Behaviormetrika, Vol. 25, No. 1, 13-23
1998

# ORTHOMAX ROTATION PROBLEM. A DIFFERENTIAL EQUATION APPROACH[1)]

Moody T. Chu*[2)] and Nickolay T. Trendafilov**[3)]

In the present paper the ORTHOMAX rotation problem is reconsidered. It is shown that its solution can be presented as a steepest ascent flow on the manifold of orthogonal matrices. A matrix formulation of the ORTHOMAX problem is given as an initial value problem for matrix differential equation of first order. The solution can be found by any available ODE numerical integrator. Thus the paper proposes a convergent method for direct matrix solution of the ORTHOMAX problem.

The well-known first order necessary condition for the VARIMAX maximizer is reestablished for the ORTHOMAX case without using Lagrange multipliers. Additionally new second order optimality conditions are derived and as a consequence an explicit second order necessary condition for further classification of the ORTHOMAX maximizer is obtained.

## 1. Introduction

In factor analysis a decomposition of the sample $n\times n$ covariance/correlation matrix $C$ is sought in the form:

$$C=AA^T+\Phi,$$

where $A$ is $n\times p(n\gg p)$ matrix of factor loadings and $\Phi$ is $n\times n$ diagonal positive definite matrix of unique variances. Such a representation is not unique. Let $Q$ be any orthogonal $p\times p$ matrix and $B=AQ$. Then $BB^T=AQQ^TA^T=AA^T$, which means that the matrix of the factor loadings $B$ yields the same covariance/correlation matrix as $A$. This indeterminicity of the factor solution leads to the problem of finding the "best" transformation (rotation) $Q$ of the factor loadings such that the rotated factors to have a structure, which to be as simple as possible for interpretation. Detailed consideration of the factor simplicity concept and many different methods for factor rotation can be found in Mulaik (1972).

The VARIMAX rotation method (Kaiser, 1958; Mulaik, 1972) is the most popular one to achieve an orthogonal simple structure solution. The core of the method is to make most of the loadings in each factor of near zero magnitude and

*Keywords*: ORTHOMAX rotation, Matrix solution, Projected gradient, Optimality conditions
* Department of Mathematics, North Carolina State University, Raleigh, North Carolina, 27695-8205 U.S.A.
** Computer Stochastic Laboratory, Institute of Mathematics, Bulgarian Academy of Sciences, P.O. Box 373, Sofia 1090, BULGARIA
[1)] Proofs should be sent to Nickolay Trendafilov, Katholieke Universiteit Leuven, Department of Electrical Engineering (ESAT), Signals, Identification, System Theory and Automation (SISTA) Kardinaal Mercierlaan 94, B-3001 Leuven, BELGIUM.
[2)] This research was supported in part by National Science Foundation under grant DMS-9422280.
[3)] This research was performed while visiting the Department of Educational Psychology, Nagoya University, JAPAN and was supported by Japan Society for the Promotion of Science.

only few of them of near unity. In VARIMAX this purpose is realized by the maximizing the sum of the within-factor variances of squared factor loadings. Let $A$ be an $n \times p$ matrix of initial factor loadings. The VARIMAX method seeks an $p \times p$ orthogonal matrix such that the matrix of transformed factor loadings $B = AQ$ maximizes the function:

$$v^2 = \sum_{j=1}^{p} v_j^2 = \sum_{j=1}^{p} \left[ \frac{1}{n} \sum_{i=1}^{n} (b_{ij}^2)^2 - \frac{1}{n^2} \left( \sum_{i=1}^{n} b_{ij}^2 \right)^2 \right]. \tag{1}$$

The first algorithm for finding the maximum of the VARIMAX function works iteratively, rotating in turn all possible pairs (planar rotation) of factors and maximizing their VARIMAX functions. In the mid sixties P. Horst and R. Sherin (Mulaik, 1972) offered independently, in the same time, a matrix formulation of the VARIMAX function and derived a SVD-based iterative algorithm for rotation of all factors simultaniously. Unfortunately, the matrix iterative process has not been given a rigorous proof of convergence. Also its computational advantages over the standard VARIMAX are quite unclear; see for details ten Berge (1984). Further simulation comparison of the two approaches showed the planar rotation to be computationaly more effecient than the matrix one (ten Berge, et al., 1988).

More elaborative matrix formulation of the VARIMAX rotation problem has been given by Neudecker (1981) (see also Magnus, J.R., & Neudecker, H., 1988). The work also includes a first order necessary condition for the VARIMAX maximizer. Ten Berge (1984) facilitated this matrix formulation of the VARIMAX problem by avoiding Hadamard products, reestablished the Neudecker's necessary condition and gave additionally second order optimality conditions. The VARIMAX rotation problem, as formulated by ten Berge (1984), is concerned with the following constrained optimization problem:

$$\text{Maximize} \quad F(Q) = n^{-3} trace \left[ \sum_{i=1}^{n} (diag(Q^T E_i Q))^2 \right] \tag{2}$$

$$\text{Subject to} \quad Q \in R^{p \times p},\ Q^T Q = QQ^T = I_p, \tag{3}$$

where $E_i = A^T A - n a_i a_i^T$, $a_i^T$ denotes the $i$-th row of $A$ and $diag\ (Q^T E_i Q)$ denotes a diagonal matrix containing the main diagonal of $Q^T E_i Q$. In these notations the VARIMAX problem is equivalent to the problem for simultaneous diagonalization of symmetric matrices. In (ten Berge, et al., 1988) this approach has been spread over the ORTHOMAX rotation family—a set of rotation methods optimizing certain criterial functions, for which it has been proven by Mulaik (1972) to form one-parameter family. It has been demonstrated that the ORTHOMAX family deals with the maximizing (2) subject to (3) for $E_i(\lambda) = \lambda A^T A - n a_i a_i^T$, where $\lambda \in [0, 1]$. For example, $\lambda = 1$ yields the VARIMAX method, $\lambda = 0$—QUARTIMAX method and e.t.c. By means of this ORTHOMAX formulation the Horst-Sherin SVD-based iterative algorithm has been generalized and reconsidered and its convergence has been proven in case of positive/negative semi-definite $E_i$ for all $i$. It has been also shown how after "repairment" of the indefinite $E_i$s the same algorithm can serve still convergently, even "painfuly slowly". Ten Berge, et al. (1988) concluded that the SVD-approach is of no practical value as rotation procedure and pointed out as its possible application the solving of INDSCAL with orthogonality constraints. Kiers (1990) proposed a majorization algorithm for solving the ORTHOMAX rotation problem which is monotone convergent regardless of the nature of the matrices $E_i$ and thus is a generalizatoin of (ten Berge, et al., 1988).

In the present paper the ORTHOMAX rotation problem is reconsidered. It is shown that its solution can be presented as a steepest ascent flow on the manifold of orthogonal matrices. This approach has been applied successfully in the analysis of certain discrete numerical methods. A short list of such considerations can be found in a recent review by Chu (1994) and detailed study—in Helmke, U. and Moore, J.B. (1994). We follow here the ideas discussed in Chu and Driessel (1990), and Chu and Trendafilov (1996). A matrix formulation of the ORTHOMAX problem is given as an initial value problem for matrix differential equation of first order. Its solution can be found by any available ODE numerical integrator (Shampine & Reichelt, 1995). The method is globally convergent, i.e. the convergence is reached independently of the starting (initial) point. Thus, the paper proposes a new *convergent* method for direct *matrix* solution of the ORTHOMAX problem, no matter what the initial $E_i$ are. Additionally we reestablish the first order condition known from the works of Neudecker (1981) and ten Berge (1984) for the VARIMAX case without using Lagrange multipliers. Next we derive new second order optimality conditions. As a consequence we obtained an explicit second order necessary condition for further classification of the ORTHOMAX maximizer. The presented method can be used for solving INDSCAL with orthogonality constraints and to remedy the convergence problems appearing when SVD-approach is applied, as reported in the simulation study in (ten Berge, et al., 1988).

We do not aim simulation comparison with the existing methods for rotation. The present method, as well as the SVD-approach (ten Berge, et al., 1988) can not compete the planar rotation from computational point of view. That is not surrprising—most of the recent methods in the computational matrix analysis are based on Housholder and Givens transformations which are also planar rotations. Our purpose here is to present a rather universal theoretically powerful method for solving least squares optimization problems subject to constraints which form a smooth manifold (e.g. the set of all orthogonal matrices, the set of all "vertical" matrices with orthonormal columns, the set of all oblique rotations and etc). As one can concieve most of the problems in the area of data analysis can be represented mathematically as problems of that kind. Reconsidering here the well-known ORTHOMAX rotation problem we want to illustrate the new approach that gives a way for both solving the problem and obtaining theoretical results for the solution behavior.

## 2. Steepest descent flow

We first introduce a topology for the set

$$\mathcal{O}(p) := \{Q \in R^{p\times p} | Q^T Q = QQ^T = I_p\} \tag{4}$$

of all $p\times p$ orthogonal matrices. It is known that $\mathcal{O}(n)$ forms a smooth manifold of dimension $p(p-1)/2$ in $R^{p\times p}$. Indeed, any vector tangent to $\mathcal{O}(p)$ at $Q\in\mathcal{O}(p)$ is necessarily of the form $QK$ for some skew-symmetric matrix $K\in R^{p\times p}$. Denote

$$S(p) := \{\text{all symmetric matrices in } R^{p\times p}\},$$

and introduce the Frobenius inner product of two matrices $X$ and $Y$:

$$\langle X, Y\rangle := \text{trace } (XY^T).$$

It follows that the tangent space $T_Q\mathcal{O}(p)$ and the normal space $N_Q\mathcal{O}(p)$ of $\mathcal{O}(p)$ at any $Q\in\mathcal{O}(p)$ are given, respectively, by:

$$T_Q\mathcal{O}(p) = QS(p)^{\perp} \tag{5}$$

$$N_Q\mathcal{O}(p) = QS(p), \tag{6}$$

where $S(p)^{\perp}$ is the orthogonal complement of $S(p)$ with respect to the Frobenius inner product and hence consists of all skew-symmetric matrices in $R^{p\times p}$.

Apparently, the ORTHOMAX problem is equivalent to the maximization of the function:

$$F_\lambda(Q) = n^{-3} trace\left[\sum_{i=1}^{n} (diag(Q^T E_i(\lambda) Q))^2\right] \tag{7}$$

over the feasible set $\mathcal{O}(p)$. Hereafter we write for short $F(Q)$ for $F_\lambda(Q)$ and $E_i$ for $E_i(\lambda)$ because the parameter $\lambda$ does not affect the further considerations. Moreover the specific form of the marices $E_i$ is also of no importance but only their symmetricity. This is a standard constrainted optimization problem, but we shall show below that we can obtain the information about the projected gradient and the projected Hessian without using the conventional Lagragian multipliers technique.

We first calculate the gradient $\nabla F(Q)$ of the objective function $F(Q)$ to be:

$$\nabla F(Q) = 4n^{-3}\sum_{i=1}^{n} E_i Q \; diag\,(Q^T E_i Q). \tag{8}$$

Suppose the projection $g(Q)$ of the gradient $\nabla F(Q)$ at a point $Q\in\mathcal{O}(p)$ onto the tangent space $T_Q\mathcal{O}(p)$ can be computed explicitly. Then the differential equation

$$\frac{dQ}{dt} = +g(Q) \tag{9}$$

naturally defines the steepest ascent flow for the function $F$ on the feasible set $\mathcal{O}(p)$. To obtain this projected gradient $g(Q)$, observe that

$$R^{p\times p} = T_Q\mathcal{O}(p) \oplus N_Q\mathcal{O}(p) = QS(p)^{\perp} \oplus QS(p). \tag{10}$$

Therefore, any matrix $X\in R^{p\times p}$ has an unique orthogonal decomposition:

$$X=Q\left[\frac{1}{2}(Q^TX-X^TQ)\right]+Q\left[\frac{1}{2}(Q^TX+X^TQ)\right]$$

as the sum of elements from $T_Q\mathcal{O}(p)$ and $N_Q\mathcal{O}(p)$. In particular, the projection $g(Q)$ of $\nabla F(Q)$ onto the tangent space $T_Q\mathcal{O}(p)$ has the form:

$$g(Q)=2n^{-3}Q\sum_{i=1}^{n}(Q^TE_iQ\ diag\ (Q^TE_iQ)-diag\ (Q^TE_iQ)Q^TE_iQ). \tag{11}$$

We therefore obtain the differential equation

$$\frac{dQ}{dt}=2n^{-3}Q\sum_{i=1}^{n}(Q^TE_iQ\ diag\ (Q^TE_iQ)-diag\ (Q^TE_iQ)Q^TE_iQ) \tag{12}$$

that defines a steepest ascent flow on the manifold $\mathcal{O}(p)$ for the objective function $F$ in (8). Starting with any point in $\mathcal{O}(p)$, say $Q(0)=I$, we may follow the flow defined by (12) by any available initial value problem solver. The flow eventually will converge to a local solution for the ORTHOMAX. This is the ready-made numerical algorithm referred to above for the ORTHOMAX.

For convenience we adopt Lie bracket notation $[X, Y]=XY-YX$ and define

$$X_i=Q^TE_iQ \tag{13}$$

for $i=1, 2, \cdots, n$. It can be easily checked that each $X_i$ must satisfy the following differential equation

$$\begin{aligned}\frac{dX_i}{dt}&=\frac{dQ^T}{dt}E_iQ+Q^TE_i\frac{dQ}{dt}\\&=2n^{-3}\left[X_i, \sum_{k=1}^{n}[X_k, diag\ (X_k)]\right].\end{aligned} \tag{14}$$

Chu (1991) have proved that the flow $X_i(t)$ stays on the *isospectral surface* $M(E_i)=\{Q^TE_iQ|Q\in\mathcal{O}(p)\}$ if it starts from some $X_i(0)\in M(E_i)$. One natural choice of the initial value will be $X_i(0)=E_i$. The system (14) is another way for solving the ORTHOMAX problem (which is computationally prefarable because does not need at any step explicit check of the orthogonality of Q). In (Chu, 1991) has been shown that the solution flow is a continuous analog of the classical Jacobi method for simultanious diagonalization of symmetric matrices.

**Example:** As reported in (ten Berge, 1984) the SVD-approach fails to rotate the matrix

$$A=\begin{bmatrix}1 & 1\\1 & 0\\0 & 1\\0 & 0\end{bmatrix}$$

and to maximize the VARIMAX function. By solving (14) that can be done, as well as the Kaiser's method does it (much faster).

## 3. Optimality conditions

The explicit formulation of the projected gradient (11) provides additional information about the first order optimality condition for a stationary point:

**Theorem 3.1.** *A necessary condition for $Q \in \mathcal{O}(p)$ to be a stationary point of the ORTHOMAX problem is that the matrix $\sum_{i=1}^{n} Q^T E_i Q \, diag\,(Q^T E_i Q)$ is symmetric.*

**Proof.** Obviously $Q$ is a stationary point only if $g(Q)=0$. The assertion then follows from (11). □

This necessary condition is well-known from Neudecker (1981) and ten Berge (1984) for the VARIMAX case ($\lambda=1$).

We also can derive a second order optimality condition to further classify the stationary points. We claim that

**Theorem 3.2.** *The action of the projected Hessian of $F$ at a stationary point $Q \in \mathcal{O}(p)$ on a tangent vector $QK$ where $K$ is skew-symmetric is given by*

$$\begin{aligned} \langle g'(Q)QK, QK \rangle = 4n^{-3} \sum_{i=1}^{n} \{ & \langle Q^T E_i Q \, diag\,(Q^T E_i Q), K^2 \rangle \\ & + \langle Q^T E_i QK \, diag\,(Q^T E_i Q), K \rangle \\ & + 2\langle Q^T E_i Q \, diag\,(Q^T E_i QK), K \rangle \}. \end{aligned} \tag{15}$$

**Proof.** From (11), observe that the Fréchet derivative of $g$ at $Q$ on a general $H$ is given by

$$\begin{aligned} g'(Q)H := 2n^{-3} \sum_{i=1}^{n} [ & H(Q^T E_i Q \, diag\,(Q^T E_i Q) \\ & - diag\,(Q^T E_i Q) Q^T E_i Q) \\ & + Q((H^T E_i Q + Q^T E_i H) \, diag\,(Q^T E_i Q) \\ & + Q^T E_i Q \, diag\,(H^T E_i Q + Q^T E_i H) \\ & - diag\,(H^T E_i Q + Q^T E_i H)(Q^T E_i Q) \\ & - diag\,(Q^T E_i Q)(H^T E_i Q + Q^T E_i H))]. \end{aligned}$$

At a stationary point, by Theorem 3.1, the quantity in the first big paratheses in the above is zero. The Hessian action on a tangent vector $H=QK$ can be calculated as follows:

$$\begin{aligned} \langle g'(Q)QK, QK \rangle = 2n^{-3} \sum_{i=1}^{n} \{ & \langle (K^T Q^T E_i Q + Q^T E_i KQ) \, diag\,(Q^T E_i Q), K \rangle \\ & + \langle Q^T E_i Q \, diag\,(K^T Q^T E_i Q + Q^T E_i KQ), K \rangle \\ & - \langle diag\,(K^T Q^T E_i Q + Q^T E_i KQ)(Q^T E_i Q), K \rangle \\ & - \langle diag\,(Q^T E_i Q)(K^T Q^T E_i Q + Q^T E_i KQ), K \rangle \}. \end{aligned}$$

The assertion follows from the adjoint property $\langle XY, Z \rangle = \langle Y, X^T Z \rangle$ and the fact that $K^T = -K$. □

The following characterization is a standard result in optimization theory. See, for example, Gill, Murray, and Wright (1981, 3.4.)

**Corollary 3.3.** *A second order sufficient (necessary) condition for a stationary point $Q \in \mathcal{O}(p)$ to be a maximizer of the ORTHOMAX problem is that*

$$\sum_{i=1}^{n}\{\langle Q^T E_i Q \; diag\,(Q^T E_i Q),\, K^2\rangle + \langle Q^T E_i QK \; diag\,(Q^T E_i Q),\, K\rangle \\ +2\langle Q^T E_i Q diag(Q^T E_i QK),\, K\rangle\} < (\leq) 0 \tag{16}$$

*for all nonzero $K \in S(p)^{\perp}$.*

**Example**: As a numerical illustration of the Corollary 3.3 we compute here two simple examples, taken from (ten Berge, 1984). Consider the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix},$$

for which the VARIMAX function attains minimum. We will check that $Q = I_2$ is a stationary point and yields a minimum. We have

$$E_1 = E_2 = \begin{bmatrix} 0 & -4 \\ -4 & 0 \end{bmatrix}$$

and

$$E_3 = E_4 = \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix}.$$

The necessary condition (Theorem 3.1) for $Q = I_2$ to be a stationary point is fulfilled. In deed, we have:

$$\sum_{i=1}^{4} Q^T E_i Q \; diag\,(Q^T E_i Q) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{17}$$

Now we check the condition (16) from Corollary 3.3. There are only two $2\times 2$ skew-symmetric matrices $K$. We consider

$$K = \begin{bmatrix} 0 & k \\ -k & 0 \end{bmatrix}.$$

The first two terms in (16) are zero. The third term gives

$$\sum_{i=1}^{4} \langle Q^T E_i Q \; diag\,(Q^T E_i QK),\, K\rangle = \left\langle \begin{bmatrix} 0 & 64k \\ -64k & 0 \end{bmatrix}, \begin{bmatrix} 0 & k \\ -k & 0 \end{bmatrix} \right\rangle \\ = 128\, k^2, \tag{18}$$

which is stritctly positive for $k \neq 0$. The same result follows for

$$K = \begin{bmatrix} 0 & -k \\ k & 0 \end{bmatrix}.$$

Therefore $\sum_{i=1}^{4} \langle Q^T E_i Q \; diag\,(Q^T E_i QK),\, K\rangle$ is positive for all nonzero $K \in S(2)^{\perp}$, i.e. from Corollary 3.3 follows that $Q = I_2$ is sufficiently minimum of the

VARIMAX function.

Now consider the rotation

$$Q=\begin{bmatrix} \sqrt{.5} & \sqrt{.5} \\ \sqrt{.5} & -\sqrt{.5} \end{bmatrix}.$$

We will check that this $Q$ is a stationary point and yields a maximum for the VARIMAX function. We have

$$Q^T E_1 Q = Q^T E_2 Q = \begin{bmatrix} -4 & 0 \\ 0 & 4 \end{bmatrix}$$

and

$$Q^T E_3 Q = Q^T E_4 Q = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}.$$

The necessary condition (Theorem 3.1) for $Q$ to be a stationary point is fulfilled. Indeed, we have:

$$\sum_{i=1}^{4} Q^T E_i Q \; diag\,(Q^T E_i Q) = \begin{bmatrix} 64 & 0 \\ 0 & 64 \end{bmatrix}. \tag{19}$$

Now we check the condition (16) from Corollary 3.3 for:

$$K=\begin{bmatrix} 0 & k \\ -k & 0 \end{bmatrix}.$$

Note that

$$\begin{aligned} Q^T E_1 QK = Q^T E_2 QK &= \begin{bmatrix} -4 & 0 \\ 0 & 4 \end{bmatrix} \times \begin{bmatrix} 0 & k \\ -k & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -4k \\ -4k & 0 \end{bmatrix} \end{aligned} \tag{20}$$

and

$$\begin{aligned} Q^T E_3 QK = Q^T E_4 QK &= \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix} \times \begin{bmatrix} 0 & k \\ -k & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 4k \\ 4k & 0 \end{bmatrix}. \end{aligned} \tag{21}$$

That means the third term in (16) is zero. The first two terms give:

$$\begin{aligned} &\sum_{i=1}^{4} \{\langle Q^T E_i Q \; diag\,(Q^T E_i Q),\, K^2\rangle + \langle Q^T E_i QK \; diag\,(Q^T E_i Q),\, K\rangle\} = \\ &\left\langle \begin{bmatrix} 64k & 0 \\ 0 & 64k \end{bmatrix}, \begin{bmatrix} -k^2 & 0 \\ 0 & -k^2 \end{bmatrix} \right\rangle + \left\langle \begin{bmatrix} 0 & -64k \\ 64k & 0 \end{bmatrix}, \begin{bmatrix} 0 & k \\ -k & 0 \end{bmatrix} \right\rangle \\ &= -256\,k^2. \end{aligned} \tag{22}$$

The same result follows for

$$K=\begin{bmatrix} 0 & -k \\ k & 0 \end{bmatrix}.$$

Therefore the left hand side of the inequality (16) is negative for all nonzero $K\in S(2)^{\perp}$, i.e. from Corollary 3.3 follows that the considered $Q$ is sufficiently maximum of the VARIMAX function.

Let the singular value decomposition of the skew-symmetric matrix $K$ be denoted by

$$K=U\Sigma W^T, \tag{23}$$

where $U$, $W\in\mathcal{O}(p)$, and $\Sigma=diag\{\sigma_1, \cdots, \sigma_p\}$ contains singular values. It follows that

$$K^2=-U\Sigma^2U^T \tag{24}$$

which in fact is the spectral decomposition of $K^2$. We know from Theorem 3.1 that $\sum_{i=1}^{n} Q^TE_iQdiag(Q^TE_iQ)$ is necessarily symmetric at any stationary point $Q$. Let

$$\sum_{i=1}^{n} Q^TE_iQ\ diag\,(Q^TE_iQ)=V\Lambda V^T, \tag{25}$$

and similarly,

$$Q^TE_iQ=W_i\Phi_iW_i^T,$$

denote, respectively, the spectral decomposition of the corresponding matrix. It is easy to see that $\Phi_i$ actually contains the eigenvalues of $E_i$ for all $i$. Noting that $tr(X\ diag\,(X))=tr((diag\,(X))^2)$ for every square matrix $X$ we have:

$$\langle Q^TE_iQ\ diag\,(Q^TE_iQK),\ K\rangle=tr((diag\,(Q^TE_iQK))^2). \tag{26}$$

Then we can rewrite

$$\begin{aligned}\langle g'(Q)QK,\ QK\rangle&=-\langle V\Lambda V^T,\ U\Sigma^2U^T\rangle+\sum_{i=1}^{n}[\langle\Phi R\Psi,\ R\rangle\\&\quad+2tr((diag\,(Q^TE_iQK))^2)]\\&=-\sum_{k=1}^{p}\lambda_k\Big(\sum_{t=1}^{p}p_{kt}^2\sigma_t^2\Big)+\sum_{i=1}^{n}\Big[\sum_{j=1}^{p}\phi_{i,j}\Big(\sum_{s=1}^{p}r_{i,js}^2\psi_{i,s}\Big)\\&\quad+2tr((diag(Q^TE_iQK))^2)\Big],\end{aligned} \tag{27}$$

where $P=(p_{kt}):=V^TU$, $R_i=(r_{i,js}):=W_i^TK$, $\Phi_i:=(\phi_{i,j})$ and $diag\,(Q^TE_iQ):=\Psi_i:=(\psi_{i,s})$. Then we can rewrite (16) as follows:

$$\begin{aligned}&\sum_{i=1}^{n}\Big[\sum_{j=1}^{p}\phi_{i,j}\Big(\sum_{s=1}^{p}r_{i,js}^2\psi_{i,s}\Big)+2tr((diag\,(Q^TE_iQK))^2)\Big]\\&<(\leq)\sum_{k=1}^{p}\lambda_k\Big(\sum_{t=1}^{p}p_{kt}^2\sigma_t^2\Big).\end{aligned} \tag{28}$$

We can make the following claim:

**Theorem 3.4.** *Suppose $Q\in\mathcal{O}(p)$ is a stationary point and suppose $\Phi_i\mathbf{1}\mathbf{1}^T\Psi_i$*

$\left(\mathbf{1}^T = \underbrace{(1, 1, \cdots, 1)}_{p}\right)$ *is a nonnegative matrix for all i. Then a second order necessary condition for Q to be a solution of the ORTHOMAX problem is that the matrix*

$$\sum_{i=1}^{n} Q^T E_i Q \, diag\,(Q^T E_i Q) \tag{29}$$

*be positive semi-definite.*

**Proof.** If $\Phi_i \mathbf{1}\mathbf{1}^T \Psi_i$ is a nonnegative matrix for all $i$ then both therms in the left hand side of the inequality (28) are non-negative. In order to mantain the inequality (28) for any skew-symmetric $K \neq 0$ it is necessary to have $\lambda_i \geq 0$ for all $i$. □

Note that the matrix $\Phi_i \mathbf{1}\mathbf{1}^T \Psi_i$ is nonnegative always when $Q^T E_i Q$ is positive or nagative semi-definite, upper or lower triangular or diagonal matrix. For example, we have the following trivial:

**Corollary 3.5.** *If Q simultaniously diagonalizes $E_1, E_2, \cdots, E_n$ in the least-square sense then*

$$\sum_{i=1}^{n} Q^T E_i Q \, diag\,(Q^T E_i Q) \tag{30}$$

*be necessarily positive semi-definite.*

In fact it is diagonal matrix with nonnegative entries.

In the present work a new method for least squares matrix optimization subject to orthogonality constraints has been demonstrated and applied for solving the well-known ORTHOMAX rotation problem. In spite of the fact that first and second order optimality conditions has been obtained by the gradient flow approach, as a conclusion, it appears that the method is not efficient computationally. This drawback is not restricted to the present application only and other integrators are under active reseach in order to avoid the direct solving of the flow equation (e.g. Helmke, 1995). Authors thank Dr. Henk Kiers, University of Groningen for the MATLAB code of VARIMAX and SVD-based rotation methods.

## References


Chu, M.T. & Driessel, K.R. (1990). The projected gradient method for least squares matrix approximations with spectral constraints, *SIAM J. Numer. Anal.*, **27**, 1050-1060.

Chu, M.T. (1991). A continuous Jacobi-like approach to the simultanious reduction of real matrices. *Linear Algebra and Its Applications*, **147**, 75-96.

Chu, M.T. (1994). A list of matrix flows with applications, *Fields Institute Communications*, **3**, 87-97.

Chu, M.T. & Trendafilov, N.T. (1996). On a differential equation approach to the weighted orthogonal procrustes problem, to appear in "Statistics & Computing".

Gill, P.E., Murray, W., & Wright, M.K. (1981). *Practical Optimization*. Florida: Academic Press.

Helmke, U. & Moore, J.B. (1994). *Optimization and Dynamical Systems*, London: Springer Verlag.

Helmke, U. (1995). Isospectral matrix flows for numerical analysis, in Moonen, M. & De Moor, B. (Eds.) *SVD and Signal Processing, III-Algorithms, Architectures and Applications*, Amsterdam: Elsevier Science B.V.

Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis, *Psychometrika*,

**23**, 187-200.

Kiers, H.A.L. (1990). Majorization as a tool for optimizing a class of matrix functions, *Psychometrika*, **55**, 417-428.

Neudecker, H. (1981). On the matrix formulation of Kaiser's Varimax criterion, *Psychometrika*, **46**, 343-345.

Magnus, J.R. & Neudecker, H. (1988). *Matrix Differential Calculus with Application in Statistics and Econometrics*, New York: Wiley.

Mulaik, S.A. (1972). *The Foundations of Factor Analysis*. New York: McGraw-Hill.

Shampine, L.F. & Reichelt, M.W. (1995). *The MATLAB ODE suite*, preprint, (also available from anonymous *ftp. mathworks. com : /pub/mathworks/toolbox/matlab/funfun.*)

ten Berge, J.M.F. (1984). A joint treatment of VARIMAX rotation and the problem of diagonalizing symmetric matrices simultaniously in the least-squares sense, *Psychometrika*, **49**, 347-358.

ten Berge, J.M.F., Knol, D.L., & Kiers, H.A.L. (1988). A treatment of the ORTHOMAX rotation family in terms of diagonalization, and a re-examination of a singular value approach to VARIMAX rotation, *Computational Statistics Quarterly*, **3**, 207-217.

(Received February, 1997)