Journal of Global Optimization **28:** 217–228, 2004.

Some optimization problems in multivariate statistics

T. RAPCSÁK
Computer and Automation Research Institute, Hungarian Academy of Sciences, H-1518 Budapest, P.O. Box 63, Hungary (e-mail:rapcsak@oplab.sztaki.hu URL: http://www.oplab.sztaki.lru/rapcsak/)

(Received and accepted in revised form 21 January 2003)

Abstract. Interesting and important multivariate statistical problems containing principal component analysis, statistical visualization and singular value decomposition, furthermore, one of the basic theorems of linear algebra, the matrix spectral theorem, the characterization of the structural stability of dynamical systems and many others lead to a new class of global optimization problems where the question is to find optimal orthogonal matrices. A special class is where the problem consists in finding, for any $2 \leqslant k \leqslant n$, the dominant k-dimensional eigenspace of an $n \times n$ symmetric matrix A in R^n where the eigenspaces are spanned by the k largest eigenvectors. This leads to the maximization of a special quadratic function on the Stiefel manifold $M_{n,k}$. Based on the global Lagrange multiplier rule developed in Rapcsák (1997) and the paper dealing with Stiefel manifolds in optimization theory (Rapcsák, 2002), the global optimality conditions of this smooth optimization problem are obtained, then they are applied in concrete cases.

Key words: Multivariate statistics, Quadratic equality constraints, Smooth optimization, Stiefel manifolds

1. Introduction

Many statistical problems in estimations, testing hypothesis, experimental designs, representation points of distributions, etc., can be formulated as optimization problems (see, e.g., Fang et al., 1996; Rao, 1993). A part of these problems can be solved by the known optimization methods, but a big part may request creating new methodology in optimization. In the paper, some statistical optimization problems are analyzed from a new class of global optimization with the aim of finding optimal orthogonal matrices.

An important procedure in statistics is principal component analysis which locates a new basis in the space of observations so that the new basis vectors be orthonormal and the sum of the variances in the directions of a given number of the new basis vectors be maximum where the variance matrix is symmetric and positive semidefinite.

If the given number is equal to 2 or 3, principal component analysis can be an efficient tool for the visualization of not necessarily square data matrices based on representations of both row- and column-objects. Row-objects from the matrix are plotted as points according to their coordinates in a two- or three-dimensional approximation. Since the columns of the data matrix are related to the basis from which row-objects derive their scores, column-objects are represented in the low-dimensional space as axes pointing in the directions which approximate their orientations in the space of the full matrix.

Gabriel (1971, 1981) and Young et al. (1993) used this technique for statistical visualization as well as Mareschal and Brans (1985) for visualizing decision objects in the interactive solution of multiattribute decision problems. This technique was developed under the name of GAIA analysis (Geometrical Analysis for Interactive Assistance) in the PROMETHEE methods devoted to solve multiattribute decision making problem and the experience with the corresponding PROMCALC & GAIA software seems to be convenient.

If the given number κ is equal to $n,$ one of the basic theorems of linear algebra, the spectral form for symmetric matrices (e.g., Bellmann, 1960) can be obtained. Singular value decomposition (SVD) is an important tool of matrix algebra and statistics that has been applied in a number of areas, for example, principal component analysis and canonical correlation in statistics, solving linear systems, least squares problems and computing the Moore–Penrose generalized inverse of a given matrix in numerical linear algebra, and low rank approximation of matrices. Its origins can be traced back to the work of French and Italian mathematicians in the 1870s. One of its largest fields of application, namely low rank matrix approximation, was first reported on by Eckart and Young (1936) in the first volume of Psychometrika. The matrix algebra and computational aspects of SVD are discussed in Kennedy and Gentle (1980), furthermore, Golub and Kahan (1965), and statistical applications are described in Greenacre (1984).

An important property of dynamical systems is that of structural stability or robustness. In heuristic terms, structural stability refers to the property that the qualitative behaviour of a dynamical system is not changed by small perturbations in its parameters (Helmke and Moore, 1994).

These problems in statistics and practical or theoretical fields can be formulated as smooth nonlinear optimization problems with the same structural properties, namely, the feasible sets are Stiefel manifolds. This class, up to now, has been insufficiently examined — in spite of its importance — either from theoretical or methodological point of view.

In the paper, the aim is to show that principal component analysis widely used in statistics and multiattribute decision making, statistical visualization and singular value decomposition, furthermore, the matrix spectral theorem, and the characterization of the structural stability of dynamical systems have a common root with the determination, for any $2 \leqslant \kappa \leqslant n,$ of the dominant k-dimensional eigenspace of an $n \times n$ symmetric matrix A in R^n where the eigenspaces are spanned by the corresponding eigenvectors. Moreover, all these problems are equivalent to the maximization of the same type of special quadratic functions on a Stiefel manifold.

Based on the global Lagrange multiplier rule developed in Rapcsák (1997) and the paper dealing with Stiefel manifolds in optimization theory (Rapcsák, 2002), the global optimality of these smooth optimization problems is characterized, then the optimality conditions are applied to problems in multivariate statistics. Since the background of the global Lagrange multiplier rule of smooth nonlinear optimization is the tensor approach in the corresponding Riemannian geometry, the proofs of the theorems in statistics are simple consequences of the dominant eigenspace theorem where no special coordinate representation is necessary.

In the Introduction, the statistical problems are briefly recalled, then the geometric background is clarified and a new representation of the Stiefel manifolds fitted to optimization theory is given. Based on this approach, necessary and sufficient, local and global optimality can be given, which lead to the characterization of dominant eigenspaces in Euclidean spaces and to handling the above statistical problems uniformly. Finally, some open questions close the paper.

2. Minimization on Stiefel Manifolds

In 1935, Stiefel introduced a new class of differentiable manifolds, called Stiefel manifolds, consisting of all the orthonormal vector systems $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\kappa \in R^n$ for every pair of positive integers (n, κ), denoted by $M_{n,\kappa}$, where R^n is the n-dimensional Euclidean space and $\kappa \leqslant n$. If $\kappa = 1$, the sphere is obtained, if $\kappa = n$, then the group of the $n \times n$ orthogonal matrices. These two special cases arise the most frequently in differential geometry.

A geometric treatment of the Stiefel manifolds appropriate for optimization theory cannot be found in standard differential geometry references. In the paper, extrinsic coordinates are used for representing the points of the Stiefel manifolds with more parameters than intrinsically are necessary, but this approach seems to be better fitted to the methodology of optimization. The Stiefel manifold $M_{n,\kappa}$ may be imbedded in the nk-dimensional Euclidean space where the use of the induced Riemannian metric is proved to be advantageous for studying optimization problems (Rapcsák, 1997). The choice of this metric differs from that of Edelman et al. (1998) where the Frobenius inner product for $n \times p$ matrices is introduced. More precisely, in the two special cases when $\kappa = 1$ and $\kappa = n$, the induced Riemannian metric and the canonical metric are the same, otherwise, they differ.

Consider the following optimization problem:

$$
\begin{aligned}
&\min f(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\kappa) \\
&\boldsymbol{x}_i^T \boldsymbol{x}_j = \delta_{ij}, \qquad 1 \leqslant i, j \leqslant \kappa \leqslant n, \\
&\boldsymbol{x}_i \in R^n, \qquad i = 1, \ldots, \kappa, \qquad n \geqslant 2,
\end{aligned}
\tag{2.1}
$$

where $f : R^{\kappa n} \to R$ is a twice continuously differentiable function and δ_{ij} is the Kronecker's delta. Since the feasible set is compact and the objective function is continuous, optimization problem (2.1) has at least one global minimum point and one global maximum point, thus several stationary points. The feasible set of problem (2.1), the Stiefel manifold denoted by $M_{n,\kappa}$ for every pair of positive integers (κ, n), can be written as

$$\begin{aligned} &\boldsymbol{x}_i^T\boldsymbol{x}_i = 1, \qquad i = 1, \dots, \kappa, \qquad (2.2a)\\ &\boldsymbol{x}_i^T\boldsymbol{x}_j = 0, \qquad i, j = 1, \dots, k, i \neq j,\\ &\boldsymbol{x}_i \in R^n, \qquad i = 1, \dots, \kappa, n \geqslant 2. \end{aligned}$$

This new and interesting global optimization problem with important theoretical and practical applications was studied in Bolla et al (1998), Edelman et al. (1998), Helmke and Moore (1994) and Rapcsák (2001, 2002). In Rapcsák (2002), the optimality conditions were obtained by the global Lagrange multiplier rule. These latter results are summarized here.

In order to study the geometric structure of the feasible set in (2.1), a new representation of the feasible set was suggested in Rapcsák (2001) providing a decomposition of the feasible set as well. Let us introduce the following notations:

$$\boldsymbol{x} = (\boldsymbol{x}_1^T, \dots, \boldsymbol{x}_\kappa^T)^T \in R^{\kappa n}, J = \{(i,j) | i, j = 1, \dots, \kappa, i < j\},$$

$$C_l = \begin{pmatrix} 0 & \dots & 0 & \dots & 0\\ \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & \dots & I_n & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & \dots & 0 & \dots & 0 \end{pmatrix}, \qquad l = 1.\dots.\kappa,$$

$$C_{ij} = \begin{pmatrix} 0 & \dots & 0 & \dots & 0 & \dots\\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots\\ 0 & \dots & 0 & \dots & I_n & \dots\\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots\\ 0 & \dots & I_n & \dots & 0 & \dots\\ 0 & \dots & 0 & \dots & 0 & \dots \end{pmatrix}, \qquad (i,j) \in J,$$

where C_l, $l = 1, \dots, \kappa$, are $kn \times kn$ blockdiagonal matrices, $C_{ij} kn \times kn$ block matrices, I^n is the identity matrix in R^n, C_l and C_{ij} contain I_n in the l-th diagonal block and in the (i, j) as well as (j, i) blocks, respectively. The $kn \times kn$ symmetric matrices C_{ij} are defined for all the pairs of different indices belonging to J, given by the $k(k$—$1)/2$ combinations of the indices 1,…, κ.

It follows that in the case of a compact Stiefel manifold, the feasible set $M_{n,k}$ given by (2.2a) and (2.2b) is equivalent to

$$\begin{aligned} &h_l(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T C_l \boldsymbol{x} - \frac{1}{2} = 0, \qquad l = 1, \dots, \kappa, \qquad (2.3)\\ &h_{ij}(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T C_{ij} \boldsymbol{x} = 0, \qquad (i,j) \in J,\\ &\boldsymbol{x} \in R^{kn}, \qquad n \geqslant 2. \end{aligned}$$

In the definition of the index set J, the restriction $i < j$ ensures that only one of the identical equalities $h_{ij}(\mathbf{x})=0$ and $h_{ji}(\mathbf{x})=0, i,j=1,\ldots,\ \kappa, i\neq j$ appears in (2.3).

Thus, the feasible set $M_{n,k}$ and its tangent space at the point $\mathbf{x}\in M_{n,\kappa}$ can be described by

$$
\begin{aligned}
&M_{n,\kappa}=\{\boldsymbol{x}\in R^{\kappa n}|h_l(\boldsymbol{x})=0, \quad l=1\ldots,k_i; \quad h_{ij}(\boldsymbol{x})=0, \quad (i,j)\in J\}, \qquad (2.4)\\
&TM_{n,\kappa}(\boldsymbol{x})=\{\boldsymbol{v}\in R^{\kappa n}|\nabla h_l(\boldsymbol{x})\boldsymbol{v}=0, \quad l=1,\ldots,\kappa; \quad \nabla h_{ij}(\boldsymbol{x})\boldsymbol{v}=0,(i,j)\in J\}=\\
&\{\boldsymbol{v}\in R^{\kappa n}|\boldsymbol{x}_l^T\boldsymbol{v}_l=0, \quad l=1,\ldots,\kappa; \quad \boldsymbol{x}_i^T\boldsymbol{v}_j+\boldsymbol{x}_j^T\boldsymbol{v}_i=0, \quad (i,j)\in J\},\\
&\qquad\qquad\qquad\qquad \boldsymbol{x}\in M_{n,\kappa},
\end{aligned}
$$

where the symbol ∇ denotes the gradient vector of a function which is a row vector. Intuitively, the tangent space at a point is the κ-dimensional vector space tangent to the submanifold with origin at the point of tangency. The normal space is the orthogonal complement. On the sphere, tangents are perpendicular to radii, and the normal space is radial.

The following statements characterize the structure of the feasible set.

THEOREM 1. (17). *The set $M_{n,\kappa}$ is a compact C^∞ differentiable manifold (Stiefel manifold) with dimension $kn-\frac{\kappa(\kappa+1)}{2}$ for every pair of positive integers (κ,n) satisfying $\kappa\leqslant n$. The Stiefel manifolds are connected if $k<n$. In cases $\kappa=n$, the Stiefel manifolds are of two components.*

By using the equality representations of the compact Stiefel manifolds $M_{n,\kappa}$, problem (2.1) is equivalent to

$$
\begin{aligned}
&\min f(\boldsymbol{x}) \qquad\qquad (2.5)\\
&h_l(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TC_l\boldsymbol{x}-\frac{1}{2}=0, \ \ l=1,\ldots,\kappa,\\
&h_{ij}(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TC_{ij}\boldsymbol{x}=0, \ \ (i,j)\in J,\\
&\boldsymbol{x}\in R^{\kappa n}, \ \ n\geqslant 2.
\end{aligned}
$$

Before stating the optimality conditions, the definition of geodesic convex sets is recalled where the geodesic is used in the classical meaning. If M is a Riemannian C^2 manifold, then a set $\mathcal{C}\subset M$ is geodesic convex if any two points of $\mathcal{C}$ are joined by a geodesic belonging to $\mathcal{C}$, moreover, a singleton is geodesic convex. Let us introduce the symmetric matrix function

$$S(\boldsymbol{x}) = \tag{2.6}$$

$$\begin{pmatrix} (\nabla f(\boldsymbol{x})C_1\boldsymbol{x})I_n & \frac{1}{2}(\nabla f(\boldsymbol{x})C_{12}\boldsymbol{x})I_n & \dots & \frac{1}{2}(\nabla f(\boldsymbol{x})C_{1\kappa}\boldsymbol{x})I_n \\ \frac{1}{2}(\nabla f(\boldsymbol{x})C_{12}\boldsymbol{x})I_n & (\nabla f(\boldsymbol{x})C_2\boldsymbol{x})I_n & \dots & \frac{1}{2}(\nabla f(\boldsymbol{x})C_{2\kappa}\boldsymbol{x})I_n \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}(\nabla f(\boldsymbol{x})C_{1\kappa}\boldsymbol{x})I_n & \frac{1}{2}(\nabla f(\boldsymbol{x})C_{2\kappa}\boldsymbol{x})I_n & \dots & (\nabla f(\boldsymbol{x})C_\kappa\boldsymbol{x})I_n \end{pmatrix},$$

$$\boldsymbol{x} = (\boldsymbol{x}_1^T, \dots, \boldsymbol{x}_\kappa^T)^T \in M_{n,\kappa}$$

THEOREM 2. (17). *If the point $\boldsymbol{x}_0 \in M_{n,\kappa}$ is a* (*strict*) *local minimum of problem* (2.1), *then*

$$\nabla f(\boldsymbol{x}_0) = \boldsymbol{x}_0^T S(\boldsymbol{x}_0), \quad and \tag{2.7}$$

$$(Hf(\boldsymbol{x}_0) - S(\boldsymbol{x}_0))_{|TM_{n,\kappa}(\boldsymbol{x}_0)} \tag{2.8}$$

is a positive semidefinite (*definite*) *matrix where the symbol $|TM_{n,\kappa}(\boldsymbol{x}_0)$ denotes the restriction to the tangent space at the point $\boldsymbol{x}_0$ and H the Hessian matrix of a function.*

If $\mathcal{C} \subseteq M_{n,\kappa}$ is an open geodesic convex set, and there exists a point $\boldsymbol{x}_0$ such that

$$\nabla f(\boldsymbol{x}_0) = \boldsymbol{x}_0^T S(\boldsymbol{x}_0), \quad and \tag{2.9}$$

$$(Hf(\boldsymbol{x}) - S(\boldsymbol{x}))_{|TM_{n,\kappa(\boldsymbol{x})}}, \quad \boldsymbol{x} \in \mathcal{C},$$

are positive semidefinite (*definite*) *matrices, then the point $\boldsymbol{x}_0$ is a* (*strict*) *global minimum of the function f on the set $\mathcal{C}$.*

3. Dominant eigenspaces in Euclidean spaces

Let us consider the problem of finding, for any $2 \leqslant \kappa \leqslant n$, the κ-dimensional eigenspace of an $n \times n$ symmetric matrix A in R^n where the eigenspaces are spanned by the κ largest eigenvectors. In the sequel, this subspace is referred to as the dominant κ-dimensional eigenspace of the matrix A for every $2 \leqslant \kappa \leqslant n$. This problem leads to the maximization of a special quadratic function $f: R^{\kappa n} \to R$ on the Stiefel manifold $M_{n,\kappa}$, i.e., under the constraint that vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_\kappa \in R^n, \kappa \leqslant n$, form an orthonormal system.

In the case of a positive definite matrix A, the dominant eigenspace problem, called the total least-squares problem, was analyzed as an optimization problem on the Grassmannian manifolds and it was proved that there is a unique global maximum in the Grassmannian manifold associated to the given number κ *iff* $\lambda_\kappa >$

$\lambda_{\kappa+1}$ where λ_κ and $\lambda_{\kappa+1}$ are the κ-th and $(\kappa+1)$-th eigenvalues of the matrix A, respectively (Byrnes and Willems, 1986). In Helmke and Moore (1994), this approach based on the Morse–Bott theory of differential topology was further developed and a similar result stated for symmetric matrices.

Here, the global optimality conditions for determining the dominant subspaces are derived from Theorem 2. Since the global Lagrange multiplier rule of smooth optimization in R^n is based on the tensor approach in the corresponding Riemannian geometry, the proofs of the theorems in statistics are simple consequences of the dominant eigenspace theorem where no special coordinate representations are necessary.

Consider the following optimization problem:

$$\max f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \mathcal{A}\boldsymbol{x} = \sum_{l=1}^{\kappa} \frac{1}{2}\boldsymbol{x}_l^T A\boldsymbol{x}_l, \tag{3.1}$$

$$\boldsymbol{x} = \left(\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_\kappa^T\right)^T \in M_{n,\kappa} \subseteq R^{\kappa n},$$

where A is the given $n \times n$ real symmetric matrix with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ and $\mathcal{A}$ is the $\kappa_n \times l_n$ blockdiagonal matrix with the diagonal blocks $A, A, \ldots, A$. Now, the dominant eigenspaces depending on κ will be characterized.

THEOREM 3. *An infinite number of global maximum points with the global maximum value* $\sum_{i=1}^{\kappa} \lambda_\kappa$ *exists in problem* (3.1), *and every global maximum point of problem* (3.1) *determines the dominant eigenspace of the matrix A for every* $2 \leqslant \kappa \leqslant n$.

Proof: If the point $\boldsymbol{x} \in M_{n,\kappa}$ is a stationary point of problem (3.1), then by Theorem 2.2,

$$\mathcal{A}\boldsymbol{x} = S(\boldsymbol{x})\boldsymbol{x}, \tag{3.2}$$

where

$$S(\boldsymbol{x}) = \tag{3.3}$$

$$\begin{pmatrix} \boldsymbol{x}_1^T A\boldsymbol{x}_1 I_n & \boldsymbol{x}_1^T A\boldsymbol{x}_2 I_n & \ldots & \boldsymbol{x}_1^T A\boldsymbol{x}_\kappa I_n \\ \boldsymbol{x}_1^T A\boldsymbol{x}_2 I_n & \boldsymbol{x}_2^T A\boldsymbol{x}_2 I_n & \ldots & \boldsymbol{x}_2^T A\boldsymbol{x}_\kappa I_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_1^T A\boldsymbol{x}_\kappa I_n & \boldsymbol{x}_2^T A\boldsymbol{x}_\kappa I_n & \ldots & \boldsymbol{x}_\kappa^T A_\kappa \boldsymbol{x}_\kappa I_n \end{pmatrix},$$

$$\boldsymbol{x} = \left(\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_\kappa^T\right)^T \in M_{n,\kappa}.$$

The first-order condition (3.2) gives that

$$A\boldsymbol{x}_i = (\boldsymbol{x}_i^T A\boldsymbol{x}_i)\boldsymbol{x}_i + \sum_{j=1, j\neq i}^{\kappa} (\boldsymbol{x}_j^T A\boldsymbol{x}_i)\boldsymbol{x}_j, \qquad i = 1, \ldots, \kappa, \tag{3.4}$$

which is equivalent to

$$((\boldsymbol{x}_i^T A\boldsymbol{x}_i)I - A)\boldsymbol{x}_i + \sum_{j=1,j\neq i}^{\kappa} (\boldsymbol{x}_j^T A\boldsymbol{x}_i)\boldsymbol{x}_j = 0, \qquad i = 1,\dots,\kappa, \tag{3.5}$$

An $n \times n$ real symmetric matrix has n real characteristic roots or real eigenvalues, the eigenvectors associated with the distinct characteristic roots are orthogonal and the eigenvectors related to the multiple characteristic roots can be chosen orthogonally. Since the eigenvectors are determined up to a scalar multiple, we can normalize them by (2.2a). By considering all the points $\boldsymbol{x} \in R^{\kappa n}$ consisting of k number of eigenvectors of A in any order as the vector components of the points $\boldsymbol{x}$, we obtain points belonging to the Stiefel manifold $M_{n,k}$. Taking the eigenvalue equations of the matrix A into account, it follows from (3.5) that all these points are stationary points in (3.1). A consequence of this fact is that the number of the stationary points given by the combinations of a complete system of eigenvectors in problem (3.1) is equal to $\binom{n}{k}$ if all the eigenvalues are distinct, or otherwise, the number of the stationary points is infinite for every $2 \leqslant \kappa \leqslant n$.

If the vector components $\boldsymbol{x}_{i,} = 1,\dots,\kappa$, of the stationary points $\boldsymbol{x}$ are of the eigenvectors of the matrix A, and the corresponding eigenvalues are $\boldsymbol{x}_i^T A\boldsymbol{x}_i, i = 1,\dots,k_1$, then by (3.2), the values of the objective function at these stationary points are equal to the sum of the corresponding eigenvalues, i.e., $\boldsymbol{x}^T \mathcal{A}\boldsymbol{x} = \sum_{i=1}^{\kappa} \boldsymbol{x}_i^T A\boldsymbol{x}_i$. By choosing a point of the Stiefel manifold $M_{n,k}$ consisting of the eigenvectors associated to the biggest κ eigenvalues, the value of the objective function is equal to $\sum_{i=1}^{\kappa} \lambda_i$. We show that this value corresponds to the subspace generated by the vector components of the given point of the Stiefel manifold $M_{n,k}$ in R^n.

The trace of a square matrix A is the sum of its diagonal elements; it is denoted by trA. Let us form the orthogonal matrices

$$X = (\boldsymbol{x}_1,\dots,\boldsymbol{x}_\kappa), \quad \boldsymbol{x} = \left(\boldsymbol{x}_1^T,\dots,\boldsymbol{x}_\kappa^T\right)^T \in M_{n,\kappa}. \tag{3.6}$$

By using (3.2) and (3.4), we have that

$$trAXX^T = tr(A\boldsymbol{x}_1,\dots,A\boldsymbol{x}_\kappa)\begin{pmatrix}\boldsymbol{x}_1^T\\ \vdots \\ \boldsymbol{x}_\kappa^T\end{pmatrix} = \tag{3.7}$$

$$tr\sum_{i=1}^{\kappa}\sum_{j=1}^{\kappa}(\boldsymbol{x}_j^T A\boldsymbol{x}_i)\boldsymbol{x}_j\boldsymbol{x}_i^T = \sum_{i=1}^{\kappa}\boldsymbol{x}_i^T A\boldsymbol{x}_i = \boldsymbol{x}^T \mathcal{A}\boldsymbol{x}$$

at every stationary point of problem (3.1), and if T is a $\kappa \times \kappa$ orthogonal matrix, then

$$trAXX^T = trAXTT^TX^T. \tag{3.8}$$

We can conclude that any other points of $M_{n,k}$ spanning the same subspace in R^n as that of a stationary point give the same value of the objective function, so it follows that the number of the global maximum points is infinite.

Now, it will be shown that

$$\boldsymbol{x}^T \mathcal{A}\boldsymbol{x} \leqslant \sum_{i=1}^{\kappa} \lambda_i, \quad \boldsymbol{x} \in M_{n,\kappa}. \tag{3.9}$$

As an orthogonal matrix T exists such that $A = T^T \Lambda T$, where the diagonal matrix $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$, we have that

$$\boldsymbol{x}^T \mathcal{A}\boldsymbol{x} = \sum_{i=1}^{\kappa} \boldsymbol{x}_i^T A \boldsymbol{x}_i = \sum_{i=1}^{\kappa} \boldsymbol{x}_i^T T^T \Lambda T \boldsymbol{x}_i, \quad \boldsymbol{x} \in M_{n,\kappa}. \tag{3.10}$$

Let $\boldsymbol{y} = (\boldsymbol{x}_1^T T^T, \ldots, \boldsymbol{x}_\kappa^T T^T)^T$. Because of the orthogonality of the matrix T, the point $\boldsymbol{y}$ belongs to $M_{n,\kappa}$. Thus,

$$\boldsymbol{x}^T t\boldsymbol{x} = \sum_{i=1}^{k} \boldsymbol{x}_i^T T^T \Lambda T \boldsymbol{x}_i = \sum_{i=1}^{k} \sum_{i=1}^{n} \lambda_j y_{ij}^2 \leqslant \sum_{i=1}^{k} \lambda_i, \quad \boldsymbol{x} \in M_{n,k} \tag{3.11}$$

where we used the orthogonality of the vector components $\boldsymbol{y}_i, i = 1, \ldots, \kappa$, and the fact that any convex combination $\sum_{j=1}^{n} \lambda_j y_{ij}^2$ fulfils the inequality

$$\sum_{j=1}^{n} \lambda_j y_{ij}^2 \leqslant max\{\lambda_j, j = 1, \ldots, n\}, \quad i = 1, \ldots, \kappa, \tag{3.12}$$

which implies the inequality (3.9).

A consequence of inequality (3.9) is that the global maximum value of problem (3.1) is equal to the sum of the k largest eigenvalues, thus the theorem is proved.

We remark that only the first-order optimality conditions are used in the proof, and no difficulty arises from the fact that the Stiefel manifold $M_{n,n}$ is of two components. It is easy to see that the matrix $A - S(\boldsymbol{x})$ is not negative semidefinite at any stationary point, so an interesting corollary of the statement is that this property is not necessary for the global maximality (see Rapcsák, 2002). The matrix function S is positive semidefinite (definite) at any stationary points iff the matrix A is positive semidefinite (definite).

4. Some optimization problems in statistics

In this part, the way how to apply Theorem 3 in concrete problems will be shown. In statistics, principal component analysis locates a new basis in the space of observations so that the new basis vectors be orthonormal and the sum of the variances in the directions of a given number of the new basis vectors be maximal where the variance matrix is symmetric and positive semidefinite.

If a complete, n component solution is specified, principal component analysis generates a model space, exactly the same as the original data space but spanned by one of the optimal systems of basis vectors. In the case of a complete principal component solution, the error space is null. If, however, a k principal component solution ($k < n$) is computed, the model space will consist of the subspace spanned by k components of the complete solution, while the subspace spanning the $(n-k)$ different components of the complete solution will constitute the error space.

In the case where the dimension of principal component model is less than the dimension of the space of observations, a measure showing how the model approximates the data matrix is given by the ratio of the sum of the first k eigenvalues to the sum of all the eigenvalues.

In the cases of $\kappa = 2, 3$, principal component analysis can be an efficient tool for the visualization of not necessarily square data matrices based on representations of both row- and column-objects. Column-objects from the matrix are plotted as points according to their coordinates in a two- or three-dimensional approximation. Since the rows of the data matrix are related to the basis from which column-objects derive their scores, row-objects are represented in the low-dimensional space as axes pointing in the directions which approximate their orientations in the space of the full matrix (Gabriel, 1971, 1981; Young et al., 1993). This technique seems to work efficiently in the PROMCALC & GAIA software, because the measure of approximation was always greater than 0.6, and except for some cases, greater than 0.8 in real-life applications (Mareshal and Brans, 1988).

THEOREM 4. *Principal component analysis, for any $2 \leqslant \kappa \leqslant n$, is equivalent to the determination of the k-dimensional dominant eigenspace with respect to the variance matrix as an $n \times \kappa$ orthogonal matrix, and then, to the extension of this orthogonal matrix to an orthonormal basis.*

Proof. Let us consider the Stiefel manifolds $M_{n,\kappa}$ for every $2 \leqslant \kappa \leqslant n$ and the real symmetric variance matrix A. By Theorem 3, an infinite number of global maximum point of problem (3.1) exists and any determines the dominant k-dimensional eigenspace of the matrix A. Any global maximum point forms an $n \times k$ orthogonal matrix X, i.e., $X^T X = I_\kappa$, where I_κ is the identity matrix in the Euclidean space R^κ, thus, this orthogonal matrix can be extended to an orthonormal basis.

Optimization problems (3.1) on the Stiefel manifolds which represent orthogonality constraints arise in the symmetric eigenvalue problem, in the least-squares problems related to some questions of system identification and parametrization in statistics where the problem is to minimize the sum of the norrns of the equation error (e.g., Byrnes and Willems, 1986), and in one of the most important decompositions in matrix computations, the singular value decomposition (SVD). The SVD of a general matrix A is a transformation into a product of three matrices, each of which has a simple special form and geometric interpretation. This SVD representation is given by the following theorem (e.g., Greenacre, 1984).

THEOREM 5. *Any real $m \times n$ matrix A with rank l, $(l \leqslant min(m,n))$, can be expressed in the form of*

$$A = UDV^T, \tag{4.1}$$

where D is a $\kappa \times \kappa$ diagonal matrix with positive diagonal elements $\alpha_1, \ldots, \alpha_\kappa$, U is an $m \times \kappa$ matrix and V is an $n \times k$ matrix such that $U^T U = I_m$, $V^T V = I_n$, i.e., the columns of U and V are orthonormal in the Euclidean sense.

DEFINITION 1. The dominant κ-dimensional left and right singular subspaces of an $m \times n$ matrix A with rank $l(l \leqslant \min(m,n))$ for $2 \leqslant \kappa \leqslant l$ are the subspaces spanned by the k largest left and right singular vectors of the singular value decomposition of A.

THEOREM 6. *The dominant κ-dimensional left and right singular subspaces of an $m \times n$ matrix A with rank l $(l \leqslant min\ (m,n))$ for $2 \leqslant \kappa \leqslant l$ are the dominant κ-dimensional eigenspaces of the matrices AA^T and $A^T A$.*

Proof: Let us apply Theorem 3 to the matrix AA^T. Then, any global maximum point $\boldsymbol{x} \in M_{n,\kappa}$ of problem (3.1) determines the κ-dimensional dominant eigenspace of the matrix AA^T in R^n. Let the $m \times \kappa$ matrix X contain the κ vector components of the global maximum point as column vectors. Thus, $X^T X = I_\kappa$, where I_κ is the identity matrix in R^κ.

An equivalent formulation of SVD in terms of diads is

$$A = \sum_{i=1}^{l} \alpha_i \mathbf{u}_i \mathbf{v}_i^T, \tag{4.2}$$

where $\mathbf{u}_1, \ldots, \mathbf{u}_l$, and $\mathbf{v}_1, \ldots, \mathbf{v}_l$ are the columns of U and V, respectively. The diagonal numbers α_i of D are called singular values, while the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$, $i = 1, \ldots l$, are termed the left and right singular vectors, respectively. The left and right singular vectors form an orthonormal basis for the columns and rows of A in m-dimensional and n-dimensional spaces, respectively.

By SVD,

$$AA^T = UD^2 U^T = \sum_{i=1}^{l} \alpha_i^2 \mathbf{u}_i \mathbf{u}_i^T, \tag{4.3}$$

where the values $\alpha_i^2, i = 1, \ldots, l$, and the column vectors of U, the left singular vectors of A are the eigenvalues and the eigenvectors of the matrix AA^T, respectively. It follows from Theorem 3 that the matrix X determines the dominant k-dimensional left singular subspace of A.

The second part of the statement can be proved in the same way. □

Open questions are how to characterize global optimality by first-order and second-order optimality conditions on the whole Stiefel manifolds, and how to solve efficiently optimization problem (3.1) on computer.

Acknowledgements

This research was supported in part by the Hungarian National Research Foundation, Grant Nos. OTKA-T043276 and OTKA-T043241.

References

1. Bellmann, R. (1960), *Introduction to Matrix Analysis*, McGraw-Hill Book Company, New York, Toronto, London.
2. Bolla, M., Michaletzky, Gy., Tusnády, G. and Ziermann, M. (1998) Extrema of sums of heterogeneous quadratic forms, *Linear Algebra and Applications* 269, 331–365.
3. Byrnes, C.I. and Willems, J C. (1986), Least-squares estimation, linear programming, and momentum: a geometric parametrization of local minima, *IMA Journal of Mathematical Control & Information* 3, 103–118.
4. Eckart, C. and Young, G. (1936), The approximation of one matrix by another of lower rank, *Psychometrika* 1, 211–218.
5. Edelman, A., Arias, T.A. and Smith, S.T. (1998), The geometry of algorithms with orthogonality constraints, *SIAM Journal on Matrix Analysis and Applications* 20, 303–353.
6. Fang, K.T., Hickernell, F.J. and Winker, P. (1996), Some global optimization algorithms in statistics, in: D.Z. Du, X.S. Zhang and Cheng K. (eds.) *Lecture Notes in Operations Research* World Publishing Corporation, pp. 14–24.
7. Gabriel, K.R. (1971), The biplot-graphic display of matrices with application to principal components analysis, *Biometrika* 58, 453–467.
8. Gabriel, K.R. (1981), Biplot display of multivariate matrices for inspection of data and diagnosis, in: V. Barnett (ed.) *Interpreting Multivariate Data*, Wiley, Chichester, UK pp. 147–174.
9. Golub, G.H. and Kahan, W. (1965), Calculating the singular values and pseudoinverse of a matrix, *SIAM Journal on Numerical Analysis* 2, 205–224.
10. Greenacre, M.J. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press, London, Orlando.
11. Helmke, U. and Moore, J.B. (1994), *Optimization and Dynamical Systems*, Springer-Verlag.
12. Kennedy, W.J. and Gentle, J.E. (1980), *Statistical Computing*, Marcel Dekker, New York, Basel.
13. Mareschal, B. and Brans, J.-P. (1988), Geometrical representations for MCDA, *European Journal of Operational Research* 34, 69–77.
14. Rao, C.R. (1993) (ed.), Computational Statistics, in: *Handbook of Statistics* 9.
15. Rapcsák, T. (1997), Smooth nonlinear optimization in R^n, Kluwer Academic Publishers, Boston, London, Dordrecht.
16. Rapcsák, T. (2001), On minimization of sums of heterogeneous quadratic functions on Stiefel manifolds, In: Migdalas, A., Pardalos, P. and Varbrand, P. (eds.), From local to global optimization, Dordrecht-Boston-London, Kluwer Academic Publishers, 277–290.
17. Rapcsák, T. (2002), On minimization on Stiefel manifolds, *European Journal of Operational Research*. 143, 365–376.
18. Stiefel, E. (1935/6), Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten, *Commentarii Mathematici Helvetici* 8, 305–353.
19. Young, F.W, Faldowski, R.A. and McFarlane, M.M. (1993), Multivariate statistical visualization, in: C.R. Rao (ed.) Handbook of Statistics: Elsevier Science Publishers, Amsterdam, pp. 959–998.