SIAM J. SCI. COMPUT.
Vol. 21, No. 4, pp. 1291–1304

THE PROCRUSTES PROBLEM FOR ORTHOGONAL STIEFEL MATRICES*

A. W. BOJANCZYK† AND A. LUTOBORSKI‡

Abstract. In this paper we consider the Procrustes problem on the manifold of orthogonal Stiefel matrices. Given matrices $\mathcal{A} \in \mathbb{R}^{m\times k}$, $\mathcal{B} \in \mathbb{R}^{m\times p}$, $m \geq p \geq k$, we seek the minimum of $\|\mathcal{A} - \mathcal{B}Q\|^2$ for all matrices $Q \in \mathbb{R}^{p\times k}$, $Q^TQ = I_{k\times k}$. We introduce a class of relaxation methods for generating sequences of approximations to a minimizer and offer a geometric interpretation of these methods. Results of numerical experiments illustrating the convergence of the methods are given.

Key words. Procrustes problem, Stiefel manifolds, projections on ellipsoids, relaxation methods

AMS subject classifications. 49M20, 65F20, 65K10

PII. S106482759630992X

1. Introduction.

We begin by defining for $p \geq k$ the set $\mathcal{OS}t(p,k)$ of orthogonal Stiefel matrices,

$$\mathcal{OS}t(p,k) = \{Q \in \mathbb{R}^{p\times k}, Q^TQ = I_{k\times k}\}, \tag{1.1}$$

which is a compact submanifold of dimension $M = pk - \frac{1}{2}k(k+1)$ of the manifold $\mathcal{O}(p)$ of all $p \times p$ orthogonal matrices and which has dimension $\frac{1}{2}p(p-1)$.

Let $\mathcal{A} \in \mathbb{R}^{m\times k}$ and $\mathcal{B} \in \mathbb{R}^{m\times p}$, where $m \geq p \geq k$ is given. Let $\|\mathcal{A}\| = (\operatorname{trace} \mathcal{A}^{\mathrm{T}}\mathcal{A})^{\frac{1}{2}}$ denote the standard Frobenius norm in $\mathbb{R}^{m\times k}$.

The Procrustes problem for orthogonal Stiefel matrices is to minimize

$$\mathcal{P}[\mathcal{A},\mathcal{B}](Q) = \|\mathcal{A} - \mathcal{B}Q\|^2 \tag{1.2}$$

for all $Q \in \mathcal{OS}t(p,k)$.

Problem (1.2) can be simplified by performing the SVD of the matrix $\mathcal{B} \in \mathbb{R}^{m\times p}$. Let $\mathcal{B} = USV^T$, where $U \in \mathcal{O}(m)$, $V \in \mathcal{O}(p)$, and $S = (\operatorname{diag}(\sigma_1, \dots, \sigma_{\mathrm{p}}), 0_{(\mathrm{m-p})\times(\mathrm{m-p})})^{\mathrm{T}}$; then

$$\begin{aligned} \mathcal{P}[\mathcal{A},\mathcal{B}](Q) &= \|U(U^T\mathcal{A} - SV^TQ)\|^2 \\ &= \|\tilde{\mathcal{A}} - S\tilde{Q}\|^2, \end{aligned} \tag{1.3}$$

where $\tilde{\mathcal{A}} = U^T\mathcal{A}$ and $\tilde{Q} = V^TQ$.

Due to the fact that the last $m - p$ rows of S are zeros we will simplify (1.3) by introducing new notations. We denote $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_p)$ and assume from now on that the problem is not degenerate or, in other words, that all $\sigma_1 \geq \cdots \geq \sigma_p > 0$. We define $A \in \mathbb{R}^{p\times k}$ to be a matrix composed of the first p rows of $\tilde{\mathcal{A}}$. Consequently the Procrustes minimization on the set of orthogonal Stiefel matrices is as follows.

*Received by the editors September 27, 1996; accepted for publication (in revised form) January 7, 1999; published electronically December 15, 1999.
http://www.siam.org/journals/sisc/21-4/30992.html
†Electrical Engineering Department, Cornell University, Ithaca, NY 14853 (adamb@ee.cornell.edu).
‡Department of Mathematics, Syracuse University, Syracuse, NY 13244-1150 (lutobor@mazur.syr.edu).

For given $A \in \mathbb{R}^{p\times k}$ and diagonal $\Sigma \in \mathbb{R}^{p\times p}$, minimize

$$\mathcal{P}[A,\Sigma](Q) = \|A - \Sigma Q\|^2 \tag{1.4}$$

for all $Q \in \mathcal{OS}t(p,k)$.

The original formulations of the Procrustes problem can be found in [1], [2]. We may write (1.4) explicitly as

$$\mathcal{P}[A,\Sigma](Q) = \operatorname{trace}(Q^T\Sigma^2 Q) - 2\operatorname{trace}(Q^T\Sigma A) + \|A\|^2. \tag{1.5}$$

The Procrustes problem has been solved analytically in the orthogonal case when $p = k$ and $\mathcal{OS}t(p,k) = \mathcal{O}(p)$; see [10]. In this case $Q \in \mathcal{O}(p)$ and we have

$$\mathcal{P}[A,\Sigma](Q) = \|A\|^2 + \|\Sigma\|^2 - 2\operatorname{trace}(Q^T\Sigma A). \tag{1.6}$$

Provided that the SVD of ΣA is $\Sigma A = P\Gamma R^T$, the minimizer in (1.6) is then

$$Q = PR^T. \tag{1.7}$$

The functional $\mathcal{P}[A,\Sigma]$ in (1.5) is a sum of two functionals in Q: the quadratic functional $\operatorname{trace}(Q^T\Sigma^2Q)$ and the linear functional $-2\operatorname{trace}(Q^T\Sigma A)$. It is well known how to minimize each of the functionals separately. The minimum value of the quadratic functional is equal to the sum of squares of the k smallest diagonal entries of Σ. This result is due to Fan [12]. The linear functional is minimized when $\operatorname{trace}(Q^T\Sigma A)$ is maximized. The maximum of this trace is given by the sum of the singular values of the matrix $\Sigma^T A$. This upper bound on the trace functional has been established by Von Neumann in [16]; see also [10].

Separate minimization of the quadratic and the linear part are well-understood methods. The analytical solution of the orthogonal Procrustes problem for Stiefel matrices is to the best of our knowledge an open problem.

It will be useful to interpret the minimization (1.4) geometrically. To do that we define an eccentric Stiefel manifold $\mathcal{OS}t[\Sigma](p,k)$ in $\mathbb{R}^{p\times k}$:

$$\mathcal{OS}t[\Sigma](p,k) = \{X \in \mathbb{R}^{p\times k} : X^T\Sigma^{-2}X = I_{k\times k}\}\ . \tag{1.8}$$

The eccentric Stiefel manifold $\mathcal{OS}t[\Sigma](p,k)$ is an image of the orthogonal Stiefel manifold $\mathcal{OS}t(p,k)$ under the linear mapping $Q \longrightarrow \Sigma Q$. The image of a sphere $\{A \in \mathbb{R}^{p\times k} : \|A\| = \sqrt{k}\}$ of radius $\sqrt{k}$ in $\mathbb{R}^{p\times k}$ under this mapping is an ellipsoid in $\mathbb{R}^{p\times k}$ of which $\mathcal{OS}t[\Sigma](p,k)$ is a subset. The eccentric Stiefel manifold is a compact set contained in a larger ball in $\mathbb{R}^{p\times k}$ centered at 0 and of radius $\sqrt{k}\sigma_1$.

Clearly

$$\mathcal{OS}t[\Sigma](p,k) = \{X \in \mathbb{R}^{p\times k} : X = \Sigma Q, \quad Q \in \mathcal{OS}t(p,k)\}. \tag{1.9}$$

We note that $\mathcal{OS}t[\Sigma](p,1)$ is a standard ellipsoid in $\mathbb{R}^p$:

$$\mathcal{OS}t[\Sigma](p,1) = \left\{ x \in \mathbb{R}^p : \frac{x_1^2}{\sigma_1^2} + \cdots + \frac{x_p^2}{\sigma_p^2} = 1 \right\} \tag{1.10}$$

and $\mathcal{OS}t[I](p,k) = \mathcal{OS}t(p,k)$. Therefore if

$$\min_{Q\in\mathcal{OS}t(p,k)} \mathcal{P}[A,\Sigma](Q) = \|A - \Sigma Q^*\|^2, \tag{1.11}$$

then a point ΣQ^* is the projection of A onto the manifold $\mathcal{OS}t[\Sigma](p,k)$. Due to the compactness of the manifold a projection ΣQ^* exists. The big difficulty which we face in the task of computing the minimizer Q^* is the fact that the manifold $\mathcal{OS}t[\Sigma](p,k)$ is not a convex set and a projection on a nonconvex set is in general nonunique.

2. Notations. Elementary plane rotation by an angle ϕ is represented by

$$G(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}. \tag{2.1}$$

Elementary plane reflection about the line with slope $\tan\phi$ is

$$R(\phi) = \begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix}. \tag{2.2}$$

For $Q \in \mathbb{R}^{p\times k}$ and $1 \le m < n \le p$ we introduce the following submatrices of Q:

$Q^{[m,n]} \in \mathbb{R}^{2\times k}$	consists of the mth and nth rows of Q,
$Q^{(m,n)} \in \mathbb{R}^{(p-2)\times k}$	consists of the rows complementary to $Q^{[m,n]}$,
$Q^{[m,n]}_{[m,n]} \in \mathbb{R}^{2\times 2}$	consists of the entries on the intersections of the mth and nth rows and columns of Q.

A plane rotation by an angle ϕ in the (k,l)-plane in $\mathbb{R}^p$, $p \ge 2$, is represented by a matrix $G_{k,l}(\phi) \in \mathbb{R}^{p\times p}$ such that

$$G_{k,l}(\phi)^{[k,l]}_{[k,l]} = G(\phi), \qquad G_{k,l}(\phi)^{(k,l)}_{(k,l)} = I_{(p-2)\times(p-2)}. \tag{2.3}$$

A plane reflection $R_{k,l}(\phi)$ in the (k,l)-plane is defined similarly by means of $R(\phi)$. $\mathcal{J}_{k,l}(p)$ is the set of all plane rotations and reflections in the (k,l)-plane. $\mathcal{J}(p)$ is the set of plane rotations and reflections in all planes. Clearly

$$\mathcal{J}_{k,l}(p) \subset \mathcal{J}(p) \subset \mathcal{O}(p).$$

3. Relaxation methods for the Procrustes problem. The Stiefel manifold $\mathcal{OS}t(p,k)$ is the admissible set for the minimizer of the functional $\mathcal{P}$ in (1.4). This manifold, however, is not a vector space, which poses severe restrictions on how the successive approximations can be obtained from the previous ones. Additive corrections are not admissible, but the Stiefel manifold is closed with respect to left multiplication by an orthogonal matrix $R \in \mathcal{O}(p)$. Thus RQ, where $Q \in \mathcal{OS}t(p,k)$, is an admissible approximation. Consequently, we restrict our considerations to a class of minimization methods which construct the approximations $\widehat{Q}$ to the minimizer Q^* by the rule

$$\widehat{Q} = RQ\ , \tag{3.1}$$

where Q and $\widehat{Q}$ denote, respectively, the current and the next approximations to the minimizer.

In what follows we will consider only relaxation minimization methods which seek for the minimizer of the functional $\mathcal{P}$, according to (3.1), with

$$R = R_N \cdots R_1, \tag{3.2}$$

where $N \ge M$ and M is the dimension of the manifold $\mathcal{OS}t(p,k)$. Each $R_i \in \mathcal{O}(p)$, $i = 1, 2, \ldots, N$, depends on a single parameter whose value results from a scalar minimization problem. We will refer to the left multiplication by R in (3.1) as a *sweep*.

Our relaxation method consists of repeated applications of sweeps which produce a nonincreasing sequence of values of the functional (1.4).

We will choose matrices R_i to be orthogonally similar to a plane rotation or reflection. Different choices of similarities will lead to different relaxation methods. We set

$$\begin{aligned} Q_0 &= Q, \quad R_0 = I, \\ Q_i &= R_{i-1} \cdots R_0 Q \ , \quad i = 1, 2, \dots , N+1, \end{aligned} \tag{3.3}$$

and define

$$R_i = R_i(\alpha) = P_i J_i(\alpha) P_i^T, \tag{3.4}$$

where $J_i \in \mathcal{J}(p)$ and where $P_i \in \mathcal{O}(p)$ may depend on the current approximation to the minimizer Q_i. It is the choice of P_i that fully determines the relaxation method (3.1)–(3.4). The selection of the parameter α in (3.4) will result from the scalar minimization

$$\|A - \Sigma(R_i(\alpha)Q_i)\| = \min_{\tilde{\alpha}} \|A - \Sigma(R_i(\tilde{\alpha})Q_i)\|. \tag{3.5}$$

The matrix R_i can be viewed as a plane rotation or reflection in a plane spanned by a pair of columns of the matrix P_i. The indices (r, s) of this pair of columns are selected according to an ordering $\mathcal{N}$ of a set of pairs $\mathcal{D}$. The ordering $\mathcal{N} : \mathcal{D} \to \{1, 2, \dots , N\}$ is a bijection, where $\mathcal{D} \supset \{(r, s) : 1 \le r \le k, \ r + 1 \le s \le p\}$. This inclusion guarantees that $\mathcal{D}$ contains at least M distinct pairs necessary to construct an arbitrary $Q \in \mathcal{OS}t(p, k)$ as a product of matrices R_i.

It is clear that relaxation methods satisfying (3.5) will always produce a nonincreasing sequence of the values $\mathcal{P}[A, \Sigma](Q_i)$.

If $P_i = I_{p\times p}$ in (3.4), then $R_i = J_i$ and the sweep (3.1) has the following particularly simple form:

$$\widehat{Q} = J_N \cdots J_1 Q. \tag{3.6}$$

The relaxation method defined by (3.6) will be referred to as a *left-sided relaxation method* (LSRM).

If $P_i = (Q_i \ , \ Q_i^{\perp})$ in (3.4), where $Q_i^{\perp}$ is the orthogonal complement of Q_i, then

$$R_i = (Q_i \ , \ Q_i^{\perp}) J_i \begin{pmatrix} Q_i^T \\ (Q_i^{\perp})^T \end{pmatrix}$$

and hence

$$Q_{i+1} = R_i Q_i = (Q_i \ , \ Q_i^{\perp}) J_i \begin{pmatrix} I_{k\times k} \\ 0_{(p-k)\times k} \end{pmatrix}.$$

Thus by induction the sweep (3.1) has the form

$$\widehat{Q} = (Q \ , \ Q^{\perp}) J_1 \cdots J_N \begin{pmatrix} I_{k\times k} \\ 0_{(p-k)\times k} \end{pmatrix}, \tag{3.7}$$

where $Q^{\perp} = Q_0^{\perp}$ represents a set of orthonormal vectors which form a basis for the orthogonal complement of the initial approximation Q_0. Once the initial $Q_0^{\perp}$ is selected, $Q_{i+1}^{\perp}$ is obtained from Q_i and $Q_i^{\perp}$ through the relation

$$\left(Q_{i+1} \ , \ Q_{i+1}^{\perp}\right) = \left(Q_i \ , \ Q_i^{\perp}\right) J_i \ .$$

The relaxation method defined by (3.7) will be referred to as a *right-sided relaxation method* (RSRM).

Our objective is to propose a geometric interpretation of the LSRM and describe its numerical implementation based on the geometric aspects of the method. We will compare our LSRM with an existing method for the Procrustes problem for orthogonal Stiefel matrices due to Park in [14]. The general description of the RSRM allows us to formulate Park's method as a relaxation method and compare it with the LSRM. The method in [14] is based on the concepts introduced earlier by Ten Berge and Knol [2], where the problem is called the unbalanced Procrustes problem and its solution is based on an iterative solution of a sequence of orthogonal Procrustes problems called balanced problems. For the study of other minimization methods on manifolds in the spaces of matrices, see [13], [15], and [3].

4. Planar Procrustes problem. We will now present the LSRM. Without loss of generality let us assume that the planes (r, s) in which transformations operate are chosen in the row cyclic order, in the way analogous to that used in the cyclic Jacobi method for the SVD computation [6]. For example, for $p = 4$ the planes (r, s) are taken in the following order:

(1,2), (1,3), (1,4), (2,3), (2,4), (3,4).

Formally, in this case, the bijection $\mathcal{N}$, $\mathcal{N} : \mathcal{D} \to \{1, \dots, \frac{1}{2}p(p-1)\}$, $\mathcal{D} = \{(r, s) : 1 \le p-1, r+1 \le s \le p\}$, is given by $\mathcal{N}(r, s) = s - r + (r-1)(p - \frac{r}{2})$ and $\hat{Q}$ in (3.6) has the following form:

$$\hat{Q} = \prod_{r=1}^{p-1} \prod_{s=1}^{r} J_{p-r,p-s+1}\, Q, \tag{4.1}$$

where $J_{r,s} \in \mathcal{J}_{r,s}(p)$. Let $Q_i = Q_{\mathcal{N}(r,s)}$ be the current approximation in the sweep. The next approximation to the minimizer is $Q_{i+1} = J_i(\alpha)Q_i$. The selection of the parameter α results from the scalar minimization

$$\|A - \Sigma\,(J_i(\alpha)Q_{r,s})\|^2 = \min_{\tilde{\alpha}} \|A - \Sigma\,(J_i(\tilde{\alpha})Q_{r,s})\|^2 . \tag{4.2}$$

Our main goal now is to show how to find α in (4.2).

Consider the functional $J \longrightarrow \|A - \Sigma(JQ)\|^2$ in (4.2), where for simplicity of notation we omitted all indices. Without loss of generality we assume that $\mathcal{N}^{-1}(i) = (r, s) = (1, 2)$ and hence

$$J = G_{1,2}(\alpha) = \operatorname{diag}\bigl(G(\alpha), I_{(p-2)\times(p-2)}\bigr),$$

where $G(\alpha)$ is a plane rotation (the case of reflection is similar and can be treated in a completely analogous way). The minimization in (4.2) is precisely the minimization of

$$f(\alpha) = \left\| \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \end{pmatrix} - \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} G(\alpha) \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1k} \\ q_{21} & q_{22} & \cdots & q_{2k} \end{pmatrix} \right\|^2$$

$$= \left\| A^{[1,2]} - \Sigma^{[1,2]}_{[1,2]} G(\alpha) Q^{[1,2]} \right\|^2 . \tag{4.3}$$

Let $U_Q \Gamma V_Q^T$ be the SVD of $Q^{[1,2]}$ such that $\Gamma = \bigl(\operatorname{diag}(\gamma_1, \gamma_2), 0_{2\times(k-2)}\bigr)$ and $\gamma_1 \ge \gamma_2 > 0$. Thus,

$$f(\alpha) = \left\| A^{[1,2]} V_Q - \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} G(\alpha) U_Q \begin{pmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \end{pmatrix} \right\|^2 . \tag{4.4}$$

Denote $B = A^{[1,2]}V_Q$. Note that the last $k-2$ columns of the matrix B in (4.4) are always approximated by zero columns, and thus the minimization of $f(\alpha)$ is equivalent to a minimization restricted to the first two columns.

Introducing a new variable $\phi = \alpha + \beta$, where β is the angle of the plane rotation (or reflection) U_Q, and setting $c = \cos\phi$, $s = \sin\phi$ we obtain the following minimization problem:

$$F(\phi) = \left\| B_{[1,2]} - \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} c & -s \\ s & c \end{pmatrix} \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} \right\|^2 + \|B_{(1,2)}\|^2. \tag{4.5}$$

We may write (4.5) explicitly as

$$F(\phi) = z_1^2 c^2 + z_2^2 s^2 - 2y_1 c - 2y_2 s + \|B\|^2, \tag{4.6}$$

where $q = (c, s)^T$ and

$$\begin{aligned} z_1^2 &= \sigma_1^2\gamma_1^2 + \sigma_2^2\gamma_2^2, \\ z_2^2 &= \sigma_1^2\gamma_2^2 + \sigma_2^2\gamma_1^2, \end{aligned} \tag{4.7}$$

$$\begin{aligned} y_1 &= b_{11}\sigma_1\gamma_1 + b_{22}\sigma_2\gamma_2, \\ y_2 &= -b_{12}\sigma_1\gamma_2 + b_{21}\sigma_2\gamma_1. \end{aligned} \tag{4.8}$$

We now denote

$$\begin{aligned} Z &= \operatorname{diag}(z_1, z_2), \qquad Y = (y_1, y_2)^T, \\ C &= Z^{-1}Y, \end{aligned} \tag{4.9}$$

where $z_1 \geq z_2 > 0$ and $C = (c_1, c_2)^T$. By completing the squares we may represent $F(\phi)$ in the following form:

$$\begin{aligned} F(\phi) &= \left(\frac{y_1}{z_1} - z_1\cos\phi\right)^2 + \left(\frac{y_2}{z_2} - z_2\sin\phi\right)^2 - \|C\|^2 + \|A^{[1,2]}\|^2 \\ &= \|C - Zq(\phi)\|^2 - \|C\|^2 + \|A^{[1,2]}\|^2 \end{aligned} \tag{4.10}$$

for $q(\phi) = (\cos\phi, \sin\phi)^T$. Thus the minimization of the functional (4.10) is equivalent to the following problem.

For given $C \in \mathbb{R}^{2\times 1}$ and diagonal $Z \in \mathbb{R}^{2\times 2}$, minimize

$$\begin{aligned} \mathcal{P}[C, Z](q) &= z_1^2\cos^2\phi + z_2^2\sin^2\phi - 2c_1z_1\cos\phi - 2c_2z_2\sin\phi + ||C||^2 \\ &= \|C - Zq\|^2 \end{aligned} \tag{4.11}$$

for all $q \in \mathcal{OS}t(2,1)$, $q = (\cos\phi, \sin\phi)^T$.

Remark. If the original problem (1.4) is degenerate, that is, when the smallest diagonal entries in Σ are zeros, then σ_2 in (4.5) may be zero. In this case the minimization of the functional $F(\phi)$ is still well defined, and the formulas corresponding to (4.6)–(4.11) are simplified. Thus, the algorithm LSRM is applicable also in degenerate cases.

The minimization problem of the type (4.11) will be called a planar Procrustes problem. This planar problem is geometrically equivalent to projecting C onto an ellipse. Such a projection has to be calculated at each step of our relaxation method. A planar Procrustes problem is a special case of a linear least squares problem with a quadratic equality constraint. Its solution can be obtained by any of the minimization techniques described in the literature; see [8], [9], [10]. These techniques, in general, do not rely on the geometric properties of an ellipse. Our geometric interpretation helps to find excellent initial approximations to the solution and provides a means for estimating approximation errors. In the next section we outline a possible way to take advantage of geometric properties of an ellipse when solving the problem (4.11).

5. Projection on an ellipse. The geometrical formulation of the planar Procrustes problem (4.11) is very simple. Given a point C and an ellipse $\mathcal{E} = \mathcal{OS}t[Z](2,1)$,

$$\mathcal{E} = \mathcal{OS}t[Z](2,1) = \left\{ (x_1, x_2)^T : \; \frac{x_1^2}{z_1^2} + \frac{x_2^2}{z_2^2} = 1 \right\}, \tag{5.1}$$

in $\mathbb{R}^2$ we want to find a point $S \in \mathcal{E}$, $S = Zq = (z_1 \cos\phi, z_2 \sin\phi)^T$ which is a projection of C onto $\mathcal{E}$.

This can be achieved in a variety of ways. The classical projection of a point onto an ellipse is due to Apollonius and is described next.

5.1. The hyperbola of Apollonius. Recall Apollonius's construction of a normal to an ellipse from a point [17]; see Figure 1. With the given ellipse $\mathcal{E}$ in (5.1) and with the point C we associate the hyperbola $\mathcal{H}$ given by

$$x_2 = \frac{m_2 x_1}{x_1 - m_1}, \tag{5.2}$$

where (m_1, m_2) is the center of $\mathcal{H}$ with coordinates

$$m_1 = \frac{c_1 z_1^2}{z_1^2 - z_2^2}, \qquad m_2 = -\frac{c_2 z_2^2}{z_1^2 - z_2^2}. \tag{5.3}$$

To find the coordinates of the projection point S we have to intersect the hyperbola of Apollonius with the ellipse that is to solve a system of two quadratic equations (5.1) and (5.2). Using (5.2) to eliminate x_2 from (5.1) we obtain a fourth order polynomial equation in x_1,

$$(x_1 - m_1)^2 (z_1^2 - x_1^2) - \left(\frac{z_1 m_2}{z_2} \right)^2 x_1^2 = 0. \tag{5.4}$$

For any specific numerical values of the coefficients this equation can be easily solved symbolically. A purely numerical alternative is to solve the system (5.1), (5.2) using Newton's method [5], [7], [8], [9].

In our implementation we have first reduced the system to a scalar trigonometric equation. Assume that $C = (c_1, c_2)^T$ is in the first quadrant and that $C \notin \mathcal{E}$. Let S be the projection of C onto $\mathcal{E}$. Then setting $(x_1, x_2) = (z_1 \cos\phi, z_2 \sin\phi)$ in (5.4) and next substituting $t = \tan\phi$ leads to the equation $g(t) = 0$ in t, where

$$g(t) = c_1 z_1 t - (z_1^2 - z_2^2) t (1 + t^2)^{-\frac{1}{2}} - c_2 z_2. \tag{5.5}$$

It is easy to see that the function $g(t)$ is convex and has one positive root. (The geometric interpretation of (5.4) makes the problem of multiple roots in (5.5) easy to deal with.) It can also be seen that for $t_0 = \frac{z_1^2 - z_2^2 + c_2 z_2}{c_1 z_1}$ we have $g(t_0) > 0$. Thus, the Newton method starting from the initial approximation t_0 will generate a decreasing,

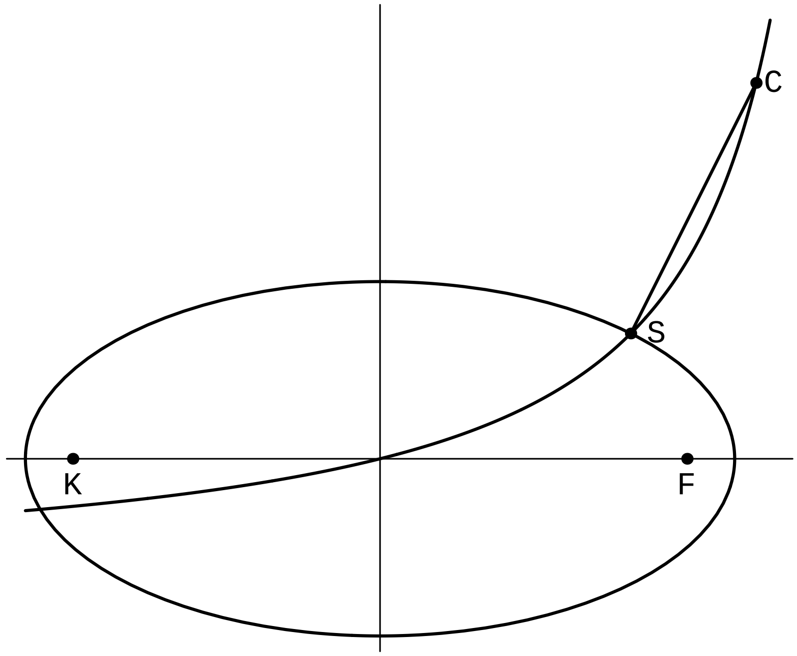

FIG. 1. *Hyperbola of Appolonius.*

convergent sequence of approximations to the root of $g(t) = 0$. We found this variant of the Newton method to be very efficient, and consequently we used it in all our numerical tests.

6. Geometric interpretation of left and right relaxation methods. Since the notion of the standard ellipsoid $\mathcal{OS}t[\Sigma](p, 1)$ in $\mathbb{R}^p$ is very intuitive we will now interpret the minimization problem (1.11) treating matrices in $\mathbb{R}^{p\times k}$ as k-tuples of vectors in $\mathbb{R}^p$. Let $A = (a_1, a_2, \ldots , a_k)$ be a given k-tuple of vectors in $\mathbb{R}^p$. Let $Q = (q_1, q_2, \ldots , q_k) \in \mathcal{OS}t(p, k)$ be the current approximation to the minimizer. Clearly the points Σq_i all belong to the ellipsoid $\mathcal{OS}t[\Sigma](p, 1)$. Thus the minimization of $\mathcal{P}[A, \Sigma](Q)$ can be interpreted as finding points Σq_i^* on the ellipsoid, where q_i^* are orthonormal vectors, that best match, as measured by $\mathcal{P}[A, \Sigma](Q)$, the given vectors a_i in $\mathbb{R}^p$.

The relaxation method described in section 3 can be interpreted as follows. Pick an orthonormal basis in $\mathbb{R}^p$. In the next sweep rotate the current set of vectors q_i as a frame, in planes spanned by all pairs of the vectors from the current basis.

In the LSRM the basis is the canonical basis and is the same for all sweeps. All relaxation steps are exactly the same, and all amount to solving a planar Procrustes problem.

In the RSRM the basis consists of two subsets and changes from sweep to sweep. The first subset of the basis consists of the columns of the current approximation Q and the second subset consists of the columns of the orthogonal complement $Q^\perp$ of Q.

Working only with the columns of Q is equivalent to the so-called balanced Procrustes problem studied by Park [14] which can be solved by means of an SVD computation. The relaxation step in [14] for the balanced problem consists of computing the SVD of the 2×2 matrix $(\Sigma q_r \ , \ \Sigma q_s)^T(a_r \ , \ a_s)$. In our relaxation setting, the relaxation step in the RSRM requires solving the scalar minimization problem (3.5),

$$\min_{c^2+s^2=1} \left\| (a_r \ , \ a_s) - \Sigma(q_r \ , \ q_s) \begin{pmatrix} c & -s \\ s & c \end{pmatrix} \right\|^2, \tag{6.1}$$

which leads to a linear equation in $\tan \alpha$ and is equivalent to the 2×2 SVD computation in [14]. Each of these 2×2 steps is a rotation of vectors Σq_r and Σq_s in the plane spanned by q_r and q_s so the rotated vectors on the ellipsoid best approximate the two given vectors a_r and a_s.

However, as the columns of Q do not span the whole space $\mathbb{R}^p$, it will usually be the case that $\text{span}\{q_1, \dots, q_k\} \neq \text{span}\{q_1^*, \dots, q_k^*\}$, and hence it may not be possible to generate a sequence of approximations that will converge to Q^*. In order to overcome this problem the matrix Q is extended by its orthogonal complement $Q^\perp = (q_{k+1}, \dots, q_p)$ so that $\text{span}(q_1^*, \dots, q_k^*) \subset \text{span}(q_1, \dots, q_p)$. The scalar minimization subproblems in [14] involving vectors from both subsets are referred to in [14] as unbalanced subproblems. These scalar minimizations have the following form:

$$\min_{c^2+s^2=1} \left\| a_r - \Sigma(q_r \ , \ q_s) \begin{pmatrix} c \\ s \end{pmatrix} \right\|^2. \tag{6.2}$$

That is, the unbalanced subproblem is to find a vector on the ellipsoid in the plane spanned by q_r and q_s closest to the given vector a_r. As the intersection of this plane and the ellipsoid is an ellipse, the unbalanced subproblem can be expressed as a planar Procrustes problem (4.11) and any of the algorithms discussed in section 5 can be used to solve this unbalanced problem.

Other choices of bases may be possible but the choices leading to the left-and right-sided relaxation methods seem to be the most natural.

7. Numerical experiments. In this section we present numerical experiments illustrating the behavior of the left and the right relaxation methods discussed in section 3. In addition, we compare performance of the relaxation methods with the MATLAB routine `constr.m` which implements the nonlinear programming minimization method known as sequential quadratic programming (SQP). The particular variant of the SQP method we included in our tests used analytical expressions for evaluating the gradient of the function and the gradient of the constraints. For performance evaluation of the SQP method see [11].

We will start by summarizing the left and the right relaxation methods given below in pseudocode.

Given $A \in \mathbb{R}^{p\times k}$, $A = (a_1, \dots, a_k)$, and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$ both algorithms construct sequences of Stiefel matrices approximating the minimizer of (1.4).

ALGORITHM LSRM.

1. *Initialization:*
   set $Maxstep$, $Q = I_{p\times k}$, $n = 0$, $r_{-1} = 0$, $r_0 = \|A - \Sigma Q\|$
2. *Iterate sweeps:*
   while $(r_n - r_{n-1}) > threshold$ and $n < Maxstep$
      for $i = 1$ to k
         for $j = i + 1$ to p
            solve planar Procrustes problem

$\min_\alpha \|A^{[i,j]} - \Sigma^{[i,j]}_{[i,j]} J(\alpha) Q^{[i,j]}\|^2$
$Q^{[i,j]} \leftarrow J(\alpha) Q^{[i,j]}$
$n \leftarrow n+1$
$r_n = \|A - \Sigma Q\|$

ALGORITHM RSRM.

1. *Initialization:*
   set $Maxstep$, $Q = I_{p\times p}$, $n = 0$, $r_{-1} = 0$, $r_0 = \|A - \Sigma Q I_{p\times k}\|$
2. *Iterate sweeps:*
   while $(r_n - r_{n-1}) > threshold$ and $n < Maxstep$
   for $i = 1$ to k
   for $j = i+1$ to k
   solve $\min_\alpha \|(A^T)^{[i,j]} - J(\alpha)(Q^T)^{[i,j]}\Sigma\|^2$
   $(Q^T)^{[i,j]} \leftarrow J(\alpha)(Q^T)^{[i,j]}$
   for $j = k+1$ to p
   solve planar Procrustes problem
   $\min_{c^2+s^2=1} \|a_j^T - (c\ s)(Q^T)^{[i,j]}\Sigma\|^2$
   $e_i^T(Q^T) \leftarrow (c\ s)(Q^T)^{[i,j]}$
   $n \leftarrow n+1$
   $r_n = \|A - \Sigma Q I_{p\times k}\|$

We measure the cost of the two relaxation methods by the number of sweeps performed by each of them.

A sweep in the LSRM method consists of $p(p+1)/2$ planar Procrustes problems. Each planar Procrustes problem requires computation of the SVD of a $2 \times k$ matrix. This can be achieved by first computing the QR decomposition followed by a 2×2 SVD problem. After the SVD is calculated, a projection on an ellipse has to be determined. The cost of a sweep is approximately $O(kp^2)$ floating-point operations.

A sweep in the RSRM method consists of $k(k+1)/2$ computations of $p \times 2$ SVD problems. In addition, there are $k(p-k)$ planar Procrustes problems, each requiring computation of the SVD of a $p \times 2$ matrix followed by computation of a projection on an ellipse. Thus, the cost of a sweep is again approximately $O(kp^2)$ floating-point operations.

Surely, the precise cost of a sweep will depend on the number of iterations needed for obtaining satisfactory projections on the resulting ellipses. For each projection, this will depend on the location of the point being projected as well as the shape of the ellipse. Computation of the projection will be most costly when the ellipse is flat.

As can be seen, sweeps in the two methods may have different costs. However, the number of sweeps performed by each of the methods will give some bases for comparing the convergence behavior of the two methods.

We measured the cost of the SQP method by the number of function and gradient evaluations as these are the most obvious units of cost for this method. The number of function and gradient evaluations indicates the speed of convergence of the SQP method. As all three units of cost are different, in all experiments, we also recorded the number of floating-point operations performed by each of the three methods as returned by the MATLAB function `flops`. Due to the peculiar way MATLAB counts operations, the number of floating-point operations returned by `flops` can be taken as only an approximation of the true count. However, this can serve as a reasonable estimate.

TABLE 1
Matrices Q generated by LSRM.

Sweep #	Q		$\|A - \Sigma Q\|$
1	-3.166246771495053e-01	5.337233877242527e-02	6.4043e-05
	-1.511766767561811e-01	-9.206530348969716e-01	
	-7.298295135502926e-01	3.298411857874910e-01	
	-5.867225130978633e-01	-2.018766329897017e-01	
5	-3.166512702480902e-01	5.342031933605162e-02	1.0450e-06
	-1.508477124262734e-01	-9.206997446627507e-01	
	-7.297521332055357e-01	3.296126838684622e-01	
	-5.868890566265494e-01	-2.020240784981074e-01	
10	-3.166512669936927e-01	5.342030989691857e-02	4.1802e-08
	-1.508494100481855e-01	-9.206948179307888e-01	
	-7.297545802980885e-01	3.296200876881795e-01	
	-5.868855792514951e-01	-2.020344537903026e-01	
15	-3.166512668678573e-01	5.342030953200504e-02	8.7804e-10
	-1.508494779430533e-01	-9.206946209057939e-01	
	-7.297546781621558e-01	3.296203837640639e-01	
	-5.868854401803947e-01	-2.020348687030837e-01	
30	-3.166512668626163e-01	5.342030951680608e-02	5.6205e-14
	-1.508494807709545e-01	-9.206946126994970e-01	
	-7.297546822383019e-01	3.296203960959194e-01	
	-5.868854343879250e-01	-2.020348859846182e-01	

We begin by illustrating the behavior of the LSRM method for finding Q^* in the Procrustes problem with $p = 4,\ k = 2,\ \Sigma = \mathrm{diag}(10^0, 10^{-1}, 10^{-2}, 10^{-3})$, and $A = \Sigma Q^*$, where

$$Q^* = \begin{pmatrix} -3.166512668626158e-01 & 5.342030951680499e-02 \\ -1.508494807711354e-01 & -9.206946126989718e-01 \\ -7.297546822385641e-01 & 3.296203960967122e-01 \\ -5.868854343875571e-01 & -2.020348859857265e-01 \end{pmatrix}. \tag{7.1}$$

The initial approximation is $Q_0 = I_{4\times 2}$. Some intermediate values of Q are listed in Table 1.

We now present comparative numerical results for the LSRM, RSRM, and SQP methods.

Recall that the functional $\mathcal{P}$ is a sum of a linear and a quadratic term. We will consider classes of examples when the functional can be approximated by its linear or the quadratic term.

In the first class of examples the linear term dominates the quadratic term, or in other words, $||A|| >> ||\Sigma||$. We deal here with a perturbed linear functional. The minimum of the functional $\mathcal{P}$ can be approximated by the sum of singular values of $\Sigma^T A$.

The second class of examples consists of cases when the quadratic term dominates the linear term, that is, when $||A|| << ||\Sigma||$. We deal here with a perturbed quadratic

TABLE 2
$A = \Sigma Q$ *and* $\frac{\sigma_1}{\sigma_p} \leq 2$.

		RSRM			LSRM			SQP		
p	k	I	ops	res	I	ops	res	I	ops	res
6	2	30	3.7e5	2.7e-7	25	4.2e5	1.5e-15	60	4.9e6	3.3e-4
	3	30	5.1e5	1.7e-5	23	4.2e5	1.3e-15	60	2.7e7	2.3e-3
	4	30	6.3e5	6.5e-6	30	6.1e5	1.4e-15	60	6.5e7	5.8e-2
	5	30	7.2e5	1.9e-5	27	6.1e5	3.8e-15	60	1.8e8	4.6e-2
9	2	30	1.4e6	1.3e-3	19	7.8e5	2.8e-16	60	1.3e7	4.2e-2
	3	30	2.0e6	2.6e-4	21	9.5e5	1.0e-15	60	6.2e7	2.7e-2
	4	30	2.6e6	2.7e-4	23	1.1e6	1.6e-15	60	1.5e8	6.3e-2
	5	30	3.2e6	3.2e-4	30	1.7e6	3.4e-15	60	5.3e8	1.1e-1
	6	30	3.8e6	3.5e-4	30	1.9e6	1.4e-12	60	1.0e9	8.3e-2
	7	30	4.4e6	1.8e-6	30	2.1e6	4.2e-10	60	2.0e9	5.3e-2
	8	30	4.9e6	5.2e-4	30	2.3e6	4.7e-10	60	3.8e9	8.7e-2

functional. Then the minimum value of the functional $\mathcal{P}$ can be approximated by the sum of the k smallest singular values of Σ.

The third class of examples consists of cases when the functional is genuinely quadratic, that is, when $A \approx \Sigma Q$ for some $Q \in \mathcal{OS}t(p,k)$. The minimum of the functional is then close to zero.

In each class of examples we pick two different matrices Σ: one corresponding to the ellipsoid being almost a sphere, that is, when $\Sigma \approx I$, and the other corresponding to the ellipsoid being very flat in one or more planes, that is, when $\frac{\sigma_1}{\sigma_p}$ is large.

The algorithms were written in MATLAB 5.0 and run on an HP9000 workstation with the machine relative precision $\epsilon = 2.2204e - 16$. We restricted the maximal number of sweeps for the relaxation methods to 30 and the maximal number of function and gradient evaluations for the SQP method to 60. We terminated iterations whenever the difference between two successive approximations was less than threshold $= 5\epsilon$. As the initial approximation we took $Q = I_{p\times k}$ for the LSRM and $(Q, Q^{\perp}) = I_{p\times p}$ for the RSRM.

Some representative results are shown in Tables 2–5. In the tables, the number of floating-point operations performed by each of the three methods as recorded by the MATLAB function `flops` is denoted by `ops`, I denotes the number of sweeps in the relaxation methods, and in the case of the SQP method I denotes the number of function evaluations. Finally, res denotes the residual error (the value of the functional $\|A - \Sigma Q\|$ for the computed minimizer).

Table 2 illustrates the behavior of the three methods when the ellipsoid is almost a sphere and when there exists Q such that $\Sigma Q = A$. That is, the bilinear and the linear terms are of comparable size. The experiments suggest that the LSRM requires fewer sweeps to obtain a satisfactory approximation to the minimizer. The SQP method performs less satisfactorily.

Table 3 illustrates the behavior of the three methods when the length of one half of the ellipsoid's axes is approximately 1.0 and the other half is approximately 0.01. In addition, there exists Q such that $\Sigma Q = A$. In this case the convergence of the RSRM is particularly slow. We observed that, at least initially, the RSRM fails to locate the minimizer in $\mathcal{OS}t(4,2)$, being unable to establish the proper signs of the entries of the matrix Q^*. The LSRM, on the other hand, approximates the minimizer correctly.

Table 4 illustrates the behavior of the three methods when the ellipsoid is almost a

TABLE 3
$A = \Sigma Q^*$ *and* $\frac{\sigma_1}{\sigma_p} \approx 10^2$.

		RSRM			LSRM			SQP		
p	k	I	ops	res	I	ops	res	I	ops	res
6	2	30	3.7e5	6.6e-2	13	2.2e5	1.0e-16	60	4.9e6	1.9e-2
	3	30	5.2e5	8.5e-3	30	5.7e5	1.5e-05	60	2.0e7	1.1e-2
	4	30	6.3e5	7.1e-3	30	6.3e5	4.7e-05	60	5.5e7	1.5e-2
	5	30	7.2e5	9.2e-3	30	7.0e5	1.9e-10	60	1.3e8	2.1e-2
9	2	30	1.4e6	1.1e-1	13	5.4e5	1.8e-16	60	1.2e7	2.4e-2
	3	30	2.0e6	6.4e-2	18	8.2e5	4.1e-16	60	5.6e7	6.6e-2
	4	30	2.6e6	4.5e-3	30	1.5e6	5.6e-11	60	1.8e8	9.9e-2
	5	30	3.2e6	6.3e-3	30	1.7e6	1.9e-04	60	4.3e8	5.9e-2
	6	30	3.8e6	6.4e-3	30	1.9e6	1.9e-06	60	5.8e8	4.1e-1
	7	30	4.4e6	6.8e-3	30	2.1e6	2.8e-07	60	2.0e9	1.0e-1
	8	30	4.9e6	6.3e-3	30	2.4e6	6.7e-06	60	3.1e9	1.1e-1

TABLE 4
$\|A\| \approx 10^{-2} \cdot \|\Sigma\|$ *and* $\frac{\sigma_1}{\sigma_p} \leq 2$.

		RSRM			LSRM			SQP		
p	k	I	ops	res	I	ops	res	I	ops	res
6	2	13	2.2e5	5.0e-1	8	1.4e5	5.0e-1	60	2.8e6	5.2e-1
	3	12	2.7e5	6.9e-1	9	1.8e5	6.9e-1	60	8.6e7	9.7e-1
	4	12	3.2e5	8.6e-1	15	3.2e5	8.6e-1	60	5.5e7	1.2e-0
	5	8	2.5e5	1.0e-0	30	7.1e5	1.0e-0	60	3.2e7	1.3e-0
9	2	11	7.4e5	2.5e-1	4	5.4e5	2.5e-1	60	1.2e7	1.1e-0
	3	19	1.7e6	6.3e-1	9	8.2e5	6.3e-1	60	5.6e7	1.2e-0
	4	16	1.8e6	8.8e-1	15	1.5e6	8.8e-1	60	1.8e8	1.6e-0
	5	15	2.0e6	1.1e-0	16	1.7e6	1.1e-0	60	4.3e8	1.8e-0
	6	18	2.8e6	1.3e-0	30	1.9e6	1.3e-0	60	4.7e8	1.9e-0
	7	30	5.5e6	1.5e-0	30	2.1e6	1.5e-0	60	2.0e9	2.0e-0
	8	11	2.4e6	1.7e-0	13	2.4e6	1.7e-0	60	3.2e9	2.0e-0

sphere, but now A is chosen so $\|A\| \approx 10^{-2}\|\Sigma\|$, that is, the quadratic term dominates the linear term. In this case the minimum of the functional can be estimated by the minimum value of the quadratic term. The experiments suggest that the LSRM and the RSRM perform a comparable number of sweeps but the LSRM requires fewer floating-point operations to obtain a satisfactory approximation to the minimizer. The SQP method does not provide a good approximation to the minimizer, at least in its initial stages.

Table 5 illustrates the behavior of the two methods when the ellipsoid is almost a sphere, but now A is chosen so $\|A\| \approx 10^2\|\Sigma\|$, that is, the linear term dominates the quadratic terms. In this case the minimum of the functional can be estimated by the minimum value of the linear term. The experiments suggest that the RSRM requires fewer sweeps to obtain a satisfactory approximation to the minimizer.

In conclusion, the LSRM often produces better approximations to the minimizer than the RSRM. In cases when the ellipsoid is flat and the minimizer is on the ellipsoid, the RSRM experiences difficulties in finding the right location of the minimizer. On the other hand, in cases when the linear term dominates the quadratic term, the RSRM method converges faster than the LSRM. The experiments also suggest that the general minimization technique SQP fails to exploit the structure of the problem and thus provides worse approximations to the minimizer than the two relaxation

TABLE 5
$\|A\| \approx 10^2 \cdot \|\Sigma\|$ *and* $\frac{\sigma_1}{\sigma_p} \leq 2$.

		RSRM			LSRM			SQP		
p	k	I	ops	res	I	ops	res	I	ops	res
6	2	5	7.5e4	1.0e2	10	1.4e5	1.0e2	57	5.3e6	1.0e2
	3	11	2.2e5	8.4e1	21	3.5e5	8.4e1	60	1.9e7	8.4e1
	4	6	1.4e5	1.6e2	11	2.0e5	1.6e2	60	4.6e7	1.6e2
	5	8	2.4e5	1.7e2	22	4.5e5	1.7e2	60	1.0e8	1.7e2
9	2	7	4.5e5	1.6e2	21	7.7e5	1.6e2	48	4.1e7	1.6e2
	3	5	4.5e5	1.9e2	22	9.0e5	1.9e2	47	4.5e7	1.9e2
	4	5	5.6e6	2.4e2	30	1.3e6	2.4e2	60	5.4e8	2.4e2
	5	16	2.1e6	2.4e2	30	1.6e6	2.4e2	60	2.9e8	2.4e2
	6	7	1.1e6	2.4e2	30	1.8e6	2.4e2	60	1.4e9	2.4e2
	7	10	1.8e6	2.9e2	19	1.2e6	2.9e2	60	1.6e9	2.9e2
	8	7	1.5e6	3.0e2	30	2.4e6	3.0e2	60	4.8e9	3.0e2

techniques.

An interesting open question is whether the considered relaxation methods are guaranteed to converge to the minimizer (in other words, whether the sequence of matrices generated by the LSRM is a minimizing sequence), and if they do converge, with what rate?

REFERENCES

[1] J.M. TEN BERGE AND K. NEVELS, *A general solution to Mosier's oblique Procrustes problem*, Psychometrika, 42 (1977), pp. 593–600.

[2] J.M. TEN BERGE AND D.L. KNOL, *Orthogonal rotations to maximal agreement for two or more matrices of different column orders*, Psychometrika, 49 (1984), pp. 49–55.

[3] A. EDELMAN, T.A. ARIAS, AND S.T. SMITH, *the Geometry of Algorithms with Orthogonality Constraints*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 303–353.

[4] L. ELDEN, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.

[5] G.E. FORSYTHE AND G.H. GOLUB, *On the stationary values of a second-degree polynomial on the unit sphere*, J. Soc. Indust. Appl. Math., 13 (1965), pp. 1050–1068.

[6] G.E. FORSYTHE AND P. HENRICI, *The cyclic Jacobi method for computing the principal values of a complex matrix*, Trans. Amer. Math. Soc., 94 (1960), pp. 1–23.

[7] W. GANDER, *Least squares with a quadratic constraint*, Numer. Math., 36 (1981), pp. 291–307.

[8] G.H. GOLUB, *Some modified matrix eigenvalue problems*, SIAM Rev., 15 (1973), pp. 318–334.

[9] G.H. GOLUB AND U. VON MATT, *Quadratically constrained least squares and quadratic problems*, Numer. Math., 59 (1991), pp. 561–580.

[10] G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1990.

[11] W. HOCK AND K. SCHITTOWSKI, *A comparative performance evaluation of* 27 *nonlinear programming codes*, Computing, 30 (1983), pp. 288–311.

[12] K. FAN, *On a theorem of Weyl concerning eigenvalues of linear transformations*, Proc. Nat. Acad. Sci. U.S.A., 35 (1949), pp. 652–655.

[13] A. LUTOBORSKI, *On the convergence of the Euler-Jacobi method*, Numer. Funct. Anal. Optim., 13 (1992), pp. 185–202.

[14] H. PARK, *A parallel algorithm for the unbalanced orthogonal Procrustes problem*, Parallel Comput., 17 (1991), pp. 913–923.

[15] S.T. SMITH, *Optimization techniques on Riemannian manifolds*, in Hamiltonian and Gradient Flows, Algorithms and Control, Fields Institute Commun. 3, Amer. Math. Soc., Providence, RI, 1994, pp. 113–136.

[16] J. VON NEUMANN, *Some matrix inequalities and metrization of the matrix space*, Tomsk Univ. Rev., (1937), pp. 286–300.

[17] H. WEBER AND J. WELLSTEIN, *Enzyklopadie der Elementar Mathematik*, B.G. Teubner, Leipzig, 1915.