpha}_2$, $\partial\varphi_2/\partial\lambda$, $\partial\varphi_2/\partial\gamma$ yield the equations

$$\begin{aligned} &2\boldsymbol{\Sigma}\,\boldsymbol{\alpha}_2 + 2\lambda\boldsymbol{\alpha}_2 + \gamma\boldsymbol{\alpha}_1 = 0,\\ &\boldsymbol{\alpha}_1^T\boldsymbol{\alpha}_2 = 1,\\ &\boldsymbol{\alpha}_1^T\boldsymbol{\alpha}_2 = 0. \end{aligned}$$

Multiplication of the first equation with $\boldsymbol{\alpha}_1^T$ (from the left side) yields $2\boldsymbol{\alpha}_1^T\boldsymbol{\Sigma}\,\boldsymbol{\alpha}_2 + \gamma = 0$. Since $\boldsymbol{\alpha}_1$ is an eigenvector of $\boldsymbol{\Sigma}$ one has $\boldsymbol{\alpha}_1^T\boldsymbol{\Sigma}\,\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_2^T\boldsymbol{\Sigma}\,\boldsymbol{\alpha}_1 = \lambda_1\boldsymbol{\alpha}_2^T\boldsymbol{\alpha}_1 = 0$ yielding $\gamma = 0$. Therefore the first equation has the form

$$\boldsymbol{\Sigma}\,\boldsymbol{\alpha}_2 = \lambda\boldsymbol{\alpha}_2 \quad \text{subject to} \qquad \boldsymbol{\alpha}_2^T\boldsymbol{\alpha}_1 = 0$$

The solution is the eigenvector $\boldsymbol{\alpha}_2$ for the second largest eigenvector λ_2.

Straightforward derivation shows that starting from eigenvector solutions $\alpha_1, \ldots, \alpha_s$ maximization of $\mathrm{var}(\boldsymbol{\alpha}^T\boldsymbol{x})$ subject to $\boldsymbol{\alpha}^T\boldsymbol{\alpha} = 1$, $\boldsymbol{\alpha}^T\boldsymbol{\alpha}_j = 0$, $j = 1, \ldots, s$ yields the eigenvector $\boldsymbol{\alpha}_{s+1}$ corresponding to the next largest eigenvalue λ_{s+1}.

In summary the solutions of the maximization problem are the eigenvectors $\alpha_1, \ldots, \alpha_p$ that correspond to eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$. The spectral decomposition theorem yields that the symmetric covariance matrix $\boldsymbol{\Sigma}$ may be written as

$$\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$$

where the columns of the orthogonal matrix $\boldsymbol{P} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p)$ are the eigenvectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$ of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega} = diag(\lambda_1, \ldots, \lambda_p)$ is a diagonal matrix which has the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ in the diagonal.

For positive definite covariance matrix $\boldsymbol{\Sigma}$ all the eigenvalues $\lambda_1, \ldots, \lambda_p$ are positive. One obtains the principal components $z_i = \boldsymbol{\alpha}_i^T\boldsymbol{x}$, $i = 1, \ldots, p$ in vector form by

$$\boldsymbol{z} = \boldsymbol{P}^T\boldsymbol{x}$$

where $\boldsymbol{z}^T = (z_1, \ldots, z_p)$. The principal components represent uncorrelated linear combinations of the variables. One obtains

$$cov(\boldsymbol{z}) = \boldsymbol{P}^T\boldsymbol{\Sigma}\boldsymbol{P} = \boldsymbol{\Lambda}$$

and therefore $\mathrm{var}(z_i) = \lambda_i$, $\mathrm{cov}(z_i, z_j) = 0$, $i \neq j$.

Principal Components

The weights of principal components $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$ are found as the columns of the spectral decomposition

$$\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T,$$

where $\boldsymbol{P} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha_p})$, $\boldsymbol{\Sigma} = diag(\lambda_1, \ldots, \lambda_p)$.

For the vector of principal components $\boldsymbol{z} = \boldsymbol{P}^T\boldsymbol{x}$ one has

$$\operatorname{cov}(\boldsymbol{z}) = \boldsymbol{\Lambda}$$

and for the variation of $\boldsymbol{x}$ and $\boldsymbol{z}$ one has

$$tr(\operatorname{cov}(\boldsymbol{x})) = tr(\operatorname{cov}(\boldsymbol{z}))$$

$$|\operatorname{cov}(\boldsymbol{x})| = |\operatorname{cov}(\boldsymbol{z})|$$

1.1.3 Variation and Explained Variation

Variation of random vectors may be measured in several ways. A simple measure of the variation in vector $\boldsymbol{x}$ is the *total variation*. For correlated variables $x_1, \ldots, x_p$ one obtains

$$tvar(\boldsymbol{x}) = \sum_{i=1}^{p} \operatorname{var}(x_i) = tr(\boldsymbol{\Sigma})$$

By considering

$$tr(\boldsymbol{\Sigma}) = tr(\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T) = tr(\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{P}) = tr(\boldsymbol{\Lambda}) = \sum_{i=1}^{p} \operatorname{var}(y_i) = tvar(\boldsymbol{y})$$

one obtains that the total variation of $\boldsymbol{x}$ is the same as the total variation of the principal components $\boldsymbol{y}$, that is,

$$\sum_{i=1}^{p} \operatorname{var}(x_i) = \sum_{i=1}^{p} \operatorname{var}(y_i).$$

A more general measure of variation in vector $\boldsymbol{y}$ is the *generalized variance* given as determined $|\operatorname{cov}(\boldsymbol{x})|$. Comparison of the generalized variance of $\boldsymbol{x}$ and the principal components y yields

$$|\operatorname{cov}(\boldsymbol{x})| = |\boldsymbol{\Sigma}| = |\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T| = |\boldsymbol{P}||\boldsymbol{\Lambda}||\boldsymbol{P}^T| = |\boldsymbol{\Lambda}| = |\operatorname{cov}(\boldsymbol{y})|$$

since $\boldsymbol{P}$ is an orthogonal matrix and therefore $|\boldsymbol{P}||\boldsymbol{P}^T| = 1$. In summary one has

$$|\operatorname{cov}(\boldsymbol{x})| = |\operatorname{cov}(\boldsymbol{y})|.$$

One may wonder if all principal components are necessary. Their ordering $\operatorname{var}(y_1) = \lambda_1 \geq \ldots \geq \operatorname{var}(y_p) = \lambda_p$ suggests to consider which part of the variation is explained by the first r principal components $z_1, \ldots, z_r$.

A simple measure for the explained variation is the *proportion of total variation*

$$t(r) = \frac{\sum_{i=1}^{r} var(y_i)}{\sum_{i=1}^{p} var(y_i)} = \frac{\lambda_1 + \ldots + \lambda_r}{\lambda_1 + \ldots + \lambda_p}.$$

Thus, if, for example, $t(2) = 0.8$ 80 percent of the total variation is explained by the first principal component.

1.1.4 Some Geometry and the Normal Distribution

The components of a point $\boldsymbol{x}^T = (x_1, \ldots, x_p)$ from $\mathbb{R}^p$ can be seen as the coordinates when the $\mathbb{R}^p$ is spanned by the unit vectors $\boldsymbol{e}_1^T = (1, 0, \ldots, 0), \ldots, \boldsymbol{e}_p^T = (0, \ldots, 0, 1)$ because $\boldsymbol{x}$ is given by

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + \cdots + x_p\boldsymbol{e}_p.$$

The corresponding vector of principal components is given by $\boldsymbol{z} = \boldsymbol{P}^T\boldsymbol{x}$, which is equivalent to $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{z}$. Therefore the point $\boldsymbol{x}$ is also represented by

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{z} = (\boldsymbol{a}_1 \ldots \boldsymbol{a}_p)\boldsymbol{z} = z_1\boldsymbol{a}_1 + \cdots + z_p\boldsymbol{a}_p.$$

Therefore, $z_1, \ldots, z_p$ are the coordinate values when the $\mathbb{R}^p$ is spanned by the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$. Thus the principal components describe the same point but use a different coordinate system. The system of basis vectors used by principal components $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ is orthogonal as the commonly used system of unit vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p$.

The multivariate normal distribution with mean zero has the density

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2} \mid \boldsymbol{\Sigma} \mid^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right\}.$$

Let us consider the spectral decomposition of $\boldsymbol{\Sigma}$ given by $\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$. Since $\boldsymbol{P}$ is orthogonal the inverse of $\boldsymbol{\Sigma}$ is given by $\boldsymbol{\Sigma}^{-1} = \boldsymbol{P}\Lambda^{-1}\boldsymbol{P}^T$. Therefore the relevant part of the density is

$$\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{P}\Lambda^{-1}\boldsymbol{P}^T\boldsymbol{x} = \boldsymbol{z}^T\Lambda^{-1}\boldsymbol{z} = \sum_{i=1}^{p}(\frac{z_i}{\sqrt{\lambda_i}})^2.$$

That means all points that have the same value of the density, that is, $\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} = c$ for some fixed value c can also be described as the points that fulfill

$$\sum_{i=1}^{p}(\frac{z_i}{\sqrt{\lambda_i}})^2 = c.$$

These points describe in the coordinates $z_1 \ldots, z_p$ an ellipsoid with lengths $\sqrt{\lambda_i c}$ on the axes.

1.2 Principal Components for Observations

For observations the problem has to be slightly modified. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be vector valued observations, $\boldsymbol{x}_i^T = (x_{i1}, \ldots, x_{ip})$, a linear transformation of the observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ by use of vector $\boldsymbol{\alpha}_j$ yields the observations

$$z_{ij} = \boldsymbol{\alpha}_j^T\boldsymbol{x}_i, i = 1, \ldots, n.$$

The empirical variance of the observations $z_{1j}, \ldots, z_{nj}$ is given by

$$s_j^2 = \frac{1}{n}\sum_j (z_{ij} - \bar{z}_j)^2 = \boldsymbol{\alpha}_j^T \boldsymbol{S}_x \boldsymbol{\alpha}_j,$$

where $\bar{z}_j = \frac{1}{n}\sum_{i=1}^n z_{ij}$ and $\boldsymbol{S}_x = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T$ is the empirical covariance matrix computed from the data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

The principal components for observations are then obtained by finding vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ such that

$$s_j^2 = \boldsymbol{\alpha}_j^T \boldsymbol{S}_x \boldsymbol{\alpha}_j \to \max_{\boldsymbol{\alpha}_j} \tag{1.2}$$

under the constraints $||\boldsymbol{\alpha}_j|| = 1, \boldsymbol{\alpha}_j^T \boldsymbol{\alpha}_s = 0, s = 1, \ldots, j-1$. Thus the only difference between (1.1) and (1.2) is that the covariance matrix $\boldsymbol{\Sigma}$ is replaced by the empirical covariance matrix $\boldsymbol{S}_x$.

Principal Components by Maximization of Empirical Variance

Find weights $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ such that for $z_j = \boldsymbol{\alpha}_j^T \boldsymbol{x}_i$

$$var(z_j) = \alpha_j^T \boldsymbol{S}_x \alpha_j \to \max_{\boldsymbol{\alpha}_j}$$

with side constraints $||\boldsymbol{\alpha}_j|| = 1, \boldsymbol{\alpha}_j^T \boldsymbol{\alpha}_s = 0, s = 1, \ldots, j$.

The problem of maximizing the empirical covariance S is formally the same as maximizing the covariance Σ. Therefore the solution is given by the spectral decomposition of S

$$\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{L}\boldsymbol{Q}^T$$

where the columns of $\boldsymbol{Q} = (\boldsymbol{q}_1 \ldots, \boldsymbol{q}_p)$ are the eigenvectors of S and $L = diag(l_1, \ldots, l_p)$ is a diagonal matrix with the eigenvalues $l_1 \geq \ldots \geq l_p$ of S. By use of $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_p$ the original data $\boldsymbol{x}_1, \ldots \boldsymbol{x}_n$ are transformed to the vector of principal components

$$\boldsymbol{z}_i = \boldsymbol{Q}^T \boldsymbol{x}_i.$$

It is easy to show that for the empirical covariance matrix for data $\boldsymbol{z}_1, \ldots \boldsymbol{z}_n$ one obtains

$$\boldsymbol{S}_z = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{z}_i - \bar{\boldsymbol{z}})(\boldsymbol{z}_i - \bar{\boldsymbol{z}})^T = \boldsymbol{L}.$$

Thus the principal components are uncorrelated and the empirical variance of the jth principal component is given by $s_j^2 = l_j$. Moreover, in analogy to the decomposition of the underlying covariance matrix $\boldsymbol{\Sigma}$ one obtains for the empirical covariance matrix:

$$tr(\boldsymbol{S}_x) = tr(\boldsymbol{S}_z), \qquad |\boldsymbol{S}_x| = |\boldsymbol{S}_z|.$$

Principal Components for Observations

The vector valued principal components are given by $\boldsymbol{z}_i = \boldsymbol{Q}^T\boldsymbol{x}_i$, $i = 1, \ldots, n$, where

$$\boldsymbol{S}_x = \boldsymbol{Q}\boldsymbol{L}\boldsymbol{Q}^T$$

is the spectral decomposition with $\boldsymbol{Q} = (\boldsymbol{q}_1 \ldots \boldsymbol{q}_p)$ containing the eigenvectors of $\boldsymbol{S}_x$ as columns and $\boldsymbol{L} = diag(l_1, \ldots, l_p)$ being the diagonal matrix of eigenvalues of $\boldsymbol{S}_x$.

$\boldsymbol{S}_z = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{z}_i - \bar{\boldsymbol{z}})(\boldsymbol{z}_i - \bar{\boldsymbol{z}})^T = \boldsymbol{L}$

$tr(\boldsymbol{S}_x = tr(\boldsymbol{S}_z))$, $|\boldsymbol{S}_x| = |\boldsymbol{S}_z|$

1.3 Estimation

Principal components in the random variable and the observation case are based on the spectral decompositions

$$\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T \qquad \text{and} \qquad \boldsymbol{S}_x = \boldsymbol{Q}\boldsymbol{L}\boldsymbol{Q}^T.$$

The corresponding eigenvectors and eigenvalues $\boldsymbol{q}_i$, l_i from $\boldsymbol{Q}$ and $\boldsymbol{L}$ may be considered as estimates of $\boldsymbol{p}_i$, λ_i from $\boldsymbol{P}$ and $\boldsymbol{\Lambda}$. If one assumes that observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are iid and normally distributed one obtains for $n\boldsymbol{S}_x$

$$n\boldsymbol{S}_x \sim W(\boldsymbol{\Sigma}, n-1).$$

If for the underlying eigenvalues $\lambda_1 > \ldots > \lambda_p$ holds, it can be shown that asymptotically $(n \to \infty)$ for the vector $\hat{\boldsymbol{\lambda}}^T = (l_1, \ldots, l_p)$ one has

$$\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \sim N(\boldsymbol{0}, 2\boldsymbol{\Lambda}^2).$$

For $\hat{\boldsymbol{\alpha}}_j = \boldsymbol{q}_j$ one obtains

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_j) \sim N(0, \lambda_j \sum_{\substack{s \\ s \neq j}}^{p} \frac{\lambda_s}{(\lambda_j - \lambda_s)^2}\boldsymbol{\alpha}_s\boldsymbol{\alpha}_s^T)$$

(see Anderson, 2003, Section 13.5).