A note on the general solution for a projection matrix in latent factor models

J. Groß K. Luebke * C. Weihs

June 2002

Universität Dortmund
Fachbereich Statistik

Abstract

In this paper the general solution for a projection matrix on latent factors fulfilling the condition $(XG)'(XG) = I_r$ is found and proved. The practical importance of this lemma is outlined.

Keywords
Latent Factor Models, Projection matrix, Generalized Inverse

1 Introduction

Latent factors are used in many multivariate statistical methods like reduced rank regression or principal component analysis.
In reduced rank regression (see Schmidli (1995) or Reinsel and Velu (1998)) the different techniques like Partial Least Squares, Canonical Correlation Analysis or Redundancy Analysis differ only in the way they project the original variables on latent factors (see Schmidli (1995), page 61). To achieve a prediction optimal projection, the mean squared error of prediction (MSEP)

*e-mail: luebke@statistik.uni-dortmund.de

can be written as a function of the projection matrix (see Weihs and Hothorn (2002), page 6). All possible projection matrices must fulfill the side-condition that the latent factors are orthonormal (see Schmidli (1995), page 55). To do a computer intensive minimization of the MSEP it is therefore useful to know the space of possible solutions of the side-condition. Within this solution space the MSEP can be minimized for example by means of simulated annealing (see Lübke (2002), page 27pp).
Knowledge about the structure of the general solution is also useful to construct an experimental design for a simulation study in order to test and compare different regression techniques. To obtain general results (and not just compare one specific example) characteristics of the underlying model and the projection matrix can be varied. This can, however, only be done if the aforementioned structure is known (see Lübke (2002), page 30pp).

This paper is organized as follows: In section 2 the reduced rank regression model is briefly introduced. Section 3 presents and proves a lemma on the general solution of the projection matrix on latent factors. In section 4 we look at a more operable form of the general solution space.

2 The latent factor model

The basic multivariate linear model looks as follows

$$Y = 1_n \mu' + XM + E \tag{1}$$

where

- $Y \in I\!R^{n \times q}$ is the data of the response variables.
- $\mu \in I\!R^{q}$ is the mean vector of the responses.
- $X \in I\!R^{n \times k}$ is the data of the explanatory variables. For simplicity it is assumed that X is mean centered.
- $M \in I\!R^{k \times q}$ is the unknown regression coefficient matrix.
- $E \in I\!R^{n \times q}$ is the matrix of the errors.

If there are many response and explanatory variables but only a small number of observations is available the number of parameters to estimate in the regression coefficient must be reduced. But even if there are 'enough' data

points available there is often the well known problem of over-fitting. Especially when the explanatory variables X are correlated there are numerical problems in estimating the regression coefficient (see Belsley et al. (1980), page 114).
Also, besides the possible statistical problems in the basic model there might be hypothetical unobservable 'latent' variables.

To solve the statistical problems it is often assumed that instead of the variables X in the model (1) so-called latent variables Z under the side-condition $Z'Z = I_r$ ($r < k$) with $Z = XG$ are used (compare for example Schmidli (1995) or Reinsel and Velu (1998)).
The model with latent factors looks as follows:

$$
\begin{aligned}
Y - 1_n\mu' = XM + E &= (XG)B + E =: ZB + E \\
\text{with the side condition} & \\
(XG)'(XG) &= I_r.
\end{aligned}
$$

Then, like with Partial Least Squares, Principal Component Regression, Canonical Correlation Analysis and Redundancy Analysis the least squares estimates of the regression coefficients of the response variables on the latent factors is the least squares estimate are of the form: $\hat{B} = Z'(Y - 1_n\hat{\mu}')$.

The above methods differ in the way they estimate the projection matrix G. To compare them in different situations the feasible projection matrices must be investigated.

3 General solution of $(XG)'(XG) = I_r$

In this section the general structure of the solution of the side-condition is revealed in the following lemma.

Lemma

Let $X \in I\!R^{n\times k}$. The general solution for $(XG)'(XG) = I_r$ is

$$G = X^+A + (I_k - X^+X)C \tag{2}$$

where $A \in I\!R^{n\times r}$ is any matrix with $span(A) \subset span(X)$ and $A'A = I_r$, and $C \in I\!R^{n\times r}$ is arbitrary.

Note that the conditions for A imply $r < k$ and $r \leq rank(X)$.
The lemma consists of two parts: The first (i) says, that any G with $G = X^+A + (I_k - X^+X)C$ and A according to the conditions is a solution of $(XG)'(XG) = I_r$. The other part (ii) says, that all solutions G can be written in the form of equation (2).

Proof

(i) G in equation (2) is a solution.
From $span(A) \subset span(X)$ immediately follows

$$\exists G \quad \text{so that} \quad XG = A. \tag{3}$$

This means that equation (3) has got a solution. The general solution is (see Harville (1997), page 141)

$$G = X^+A + (I_k - X^+X)C. \tag{4}$$

From (3) it follows that

$$(XG)'(XG) = A'A.$$

But according to the lemma $A'A = I_r$ so:

$$(XG)'(XG) = A'A = I_r.$$

(ii) Every solution G can be written as $\tilde{G}$ with $\tilde{G} = X^+A + (I_k - X^+X)C$.
Let $A := XG$ and $C := G$.
Now the conditions on A are fulfilled:

$$span(A) = span(XG) \subset span(X).$$

Also

$$A'A = (XG)'(XG) = I_r.$$

And $\tilde{G} = G$, since:

$$\begin{aligned} \tilde{G} &= X^+XG + (I_k - X^+X)G \\ &= X^+XG + G - X^+XG = G. \end{aligned}$$

This completes the proof. $\square$

When X is of full colum rank the general solution becomes $G = X^+A$, since $X^+X = (X'X)^{-1}X'X = I_k$.

4 Practical Considerations

Knowing the structure of the general solution of the side-condition one only needs to implement this either in an optimization algorithm or in an experimental design for the latent factor model. To make the assumptions in the lemma on A more practical one can use $A = U_X F$ where $U_X \in I\!R^{n\times p}$ (p being the rank of X) stems from the singular value decomposition of $X = U_X L_X V_X'$, and $F \in I\!R^{p\times r}$ an arbitrary orthonormal matrix. Then the assumptions on A are given as

$$\begin{aligned} span(A) &= span(U_X F) \\ &\subset span(U_X) \\ &= span(X) \end{aligned}$$

and

$$\begin{aligned} A'A &= (U_X F)'(U_X F) \\ &= F'U_X'U_X F \\ &= F'I_p F \\ &= I_r. \end{aligned}$$

With this it is easy to fulfill the side condition $Z'Z = I_r$: One only needs an arbitrary orthornomal matrix F. When r is smaller than p it is always possible to (post-)ortho-normalize a full rank $p \times r$ matrix F for example by the Gram-Schmidt method. The condition of $r < p$ is fulfilled in the common problem when X is of full-column rank and one looks for a latent factor model with fewer latent factors than there are explanatory variables. An application of this method for the case that X is of full column rank is given in depth by Lübke (2002).

Acknowledgment

This work has been supported by the Collaborative Research Center 'Reduction of Complexity for Multivariate Data Structures' of the German Research Foundation (DFG).

References

David A. Belsley, Edwin Kuh, and Roy E. Welsch. *Regression diagnostics.* Wiley, 1980.

David A. Harville. *Matrix Algebra From a Statisticians's Perspective.* Springer, 1997.

Karsten Lübke. Optimale Prognose mit latenten Faktoren am Beispiel der Differenzierungsgüte verschiedener chemotaxonomischer Marker in komplexen biologischen Matrizen, Diplomarbeit, Universität Dortmund, 2002.

Gregory C. Reinsel and Raja P. Velu. *Multivariate Reduced-Rank Regression, Theory and Applications.* Springer, 1998.

Heinz Schmidli. *Reduced Rank Regression.* Physica Verlag, 1995.

Claus Weihs and Torsten Hothorn. Determination of optimal prediction oriented multivariate latent factor models using loss functions. Technical Report 15, Sonderforschungsbereich 475, Univerität Dortmund, 2002.