MATRIX APPROXIMATION BY POLYNOMIAL FACTOR ANALYSIS

JAN DE LEEUW AND KEKONA SORENSON

ABSTRACT. A technique is proposed to fit multivariate polynomials of a number of orthonormal factors to a data matrix.

1. PROBLEM

Suppose Y is an $n \times m$ data matrix, with n measurements on m variables. The problem we study in this paper is minimization of the loss function

$$\sigma(A, X) = \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - P_j(x_{i1}, \cdots, x_{ir}))^2 \tag{1}$$

where the P_j are multivariate polynomials of r variables, and X is an $n \times r$ orthonormal matrix of *factor scores*, i.e. $X'X = I$.

Multivariate polynomials are of the form

$$P(x_1, \cdots, x_r) = \sum_{i_1=0}^{d}\sum_{i_2=0}^{d}\cdots\sum_{i_r=0}^{d} a_{i_1 i_2 \cdots i_r}\prod_{k=1}^{r} x_k^{i_k},$$

where the convention is that $0^0 = 1$. Thus we see that each polynomial can be identified with an r-dimensional array $\mathcal{A}$, as in Hankin [2009]. The array is taken to be square, i.e. it has $(d + 1)$ rows,

Date: Friday 11th November, 2011 — 15h 41min — Typeset in LUCIDA BRIGHT.

Key words and phrases. Factor Analysis, Principal Component Analys, Matrix Approximation.

columns, slices, and so on. This does not cause any loss of generality, since an arbitrary number of elements of $\mathcal{A}$ can be zero.

If we define vectors

$$\overline{x}_k = \begin{bmatrix} 1 \\ x_k \\ x_k^2 \\ \vdots \\ x_k^d \end{bmatrix}$$

then P can also be written as the multilinear form $\mathcal{A}(\overline{x}_1, \cdots, \overline{x}_r)$.

Loss function (1) says we approximate the data values y_{ij} by the values of m polynomials P_j at the factor scores $(x_{i1}, \cdots, x_{ir})$. For each variable j we generally require different values of the coefficients (i.e. elements of the array $\mathcal{A}_j$) to be equal to zero. We have to minimize our loss functions, as indicated, over the orthonormal factor scores X and over the polynomial coefficients, of which some may be constrained to be zero.

2. Factors

We say that a polynomial P of n variables $x_1, \cdots, x_n$ *depends on* variable x_k if the partial derivative $\mathcal{D}_k P$ is not identically equal to zero. Using this definition we can divide the n factor variables into two groups. An index k defines a *common factor* if more than one of the P_j depend on x_k, it is a *unique factor* if exactly one of the P_j depends on x_k.

In *polynomial common factor analysis* or *PCFA* there are p common factors and m unique factors. Without loss of generality we can let these groups correspond with the first p and the next m columns of X. In *polynomial principal component analysis* or *PPCA* there are p common factors and no unique factors.

The problem of minimizing (1) without assuming orthogonality of the factors, or with the assumption that only some factors are orthogonal to others, is also of interest, but it requires different techniques and will not be discussed further here.

In our form of factor analysis the factor scores are fixed, similar to what is done in the nonlinear factor analysis techniques of **?**McDonald [1967a,b]; Etezadi-Amoli and McDonald [1983]. Nonlinear random factor score models have been proposed by Mooijaart and Bentler [1985] and by Amemiya and his students Yalcin [1995]; Wall [1998]; Wall and Amemiya [2000]; Yalcin and Amemiya [2001]. Least squares fitting of the linear factor model with fixed scores was discussed earlier, with extensions to general linear structural equation models, in De Leeuw [2004].

3. Examples

In *linear principal component analysis* there is a $r \leq m$ such that

$$P_j(x_{i1}, \cdots, x_{ir}) = \sum_{s=1}^{r} a_{js} x_{is},$$

and in *polynomial principal component analysis*

$$P_j(x_{i1}, \cdots, x_{ir}) = Q_j(x_{i1}, x_{i2}, \cdots, x_{ir}),$$

where the Q_j are polynomials of r variables.

In *linear common factor analysis* there is a $p \leq m$ such that $r = p + m \leq n$ and

$$P_j(x_{i1}, \cdots, x_{ir}) = \sum_{s=1}^{p} a_{js} x_{is} + a_{j,p+j} x_{i,p+j}.$$

In *polynomial common factor analysis*

$$P_j(x_{i1}, \cdots, x_{ir}) = Q_j(x_{i1}, x_{i2}, \cdots, x_{ip}, x_{i,p+j}),$$

where the Q_j are polynomials in $p+1$ variables. The $x_{i1}, x_{i2}, \cdots, x_{ip}$ are the *common factor scores*, the $x_{i,p+j}$ are the *unique factor scores* for variable j. In the linear case the a_{js} with $s \leq p$ are the *common*

factor loadings, and $a_{j,p+j}$ is the *unique factor loading* for variable j. The square $a^2_{j,p+j}$ is the *unique variance* or the *uniqueness*. In the nonlinear case there are factor loadings for all terms in the polynomials Q_j.

One simple polynomial common factor model, with two common factors, for example, is

$$\begin{aligned} P_j(x_{i1},\cdots,x_{ir}) &= \\ &= a_{j1} + a_{j2}x_{i1} + a_{j3}x_{i2} + \alpha_{j4}x_{i1}^2 + a_{j5}x_{i2}^2 + a_{j6}x_{i1}x_{i2} + \\ &\qquad + a_{j,6+j}x_{i,2+j} \end{aligned}$$

In this model there are 6 common factor loadings, corresponding with the 6 terms of a general bivariate quadratic.

4. Algorithm

It would be interesting to find out if techniques from computational algebraic geometry, symbolic computation, or semi-definite programming can be used to solve the problem of minimizing $\sigma(A,X)$, but in this paper we take a more simple-minded algorithmic approach. The algorithm is of the *Alternating Least Squares* type, in which minimization over A for given X and minimization over X for given A are alternated. It probably makes sense to compute an initial X using the singular value decomposition.

Minimization over A for given orthonormal X is a simple linear least squares problem, which can be solved for each j separately, even if some of the coefficients are constrained to be zero.

Minimization over X for given A is done by the majorization technique explained by De Leeuw and Groenen [2011, Section]. It suffices to observe that for given A the loss function (1) is itself a polynomial on $n \times r$ variables, which we have to minimize over $X'X = I$.

To apply majorization we need an expressions for the partial derivatives for respect to X. DefineClearly

$$(D)_1\sigma(X,A) = -2\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - P_j(x_i))\mathcal{D}P_j(X)$$

4.1. **Interactions.** See McDonald [1967c].

4.2. **Difficulty Factors.** See McDonald [1965]; McDonald and Ahlawat [1974].

5. Examples

6. Observed Properties of the Algorithm

6.1. **Loss as a function of θ.** Give some graphs of the functions optimized in the Jacobi step. Unimodal ?

6.2. **Convergence.** Describe convergence in qualitative terms.

6.3. **Local Minima.** From different starting points. From the corresponding linear model.

6.4. **Optimal Cycling through Substeps.** The question: when do we optimize over A. After each Jacobi step ? After update all common factors ? After updating all factors ?

7. Generalizations

7.1. **Weights.**

$$\sigma(A,X) = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}(y_{ij} - P_j(x_{i1},\cdots,x_{in}))^2, \tag{2}$$

where $w_{ij} \geq 0$ are given weights.

7.2. **Optimal Scaling.** This combines the algorithm for fitting X and A with optimal scaling (for example, monotone transformations) of the columns of Y. Thus we extend the optimal scaling techniques for nonlinear principal component analysis [De Leeuw, 2006a] to polynomial factor analysis. Note that the linear special case is FACTALS, at least the version of FACTALS that directly fits the model to the data matrix, and not the covariance matrix.

7.3. **Aspects.** In the recent literature there have been some proposals to minimize

$$\log\det(R'R)$$

where $R\overset{\Delta}{=}Y-P(X)$ are the residuals. This can be tackled along the lines of the aspect approach [De Leeuw, 1990; Mair and De Leeuw, 2010], using the majorization

$$\log\det(C)\leq\log\det(\tilde{C})+\mathbf{tr}\ \tilde{C}^{-1}(C-\tilde{C}).$$

Thus in a majorization step we minimize the weighted least squares function

$$\mathbf{tr}\ R(\tilde{R}'\tilde{R})^{-1}R'.$$

7.4. **Majorization for Logit and Probit Models.** For binary data this uses the loss function

$$\mathcal{L}=\prod_{i=1}^{n}\prod_{j=1}^{m}\pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}},$$

where

$$\pi_{ij}=F(P_j(x_{i1},\cdots,x_{in}))$$

and $F()$ is either the probit or the logit function. We then use majorization, as in De Leeuw [2006b]; De Leeuw and Lange [2009].

Appendix A. Code

References

J. De Leeuw. Multivariate Analysis with Optimal Scaling. In S. Das Gupta and J. Sethuraman, editors, *Progress in Multivariate Analysis*, Calcutta, India, 1990. Indian Statistical Institute.

J. De Leeuw. Least Squares Optimal Scaling of Partially Observed Linear Systems. In K. van Montfort, J. Oud, and A. Satorra, editors, *Recent Developments on Structural Equation Models*, chapter 7. Kluwer Academic Publishers, Dordrecht, Netherlands, 2004. URL http://preprints.stat.ucla.edu/360/lserr.pdf.

J. De Leeuw. Nonlinear Principal Component Analysis and Related Techniques. In M. Greenacre and J. Blasius, editors, *Multiple Correspondence Analysis and Related Methods*. Chapman and Hall, 2006a.

J. De Leeuw. Principal Component Analysis of Binary Data by Iterated Singular Value Decomposition. *Computational Statistics and Data Analysis*, 50(1):21–39, 2006b.

J. De Leeuw and P. Groenen. Majorizing a Multivariate Polynomial over the Unit Sphere, with Applications. Preprint Series 630, UCLA Department of Statistics, Los Angeles, CA, 2011. URL http://preprints.stat.ucla.edu/630/polymaj.pdf.

J. De Leeuw and K. Lange. Sharp Quadratic Majorization in One Dimension. *Computational Statistics and Data Analysis*, 53:2471–2484, 2009.

J. Etezadi-Amoli and R.P. McDonald. A Second Generation Nonlinear Factor Analysis. *Psychometrika*, 48:315–342, 1983.

R.K.S. Hankin. *multipol: multivariate polynomials*, 2009. URL http://CRAN.R-project.org/package=multipol. R package version 1.0-4.

P. Mair and J. De Leeuw. A General Framework for Multivariate Analysis with Optimal Scaling: The R Package aspect. *Journal of*

Statistical Software, 32(9):1–23, 2010.

R. P. McDonald. Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika*, 32:77–112, 1967a.

R.P. McDonald. Difficulty Factors and Nonlinear Factor Analysis. *British Journal of Mathematical and Statistical Psychology*, 18:11–23, 1965.

R.P. McDonald. *Nonlinear Factor Analysis*. Number 15 in Psychometric Monographs. Psychometric Society, 1967b.

R.P. McDonald. Factor Interaction in Nonlinear Factor Analysis. *British Journal of Mathematical and Statistical Psychology*, 20: 209–215, 1967c.

R.P. McDonald and K.S. Ahlawat. Difficulty Factors in Binary Data. *British Journal of Mathematical and Statistical Psychology*, 27:82–99, 1974.

A. Mooijaart and P. Bentler. Random Polynomial Factor Analysis. In E. Diday et al., editor, *Data Analysis and Informatics*, volume IV, pages 241–250, 1985.

M. Wall. *On Nonlinear Structural Equation Analysis*. PhD thesis, Iowa State University, December 1998.

M. Wall and Y. Amemiya. Estimation for Polynomial Structural Equation Models. *Journal of American Statistical Association*, 95: 929–940, 2000.

I. Yalcin. *Nonlinear Factor Analysis*. PhD thesis, Iowa State University, 1995.

I. Yalcin and Y. Amemiya. Nonlinear Factor Analysis as a Statistical Method. *Statistical Science*, 16:275–294, 2001.

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu

E-mail address, Kekona Sorenson: kekona@stat.ucla.edu

URL, Kekona Sorenson: http://www.stat.ucla.edu/~kekona