RR – 85-18

Principal Component Analysis with a singular weight matrix

Jan de Leeuw

1: Problem

Suppose Y is a given $n \times m$ matrix, A is a given $m \times m$ positive semidefinite matrix. Keller and Wansbeek study the minimization of

$$\sigma(X) = \operatorname{tr} (X - Y)A(X - Y)',$$

over all X that satisfy $\operatorname{rank}(X) \leq p$ and $(X - Y)AA^{+} = X - Y$. Here p is a given integer, and A^{+} is the Moore-Penrose inverse of A. The motivation for this problem is explained by Keller and Wansbeek, it derives from maximum likelihood estimation in a linear errors-in-variables model.

Keller and Wansbeek use parametric versions of the rank-constraints, in combination with undetermined multipliers and matrix derivative calculus. The rank constraints can be formulated either as $XB = 0$, with B an $m \times (m - p)$ matrix of rank $m - p$, or as $X = GH'$, with G an $n \times p$ matrix and H an $m \times p$ matrix. The two formulations lead to two conditional minimization problems, and Keller and Wansbeek show after a lot of algebra that the solutions to these two problems are indeed the same. Nevertheless their treatment of the problem does not seem to be completely satisfactory. Questions of existence are not settled clearly, the full-rank constraint on B is more or less ignored, only necessary conditions for an extreme value are actually used. The two parametrizations of the rank-constraint on X lead to two different and lengthy proofs, which are purely mechanical and not very illuminating.

In this note we propose a different method to study the same problem, which does not seem to have the disadvantages of the tools used by Keller and Wansbeek.

2: Notation

Define $r = \operatorname{rank}(A)$ and $s = \operatorname{rank}(Y)$. The matrix K contains eigenvectors corresponding with nonzero eigenvalues of A, thus K is $m \times r$, and we choose K such that $K'K = I$. The nonzero eigenvalues are collected in the $r \times r$ diagonal matrix Λ^2, they are ordered in such a way that $AK = K\Lambda^2$. We also use L, which is an orthonormal basis for the null-space of A, i.e. $K'L = 0$ and $L'L = I$, while L is $m \times (m - r)$. Also define $U = YL$, $t = \operatorname{rank}(U)$, and $V = YK\Lambda$.

To define our problem we let

$\sigma_* = \inf \{\sigma(X) \mid X \in \mathfrak{J}\}$,

with

$\mathfrak{J} = \{X \in \mathbf{R}^{n\times m} \mid \text{rank}(X) \leq p \;\&\; XL = U\}$,

and thus $\sigma_* = +\infty$ if $\mathfrak{J}$ is empty. Also define

$\mathfrak{J}_* = \{X \in \mathfrak{J} \mid \sigma(X) = \sigma_*\}$.

Our problem is to compute σ_* and $\mathfrak{J}_*$.

3: Existence

Theorem 1: $\mathfrak{J}$ is nonempty if and only if $p \geq t$.

Proof: The general solution to $XL = U$ is $X = ZK' + UL'$, with $Z \in R^{n\times r}$ and otherwise arbitrary. For any X of this form we have $\text{rank}(X) = \text{rank}(Z \mid U)$. By varying Z we can make rank(X) equal to any value between t and $\min(t + r,n)$. Thus $\text{rank}(X) \leq p \;\&\; XL = U$ is possible if and only if $p \geq t$. Q.E.D.

Theorem 2: $\mathfrak{J}_*$ is nonempty if and only if $p \geq t$.

Proof: Necessity is obvious from theorem 1. To prove sufficiency we define, for some arbitrary $\varepsilon > 0$,

$\sigma_\varepsilon(X) = \text{tr } (X - Y)(A + \varepsilon^2 LL')(X - Y)'$.

If $X \in \mathfrak{J}$ then $\sigma_\varepsilon(X) = \sigma(X)$. Thus $\mathfrak{J}_*$ also is the set of minimizers of $\sigma_\varepsilon(X)$ on $\mathfrak{J}$. Moreover $YLL' \in \mathfrak{J}$ because $\text{rank}(YLL') = \text{rank}(U) = t \leq p$, and $(YLL')L = U$. Because $\sigma_\varepsilon(YLL') = \sigma(YLL') = \text{tr } YAY'$, the set $\mathfrak{J}_*$ is also the set of minimizers of $\sigma_\varepsilon(X)$ on $\mathfrak{J} \cap \{\sigma_\varepsilon(X) \leq \text{tr } YAY'\}$, which is a compact neighborhood of Y, showing that $\mathfrak{J}_*$ is nonempty. Q.E.D.

4: Construction

Define (assuming from now on that $p \geq t$)

$\sigma_*(\varepsilon) = \inf \{\sigma_\varepsilon(X) \mid \text{rank}(X) \leq p\}$,

$\mathfrak{J}_*(\varepsilon) = \{X \in \mathbf{R}^{n\times m} \mid \text{rank}(X) \leq p \;\&\; \sigma_\varepsilon(X) = \sigma_*(\varepsilon)\}$.

Suppose $\varepsilon_1, \varepsilon_2, \ldots$ is an increasing unbounded sequence of positive numbers, and set $\sigma_k = \sigma_*(\varepsilon_k)$ and $\mathfrak{J}_k = \mathfrak{J}_*(\varepsilon_k)$.

Theorem 3: a: $\lim_{k\to\infty} \sigma_k = \sigma_*$.

b: If $X_k \in \mathfrak{J}_k$ for all k, and X_* is a subsequential limit of $X_1, X_2, \ldots$, then $X_* \in \mathfrak{J}_*$.

Proof: It is clear that σ_k increases, and that $\sigma_k \leq \sigma_*$ for all k. Thus

σ_k converges to, say, $\sigma_{**} \leq \sigma_*$.

Define $\tau_k = \mathrm{tr}\,(X_k - Y)LL'(X_k - Y)'$. Thus $\sigma_k = \sigma(X_k) + \varepsilon_k^2\tau_k \leq \sigma_*$, or $\varepsilon_k^2\tau_k \leq \sigma_* - \sigma_k$. Taking the subsequential limit shows that $\varepsilon_k^2\tau_k$ remains bounded by $\sigma_* - \sigma_{**}$, which implies that τ_k converges to zero, and thus that $X_* \in \mathcal{J}$. Because $\sigma(X_*) = \sigma_{**} \leq \sigma_*$ this means that $X_* \in \mathcal{J}_*$, and $\sigma_{**} = \sigma_*$. Q.E.D.

5: Computation

To compute the limits we need some extra notation. We write $(Z)_p$ for a best rank p approximation of the matrix Z in the ordinary least squares norm. Thus $(Z)_p$ minimizes $\mathrm{tr}\,(X - Z)(X - Z)'$ over all X with $\mathrm{rank}(X) \leq p$. If $\mathrm{rank}(Z) \leq p$, then obviously $(Z)_p = Z$. If $\mathrm{rank}(Z) > p$, then $\mathrm{rank}((Z)_p) = p$, but $(Z)_p$ is not necessarily unique. We compute $(Z)_p$ by 'truncating' the singular value decomposition of Z, keeping only the p largest singular values. If the singular values of Z are ordered as $\psi_1 \geq \psi_2 \geq \ldots \geq \psi_m$, and if $\mathrm{rank}(Z) > p$, then $(Z)_p$ is unique if and only if $\psi_p > \psi_{p+1}$. These are all familiar results, which can be found, for example, in review papers by Corsten (1976) and Rao (1980).

The results can be used directly to compute $\sigma_*(\varepsilon)$ and $\mathcal{J}_*(\varepsilon)$. We find

$$\sigma_*(\varepsilon) = \sum_{q=p+1}^{m} \psi_q^2(V \mid \varepsilon U) = \varepsilon^2 \sum_{q=p+1}^{m} \lambda_q(UU' + \frac{1}{\varepsilon^2} VV'),$$

where we use $\psi_q(\ )$ for the ordered singular values, and $\lambda_q(\ )$ for the ordered eigenvalues of a matrix. We continue to assume that $p \geq t$, we also assume that $s > p$ (otherwise the problem has the obvious solution $X = Y$). Define the $n \times (n - t)$ matrix $\mathbf{R} = (G|H)$ which satisfies $\mathbf{R}'\mathbf{R} = I$ and $\mathbf{R}'U = 0$. The partitioning of $\mathbf{R}$ is such that in addition $H'Y = 0$ and H is $n \times (n - s)$. Thus G is $n \times (s - t)$.

Perturbation theory (Kato, 1966, Baumgärtel, 1972) now makes it possible to describe eigenvalues of $UU' + \varepsilon^{-2}VV'$ in somewhat more detail. The t largest eigenvalues are $\lambda_q(UU') + O(\varepsilon^{-2})$, the next s - t eigenvalues are $\varepsilon^{-2}\lambda_q(G'VV'G) + o(\varepsilon^{-2})$, and the final n - s eigenvalues are zero. It follows that

$$\sigma_* = \sum_{q=1+p-t}^{s-t} \lambda_q(G'VV'G).$$

The elements of $\mathfrak{J}_*(\varepsilon)$ are of the form

$$(\tfrac{1}{\varepsilon}V \mid U)_p \begin{pmatrix} \varepsilon\Lambda^{-1} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} K' \\ L' \end{pmatrix}.$$

This follows directly for ordinary rank-p matrix approximation theory with nonsingular weighting matrices (for example Rao, 1980, section 3). We now need limit theory for eigenvectors or eigenprojections, which is a bit more complicated than the corresponding theory for eigenvalues. Nevertheless all the necessary results are readily available (Kato, 1966, chapter 2).

We use the fact that $(Z)_p = T_pT_p'Z$, where T_p are eigenvectors corresponding with the p largest eigenvalues of ZZ' (observe that T_p may not be unique). The eigenvectors corresponding with the t largest eigenvalues of $UU' + \varepsilon^{-2}VV'$ are asymptotically the eigenvectors corresponding with the nonzero eigenvalues of UU'. The remaining p - t eigenvectors are of the form GM, where M is an $(s - t) \times (p - t)$ matrix of eigenvectors corresponding with the p - t largest eigenvalues of G'VV'G. Again we have to be a bit careful here, because M may not be uniquely defined if there are multiple eigenvalues. In this case the p - t eigenvectors and their corresponding projection may not converge, and we use the fact that the limit of any convergent subsequence is of the form GM. For such a subsequence

$$(\tfrac{1}{\varepsilon}V \mid U)_p = (UU^+ + GMM'G')(\tfrac{1}{\varepsilon}V \mid U) + o(\varepsilon^{-1}),$$

and thus

$$X = (UU^+ + GMM'G')Y$$

is in $\mathfrak{J}_*$. Because $UU^+ = I - GG' - HH'$ it also follows that

$$X = (I - GNN'G')Y,$$

where N is an $(s - t) \times (s - p)$ matrix of eigenvectors corresponding with the s - p smallest eigenvalues of G'VV'G.

6 Dual derivation

In stead of using $(Z)_p = T_pT_p'Z$, with T_p the eigenvectors corresponding with the p largest eigenvalues of ZZ', we can also use $(Z)_p = ZS_pS_p'$, with S_p eigenvectors corresponding with the p largest eigenvalues of Z'Z. If we apply this we find

$$(\tfrac{1}{\varepsilon}V \mid U)_p = (\tfrac{1}{\varepsilon}V \mid U)\{\begin{pmatrix} EE' & 0 \\ 0 & DD' \end{pmatrix} + \tfrac{1}{\varepsilon}\begin{pmatrix} 0 & FF'V'(U^+)' \\ U^+VFF' & 0 \end{pmatrix} + o(\varepsilon^{-1})\}.$$

This is considerably more complicated as the corresponding expression in the previous section, which is due to the row-column asymmetry of our problem. In the formula E is an $r \times (p - t)$ matrix of eigenvectors corresponding with the $p - t$ largest eigenvalues of $V'GG'V$, F is an $r \times (r - (p - t))$ matrix of eigenvectors corresponding with the $r - (p - t)$ smallest eigenvalues, $E'F = 0$, and D is an $(m - r) \times t$ matrix of eigenvectors corresponding with the nonzero eigenvalues of $U'U$ (i.e. $DD' = U^+U$). The same comments about multiple eigenvalues apply as in the previous section. In the previous section we assumed that $s > p$ to avoid trivial complications, in this section we assume for the same reason that $p - t < r$ (again, if $p - t \geq r$, then $X = Y$ is a solution, because $s \leq r + t$).

It follows that

$$X = Y(I - K\Lambda FF'\Lambda^{-1}K' + LU^+VFF'\Lambda^{-1}K'),$$

is in $\mathcal{J}_*$. In this connection it is of some interest to see that $U^+V = 0$ if Y is of full column rank, in which case the expression for X simplifies to

$$X = Y(I - K\Lambda FF'\Lambda^{-1}K').$$

Both in the general and in the special case the matrix with which we postmultiply Y to get X is idempotent, but not necessarily symmetric. It is symmetric if $\Lambda = I$, i.e. if A is a projector.

We also remark that $V'GG'V = V'(I - HH' - UU^+)V = V'(I - UU^+)V = 0$. Thus if Y is of full column rank, then $V'GG'V = V'V$. In general it is consequently not necessary to compute G if we apply the method of this section, if $s = m$ it is also not necessary to compute U.

7 Relationship with simultaneous diagonalization

We use the results of De Leeuw (1981) here. Define $W = (W_1 \mid W_2)$, with W_1 of order $m \times r$ and W_2 of order $m \times (m - r)$, by

$$W_1 = (K\Lambda - LU^+V)(E \mid F),$$

$$W_2 = LQ,$$

where Q is a complete set of eigenvectors of $U'U$. Then

$$W_1'A^+W_1 = I,$$

$$W_2'A^+W_2 = 0,$$

$$W_1'Y'YW_1 = \Omega,$$

$$W_2'Y'YW_2 = \Xi,$$

$$W_1'Y'YW_2 = 0,$$

$$W_2'Y'YW_1 = 0,$$

where Ω are the eigenvalues of $V'G'GV$ and Ξ are the eigenvalues of $U'U$, both collected in diagonal matrices. Thus W diagonalizes both A^+ and $Y'Y$, moreover W is nonsingular (it is not necessarily the unique matrix with these properties, cf De Leeuw, 1981).

Using the formula for X derived in the previous section we find

$$XW_1 = (I - UU^+)V(I - FF')(E \mid F) = (I - UU^+)V(E \mid 0),$$

$$XW_2 = UQ.$$

Thus XW has $(r - (p - t)) + ((m - r) - t) = m - p$ columns equal to zero, and possibly more if $(I - UU^+)VE$ has zero columns. More precisely, assuming that $p - t < r$ as usual, the number of zero columns in XW is $(m - p) + ((p - t) - \text{rank}(GV)) = m - s$.

Because

$$W_1'X'XW_1 = \begin{pmatrix} \Omega & 0 \\ 0 & 0 \end{pmatrix},$$

$$W_1'X'XW_2 = 0,$$

$$W_2'X'XW_1 = 0,$$

$$W_2'X'XW_2 = \Xi,$$

with $\underline{\Omega}$ the $p - t$ largest eigenvalues of $V'G'GV$, it follows that W also diagonalizes $X'X$.

Thus the results of section 6 make it possible to give the principal-relations representation of X in the sense of Keller and Wansbeek directly. Not surprisingly the results of section 5 can be used to find a principal components representation. We write $UU^+ = JJ'$, where J is the $n \times t$ matrix of eigenvectors corresponding with nonzero eigenvalues of UU'. Thus

$$X = (J \mid GM)\begin{pmatrix} J'Y \\ M'G'Y \end{pmatrix}.$$

In this decomposition of X the matrix on the left is orthonormal, the matrix on the right is A-orthogonal.

8 Further results

Examples, rate of convergence results, and some generalizations will be published in a later version of this paper.

References:

W.J. Keller & T. Wansbeek: Multivariate methods for quatitative and qualitative data.
Rep. 4834-81-M1, Central Bureau of Statistics, Department for statististical methods, Voorburg, The Netherlands

C.R. Rao: Matrix approximation and reduction of dimensionality in multivariate statistical analysis.
In P.R. Krishnaiah (ed): Multivariate analysis V Amsterdam, North Holland Publishing Co, 1980.

L.A.C. Corsten: Matrix approximation: a key to multivariate methods.
Paper presented at Biometric conference, Boston, 1976.

J. de Leeuw: Generalized eigenvalue problems with positive semidefinite matrices.
Paper accepted for publication in Psychometrika.

T. Kato: Perturbation theory for linear operators.
Berlin, Springer, 1966[1], 1976[2].

H. Baumgärtel: Endlichdimensionale analytische Störungstheorie.
Berlin, Akademie Verlag, 1972.