THE PROCRUSTES PROBLEM FOR ORTHOGONAL STIEFEL MATRICES

A. W. BOJANCZYK AND A. LUTOBORSKI

CTC96TR255, 8/96.

ABSTRACT. In this paper we consider the Procrustes problem on the manifold of orthogonal Stiefel matrices. That is, 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 minimizing sequences and offer a geometric interpretation of these methods. Results of numerical experiments illustrating the convergence of the methods are given.

1. INTRODUCTION

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

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

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 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 $m \geq p \geq k$ be given. Let $\|\mathcal{A}\| = (trace\ \mathcal{A}^T\mathcal{A})^{\frac{1}{2}}$ 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 singular value decomposition 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 $S = (diag(\sigma_1,\ldots,\sigma_p), O_{(m-p)\times(m-p)})^T$, 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 $\Sigma = diag(\sigma_1,\ldots,\sigma_p)$ and assuming from now on 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 :

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

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

The Procrustes problem has been solved analytically in the orthogonal case when $p = k$ and $\mathcal{OS}t(p,k) = \mathcal{O}(p)$, see [8]. In this case $Q \in \mathcal{O}(p)$ and we have

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

Provided that the singular value decomposition 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 bilinear functional $trace(Q^T\Sigma^2 Q)$ and the linear functional $-2trace(Q^T\Sigma A)$. It is well known how to minimize each of the functionals separately.

The minimum value of the bilinear functional is equal to the sum of squares of the k smallest diagonal entries of Σ. This result is due to Ky Fan, [9]. The linear functional is minimized when $trace(Q^T\Sigma A)$ 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 J.Von Neumann in [13], see also [8].

Separate minimization of the quadratic and the linear part are well understood since we know both the analytical solutions and robust numerical methods. The analytical solution of the orthogonal Procrustes problem for Stiefel matrices is not known to the best of our knowledge and constitutes a major challenge.

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 $\mathbb{R}^p$:

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

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 eccentric Stiefel 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.

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 :

$Q^{[m]} \in \mathbb{R}^{1\times k}$ consists of the m-th row of Q

$Q^{[m,n]} \in \mathbb{R}^{2\times k}$ consists of the m and n-th 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 m-th and n-th rows and columns of Q.

A plane rotation by an angle ϕ in the (k,l)-plane in $\mathbb{R}^p$ is represented by a matrix $G_{k,l}(\phi)$ 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 similarily 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 succesive 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 $\hat{Q}$ to the minimizer Q^* by the rule

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

where Q and $\hat{Q}$ 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 to a sweep. Our relaxation method consists of repeated applications of sweeps which produce a minimizing sequence for the problem (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

$$\begin{aligned} Q_0 =& Q \\ Q_i =& R_{i-1} \cdots R_1 Q \ , \quad i = 2, \dots , N+1 \ , \end{aligned} \tag{3.3}$$

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 $P_i \in \mathcal{O}(p)$ 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),(3.3),(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 $\mathcal{N}$: $\mathcal{D} \to \{1, 2, ..., N\}$ is an 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

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

The relaxation method defined by (3.6) will be refered to as a *left-sided relaxation* method or 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

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

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

A specific type of a right-sided relaxation method was investigated by H. Park in [11]. The method in [11] is based on the concepts discussed earlier by Ten Berge and Knol [2] where the Procrustes problem for orthogonal Stiefel matrices, called there unbalanced problem, is solved by means of iteratively solving a sequence of orthogonal Procrustes problems, called balanced problems. The relaxation approach reveals new left-sided relaxation method which is the topic of our paper.

Another interesting aspect of this approach is a clear geometric interpretation of the relaxation step. For the study of other minimization methods on submanifolds of spaces of matrices see [10] and [12] .

4. Planar Procrustes problem

We will now present the left-sided relaxation method. 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, see [5].

In this case $\mathcal{N} : \mathcal{D} \to \{1, \ldots, \frac{1}{2}p(p-1)\}$, $\mathcal{D} = \{(r,s) : 1 \leq p-1, r+1 \leq s \leq p\}$ 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

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

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

$$\begin{aligned} f(\alpha) &= \left\| \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \end{pmatrix} - \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} G(\alpha) \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1k} \\ q_{21} & q_{22} & \cdots & q_{2k} \end{pmatrix} \right\|^2 \\ &= \left\| A^{[1,2]} - \Sigma_{[1,2]}^{[1,2]} G(\alpha) Q^{[1,2]} \right\|^2 . \end{aligned} \tag{4.3}$$

Let $U_Q \Gamma V_Q^T$ be the SVD decomposition of $Q^{[1,2]}$ such that $\Gamma = diag(\gamma_1, \gamma_2)$ and $\gamma_1 \geq \gamma_2 > 0$. 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 $p-2$ columns of the matrix B in (4.4) are always approximated by zero columns, and thus the minimization of $f(\alpha)$ is equivalent to minimization restricted to the first two columns, that is

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

where $c = \cos\phi$, $s = \sin\phi$, $\phi = \alpha + \beta$ and β is the angle of the plane rotation (or reflection) U_Q.

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

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

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:

$$\begin{aligned} F(\phi) &= \left(\frac{y_1}{z_1} - z_1\cos\phi\right)^2 + \left(\frac{y_2}{z_2} - z_2\sin\phi\right)^2 - ||C||^2 + ||A^{[1,2]}||^2 \\ &= ||C - Zq(\phi)||^2 - ||C||^2 + ||A^{[1,2]}||^2 \end{aligned} \tag{4.9}$$

for $q(\phi) = (\cos\phi, \sin\phi)^T$. Thus the minimization of the functional (4.9) 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

$$\begin{aligned} \mathcal{P}[C, Z](q) &= z_1^2\cos^2\phi + z_2^2\sin^2\phi - 2c_1 z_1\cos\phi - 2c_2 z_2\sin\phi + ||C||^2 \\ &= ||C - Zq||^2 \end{aligned} \tag{4.10}$$

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

The minimization problem of the type (4.10) will be called a planar Procrustes problem. Such problem has to be solved on each step of our relaxation method and is geometrically equivalent to projecting C onto an ellipse. In the next section we consider two different iterative methods for finding the projection.

5. Projection on an ellipse

The geometrical formulation of the planar Procrustes problem (4.10) 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. We describe the classical projection of a point onto an ellipse due to Appolonius leading to a scalar fourth order algebraic equation and a method of iterated reflections based on the reflection property of the ellipse.

5.1. **The hyperbola of Appolonius.** Recall the construction of a normal to an ellipse from a point, see [14], due to Appolonius. With the given ellipse $\mathcal{E}$ in (5.1) and with the point C we associate the hyperbola $\mathcal{H}$ in the following way. The equation of the normal to the ellipse at $S = (s_1, s_2)^T$ is

$$\frac{s_2}{z_2^2}(\xi_1 - s_1) - \frac{s_1}{z_1^2}(\xi_2 - s_2) = 0. \tag{5.2}$$

Clearly CS is the normal to $\mathcal{E}$ iff the point $C = (c_1, c_2)^T$ satisfies (5.2) for $\xi_1 = c_1$ and $\xi_2 = c_2$. Equivalently CS is the normal to $\mathcal{E}$ iff the point S satisfies

$$\frac{x_2}{z_2^2}(c_1 - x_1) - \frac{x_1}{z_1^2}(c_2 - x_2) = 0 \tag{5.3}$$

for $x_1 = s_1$ and $x_2 = s_2$. (5.3) is a quadratic, in x_1 and x_2, equation of the hyperbola $\mathcal{H}$ which can be written as

$$x_2 = \frac{m_2 x_1}{x_1 - m_1} \tag{5.4}$$

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

$$m_1 = \frac{c_1 z_1^2}{z_1^2 - z_2^2} \qquad m_2 = -\frac{c_2 z_2^2}{z_1^2 - z_2^2}. \tag{5.5}$$

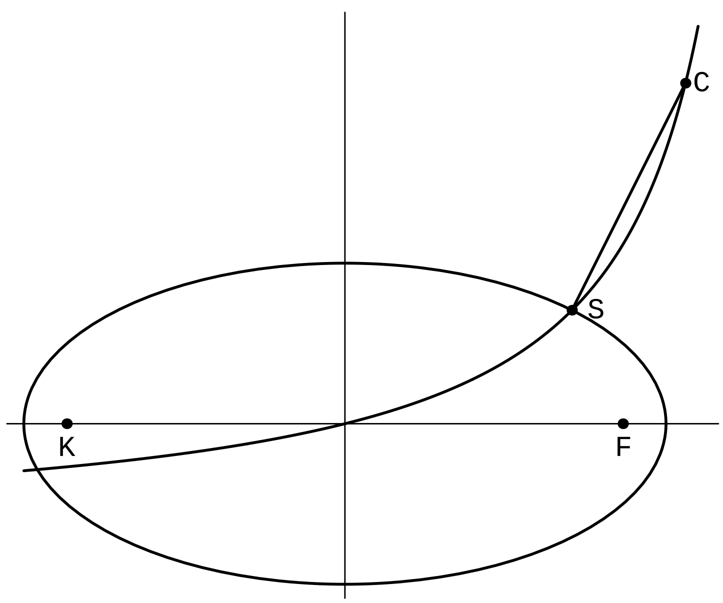

Figure 1. Hyperbola of Appolonius

The hyperbola $\mathcal{H}$ has asymptotes parallel to the axes of the ellipse and passes through the origin and the point C. $\mathcal{H}$ degenerates to a line when the point C is on one of the axes of the ellipse. This nongeneric, trivial case will not be analyzed here. $\mathcal{H}$ can also be characterized as a locus of the centers of the conics in the pencil generated by the ellipse and an arbitrary circle centered at C .

To find the coordinates of the projection point S we have to intersect the hyperbola of Appolonius with the ellipse that is to solve a system of two quadratic equations (5.1) and (5.4). Using (5.4) to eliminate x_2 from (5.1) we obtain a fourth order polynomial equation in x_1

$$(x_1 - m_1)^2 (z_1^2 - x_1^2) - \left(\frac{z_1 m_2}{z_2}\right)^2 x_1^2 = 0. \tag{5.6}$$

For any specific numerical values of the coefficients this equation can be easily solved symbolically. A simpler, purely numerical alternative is to solve the system (5.1) , (5.4) using Newton's method.

Another alternative is to reduce the system to a scalar 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 (5.3) and next substituting $t = \tan\phi$ leads to the equation $g(t) = 0$ in t, where

$$g(t) = c_1 z_1 t - (z_1^2 - z_2^2) t (1 + t^2)^{-\frac{1}{2}} - c_2 z_2. \tag{5.7}$$

It is easy to see that

$$g''(t) = 3(z_1^2 - z_2^2) t (1 + t^2)^{-\frac{5}{2}} > 0$$

and since $g(0) = -c_2 z_2$ and $\lim_{t\to\infty} g(t) = \infty$, then the function $g(t)$ is convex and has one positive root. 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, Newton method starting from the initial approximation t_0 will generate a decreasing, convergent sequence of approximations to the root of $g(t) = 0$.

5.2. **Iterated reflections.** Assume, as before, that C is in the first quadrant and that $C \notin \mathcal{E}$, see Fig.2. Every other case reduces to this one through reflections.

Let

$$C = (c_1, c_2)^T \qquad B = \left(\frac{1}{z_1}(z_1^2 - z_2^2), 0\right)^T \tag{5.8}$$

$$F = \left(\sqrt{z_1^2 - z_2^2}, 0\right)^T \qquad K = -F \tag{5.9}$$

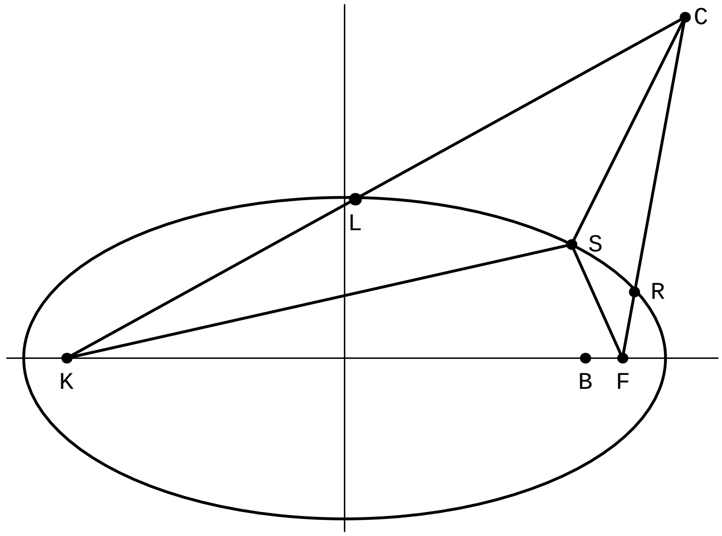

Figure 2. Iterated reflections

The reflection property of $\mathcal{E}$ says that (for C both inside and outside $\mathcal{E}$) S is characterized by :

$$\not< KSC = \not< FSC$$

Let $L, R \in \mathcal{E}$, as on Figure 2, be given by

$$L = (z_1 \cos\phi_L, z_2 \sin\phi_L)^T \qquad R = (z_1 \cos\phi_R, z_2 \sin\phi_R)^T$$

where for C outside $\mathcal{E}$:

$$\tan\phi_L = \frac{z_1 c_2}{z_2 c_1} \tag{5.10}$$

$$\cos\phi_R = \frac{m^2 z_1^2 \sqrt{z_1^2 - z_2^2} + \sqrt{(1+m^2) z_1^2 z_2^4}}{m^2 z_1^3 + z_1 z_2^2} \qquad m = \frac{c_2}{c_1 - \sqrt{z_1^2 - z_2^2}} \tag{5.11}$$

(In the case C inside $\mathcal{E}$, the coordinates of the points L and R can be computed similarly).

Clearly $\phi_R < \phi < \phi_L$. Analogously as in the bisection method, we compute $M = (z_1 \cos\phi_M, z_2 \sin\phi_M)$ (an intermediate point between L and R) from L and R by setting $\phi_M = \frac{1}{2}(\phi_L + \phi_R)$. If

$$\frac{||F - M||}{||K - M||} < \frac{\langle F - M, C - M\rangle}{\langle K - M, C - M\rangle} \tag{5.12}$$

(where $\langle \cdot, \cdot \rangle$ denotes the inner product), then $\phi_R \le \phi \le \phi_M$ and we set $L_1 = M$ and $R_1 = R$ so that $\phi_{R_1} \le \phi \le \phi_{L_1}$. If (5.12) doesn't hold we set $L_1 = L$ and $R_1 = M$. We thus construct a sequence $\{L_n\} \subset \mathcal{E}$ such that $\lim_{n\to\infty} L_n = S$.

5.3. **Some remarks on the general and planar Procrustes problems.** The planar Procrustes problem has several features which the general problem (1.4) of projection onto the eccentric Stiefel manifold does not posses.

- $\mathcal{E}$ has the reflection properties and Appolonius normal.

The reflection properties of the ellipse do not extend to the eccentric Stiefel manifold and in particular not even to ellipsoids in $\mathbb{R}^p$. The construction of an Appolonius normal to the ellipse based on the orthogonality of the ellipse and an associated hyperbola which results in a scalar equation (5.6) is also particular to the planar problem. As a result in the case $p > k > 1$ our relaxation step, which amounts to solving a planar Procrustes problem, cannot be directly generalized to a higher dimensional problem.

• A point not belonging to $\mathcal{E}$ has either a unique projection onto $\mathcal{E}$ or a finite number of projections.

Hence if $C \notin conv(\mathcal{E})$ then there exits a unique projection S of C onto $\mathcal{E}$ characterized by

$$\langle C - S, F - S\rangle \leq 0$$

for all $F \in conv(\mathcal{E})$. A point C inside the ellipse $\mathcal{E}$ has a non-unique projection S onto $\mathcal{E}$ iff C is on the major axis between points B and $-B$. Various locations of C and its projection(s) S are shown on Fig. 3. along with the upper part of the evolute of the ellipse which is the locus of the centers of curvature of $\mathcal{E}$. The number of normals 2,3 or 4 that can be drawn from C depends on the position of C relative to the evolute.

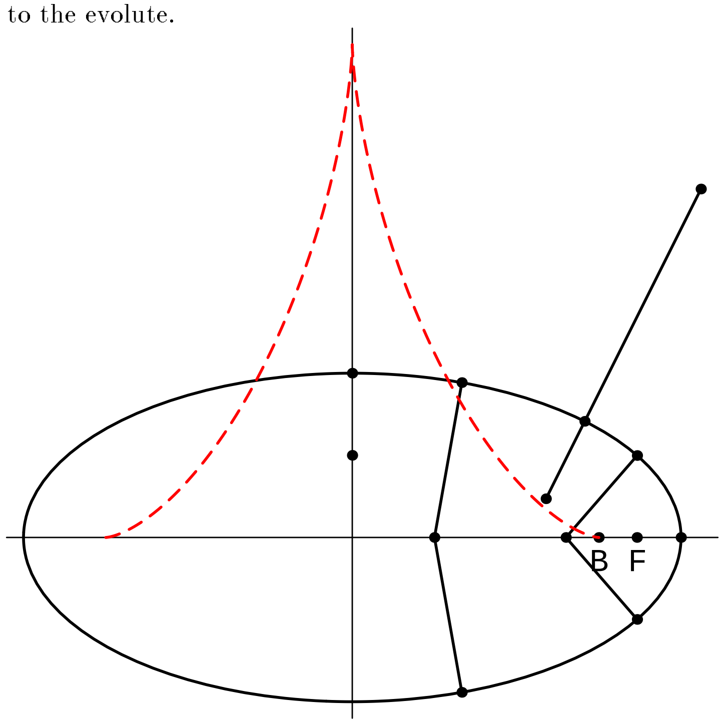

Figure 3. Points in the first quadrant and their projections on the ellipse.

The non-uniqueness of S for C on the segment $[-B, B]$ is reflected in non-differentiability of the function $d(C) = \min_q \|C - Zq\|$ shown on Fig. 4. In general solutions to the Procrustes problem may form a submanifold of the Stiefel manifold.

Finally we observe that $\mathcal{E}$ cuts the plane into two components. However if

$$C = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{5.13}$$

then C is on a sphere of radius $\sqrt{2}$ in $\mathbb{R}^{3\times 2}$ containing the Stiefel manifold $\mathcal{OS}t[I](3,2)$ but the point kC for any $k \in \mathbb{R}$ cannot be connected to 0 with a segment intersecting $\mathcal{OS}t(3,2)$.

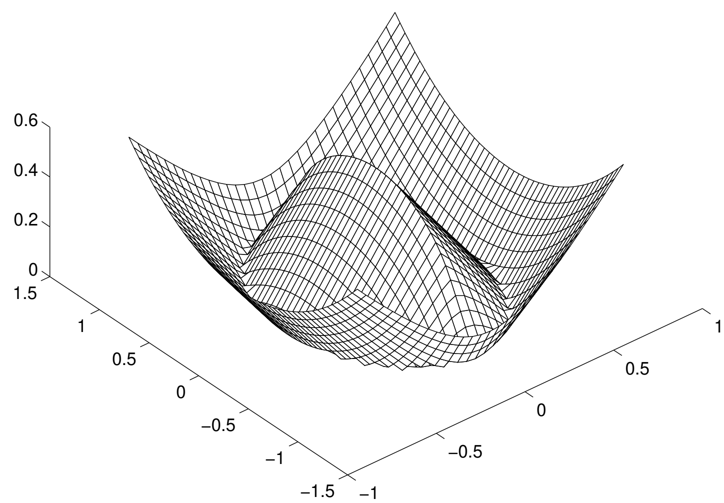

Figure 4. Graph of $d(C) = \min_{||q||=1} ||C - Zq||$

Both observations concerning the analytical problem are reflected in the computations and have computational implications.

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 vecors 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 $Q = (q_1, q_2, \ldots, q_k)$ 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 left-sided relaxation method the basis is the cannonical 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 right-sided relaxation method 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 in [11] which can be solved by means of an

SVD computation. The relaxation step in [11] 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 right-sided relaxation method 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 tangent of α and is equivalent to the 2×2 SVD computation in [11]. 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 might happen that $span\{q_1, \dots, q_k\} \neq span\{q_1^*, \dots, q_k^*\}$, and hence it might 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 $span(q_1^*, \dots, q_k^*) \subset span(q_1, \dots, q_p)$. The scalar minimization subproblems in [11] involving vectors from both subsets are referred to in [11] 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 (4.10) 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 the 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. We will start by summarizing the left and the right relaxation methods given below in pseudocodes.

Given $A \in \mathbb{R}^{p \times k}$, $A = (a_1, ..., a_k)$, and $\Sigma = \mathrm{diag}(\sigma_1, ..., \sigma_p)$ both algorithms construct of 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 methods by the number of sweeps performed by each of the two methods.

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 adition, 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 performend by each of the methods will give some bases for comparing the convergence behavior of the two methods.

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

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

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

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	

Table 1. Matrices Q in a minimizing sequence generated by LSRM.

We will now present comperative numerical results for the LSRM and RSMR methods.

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

In the first class of examples the linear term dominates the bilinear term or in other words when $||A|| >> ||\Sigma||$. 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 bilinear 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 exampless consists of cases when the functional is 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 $\Sigma \approx I$, 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 4.2 and run on an HP9000 workstation with the machine relative precision $\epsilon = 2.2204e-16$. We set $Maxstep = 30$ and $threshold = 5\epsilon$. 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. The planar Procrustes solver used was based on the hyperbola of Appolonius (the iterated reflections solver was giving numerically equivalent results). Some representative results are shown in Tables 2-6.

		RSRM			LSRM		
p	k	# sweeps	$\|A-\Sigma Q\|$	$\|Q^*-Q\|$	# sweeps	$\|A-\Sigma Q\|$	$\|Q^*-Q\|$
6	2	26	2.38e-15	5.30e-15	19	9.22e-16	2.37e-15
	3	30	4.41e-12	1.33e-11	25	1.48e-15	2.53e-15
	4	30	5.21e-11	1.61e-10	26	1.99e-15	6.30e-15
	5	24	5.86e-15	1.07e-14	30	6.00e-11	1.35e-10
9	2	30	6.85e-09	5.79e-08	30	9.78e-06	6.11e-05
	3	30	8.02e-06	6.89e-05	30	7.01e-07	3.18e-06
	4	30	3.66e-03	3.87e-02	30	2.07e-05	1.28e-04
	5	30	5.60e-03	5.93e-02	30	5.06e-07	1.98e-06
	6	30	4.57e-04	4.68e-03	30	5.32e-12	1.16e-11
	7	30	1.62e-03	1.08e-02	30	1.77e-13	3.05e-13
	8	30	2.46e-03	2.21e-02	30	2.98e-12	5.34e-12

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

		RSRM			LSRM		
p	k	# sweeps	$\|A-\Sigma Q\|$	$\|Q^*-Q\|$	# sweeps	$\|A-\Sigma Q\|$	$\|Q^*-Q\|$
6	2	30	9.00e-03	1.53e+00	26	2.08e-16	4.69e-15
	3	30	7.49e-03	1.55e+00	18	4.47e-16	1.62e-14
	4	30	7.66e-03	1.73e+00	16	5.56e-16	2.86e-15
	5	30	3.10e-03	8.62e-01	30	8.99e-16	7.98e-15
9	2	30	6.82e-03	1.70e+00	30	9.08e-11	2.62e-08
	3	30	5.04e-03	1.41e+00	30	9.71e-07	2.84e-04
	4	30	4.67e-03	1.35e+00	30	1.19e-04	3.39e-02
	5	30	7.04e-03	2.21e+00	30	1.18e-03	3.01e-01
	6	30	6.45e-03	1.96e+00	30	3.82e-06	1.48e-05
	7	30	4.93e-03	1.49e+00	30	3.53e-06	1.20e-05
	8	30	4.87e-03	1.54e+00	30	8.44e-10	1.70e-09

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

Table 2 illustrates the behavior of the two 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 less sweeps to obtain a satisfactory approximation to the minimizer.

Table 3 illustrates the behavior of the two methods when the length of the 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 convergance 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.

		RSRM				LSRM			
p	k	# sweeps	$esterror$	$\|A-\Sigma Q\|$	$sweepcorr$	# sweeps	$esterror$	$\|A-\Sigma Q\|$	$sweepcorr$
6	2	30	-1.70e-02	2.12e-02	1.89e-08	6	-1.70e-02	2.12e-02	6.10e-16
	3	30	-2.64e-02	3.38e-02	3.93e-08	13	-2.64e-02	3.38e-02	3.81e-16
	4	30	-5.03e-02	6.10e-02	1.10e-07	10	-5.03e-02	6.10e-02	3.53e-16
	5	30	2.55e-02	9.04e-01	3.32e-09	30	2.55e-02	9.04e-01	1.11e-15
9	2	30	-4.41e-02	4.53e-02	8.44e-08	9	-4.41e-02	4.53e-02	6.93e-18
	3	30	-3.61e-02	3.86e-02	4.26e-08	14	-3.61e-02	3.86e-02	5.41e-16
	4	30	-4.05e-02	4.60e-02	4.98e-06	11	-4.05e-02	4.59e-02	5.75e-16
	5	30	-5.44e-02	6.30e-02	1.06e-06	15	-5.41e-02	6.27e-02	7.35e-16
	6	30	3.49e-02	7.10e-01	1.00e-06	9	3.49e-02	7.10e-01	9.99e-16
	7	30	4.89e-02	1.04e+00	2.67e-07	15	4.89e-02	1.04e+00	2.22e-16
	8	30	5.53e-02	1.34e+00	1.88e-08	14	5.53e-02	1.34e+00	6.66e-16

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

		RSRM				LSRM			
p	k	# sweeps	$esterror$	$\|A-\Sigma Q\|$	$sweepcorr$	# sweeps	$esterror$	$\|A-\Sigma Q\|$	$sweepcorr$
6	2	6	-3.32e-02	1.36e+01	7.10e-15	30	-3.32e-02	1.36e+01	5.27e-13
	3	5	-1.51e-02	3.01e+01	3.55e-15	20	-1.51e-02	3.01e+01	7.10e-15
	4	4	-1.81e-02	2.51e+01	7.10e-15	30	-1.81e-02	2.51e+01	3.55e-15
	5	4	-1.43e-02	3.16e+01	3.55e-15	28	-1.43e-02	3.16e+01	0.00e+00
9	2	4	-1.62e-02	1.60e+01	0.00e+00	6	-1.62e-02	1.60e+01	0.00e+00
	3	6	-2.37e-02	2.75e+01	1.06e-15	20	-2.37e-02	2.75e+01	1.77e-15
	4	7	-3.40e-02	3.66e+01	2.13e-15	8	-3.40e-02	3.66e+01	0.00e+00
	5	5	-3.26e-02	3.82e+01	1.42e-15	14	-3.26e-02	3.82e+01	1.42e-15
	6	5	-2.87e-02	4.33e+01	0.00e+00	11	-2.87e-02	4.33e+01	0.00e+00
	7	5	-2.69e-02	4.62e+01	7.10e-15	14	-2.69e-02	4.62e+01	7.10e-15
	8	5	-2.70e-02	4.61e+01	2.13e-15	22	-2.70e-02	4.61e+01	7.10e-15

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

Table 4 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 bilinear term dominates the linear term. In this case the minimum of the functional can be estimated by the minimum value of the bilinear term. In Table 4 $esterror$ denotes the difference between the minimum value of the bilinear term and the computed value of the functional, and $sweepcorr = ||A - \Sigma Q|| - ||A - \Sigma \hat{Q}||$ where Q and $\hat{Q}$ are the last and penultimate approximations to the minimizer. The experiments suggest that the LSRM requires less sweeps to obtain a satisfactory approximation to the minimizer.

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. In Table 5 $esterror$ denotes the difference between the minimum value of the linear term and the computed value of the functional. The experiments suggest that the RSRM requires less sweeps to obtain a satisfactory approximation to the minimizer.

8. Remarks on Constrained Linear Least Squares Problems

The quadratically constrained linear least squares problem

$$\min_{||x||=\alpha} ||b - Ax|| \tag{8.1}$$

arises in many applications [6], [7], [3]. By changing variables this problem can be transformed into a special Procrustes problem. The Procrustes problem is

$$\min_{||q||=1} ||a - \Sigma q||^2 \tag{8.2}$$

where $A = U\Sigma V^T$, $a = U^T b/\alpha$ and $q = V^T x/\alpha$. This Procrustes problem is equivalent to projecting the point a onto the ellipsoid $\mathcal{OS}t[\Sigma](p,1)$.

Let $||q|| = 1$. It is clear that the vector $n = \Sigma^{-1}q$ has the direction of the normal vector to the ellipsoid at the point Σq. Thus if Σq^* is the projection of a on the ellipsoid then the vector $a - \Sigma q^*$ is parallel to the vector n. Thus there exists a scalar β so

$$a_i - \sigma_i q_i = \beta \frac{q_i}{\sigma_i} \tag{8.3}$$

where $\Sigma = diag(\sigma_i)$. As $||q|| = 1$ one can obtain an equation for β

$$1 = \sum_{i=1}^{p} \left(\frac{a_i \sigma_i}{\beta + \sigma_i^2} \right)^2 . \tag{8.4}$$

The parameter β can be computed by solving this equation. Then the components of q are given by

$$q_i = \frac{a_i}{\sigma_i + \frac{\beta}{\sigma_i}} .$$

The equation (8.4) is the so-called secular equation which characterizes the critical points of the Lagrangian

$$h(\lambda, q) = ||\Sigma q - a||^2 + \lambda(||q||^2 - 1), \tag{8.5}$$

see [8]. Thus the multiplier β in (8.3) is the Lagrange multiplier λ in (8.5).

References

[1] J.M. Ten Berge and K. Nevels, *A general solution to Mosier's oblique Procrustes problem*, Psychometrika **42** (1977), 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), 49–55.

[3] L. Elden, *Algorithms for the regularization of ill-conditioned least squares problems* , **17** (1977), 134-145.

[4] G.E. Forsythe and G.H. Golub, *On the stationary values of a second-degree polynomial on the unit sphere* , SIAM **4** (1965), 1050–1068.

[5] G.E. Forsythe and P. Henrici, *The cyclic Jacobi method for computing the principal values of a complex matrix*, Trans. AMS **94** (1960), 1–23.

[6] W. Gander, *Least Squares with a Quadratic Constraint* , Numer. Math. **36** (1981), 291-307.
[7] G.H. Golub and U. von Matt, *Quadratically constrained least squares and quadratic problems*, Numer. Math. **59** (1991), 561–580.
[8] G.H. Golub and C.F. Van Loan, *Matrix Computations* Johns Hopkins, Baltimore,1990.
[9] Ky Fan, *On a theorem of Weyl concerning eigenvalues of linear transformations*, Proc. N.A.S. **35** (1949), 652–655.
[10] A. Lutoborski, *On the convergence of the Euler-Jacobi method*, Numer. Funct. Anal. and Optimiz. **13** (1992), 185–202.
[11] H. Park, *A parallel algorithm for the unbalanced orthogonal Procrustes problem*, Parallel Computing **17** (1991), 913–923.
[12] S.T. Smith, *Optimization techniques on Riemannian manifolds*, Fields Institute Comm. **3** (1994), 113–136.
[13] J. Von Neumann, *Some matrix inequalities and metrization of the matrix space*, Tomsk Univ. Rev. (1937), 286–300.
[14] H. Weber and J.Wellstein, *Enzyklopadie der Elementar Mathematik* B.G. Teubner, Leipzig, 1915.

Electrical Engineering Department, Cornell University, Ithaca, N.Y. 14853
E-mail address: `adamb@ee.cornell.edu`

Department of Mathematics, Syracuse University, Syracuse, N.Y. 13244-1150
E-mail address: `lutobor@mazur.syr.edu`