ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
21,2 (1991), pp. 257–263

K. K. LAU (Manchester) and K. ZIĘTAK (Wrocław)

A NOTE ON LINEAR APPROXIMATIONS OF MATRICES

Abstract. We consider the linear approximation of matrices, and present some particular cases for which the best approximation may be easily computed.

1. Introduction. Let $\mathbf{M}_{mn}$ denote the linear space of real $m \times n$ matrices endowed with the norm $\|\cdot\|$. Let $\mathbf{M}_1$ be a linear subspace of $\mathbf{M}_{mn}$ and let $\mathbf{M}_2$ be a linear subspace of $\mathbf{M}_1$. We consider the following problem:

For a given matrix $A \in \mathbf{M}_1$, find $\widetilde{B} \in \mathbf{M}_2$ such that

$$\|A - \widetilde{B}\| = \min_{B \in \mathbf{M}_2} \|A - B\|. \tag{1.1}$$

The matrix $\widetilde{B}$ is called a *best* $\|\cdot\|$*-approximation* of A by matrices from $\mathbf{M}_2$. The *approximation error* is denoted by δ, where

$$\delta = \|A - \widetilde{B}\|.$$

Let $\dim \mathbf{M}_2 = s$ and let $\mathbf{M}_2$ be spanned by the linearly independent matrices $B_1, \ldots, B_s$, i.e. $\mathbf{M}_2 = \operatorname{span}\{B_1, \ldots, B_s\}$. Then (1.1) can be written in the form

$$\Big\|A - \sum_{i=1}^{s} \widetilde{\alpha}_i B_i\Big\| = \min_{\alpha_i \in \mathbf{R}^s} \Big\|A - \sum_{i=1}^{s} \alpha_i B_i\Big\|. \tag{1.2}$$

Best approximations for (1.1) are characterized by the following theorem, derived from Theorem 1.1 in [4] (see [5]):

THEOREM. *$\widetilde{B} \in M_2$ is a best $\|\cdot\|$-approximation of $A \in \mathbf{M}_1 \setminus \mathbf{M}_2$ if and only if there exists V such that*

$$\langle A - \widetilde{B}, V\rangle = \|A - \widetilde{B}\|, \quad \|V\|^* = 1, \quad V \in \mathbf{M}_2^{\perp}, \tag{1.3}$$

where

1991 *Mathematics Subject Classification*: Primary 15A60, 65D99.
Key words and phrases: linear approximation of matrices, unitarily invariant norms.

(i) $\|\cdot\|^*$ *is the dual norm to the norm* $\|\cdot\|$, *and is defined by*

$$\|A\|^* = \max_{X\in \mathbf{M}_1, \|X\|\le 1} \langle A, X\rangle ,$$

where

$$\langle A, X\rangle = \mathrm{trace}(A^T X) = \sum_{i,j} a_{ij}x_{ij} ,$$

(ii) $\mathbf{M}_2^{\perp}$ *is the orthogonal complement of* $\mathbf{M}_2$ *with respect to* $\mathbf{M}_1$, *i.e.* $\mathbf{M}_1 = \mathbf{M}_2 \oplus \mathbf{M}_2^{\perp}$.

2. Best Frobenius approximations. Let us consider the problem (1.2) for the Frobenius norm, which is given by

$$\|A\|_F = \Big(\sum_{i,j} |a_{ij}|^2\Big)^{1/2} .$$

It is a particular case of unitarily invariant norms (e.g. see [2], [3]). A *unitarily invariant norm* is one which satisfies

$$\|A\| = \|UAV\|$$

for arbitrary unitary matrices U and V. For example the norms

$$\|A\|_p = \Big(\sum_i \sigma_i^p(A)\Big)^{1/p}, \quad 1 \le p \le \infty , \tag{2.1}$$

where $\sigma_i(A)$ denotes the ith singular value of A, with

$$\sigma_1(A) \ge \ldots \ge \sigma_t(A) \ge 0, \quad t = \min\{m, n\} ,$$

are unitarily invariant. For $p = 2$ we have the Frobenius norm.

Now let $A \in \mathbf{M}_1$ be expressed in the form

$$A = A^{(1)} + A^{(2)} \tag{2.2}$$

with $A^{(1)} \in \mathbf{M}_2$ and $A^{(2)} \in \mathbf{M}_2^{\perp}$. This expression is unique. Moreover, it is easy to verify that $A^{(1)}$ is in fact equal to the best Frobenius approximation $\widetilde{B}_F$ of A, which is given by

$$\widetilde{B}_F = \sum_{i=1}^{s} \widetilde{\alpha}_i B_i ,$$

where $\widetilde{\alpha}_i$, $i = 1, \ldots, s$, are the solution of the system (see [5])

$$\sum_{k=1}^{s} \langle B_k, B_j\rangle \alpha_k = \langle A, B_j\rangle, \quad j = 1, \ldots, s . \tag{2.3}$$

The matrix of this system is nonsingular because we have assumed that the matrices $B_1, \ldots, B_s$ are linearly independent.

Using the fact that $A^{(1)} = \widetilde{B}_F$, the following corollary states when the best Frobenius approximation of A is also a best $\|\cdot\|$-approximation of A, for an arbitrary norm $\|\cdot\|$ (see (1.3)):

COROLLARY. *Let $A \in \mathsf{M}_1$ be expressed in the form (2.2) and let $\|\cdot\|$ be an arbitrary norm in M_1. Then the matrix $A^{(1)}$, i.e. the best Frobenius approximation of A, is a best $\|\cdot\|$-approximation of A if and only if there exists a matrix V such that*

$$(2.4) \qquad \langle A^{(2)}, V\rangle = \|A^{(2)}\|, \quad \|V\|^* = 1, \quad V \in \mathsf{M}_2^{\perp}.$$

We now briefly discuss three special cases of problem (1.1) in which best Frobenius approximations play a key role. We shall use the same notation for the matrix norm (2.1) and for the l_p-norm in the vector space $\mathbf{R}^t$, since

$$(2.5) \qquad \|A\|_p = \|\mathbf{s}\|_p,$$

where

$$\mathbf{s} = (\sigma_1(A), \ldots, \sigma_t(A))^T, \quad t = \min\{m, n\}.$$

Note also that the dual norm to the l_p-norm in the space $\mathbf{R}^t$ is the l_q-norm where

$$(2.6) \qquad 1/p + 1/q = 1.$$

The same holds for the norms (2.1) in the space M_{mn}, and in particular the space of symmetric matrices.

Case (i). Let

$$\dim(\mathsf{M}_2) = \dim(\mathsf{M}_1) - 1$$

in (1.1). This problem is considered in [5] for an arbitrary norm $\|\cdot\|$, and it is shown that in this case the best $\|\cdot\|$-approximations B of A are given explicitly in terms of the residuum matrix $R_F = A - \widetilde{B}_F$ where $\widetilde{B}_F$ is the best Frobenius approximation of A. Furthermore, $\widetilde{B}_F$ itself is given by $\widetilde{\alpha}_i$, $i = 1, \ldots, s$, which form the solution of the linear system (2.3).

Case (ii). Let M_1 be the space of symmetric matrices of order n. Let M_2 be spanned by symmetric matrices $B_k = (b_{ij}^{(k)})$, $k = 1, \ldots, s$, such that

$$b_{jj}^{(k)} = 0 \quad \text{for } j = j_1, \ldots, j_r \text{ and } k = 1, \ldots, s$$

with $s = \frac{n}{2}(n+1) - r$ (we assume that the j_i's are distinct).

We consider the problem (1.1) with the norm (2.1). Without loss of generality we assume $j_i = i$, for $i = 1, \ldots, r$. Then every matrix in $\mathsf{M}_2^{\perp}$ has the form

$$\operatorname{diag}(d_1, \ldots, d_r, 0, \ldots, 0).$$

Let $A \in \mathsf{M}_1$ be expressed in the form (2.2). Then

$$A^{(2)} = \operatorname{diag}(a_{11}, \ldots, a_{rr}, 0, \ldots, 0).$$

We assume that $A \notin \mathsf{M}_2$, so it is not a trivial case and $A^{(2)} \neq 0$. We take

$$V = \operatorname{diag}(a'_{11}, \ldots, a'_{rr}, 0, \ldots, 0)$$

such that the vector

$$\mathbf{a}' = (a'_{11}, \ldots, a'_{rr}, 0, \ldots, 0)^T \in \mathbf{R}^n$$

satisfies

$$\langle \mathbf{a}, \mathbf{a}' \rangle = \|\mathbf{a}\|_p, \quad \|\mathbf{a}'\|_q = 1\,,$$

where p, q satisfy (2.6) and $\mathbf{a} = (a_{11}, \ldots, a_{rr}, 0, \ldots, 0)^T$. Such a vector $\mathbf{a}'$ always exists ($\mathbf{a}'$ is the dual vector to $\mathbf{a}$). For $1 < p < \infty$ the vector $\mathbf{a}'$ has the components

$$a'_{ii} = \operatorname{sign}(a_{ii})\,(|a_{ii}|/\|\mathbf{a}\|_p)^{p-1}\,, \tag{2.7}$$

for $p = 1$ we may take

$$a'_{ii} = \begin{cases} \operatorname{sign}(a_{ii}) & \text{if } a_{ii} \neq 0, \\ h_i & \text{if } a_{ii} = 0, \end{cases} \tag{2.8}$$

$i = 1, \ldots, r$, where $|h_i| \leq 1$, and for $p = \infty$

$$a'_{ii} = \begin{cases} \operatorname{sign}(a_{ii}) g_i & \text{if } |a_{ii}| = \|\mathbf{a}\|_\infty, \\ 0 & \text{if } |a_{ii}| < \|\mathbf{a}\|_\infty, \end{cases} \tag{2.9}$$

$i = 1, \ldots, r$, where $g_i \geq 0$ and $\sum_i g_i = 1$. Hence we have

$$\begin{aligned} \|A^{(2)}\|_p &= \|\mathbf{a}\|_p = \langle \mathbf{a}, \mathbf{a}' \rangle = \langle A^{(2)}, V \rangle\,, \\ \|V\|^* &= \|V\|_q = \|\mathbf{a}'\|_q = 1\,, \quad V \in \mathsf{M}_2^\perp\,. \end{aligned} \tag{2.10}$$

Therefore Corollary implies that $A^{(1)} = \widetilde{B}_F$ is a best $\|\cdot\|_p$-approximation of A, for any $1 \leq p \leq \infty$, and for each value of p the approximation error is equal to

$$\delta = \|\mathbf{a}\|_p = \|A^{(2)}\|_p\,.$$

Case (iii). Again, let M_1 be the space of symmetric matrices of order n. Let M_2 be spanned by symmetric matrices $B_k = (b_{ij}^{(k)})$, $k = 1, \ldots, s$, such that

$$b_{lj}^{(k)} = b_{jl}^{(k)} = 0 \qquad \text{for } j \neq l,\ k = 1, \ldots, s,$$

with $s = \frac{n}{2}(n+1) - (n-1)$, where the natural number l is given.

Every matrix X from $\mathbf{M}_2^{\perp}$ has the form

$$X = \begin{pmatrix} & & & x_1 & & & \\ & 0 & & \vdots & & 0 & \\ & & & x_{l-1} & & & \\ x_1 & \dots & x_{l-1} & 0 & x_{l+1} & \dots & x_n \\ & & & x_{l+1} & & & \\ & 0 & & \vdots & & 0 & \\ & & & x_n & & & \end{pmatrix} \equiv \mathbf{x}\mathbf{e}_l^T + \mathbf{e}_l\mathbf{x}^T$$

with $\mathbf{x} = (x_1, \dots, x_{l-1}, 0, x_{l+1}, \dots, x_n)^T$ and $\mathbf{e}_l$ the lth unit vector. Let $X \neq 0$. Then $\operatorname{rank}(X) = 2$. It is easy to verify that the nonzero eigenvalues of X are $\pm\|\mathbf{x}\|_2$ and the vectors

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}\left(\mathbf{e}_l + \frac{1}{\|\mathbf{x}\|_2}\mathbf{x}\right), \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}}\left(\mathbf{e}_l - \frac{1}{\|\mathbf{x}\|_2}\mathbf{x}\right)$$

are the normalized orthogonal eigenvectors of X corresponding to $\|\mathbf{x}\|_2$ and $-\|\mathbf{x}\|_2$ respectively. Therefore X has the form

$$\begin{aligned} X &= \|\mathbf{x}\|_2\mathbf{u}_1\mathbf{u}_1^T - \|\mathbf{x}\|_2\mathbf{u}_2\mathbf{u}_2^T \\ &= U\begin{pmatrix} \|\mathbf{x}\|_2 & & & & 0 \\ & -\|\mathbf{x}\|_2 & & & \\ & & 0 & & \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix}U^T, \end{aligned} \tag{2.11}$$

where an orthogonal matrix U has $\mathbf{u}_1$ and $\mathbf{u}_2$ as its first and second column respectively. Let $A \in \mathbf{M}_1$ be expressed in the form (2.2). Then $A^{(2)}$ has the form (see (2.11))

$$A^{(2)} = Q\begin{pmatrix} \lambda & & & & 0 \\ & -\lambda & & & \\ & & 0 & & \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix}Q^T,$$

where Q is an orthogonal matrix. Let $\mathbf{a} = (\lambda, -\lambda, 0, \dots, 0)^T$. Then there exists a dual vector $\mathbf{a}'$ to $\mathbf{a}$ having the form (see (2.7)–(2.9))

$$\mathbf{a}' = (\mu, -\mu, 0, \dots, 0)^T$$

with an appropriate μ. Therefore

$$V = Q \begin{pmatrix} \mu & & & & 0 \\ & -\mu & & & \\ & & 0 & & \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix} Q^T$$

belongs to $\mathbf{M}_2^\perp$ (see (2.11)) and V satisfies conditions (1.3) for $\widetilde{B} = A^{(1)}$, so again $\widetilde{B}_F = A^{(1)}$ is a best $\|\cdot\|_p$-approximation of A, for any $1 \le p \le \infty$, and for each value of p

$$\delta = \|A^{(2)}\|_p .$$

3. Commuting matrices. Finally, we identify another special case of (1.1). It is, however, not treated in the same manner as the previous cases.

Let $\|\cdot\|$ be an arbitrary unitarily invariant norm and let $\mathbf{M}_1$ be the space of symmetric matrices of order n. Let $A \in \mathbf{M}_1$ and let $\mathbf{M}_2$ be spanned by matrices $B_1, \ldots, B_s$ which commute with A, so

$$AB_j = B_jA, \quad j = 1, \ldots, s .$$

In particular, we may choose $B_j = C^j$, for some C. In addition, we assume that B_i commutes with B_j for $i \neq j$. Therefore there exists an orthogonal matrix Q such that the matrices Q^TAQ and Q^TB_jQ ($j = 1, \ldots, s$) are diagonal. Let

$$(3.1) \qquad Q^TAQ = \mathrm{diag}(\lambda_k), \quad Q^TB_jQ = \mathrm{diag}(\lambda_k^{(j)}), \quad j = 1, \ldots, s .$$

It is known that if $\|\cdot\|$ is a unitarily invariant norm then there exists a symmetric gauge function ϕ such that (see [2, 3])

$$\|A\| = \phi(\sigma(a)) \equiv \phi(\sigma_1(A), \ldots, \sigma_n(A)) .$$

The symmetric gauge function is a norm in $\mathbf{R}^n$. From (3.1) and the properties of the symmetric gauge function ϕ we obtain

$$\Big\|A - \sum_j \alpha_j B_j\Big\| = \phi\Big(\lambda - \sum_j \alpha_j \lambda^{(j)}\Big) ,$$

where $\lambda = (\lambda_1, \ldots, \lambda_n)^T$ and $\lambda^{(j)} = (\lambda_1^j, \ldots, \lambda_n^j)^T$. Thus problem (1.1) reduces to the solution of the overdetermined linear system

$$G\alpha = \lambda ,$$

where the matrix G has the form

$$G = (\lambda^{(1)}, \ldots, \lambda^{(s)}) ,$$

in the norm ϕ. If we choose the norm (2.5) then we find the l_p-solution of this system. For $p = \infty$ we solve in the Chebyshev sense.

References

[1] C. S. Duris and V. P. Sreedharan, *Chebyshev and l_1-solutions of linear equations using least squares solutions*, SIAM J. Numer. Anal. 5 (1968), 491–505.

[2] C.-K. Li and N.-K. Tsing, *On the unitarily invariant norms and some related results*, Linear and Multilinear Algebra 20 (1987), 107–119.

[3] L. Mirsky, *Symmetric gauge functions and unitarily invariant norms*, Quart. J. Math. Oxford Ser. 11 (1960), 50–59.

[4] I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces*, Springer, 1970.

[5] K. Ziętak, *On the linear approximation of matrices*, Report N-189, Institute of Computer Science, University of Wroclaw, 1988.

KUNG-KIU LAU
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF MANCHESTER
OXFORD ROAD
MANCHESTER M13 9PL, ENGLAND

KRYSTYNA ZIĘTAK
INSTITUTE OF COMPUTER SCIENCE
UNIVERSITY OF WROCLAW
UL. PRZESMYCKIEGO 20
51-151 WROCLAW, POLAND

Received on 10.1.1990