BIT 28 (1988), 133-143

THE SYMMETRIC PROCRUSTES PROBLEM

NICHOLAS J. HIGHAM

Department of Mathematics, University of Manchester, Manchester M13 9PL, England

Abstract.

The following "symmetric Procrustes" problem arises in the determination of the strain matrix of an elastic structure: find the symmetric matrix X which minimises the Frobenius (or Euclidean) norm of $AX-B$, where A and B are given rectangular matrices. We use the singular value decomposition to analyse the problem and to derive a stable method for its solution. A perturbation result is derived and used to assess the stability of methods based on solving normal equations. Some comparisons with the standard, unconstrained least squares problem are given.

1980 Mathematics Subject Classification: Primary 65F20.

Keywords: Least squares, symmetric Procrustes problem, singular value decomposition, normal equations.

1. Introduction.

Consider the class of constrained least squares approximation problems: Find

$$\min_{X \in P} \|AX - B\|, \quad A, B \in \mathbb{R}^{m \times n}, \tag{1.1}$$

where $P \subseteq \mathbb{R}^{n \times n}$, and $\|\cdot\|$ denotes the Frobenius (or Euclidean) norm,

$$\|Y\| = \left(\sum_{i,j} y_{ij}^2\right)^{1/2}.$$

A variety of matrix approximation problems is contained in the class (1.1). With no constraints on X ($P = \mathbb{R}^{n \times n}$) a standard least squares problem is obtained, having a solution $X = A^+B$, where A^+ is the pseudo-inverse of A (see, for example, [4, Ch. 6]).

Taking for P the set of orthogonal matrices yields the orthogonal Procrustes*

Received January 1987. Revised June 1987.

* Procrustes: an ancient Greek robber who tied his victims to an iron bed, stretching their legs if too short for it, and lopping them if too long.

problem, which arises in factor analysis [5, 14]. A solution is $X = U$ where $B^T A = UH$ is a polar decomposition [8].

When $m = n$ and $A = I$, (1.1) is a matrix nearness problem: find the nearest matrix to B in the set P. An important nearness problem is to find the nearest normal matrix; see [13].

In this paper we explore the problem obtained from (1.1) when one imposes what is perhaps the simplest constraint on X, that of symmetry. This problem received attention in the literature two decades ago [2, 11], but seems to have been neglected since. We will refer to this problem as the "symmetric Procrustes" (SP) problem, by analogy with the orthogonal version mentioned above. The problem arises in the investigation of elastic structures wherein vectors f_i of observed forces are postulated to be related to vectors d_i of observed displacements according to $Xf_i = d_i$, where X is the symmetric strain (or flexibility) matrix [2, 11] (see also [15, Sec. 3.5]). X is to be estimated from the experimental data in a least squares sense.

In section 2 we analyse the SP problem using the singular value decomposition (SVD). The general solution is derived and its properties investigated. A theorem is derived which describes the sensitivity of the SP problem to perturbations in the data.

In section 3 we consider computation of the solution. A stable method based on the SVD is proposed and compared with other methods based on solving normal equations.

Finally, in section 4 we give a numerical example and compare the symmetric Procrustes solution with the unconstrained least squares solution.

This work extends the treatments in [2] and [11], the main contributions being use of the SVD to solve and analyse the SP problem, and examination of the stability of methods for numerical solution.

We remark that the theory and algorithms presented here are easily adapted to the "skew-symmetric Procrustes" problem, in which the constraint in (1.1) is taken to be that of skew symmetry.

2. Analysis.

In this section we investigate the mathematical properties of the SP problem

$$\min_{X = X^T} \|AX - B\|, \quad A, B \in \mathbb{R}^{m \times n}, \quad m \geqslant n. \tag{2.1}$$

As might be expected, there are many similarities to the standard least squares (LS) problem. In particular, the appropriate tool for analysis is the SVD.

Let A have the SVD

$$A = P \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} Q^T, \tag{2.2}$$

where $P \in \mathbb{R}^{m \times n}$ and $Q \in \mathbb{R}^{n \times n}$ are orthogonal and

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n), \quad \sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant 0.$$

Using the invariance of the Frobenius norm under orthogonal transformations we have

$$\begin{aligned} \|AX - B\|^2 &= \left\| P \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} Q^T X - B \right\|^2 = \left\| \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} Q^T X Q - P^T B Q \right\|^2 \\ &= \left\| \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} Y - C \right\|^2 = \|\Sigma Y - C_1\|^2 + \|C_2\|^2, \end{aligned}$$

where

$$Y = Q^T X Q,$$

$$C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = P^T B Q, \quad C_1 = (c_{ij}) \in \mathbb{R}^{n \times n}.$$

Thus the SP problem reduces to minimising the quantity

$$\|\Sigma Y - C_1\|^2 = \sum_{i=1}^{n} (\sigma_i y_{ii} - c_{ii})^2 + \sum_{j>i} ((\sigma_i y_{ij} - c_{ij})^2 + (\sigma_j y_{ij} - c_{ji})^2), \tag{2.3}$$

where we have used the required symmetry of Y. The variables $y_{ij} (j \geqslant i)$ in (2.3) are uncoupled, so it suffices to minimise independently each of the terms $(\sigma_i y_{ii} - c_{ii})^2$ and $(\sigma_i y_{ij} - c_{ij})^2 + (\sigma_j y_{ij} - c_{ji})^2$, $(j > i)$. The general solution is easily found to be, for $1 \leqslant i \leqslant j \leqslant n$,

$$y_{ij} = \begin{cases} \dfrac{\sigma_i c_{ij} + \sigma_j c_{ji}}{\sigma_i^2 + \sigma_j^2}, & \sigma_i^2 + \sigma_j^2 \neq 0, \\ \text{arbitrary}, & \text{otherwise}. \end{cases} \tag{2.4}$$

The required solution X is given by $X = QYQ^T$.

Note that if the symmetry constraint were not present, then, assuming $\operatorname{rank}(A) = n$, we could take $Y = \Sigma^{-1} C_1$ and obtain the residual

$$\varrho_{LS} = \|AX_{LS} - B\| = \|C_2\|. \tag{2.5}$$

For $\operatorname{rank}(A) = n$, the residual for the SP problem can be shown to be

$$\begin{aligned} \varrho_{SP}^2 &= \|AX_{SP} - B\|^2 \\ &= \sum_{j>i} \frac{(\sigma_i c_{ji} - \sigma_j c_{ij})^2}{\sigma_i^2 + \sigma_j^2} + \|C_2\|^2. \end{aligned} \tag{2.6}$$

Expressions (2.5) and (2.6) quantify the effect on the residual of requiring X to be symmetric. The summation term in (2.6) has no obvious interpretation, but of course it must be zero if $X_{LS} = A^+B$ is symmetric

Several interesting properties of the set of minimisers

$$S = \left\{ X \in \mathbb{R}^{n \times n} : X = X^T \quad \text{and} \quad \|AX - B\| = \min_{Z = Z^T} \|AZ - B\| \right\}$$

are displayed in the following lemma (cf. the corresponding properties for the LS problem [4, p. 138]).

LEMMA 1.

(a) $X \in S \Leftrightarrow X = X^T$ *and* $A^TAX + XA^TA = A^TB + B^TA$.

(b) S *is convex.*

(c) S *has a unique element* X_{SP} *of minimal Frobenius norm.*

(d) $S = \{X_{SP}\} \langle = \rangle \operatorname{rank}(A) = n$.

PROOF (Sketch).

(a): The matrix equation is obtained by manipulating (2.4). An alternative derivation from first principles is possible by considering the expression $f(X+E) \geqslant f(X)$, for $E = E^T$, where $f(X) = \operatorname{trace}((AX-B)^T(AX-B)) = \|AX-B\|^2$.
(c): X_{SP} is obtained by setting the arbitrary elements in (2.4) to zero. (b) follows from (a), and (d) from (2.4). ■

We will refer to the equations

$$A^TAX + XA^TA = A^TB + B^TA \tag{2.7}$$

as the *normal equations* for the SP problem, since they are analogous to the normal equations for the LS problem.

A special case of the SP problem is to find the nearest symmetric matrix to a given square matrix B ($m = n$ and $A = I$). The normal equations (2.7) yield immediately the solution $X = (B + B^T)/2$, which is unique by Lemma 1(d). This solution is well-known [3].

In stating the SP problem Brock [2] actually requires that X be symmetric *positive definite*. However, his method of solution, like the one in [11], consists of deriving the normal equations and describing a method for solving them. The definiteness of the solution is not discussed. Clearly, X_{SP} will not, in general, be positive definite. (Consider, for example, the nearest symmetric matrix problem above.) However, a sufficient condition for the definiteness of X_{SP} is easily derived.

LEMMA 2. *If A is of full rank and $B^TA + A^TB$ is positive (semi-)definite then X_{SP} is positive (semi-)definite.*

PROOF. A^TA has positive eigenvalues, so the unique solution X_{SP} to the normal equations (2.7) may be expressed in the form [10, p. 414]

$$X_{SP} = \int_0^\infty e^{-A^TAt}(B^TA + A^TB)e^{-A^TAt}\,dt.$$

The result follows by considering the quadratic form $y^TX_{SP}y$. ■

We note that the condition in Lemma 2 is not necessary—an example in section 4 shows that X_{SP} can be positive definite when $A^TB + B^TA$ is indefinite.

A natural specialisation of the SP problem is (1.1) with P the set of symmetric positive semi-definite matrices. (We must allow X to have zero eigenvalues for the problem to be well-posed.) For $m = n$ and $A = I$, solutions are known in the case of the 2-norm [6, 7], but the general problem appears to be unsolved.

We conclude this section by examining the sensitivity of the SP problem to perturbations in the data. Our finding will be used in the next section to assess the stability of numerical methods for solving the SP problem. In the following we use the condition number $\kappa_2(A) = \sigma_1/\sigma_r$, where $r = \text{rank}(A)$.

THEOREM 1. *Let $A \in \mathbb{R}^{m\times n}$, $m \geq n$, and let X solve (2.1). Let $\hat{X}$ solve*

$$\min_{X = X^T} \|(A+\delta A)X - (B+\delta B)\|, \tag{2.8}$$

where δA, $\delta B \in \mathbb{R}^{m\times n}$, and define

$$R = AX - B,$$

$$\hat{R} = (A+\delta A)\hat{X} - (B+\delta B),$$

$$\varepsilon_A = \|\delta A\|_2/\|A\|_2,$$

$$\kappa = \begin{cases} \kappa_2(A), & \text{rank}(A) = n, \\ \sqrt{2}\kappa_2(A), & \text{rank}(A) < n. \end{cases} \tag{2.9}$$

Assume that $\text{rank}(A) = \text{rank}(A+\delta A)$ *and* $\kappa\varepsilon_A < 1$. *Then*

$$\|X - \hat{X}\| \leq \frac{\kappa}{1-\kappa\varepsilon_A}\left(\varepsilon_A\|X\| + \frac{\|\delta B\|}{\|A\|_2} + \kappa\varepsilon_A\frac{\|R\|}{\|A\|_2}\right) + \kappa\varepsilon_A\|X\|, \tag{2.10}$$

and

$$\|R - \hat{R}\| \leq \varepsilon_A\|X\|\,\|A\|_2 + \|\delta B\| + \kappa\varepsilon_A\|R\|. \tag{2.11}$$

The last term in (2.10) *can be omitted if* $\text{rank}(A) = n$.

PROOF. Our approach is to reduce the SP problem to an unconstrained LS problem, and then to apply standard perturbation theory.

Let $p = mn$ and $q = n^2$. Using the vec operator, which stacks the columns of a matrix into one long vector, we can write

$$\|AX-B\| = \|\text{vec}(AX-B)\|_2 = \|(I_n \otimes A)x - b\|_2,$$

where $I_n \otimes A = (\delta_{ij}A) \in \mathbb{R}^{p\times q}$, $x = \text{vec}(X) \in \mathbb{R}^q$, and $b = \text{vec}(B) \in \mathbb{R}^p$. The symmetry of X restricts x to an s-dimensional subspace of $\mathbb{R}^q$, where $s = n(n+1)/2$; let S_n be a matrix whose columns form an orthonormal basis for this subspace.

On setting $x = S_n y$, and defining $G = (I_n \otimes A)S_n$, we see that the SP problem (2.1) is equivalent to the unconstrained LS problem

$$\min_y \|Gy-b\|_2. \tag{2.12}$$

Similarly, the perturbed problem (2.8) may be identified with perturbations $G \to G+\delta G$, $b \to b+\delta b$, $y \to \hat{y}$ in (2.12), where

$$\delta G \equiv (I_n \otimes \delta A)S_n, \quad \delta b = \text{vec}(\delta B), \quad S_n\hat{y} = \text{vec}(\hat{X}).$$

Now the singular values of G are related to those of A according to

$$\sigma(G) = \left\{\sqrt{\left(\frac{1}{2}\left[\sigma_i^2(A)+\sigma_j^2(A)\right]\right)}: \quad 1 \le i \le j \le n\right\}.$$

This can be proved using the variational characterisation of singular values [4, p. 286], in conjunction with the relation $Gy = \text{vec}(AY)$ where $Y = Y^T$ and $y = \text{vec}(Y)$. It follows that $\kappa_2(G) = \kappa$ in (2.9), and $\|\delta G\|_2 = \|\delta A\|_2$; hence $\kappa_2(G)\varepsilon_G := \kappa_2(G)\|\delta G\|_2/\|G\|_2 = \kappa\varepsilon_A < 1$. Also, $\text{rank}(A) = \text{rank}(A+\delta A)$ implies $\text{rank}(G) = \text{rank}(G+\delta G)$. Thus we can apply Theorem 5.1 of [16] to (2.12) and its perturbation, to obtain

$$\|y-\hat{y}\|_2 \le \frac{\kappa_2(G)}{1-\kappa_2(G)\varepsilon_G}\left(\varepsilon_G\|y\|_2 + \frac{\|\delta b\|_2}{\|G\|_2} + \kappa_2(G)\varepsilon_G\frac{\|r\|_2}{\|G\|_2}\right) + \kappa_2(G)\varepsilon_G\|y\|_2,$$

where $r = Gy-b$, and where the last term may be omitted if G has full rank. This is easily seen to be equivalent to (2.10). Similarly, a bound from Theorem 5.1 of [16] for the change in the residual yields (2.11). ■

As the proof shows, the bounds in Theorem 1 have the same form as those for the LS problem and their interpretation is identical. For a given problem if the residual is relatively small then $\kappa_2(A)$ measures the sensitivity of X_{SP} to perturbations in the data; otherwise the solution sensitivity is measured by $\kappa_2(A)^2$. Residual sensitivity is always measured by $\kappa_2(A)$.

3. Computation.

The derivation of X_{SP} in section 2 leads to the following computational procedure for solving the SP problem (2.1).

Algorithm SP.

1. Compute the SVD $A = P\begin{bmatrix}\Sigma \\ 0\end{bmatrix}Q^T$, using the Golub-Reinsch SVD algorithm [4, p. 293].
2. Form $C_1 = (c_{ij}) \in \mathbb{R}^{n\times n}$ where $\begin{bmatrix}C_1 \\ C_2\end{bmatrix} = P^TBQ$.
3. Compute the upper triangle of the symmetric matrix $Y \in \mathbb{R}^{n\times n}$ according to

$$y_{ij} = \begin{cases} c_{ii}/\sigma_i, & i = j \leqslant \operatorname{rank}(A), \\ (\sigma_i c_{ij} + \sigma_j c_{ji})/(\sigma_i^2 + \sigma_j^2), & j > i \quad \text{and} \quad i \leqslant \operatorname{rank}(A), \\ 0, & \text{otherwise.} \end{cases}$$

4. Compute the upper triangle of the symmetric matrix $X = QYQ^T$.

Cost. (At most) $2m^2n + 6mn^2 + 37n^3/6$ flops.

Of course, any stable algorithm for computing the SVD could be used in step 1. The issues in the implementation of steps 1 and 2 are essentially the same as those associated with solving the LS problem using the SVD and are discussed in [4, p. 173–175]. For example, if $m \gg n$ it is more efficient not to form P^T explicitly but rather to apply it to B in factored form as it is developed.

Making use of the backward error analysis for the Golub-Reinsch SVD algorithm [4, p. 174] one can prove the following stability result. The computed solution $\hat{X}$ from Algorithm SP satisfies

$$\hat{X} = Z + \delta Z, \quad \|\delta Z\| \leqslant \varepsilon\|Z\|,$$

where Z is the solution to

$$\min_{Z = Z^T} \|(A + \delta A)Z - (B + \delta B)\|,$$

where

$$\|\delta A\| \leqslant \varepsilon\|A\|, \quad \|\delta B\| \leqslant \varepsilon\|B\|,$$

and where ε is a small multiple of the machine precision. Thus $\hat{X}$ is close to the true solution of a problem with slightly perturbed data, and this is all that can be expected of an algorithm for solving the *SP* problem in finite precision arithmetic.

Now we discuss alternative methods for solving the *SP* problem. First, we describe the method of Larson [11]; Brock [2] gives a very similar method. By differentiating $\|AX-B\|^2$ with respect to the elements in the lower triangle of X Larson [11] derives the equations

$$H^T Hx = H^T b, \quad x \in \mathbb{R}^t, \quad H \in \mathbb{R}^{mn \times t}, \quad b \in \mathbb{R}^{mn}, \tag{3.1}$$

where $t = n(n+1)/2$. Here, $b = \mathrm{vec}(B)$, x is $\mathrm{vec}(X)$ with the redundant elements coming from the strictly upper triangular part of X removed, and H is a modified version of the Kronecker product $I_n \otimes A = (\delta_{ij}A)$, of the form illustrated for $m = n = 3$ by

$$Hx = \left[\begin{array}{ccc|cc|c} a_{11} & a_{12} & a_{13} & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 0 \\ \hline 0 & a_{11} & 0 & a_{12} & a_{13} & 0 \\ 0 & a_{21} & 0 & a_{22} & a_{23} & 0 \\ 0 & a_{31} & 0 & a_{32} & a_{33} & 0 \\ \hline 0 & 0 & a_{11} & 0 & a_{12} & a_{13} \\ 0 & 0 & a_{21} & 0 & a_{22} & a_{23} \\ 0 & 0 & a_{31} & 0 & a_{32} & a_{33} \end{array}\right] \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ x_{22} \\ x_{23} \\ x_{33} \end{bmatrix}$$

Note that $\|AX-B\| = \|Hx-b\|_2$ and equations (3.1) are just the normal equations for the *LS* problem $\min_x \|Hx-b\|_2$ (cf. (2.12)). In fact, equations (3.1) are essentially the same as the normal equations (2.7), being obtained from them by using symmetry considerations to reduce the order.

If A has full rank, then $H^T H$ is symmetric positive definite and one could use a Choleski factorisation of $H^T H$ to solve (3.1). However, any method based on forming the normal equations, whether in the form (3.1) or (2.7), has two serious drawbacks; these are well-known in the case of the *LS* problem [4, p. 142]. First, loss of information may occur when the products $A^T A$ and $A^T B$, or $H^T H$ and $H^T b$, are formed in floating point arithmetic. For example, even if A has full rank the computed $A^T A$ may be singular. Second, the normal equations bring about a well-known "condition squaring" effect. To be precise, if (2.7) is solved by a method for equations of the general form $CX + XC = M$, where C is symmetric positive definite and M is symmetric, or if (3.1) is solved by a method for general positive definite linear equations $Cx = b$, then the errors in the computed solutions will be proportional to $\kappa_2(A)^2$ for (2.7) and to $\kappa_2(H)^2$ for

(3.1). Yet, as Theorem 1 shows, the mathematical sensitivity of the *SP* problem is measured by $\kappa_2(A)$ when the residual is small. Thus in small residual problems with an ill-conditioned A, methods based on the normal equations will generally yield less accurate solutions than could be obtained using a stable method such as Algorithm *SP*.

To avoid forming the normal equations (3.1) one can consider using a *QR* factorisation of H to solve directly the *LS* problem $\min_x \|Hx-b\|_2$. However, H is $mn \times n(n+1)/2$, and to be competitive with Algorithm *SP* the *QR* factorisation would have to be accomplished in $O(m^2n)$ flops and $O(mn)$ storage. It appears that fill-in precludes this order of efficiency.

Finally, we mention two methods for solving the "big" normal equations (2.7). Since these are a special form of the Sylvester equation $AX+XB=C$ one could apply the orthogonal reduction technique of Bartels and Stewart [1]. This involves computing Schur decompositions of A and B in the Sylvester equation. However, in (2.7) "A" = "B" = A^TA, and if A^TA is formed explicitly this approach reduces to an unstable variant of Algorithm *SP* that computes the singular values and right singular vectors of A via a spectral decomposition of A^TA. Of course, one can instead use an *SVD* of A, in which case this approach is essentially the same as Algorithm *SP*.

Alternatively, matrix sign function techniques enable (2.7) to be solved using an iterative process involving only matrix multiplications and inversions [9].

In conclusion, Algorithm *SP* is the method of choice for solving the *SP* problem because of its very desirable stability properties. It also has the advantage of being able to cope with rank-deficient problems. We will not attempt to compare the computational costs of the methods discussed in this section. We note, however, that it appears that in the applications in [2] and [11] m and n are relatively small (e.g. $n = 3$), in which case computational cost is unlikely to be a major consideration in the choice of method.

4. Example and discussion.

As an example of an *SP* problem consider the following problem given in [2],

$$A = \begin{bmatrix} 5 & 3 & 2 \\ 1 & 2 & 4 \\ 6 & 0 & 3 \\ -1 & 2 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 15 & 10 & -3 \\ 1 & 5 & 3 \\ 15 & 6 & -3 \\ 2 & 3 & -2 \end{bmatrix} \quad (A^TB+B^TA \text{ is indefinite}).$$

Using a MATLAB [12] implementation of Algorithm *SP*, with machine precision $\approx 10^{-15}$, we obtained the (positive definite) computed solution whose upper triangle is given to four decimal places by

$$X_{SP} = \begin{bmatrix} 2.9339 & .9203 & -.9896 \\ & 1.8791 & .0315 \\ & & .9838 \end{bmatrix}.$$

For comparison, the solution to the unconstrained LS problem $\min_X \|AX-B\|$ is

$$X_{LS} = \begin{bmatrix} 2.9305 & .9305 & -1.0000 \\ .8662 & 1.8662 & .0000 \\ -.9605 & .0395 & 1.0000 \end{bmatrix}.$$

We give also the relative residuals $\varrho(X) = \|AX-B\|/(\|A\|\,\|X\|)$ and the 2-norm condition numbers of the solutions.

	$\varrho(X)$	$\kappa_2(X)$
SP:	1.95E−2	8.38
LS:	1.84E−2	7.82

Note that the relative residual for the SP solution is only slightly larger than that for the LS solution, and the condition numbers of the two solutions are of the same order of magnitude.

We have observed the same behaviour on several randomly generated test problems, with various dimensions m and n, and with different A of varying condition. Of course it is easy to construct problems where the LS solution yields an arbitrarily smaller residual, and problems where one of the LS solution and the SP solution is arbitrarily better conditioned than the other (for example, consider $m = n$ and $A = I$). Our limited experiments lead us to suggest that in applications where there are no a priori constraints on X, X_{SP} may be worth considering as an alternative to X_{LS} if the reduction in storage and subsequent computation brought about by a symmetric, and potentially positive definite, solution are desirable. Indeed if X_{LS} is computed via the SVD then X_{SP} can be computed, using Algorithm SP, at relatively little extra cost.

Acknowledgements.

I thank Ian Gladwell and Des Higham for their comments on the manuscript. Part of this work was done during a visit to Stanford University in summer 1986. I thank Gene Golub for providing financial support and a stimulating working environment. I also thank the referee for strengthening Theorem 1 and shortening the proof.

REFERENCES

1. R. H. Bartels and G. W. Stewart, *Solution of the matrix equation* $AX+XB = C$, Comm. ACM, 15 (1972), pp. 820–826.
2. J. E. Brock, *Optimal matrices describing linear systems*, AIAA Journal, 6 (1968), pp. 1292–1296.
3. K. Fan and A. J. Hoffman, *Some metric inequalities in the space of matrices*, Proc. Amer. Math. Soc., 6 (1965), pp. 111–116.
4. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 1983.
5. B. F. Green, *The orthogonal approximation of an oblique structure in factor analysis*, Psychometrika, 17 (1952), pp. 429–440.
6. P. R. Halmos, *Positive approximants of operators*, Indiana Univ. Math. J., 21 (1972), pp. 951–960.
7. N. J. Higham, *Computing a nearest symmetric positive semi-definite matrix*, to appear in Linear Algebra and Appl.
8. N. J. Higham, *Computing the polar decomposition – with applications*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 1160–1174.
9. W. D. Hoskins, D. S. Meek and D. J. Walton, *The numerical solution of the matrix equation* $XA+AY = F$, BIT, 17 (1977), pp. 184–190.
10. P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Second Edition, Academic Press, New York, 1985.
11. H. J. Larson, *Least squares estimation of the components of a symmetric matrix*, Technometrics, 8 (1966), pp. 360–362.
12. C. B. Moler, *MATLAB users' guide*, Technical Report CS81-1 (revised), Department of Computer Science, University of New Mexico, Albuquerque, 1982.
13. A. Ruhe, *Closest normal matrix finally found!*, BIT 27, (1987), 585–598.
14. P. H. Schonemann, *A generalized solution of the orthogonal Procrustes problem*, Psychometrika, 31 (1966), pp. 1–10.
15. G. Strang, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, Wellesley, Massachusetts, 1986.
16. P.-Å. Wedin, *Perturbation theory for pseudo-inverses*, BIT, 13 (1973), pp. 217–232.