Mathematical
Inequalities
& Applications
Volume 1, Number 3 (1998), 305–318

ON A CONJECTURE ON THE CLOSEST NORMAL MATRIX

ANDERS BARRLUND

(*communicated by L.-E. Persson*)

Abstract. Let A be a complex $n \times n$ matrix and let $\mathcal{N}_n$ be the set of normal $n \times n$ matrices. A conjecture is that

$$\|A - \mathcal{N}_n\|_F^2 \leqslant \frac{n-1}{n}\mathrm{dep}^2(A),$$

where $\mathrm{dep}^2(A) = \|A\|_F^2 - \sum_{i=1}^n \lambda_i^2(A)$ and $\lambda_i(A), i = 1, \ldots, n$ are the eigenvalues of A. We prove that the conjecture is correct for all even n and for $n = 3, 5, 7$. However, for the dimensions, $n = 3, 5, 6, 7$, and presumably also other problem dimensions it is possible to derive sharper bounds. We also prove a bound for odd n which converges to the bound in the conjecture when n tends to infinity. The main idea in the proofs is to use LP problems with constraints based on different ways to approximate A with normal matrices.

1. Introduction

The following notation will be used:

1. $\mathcal{M}_n$ is the set of complex $n \times n$ matrices.
2. $\mathcal{N}_n$ is the set of $n \times n$ normal matrices. A matrix A is normal if and only if $A^H A = AA^H$.
3. $\mathcal{U}_n$ is the set of $n \times n$ upper triangular matrices.
4. $\|A\|_F$ is the Frobenius norm of $A \in \mathcal{M}_n$.
5. $\mathrm{dep}(A) = \sqrt{\|A\|_F^2 - \|\Gamma\|_F^2}$ is the departure from normality as defined by Henrici with $\Gamma = \mathrm{diag}(\lambda_i)$ being the diagonal matrix of the eigenvalues of $A \in \mathcal{M}_n$.
6. $\bar{a}$ denotes the conjugate of a.
7. $\|A - \mathcal{N}_n\|_F = \inf_{N \in \mathcal{N}_n} \|A - N\|_F$.
8. $\mathrm{diag}(A)$ is the diagonal part of A.

In this paper we study the possibility to find bounds of the form

$$\|A - \mathcal{N}_n\|_F^2 \leqslant c_n \mathrm{dep}^2(A). \tag{1.1}$$

where c_n is a number which depends on n.

Mathematics subject classification (1991): 15A45.
Key words and phrases: Normal matrix, LP-problem.

Paper MIA-01-30

Let us begin with one of the main ideas when we study the problem to approximate a matrix A with normal matrices. Let $A = URU^H$ be the Schur decomposition of A. If N is normal then also U^HNU is normal and

$$\|R - U^HNU\|_F^2 = \|A - N\|_F^2.$$

This implies that a bound of $\|A - \mathscr{N}_n\|_F$ is correct for all $A \in \mathscr{M}_n$ if and only if it is correct for all $A \in \mathscr{U}_n$. Without loss of generality we will consider upper triangular matrices in all proofs in this paper.

It is of interest to see how small c_n can be selected in (1.1). It is trivial that we can select $c_n = 1$, that is $\|A - \mathscr{N}_n\|_F^2 \leqslant \operatorname{dep}^2(A)$, since for $A \in \mathscr{U}_n$ the diagonal matrix $\operatorname{diag}(A)$ is normal and $\|A - \operatorname{diag}(A)\|_F^2 = \operatorname{dep}^2(A)$. It has turned out to be possible but not trivial to derive bounds which are slightly sharper than this trivial bound. All methods we use to get sharper bounds is based on the idea to use a normal N with the same diagonal as the upper triangular matrix A and some off-diagonal elements nonzero such that we get an N with $\|A - N\|_F^2 < \operatorname{dep}^2(A)$.

Professor Lajos László [5] has presented the following conjecture:

CONJECTURE 1. [5] *If $A \in \mathscr{M}_n$, then*

$$\|A - \mathscr{N}_n\|_F^2 \leqslant \frac{n-1}{n}\operatorname{dep}^2(A). \tag{1.2}$$

It is well known that the conjecture (1.2) is correct for $n = 2$. If

$$A = \begin{bmatrix} \lambda_1 & a_{12} \\ 0 & \lambda_2 \end{bmatrix},$$

then we can select

$$N = \begin{bmatrix} \lambda_1 & a_{12}/2 \\ u & \lambda_2 \end{bmatrix}, \tag{1.3}$$

where

$$u = \begin{cases} -(\lambda_1 - \lambda_2)/(\bar{\lambda}_1 - \bar{\lambda}_2) \cdot \bar{a}_{12}/2 & \text{if } \lambda_1 \neq \lambda_2 \\ -\bar{a}_{12}/2 & \text{if } \lambda_1 = \lambda_2. \end{cases}$$

This gives $\|A - N\|_F^2 = \frac{1}{2}\operatorname{dep}^2(A)$.

Professor Kh. D. Ikramov [3, 4] has proved that for any upper triangular 3×3 matrix

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{bmatrix}$$

it is possible to find complex numbers n_{21}, n_{31} and n_{32} such that

$$N(B) = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ n_{21} & b_{22} & b_{23} \\ n_{31} & n_{32} & b_{33} \end{bmatrix}, \tag{1.4}$$

is normal. It is known that this implies that the conjecture (1.2) is correct for $n = 3$ [5]. In this paper we use (1.4) to derive a sharper bound for the case $n = 3$.

Professor L. Elsner and Kh. D. Ikramov [1] have proved that if László's conjecture (1.2) is correct for all $n \leqslant m$ then the slightly weaker inequality

$$\|A - \mathscr{N}_n\|_F^2 \leqslant \frac{n - (1 - 1/m)}{n} \mathrm{dep}^2(A), \tag{1.5}$$

is satisfied for arbitrary n. Before this paper,

$$\|A - \mathscr{N}_n\|_F^2 \leqslant \frac{n - 2/3}{n} \cdot \mathrm{dep}^2(A) \tag{1.6}$$

was the sharpest bound we could get from their result.

In this paper we prove that László's conjecture (1.2) is correct for all even n, $n = 3$, $n = 5$ and $n = 7$. However, for $n = 3, 5, 6, 7$ we have proved sharper bounds than (1.2). We also prove a bound which is close to the conjecture when n is odd.

The paper is outlined as follows. In Section 2 we prove a sharper bound in the case $n = 3$. A Lemma is presented in Section 3, which is used to prove the conjecture (1.2) for even n in section 4. Section 5 presents a bound close to the conjecture (1.2) for odd n. In Section 6 we describe a general method to find constraints which is used in Sections 7-8 to prove sharper bounds for $n = 5,\ 6$ and 7.

2. The case $n = 3$

In the case $n = 3$, we can prove the following upper bound for the distance to the closest normal matrix, which is sharper than the bound in László's conjecture (1.2).

THEOREM 1. *If $A \in \mathscr{M}_3$, then*

$$\|A - \mathscr{N}_3\|_F^2 \leqslant (1 - \frac{3}{8})\mathrm{dep}^2(A). \tag{2.1}$$

Proof. The idea in the proof is to derive a *Linear Program* (LP-problem), where the inequalities are given by different ways to approximate A with normal matrices. The unknowns in the LP problem are

$$[x_1,\ x_2,\ x_3,\ x_4]^T \equiv [|a_{12}|^2,\ |a_{13}|^2, |a_{23}|^2, \mathrm{dep}^2(A) - \|A - \mathscr{N}_3\|_F^2]^T / \mathrm{dep}^2(A).$$

Let

$$B = \begin{bmatrix} a_{11} & a_{12}/2 & a_{13}/3 \\ 0 & a_{22} & a_{23}/2 \\ 0 & 0 & a_{33} \end{bmatrix},$$

and let $N(B)$ be the completion (1.4) of the matrix B. From the diagonal elements of $N^H(B)N(B) = N(B)N^H(B)$ we get

$$|b_{12}|^2 + |b_{13}|^2 = |n_{21}|^2 + |n_{31}|^2, \tag{2.2}$$

$$|n_{21}|^2 + |b_{23}|^2 = |b_{12}|^2 + |n_{32}|^2,$$
$$|n_{31}|^2 + |n_{32}|^2 = |b_{13}|^2 + |b_{23}|^2. \tag{2.3}$$

By combing (2.2) and (2.3) we get

$$2|n_{31}|^2 + |n_{21}|^2 + |n_{32}|^2 = |b_{12}|^2 + 2|b_{13}|^2 + |b_{23}|^2,$$

which implies that

$$|n_{31}|^2 + |n_{21}|^2 + |n_{32}|^2 \leqslant |b_{12}|^2 + 2|b_{13}|^2 + |b_{23}|^2, \tag{2.4}$$

and

$$\begin{aligned}\|A - N(B)\|_F^2 &= |n_{31}|^2 + |n_{21}|^2 + |n_{32}|^2 + \frac{1}{4}|a_{12}|^2 + \frac{4}{9}|a_{13}|^2 + \frac{1}{4}|a_{23}|^2 \\ &\leqslant \frac{1}{4}|a_{12}|^2 + \frac{2}{9}|a_{13}|^2 + \frac{1}{4}|a_{23}|^2 + \frac{1}{4}|a_{12}|^2 + \frac{4}{9}|a_{13}|^2 + \frac{1}{4}|a_{23}|^2 \\ &= \frac{1}{2}|a_{12}|^2 + \frac{2}{3}|a_{13}|^2 + \frac{1}{2}|a_{23}|^2 \\ &= \operatorname{dep}^2(A) - \frac{1}{2}|a_{12}|^2 - \frac{1}{3}|a_{13}|^2 - \frac{1}{2}|a_{23}|^2.\end{aligned}$$

Consequently,

$$-x_1/2 - x_2/3 - x_3/2 + x_4 \geqslant 0.$$

It is also possible to use the idea in (1.3) to find a normal matrix N with the structure

$$N = \begin{bmatrix} a_{11} & 0 & a_{13}/2 \\ 0 & a_{22} & 0 \\ u & 0 & a_{33} \end{bmatrix}, \tag{2.5}$$

where $|u| = |a_{13}|/2$. This N satisfies

$$\|A - N\|_F^2 = \operatorname{dep}^2(A) - |a_{13}|^2/2,$$

which together with the obvious inequality $\|A - \mathscr{N}_3\|_F \leqslant \|A - N\|_F$ implies that

$$-x_2/2 + x_4 \geqslant 0. \tag{2.6}$$

Consequently, x_4 must be larger than the solution of the problem

$$\min x_4$$

subject to

$$\begin{gathered} -x_1/2 - x_2/3 - x_3/2 + x_4 \geqslant 0, \\ -x_2/2 + x_4 \geqslant 0, \\ x_1 + x_2 + x_3 = 1, \\ x_i \geqslant 0, \quad i = 1, \ldots, 4. \end{gathered}$$

This LP-problem can be solved with conventional methods for LP problems, see e. g. [2]. Its solution is not unique. One minimizer is

$$[x_1, x_2, x_3, x_4]^T = [1/4,\ \ 3/4,\ \ 0,\ \ 3/8]^T,$$

and another one is

$$[x_1, x_2, x_3, x_4]^T = [0,\ \ 3/4,\ \ 1/4,\ \ 3/8]^T.$$

However, the minimum is unique and $3/8$, which proves that

$$\|A - \mathscr{N}_3\|_F^2 \leqslant (1 - \frac{3}{8})\mathrm{dep}^2(A). \quad \square$$

Note that all constraints are active in the solution of the LP problem in the above proof. It implies that if we formulate an LP problem where any of the constraints is left out then we get a weaker bound.

3. A first general bound

It is possible to combine the idea in the proof of Elsner and Ikramov's bound (1.5) with Theorem 1 to get a sharper bound than (1.6). The result is proved in the following lemma. (The reason why we present it as a lemma is that we will use it to prove even sharper bounds in the next sections.)

LEMMA 1. *If $A \in \mathscr{M}_n$, then*

$$\|A - \mathscr{N}_n\|_F^2 \leqslant (1 - \frac{3}{4n})\mathrm{dep}^2(A). \tag{3.1}$$

Proof. (Basically the same idea as in the proof in [1]). We consider an $A \in \mathscr{U}_n$. Let $k = \lfloor n/3 \rfloor$ and let $r = \lfloor (n - 3k)/2 \rfloor$, (which implies that $r = 1$ if $n \equiv 2 (\mathrm{mod}\ 3)$, $r = 0$ otherwise). Construct a set of disjoints sets of the numbers $1, \ldots, n$ such that k sets are of size 3 and r sets are of size 2. That is, we get $S_1 = (i_{11}, i_{12}, i_{13})$, $S_2 = (i_{21}, i_{22}, i_{23}), \ldots, S_k = (i_{k1}, i_{k2}, i_{k3})$ where $i_{j1} < i_{j2} < i_{j3}$, $j = 1, \ldots k$, and if $r = 1$, $S_{k+1} = (i_{k+1,1}, i_{k+1,2})$, with $i_{k+1,1} < i_{k+1,2}$. Let $A_j \in \mathscr{U}_3 \cup \mathscr{U}_2$ be the submatrix of A consisting of rows and columns with index in set S_j, $j = 1, \ldots, k + r$, and let N_j be the corresponding submatrix of a normal N. The elements of N which are not on the diagonal and not covered by any N_j are zero. There exists a permutation matrix P such that $PNP^T = \mathrm{diag}(N_1, N_2, \ldots)$, where $\mathrm{diag}(N_1, N_2, \ldots)$ is block diagonal, which implies that N is normal if all N_j are normal. The diagonal of N is identical to the diagonal of A. The rest of the elements of N are selected such that N_j is normal and

$$\|A_j - N_j\|_F^2 \leqslant \mathrm{dep}^2(A_j) \cdot (1 - \frac{3}{8}), \quad j = 1, \ldots, k,$$

and if $r = 1$,

$$\|A_j - N_j\|_F^2 \leqslant \mathrm{dep}^2(A_j) \cdot (1 - \frac{1}{2}), \quad j = k + 1.$$

This is possible by Theorem 1 and (1.3). Then

$$\|A-N\|_F^2 \leqslant \operatorname{dep}^2(A) - \frac{3}{8}\cdot\sum_{j=1}^{k}\left(|a_{i_{j1},i_{j2}}|^2+|a_{i_{j1},i_{j3}}|^2+|a_{i_{j2},i_{j3}}|^2\right) - \frac{1}{2}\sum_{j=1}^{r}|a_{i_{k+r,1},i_{k+r,2}}|^2.$$

Let

$$\gamma = \frac{1}{n(n-1)/2}\sum_{i=1}^{n-1}\sum_{j=1}^{n}|a_{ij}|^2.$$

The average of all bounds which can be derived with the above method is

$$\operatorname{dep}^2(A) - \frac{3}{8}\sum_{j=1}^{k}(\gamma+\gamma+\gamma) - \frac{1}{2}\sum_{j=1}^{r}\gamma.$$

Let N_b be the best approximation of all N which can be constructed in this way. Then it is at least as good as the average of all bounds derived in this way. The average gives

$$\|A-N_b\|_F^2 \leqslant \operatorname{dep}^2(A) - \frac{3}{8}\cdot 3k\gamma - \frac{1}{2}r\gamma = \operatorname{dep}^2(A)\cdot\left(1-\frac{3k\cdot\frac{3}{8}+r\cdot\frac{1}{2}}{n\cdot(n-1)/2}\right).$$

Since,

$$\frac{3k\cdot\frac{3}{8}+r\cdot\frac{1}{2}}{(n-1)/2} = \frac{\frac{9k}{4}+r}{n-1} \geqslant \frac{3}{4},$$

for all $n \geqslant 4$ the lemma is proved. □

4. A proof of László's conjecture for even n

We prove László's conjecture (1.2) in the cases where n is even.

THEOREM 2. *If $A \in \mathscr{M}_n$, where n is an even number then*

$$\|A-\mathscr{N}_n\|_F^2 \leqslant \frac{n-1}{n}\operatorname{dep}^2(A). \tag{4.1}$$

Proof. Consider an $A \in \mathscr{U}_n$. From (3.1) we see that we can find a normal matrix of the form

$$N = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix},$$

where $N_1, N_2 \in \mathscr{N}_{n/2}$, such that

$$\|A-N\|_F^2 \leqslant \operatorname{dep}^2(A) - \frac{3}{4(n/2)}\left(\sum_{i=1}^{n/2-1}\sum_{j=i+1}^{n/2}|a_{ij}|^2 + \sum_{i=n/2+1}^{n-1}\sum_{j=i+1}^{n}|a_{ij}|^2\right).$$

However, this inequality can be sharpened. It is possible to find a unitary matrix of the form

$$Q = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix},$$

where $Q_1, Q_2 \in \mathscr{M}_{n/2}$ are unitary, such that

$$QNQ^H = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \quad QAQ^H = \tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix},$$

where all blocks $\in \mathscr{M}_{n/2}$, D_1, D_2 are diagonal, and $\|\tilde{A}_{12}\|_F$ is equal to the Frobenius norm of the corresponding block of A. Let $(i_1, i_2, \ldots, i_{n/2})$ be one way to sort the numbers $(1, \ldots, n/2)$ and let $(j_1, \ldots j_{n/2})$ be one way to sort the numbers $(n/2 + 1, \ldots, n)$. For each such sorting we can find a normal $\tilde{N}$ which is equal to QNQ^H except in the elements (i_k, j_k), (j_k, i_k), $k = 1, \ldots, n/2$. With the idea in (2.5) these elements are selected such that

$$\|\tilde{A} - \tilde{N}\|_F^2 = \|\tilde{A} - QNQ^H\|_F^2 - \frac{1}{2}\sum_{k=1}^{n/2} |\tilde{a}_{i_k j_k}|^2.$$

(For example, in the case $n = 4$, we can use $i_1 = 1$, $i_2 = 2$, $j_1 = 3$, $j_2 = 4$ and a normal matrix $\tilde{N}$ with the structure

$$\tilde{N} = \begin{bmatrix} \tilde{a}_{11} & 0 & \tilde{a}_{13}/2 & 0 \\ 0 & \tilde{a}_{22} & 0 & \tilde{a}_{24}/2 \\ u_1 & 0 & \tilde{a}_{33} & 0 \\ 0 & u_2 & 0 & \tilde{a}_{44} \end{bmatrix}$$

with $|u_1| = |\tilde{a}_{13}/2|$ and $|u_2| = |\tilde{a}_{24}/2|$ such that

$$\|\tilde{A} - \tilde{N}\|_F^2 = \|\tilde{A} - QNQ^H\|_F^2 - \frac{1}{2}(|\tilde{a}_{13}|^2 + |\tilde{a}_{24}|^2).$$

We can also use $i_1 = 1$, $i_2 = 2$, $j_1 = 4$, $j_2 = 3$ and a normal matrix $\tilde{N}$ with the structure

$$\tilde{N} = \begin{bmatrix} \tilde{a}_{11} & 0 & 0 & \tilde{a}_{14}/2 \\ 0 & \tilde{a}_{22} & \tilde{a}_{23}/2 & 0 \\ 0 & u_2 & \tilde{a}_{33} & 0 \\ u_1 & 0 & 0 & \tilde{a}_{44} \end{bmatrix}$$

with $|u_1| = |\tilde{a}_{14}/2|$ and $|u_2| = |\tilde{a}_{23}/2|$ such that

$$\|\tilde{A} - \tilde{N}\|_F^2 = \|\tilde{A} - QNQ^H\|_F^2 - \frac{1}{2}(|\tilde{a}_{14}|^2 + |\tilde{a}_{23}|^2).$$

)

Let $\tilde{N}_b$ be the best approximation of all these $\tilde{N}$. Then it must be at least as good as the average of all these inequalities which gives

$$\|\tilde{A} - \tilde{N}_b\|_F^2 \leqslant \|\tilde{A} - QNQ^H\|_F^2 - \frac{1}{n}\|\tilde{A}_{12}\|_F^2.$$

This implies that

$$\|A - Q^H \tilde{N}_b Q\|_F^2 \leqslant \tag{4.2}$$

$$\operatorname{dep}^2(A) - \frac{3}{4(n/2)}\left(\sum_{i=1}^{n/2-1}\sum_{j=i+1}^{n/2}|a_{ij}|^2 + \sum_{i=n/2+1}^{n-1}\sum_{j=i+1}^{n}|a_{ij}|^2\right) - \frac{1}{n}\sum_{i=1}^{n/2}\sum_{j=n/2+1}^{n}|a_{ij}|^2.$$

Let

$$\begin{aligned} x_1 &= \left(\sum_{i=1}^{n/2-1}\sum_{j=i+1}^{n/2}|a_{ij}|^2 + \sum_{i=n/2+1}^{n-1}\sum_{j=i+1}^{n}|a_{ij}|^2\right)/\operatorname{dep}^2(A), \\ x_2 &= \sum_{i=1}^{n/2}\sum_{j=n/2+1}^{n}|a_{ij}|^2/\operatorname{dep}^2(A), \\ x_3 &= (\operatorname{dep}^2(A) - \|A - \mathcal{N}_n\|_F^2)/\operatorname{dep}^2(A), \end{aligned}$$

then the inequality (4.2) gives

$$-\frac{3}{2n}x_1 - \frac{1}{n}x_2 + x_3 \geqslant 0. \tag{4.3}$$

It is also obvious that

$$x_1 + x_2 = 1, \quad x_1 \geqslant 0, \quad x_2 \geqslant 0. \tag{4.4}$$

Obviously, x_3 is larger than the solution of the LP problem where we minimize x_3 subject to (4.3) and (4.4). This LP problem has the solution

$$[x_1, x_2, x_3]^T = [0 \;\; 1 \;\; 1/n]^T$$

which proves the theorem. □

5. A bound for odd n

For odd n Theorem 2 can be used to prove the following bound which is slightly weaker than the bound in László's conjecture (1.2).

THEOREM 3. *If $A \in \mathcal{M}_n$, where n is an odd number then*

$$\|A - \mathcal{N}_n\|_F^2 \leqslant \left(1 - \frac{1}{n}\cdot\frac{n - 7/4}{n-1}\right)\operatorname{dep}^2(A). \tag{5.1}$$

Proof. Consider the possible ways to divide the numbers $(1, \ldots, n)$ in one set $(i_1, i_2, \ldots i_{n-3})$ with $n-3$ numbers and one set (j_1, j_2, j_3) with three numbers. The sets are sorted such that $i_1 < \ldots < i_n$ and $j_1 < j_2 < j_3$. For each such partition it is possible to find a normal matrix of the form

$$N = P\begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}P^T,$$

where $N_1 \in \mathscr{M}_{n-3}$, $N_2 \in \mathscr{M}_3$ and P is a permutation matrix such that rows and columns $(i_1, \ldots i_{n-3})$ come first. From (2.1) and (4.1) we see that we can select N_1 and N_2 such that

$$\|A - N\|_F^2 \leqslant \operatorname{dep}^2(A) - \left(\frac{1}{n-3} \sum_{k=1}^{n-4} \sum_{l=k+1}^{n-3} |a_{i_k,i_l}|^2 + \frac{3}{8} \sum_{k=1}^{2} \sum_{l=k+1}^{3} |a_{j_k,j_l}|^2 \right). \qquad (5.2)$$

Let N_b be the best approximation of all these N then it must be at least as good as the average of the right hand-sides in (5.2) for all different divisions. It gives

$$\begin{aligned} \|A - N_b\|_F^2 &\leqslant \operatorname{dep}^2(A) \left(1 - \frac{1}{(n-3)} \cdot \frac{(n-3)(n-4)}{n(n-1)} - \frac{3}{8} \cdot \frac{3 \cdot 2}{n(n-1)} \right) \\ &= \operatorname{dep}^2(A)(1 - \frac{1}{n} \cdot \frac{n - 7/4}{n-1}). \quad \square \end{aligned}$$

6. General method to derive constraints

It is possible to derive sharper bounds if we work with constraints on the element level. Here we describe the general method that is used to derive such constraints and that is used in Sections 7-8 to derive sharper bounds for $n = 5, 6, 7$.

There are basically two kinds of constraints that will be used in the rest of this paper. Let us begin with the *first kind*. Consider the case where we want to find a bound in the case $n = m$, where $m \geqslant 4$. Let m_1 and m_2 be two integers such that $m_1 + m_2 = m$. There exist several ways to divide the numbers $(1, 2, \ldots m)$ in two sets $(i_1, i_2, \ldots, i_{m_1})$ and $(j_1, j_2, \ldots j_{m_2})$, with $i_1 < i_2 < \ldots < i_{m_1}$, $j_1 < j_2 < \ldots < j_{m_1}$. For each such division there exists a permutation matrix P such that

$$P^T A P = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix},$$

where $\tilde{A}_{11}$ consists of rows and columns $i_1, \ldots, i_{m_1}$, (that is the (k,l) element of $\tilde{A}_{11}$ is equal to the (i_k, i_l) element of A) and $\tilde{A}_{22}$ consists of rows and columns $j_1, \ldots, j_{m_2}$. If we in the case $n = m_1$ have a bound of the form

$$\|A - N\|_F^2 \leqslant \operatorname{dep}^2(A) - \sum_{k=1}^{m_1-1} \sum_{l=k+1}^{m_1} c_{kl}^{(1)} |a_{kl}|^2,$$

where $c_{kl}^{(1)}, k = 1, \ldots, m_1 - 1, \;\; l = k+1, \ldots m_1$ are different real positive numbers, then we can find a normal $N_1 \in \mathscr{M}_{m_1}$ such that

$$\|\tilde{A}_{11} - N_1\|_F^2 \leqslant \operatorname{dep}^2(\tilde{A}_{11}) - \sum_{k=1}^{m_1-1} \sum_{l=k+1}^{m_1} c_{kl}^{(1)} |a_{i_k,i_l}|^2.$$

Similarly, if we in the case $n = m_2$ have a bound of the form

$$\|A - N\|_F^2 \leqslant \mathrm{dep}^2(A) - \sum_{k=1}^{m_2-1} \sum_{l=k+1}^{m_2} c_{kl}^{(2)} |a_{kl}|^2,$$

then we can find a normal $N_2 \in \mathscr{M}_{m_2}$ such that

$$\|\tilde{A}_{22} - N_2\|_F^2 \leqslant \mathrm{dep}^2(\tilde{A}_{22}) - \sum_{k=1}^{m_2-1} \sum_{l=k+1}^{m_2} c_{kl}^{(2)} |a_{j_k,j_l}|^2.$$

Then , in the case $n = m_1 + m_2$, the matrix

$$N = P \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix} P^T,$$

satisfies

$$\|A - N\|_F^2 \leqslant \mathrm{dep}^2(A) - \sum_{k=1}^{m_1-1} \sum_{l=k+1}^{m_1} c_{kl}^{(1)} |a_{i_k,i_l}|^2 - \sum_{k=1}^{m_2-1} \sum_{l=k+1}^{m_2} c_{kl}^{(2)} |a_{j_k,j_l}|^2.$$

The constraint (2.6) is an example that is derived in this way with $(i_1, i_2) = (1, 3)$.

We continue with the *second kind*. In the case $(i_1, \ldots, i_m) = (1, \ldots, m)$ we get $P = I$ and can derive sharper bounds. Here we first select normal matrices $N_1 \in \mathscr{M}_{m_1}$ and $N_2 \in \mathscr{M}_{m_2}$ such that

$$N = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}$$

satisfies

$$\|A - N\|_F^2 \leqslant \mathrm{dep}^2(A) - \sum_{k=1}^{m_1-1} \sum_{l=k+1}^{m_1} c_{kl}^{(1)} |a_{kl}|^2 - \sum_{k=1}^{m_2-1} \sum_{l=k+1}^{m_2} c_{kl}^{(2)} |a_{k+m_1,l+m_1}|^2.$$

It is possible to find a unitary matrix Q of the form

$$Q = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix},$$

such that

$$QNQ^H = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \quad QAQ^H = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix},$$

where $Q_1 \in \mathscr{M}_{m_1}$ and $Q_2 \in \mathscr{M}_{m_2}$ are unitary, D_1 and D_2 are diagonal matrices, and $\|\tilde{A}_{12}\|_F^2 = \|A_{12}\|_F^2$. Let $m_{min} = \min(m_1, m_2)$ and $m_{max} = \max(m_1, m_2)$. We can select two sets of m_{min} numbers $(i_1, i_2, \ldots, i_{m_{min}})$ and $(j_1, j_2, \ldots, j_{m_{min}})$ such that all numbers are distinct and all $i_k \leqslant m_1$ and all $j_k > m_1$. We can select a normal $\tilde{N}$ which is

identical to QNQ^H except in the elements (i_k,j_k), (j_k,i_k), $k=1,\ldots,m_{min}$. These elements are selected such that

$$\|A-\tilde{N}\|_F^2 = \|A-N\|_F^2 - \frac{1}{2}\sum_{k=1}^{m_{min}} |\tilde{a}_{i_k,j_k}|^2$$

with the idea (1.3). Let $\tilde{N}_b$ be the best approximation of these $\tilde{N}$ then it will satisfy

$$\|A-Q^H\tilde{N}_bQ\|_F^2 \leqslant \|A-N\|_F^2 - \frac{m_{min}}{2m_1\cdot m_2}\|\tilde{A}_{12}\|_F^2 =$$

$$\mathrm{dep}^2(\mathrm{A}) - \sum_{k=1}^{m_1-1}\sum_{l=k+1}^{m_1} c_{kl}^{(1)}|a_{kl}|^2 - \sum_{k=1}^{m_2-1}\sum_{l=k+1}^{m_2} c_{kl}^{(2)}|a_{k+m_1,l+m_1}|^2 - \frac{1}{2\cdot m_{max}}\sum_{k=1}^{m_1}\sum_{l=1}^{m_2}|a_{k,l+m_1}|^2$$

which gives an inequality to an LP problem. The constraint (4.3) is an example that is derived in this way.

When we derived the results in Sections 7-8, we have first derived all constraints which can be derived by the above methods. Next, an LP problem with all these constraints is solved with matlab. We only present the constraints which are active in the solution points in the proofs in Sections 7-8, since the inactive constraints does not effect the solution.

7. The case $n=5$

In the case $n=5$ we derive the following bound. Since $71/342 > 0.2076 > 0.2 = 1/5$, the bound is sharper than the bound in László's conjecture (1.2).

THEOREM 4. *If $A\in\mathscr{M}_5$, then*

$$\|A-\mathscr{N}_5\|_F^2 \leqslant (1-\frac{71}{342})\mathrm{dep}^2(A). \tag{7.1}$$

Proof. Let $x=[x_{12},x_{13},x_{14},x_{15},x_{23},x_{24},x_{25},x_{34},x_{35},x_{45},x_b]$ where

$$x_{ij}=|a_{ij}|^2/\mathrm{dep}^2(A), \qquad x_b=(\mathrm{dep}^2(A)-\|A-N\|_F^2)/\mathrm{dep}^2(A). \tag{7.2}$$

(We label the elements in x in this way to make it easier to understand how we get the constraints). The following inequalities can be satisfied for various $N\in\mathscr{N}_5$.

$$-(x_{12}+x_{25}+x_{34})/2 - x_{15}/3 + x_b \geqslant 0, \tag{7.3}$$

$$-(x_{13}+x_{25}+x_{34})/2 - x_{14}/3 + x_b \geqslant 0, \tag{7.4}$$

$$-(x_{13}+x_{24}+x_{35})/2 - x_{15}/3 + x_b \geqslant 0, \tag{7.5}$$

$$-(x_{14}+x_{23}+x_{45})/2 - x_{15}/3 + x_b \geqslant 0, \tag{7.6}$$

$$-(x_{15}+x_{23}+x_{34})/2 - x_{24}/3 + x_b \geqslant 0, \tag{7.7}$$

$$-(x_{14}+x_{23}+x_{35})/2 - x_{25}/3 + x_b \geqslant 0, \tag{7.8}$$

$$-(x_{12}+x_{45})/2-\frac{1}{4}\sum_{i=1}^{2}\sum_{j=4}^{5}x_{ij}+x_b\geqslant 0, \tag{7.9}$$

$$-(x_{12}+x_{34})/2-\frac{1}{4}\sum_{i=1}^{2}\sum_{j=3}^{4}x_{ij}-\frac{1}{8}\sum_{i=1}^{4}x_{i5}+x_b\geqslant 0, \tag{7.10}$$

$$-(x_{12}+x_{23}+x_{45})/2-x_{13}/3-\frac{1}{6}\sum_{i=1}^{3}\sum_{j=4}^{5}x_{ij}+x_b\geqslant 0, \tag{7.11}$$

$$-(x_{12}+x_{34}+x_{45})/2-x_{35}/3-\frac{1}{6}\sum_{i=1}^{2}\sum_{j=3}^{5}x_{ij}+x_b\geqslant 0, \tag{7.12}$$

Let us see how the constraints can be derived with the general method in section 6. The constraints (7.3)-(7.8) are of the first kind with $m_1=3$ and $m_2=2$. The sets (i_1,i_2,i_3) are (1,2,5), (1,3,4), (1,3,5), (1,4,5), (2,3,4) and, (2,3,5), respectively. The constraint (7.9) is of the first kind with $m_1=4$, $m_2=1$ and $(i_1,i_2,i_3,i_4)=(1,2,4,5)$. The constraints (7.10)-(7.12) are of the second kind with $m_1=4,3$ and 2, respectively.

If we minimize x_b subject to these constraints and the obvious constraints

$$x_{ij}\geqslant 0,\quad i=1,\dots,4,\ \ j=i+1,\dots,5,$$

$$\sum_{i=1}^{4}\sum_{j=i+1}^{5}x_{ij}=1,$$

then one of infinitely many solutions is

$$x=[416\ \ 844\ \ 2466\ \ 3108\ \ 540\ \ 1554\ \ 2196\ \ 570\ \ 784\ \ 176\ \ 2627]^T/12654.$$

Since $x_b=2627/12654=71/342$ the theorem is proved. □

8. The cases $n=6,7,\dots$

In the case $n=6$ we derive the following bound. Since $217/1184>0.1832>1/6$, the bound is sharper than the bound in László's conjecture (1.2).

THEOREM 5. *If $A\in\mathscr{M}_6$, then*

$$\|A-\mathscr{N}_6\|_F^2\leqslant(1-\frac{217}{1184})\mathrm{dep}^2(A). \tag{8.1}$$

Proof. Let $x=[x_{12},\dots,x_{56},x_b]$ where x_{ij} and x_b are defined as in (7.2). The following inequalities are satisfied for various $N\in\mathscr{N}_6$.

$$-\frac{1}{2}(x_{12}+x_{34}+x_{56})-\frac{1}{4}(x_{13}+x_{14}+x_{23}+x_{24})-\frac{1}{8}\sum_{i=1}^{4}\sum_{j=5}^{6}x_{ij}+x_b\geqslant 0, \tag{8.2}$$

$$-\frac{1}{2}(x_{12}+x_{34}+x_{56})-\frac{1}{4}(x_{35}+x_{36}+x_{45}+x_{46})-\frac{1}{8}\sum_{i=1}^{2}\sum_{j=3}^{6}x_{ij}+x_b\geqslant 0, \quad (8.3)$$

$$-\frac{1}{2}(x_{12}+x_{23}+x_{45}+x_{56})-\frac{1}{3}(x_{13}+x_{46})-\frac{1}{6}\sum_{i=1}^{3}\sum_{j=4}^{6}x_{ij}+x_b\geqslant 0, \quad (8.4)$$

$$-\frac{1}{2}(x_{12}+x_{26}+x_{34}+x_{45})-\frac{1}{3}(x_{16}+x_{35})+x_b\geqslant 0, \quad (8.5)$$

$$-\frac{1}{2}(x_{13}+x_{24}+x_{35}+x_{46})-\frac{1}{3}(x_{15}+x_{26})+x_b\geqslant 0, \quad (8.6)$$

$$-\frac{1}{2}(x_{13}+x_{24}+x_{36}+x_{45})-\frac{1}{3}(x_{16}+x_{25})+x_b\geqslant 0, \quad (8.7)$$

$$-\frac{1}{2}(x_{14}+x_{23}+x_{36}+x_{45})-\frac{1}{3}(x_{15}+x_{26})+x_b\geqslant 0, \quad (8.8)$$

$$-\frac{1}{2}(x_{14}+x_{23}+x_{35}+x_{46})-\frac{1}{3}(x_{16}+x_{25})+x_b\geqslant 0, \quad (8.9)$$

$$-\frac{1}{2}(x_{15}+x_{23}+x_{34}+x_{56})-\frac{1}{3}(x_{16}+x_{24})+x_b\geqslant 0, \quad (8.10)$$

$$-\frac{1}{2}(x_{12}+x_{24}+x_{56})-x_{14}/3-\frac{1}{8}\sum_{i=1}^{4}\sum_{j=5}^{6}x_{ij}+x_b\geqslant 0, \quad (8.11)$$

$$-\frac{1}{2}(x_{12}+x_{25}+x_{34})-x_{15}/3-\frac{1}{6}(x_{16}+x_{26}+x_{56})+x_b\geqslant 0, \quad (8.12)$$

$$-\frac{1}{2}(x_{25}+x_{34}+x_{56})-x_{26}/3-\frac{1}{6}(x_{12}+x_{15}+x_{16})+x_b\geqslant 0, \quad (8.13)$$

$$-\frac{1}{2}(x_{12}+x_{35}+x_{56})-x_{36}/3-\frac{1}{8}\sum_{i=1}^{2}\sum_{j=3}^{6}x_{ij}+x_b\geqslant 0. \quad (8.14)$$

The ways these constraints are derived are similar to the ways the constraints in the previous proof were derived. Therefore we omit to explain the constraints here.

If we minimize x_b subject to these constraints and the obvious constraints

$$x_{ij}\geqslant 0, \quad i=1,\ldots,5,\ j=i+1,\ldots,6,$$

$$\sum_{i=1}^{5}\sum_{j=i+1}^{6}x_{ij}=1,$$

then we get the unique solution

$$x=[368\ 208\ 1812\ 2490\ 3025\ 0\ 1604\ 1955\ 2490\ 1000\ 1604\ 1812\ 0\ 208\ 368\ 3472]^T/18944.$$

Since $x_b=3472/18944=217/1184$ the theorem is proved. □

We have also performed the calculations in the case $n = 7$ and got the result in Theorem 6. The proof of Theorem 6 is based on the same idea as the proof of Theorem 5. However, since the proof of Theorem 6 is very messy it is omitted here. Since $9393/64921 > 0.1446 > 1/7$, the bound is sharper than the bound in László's conjecture (1.2).

THEOREM 6. *If $A \in \mathscr{M}_7$, then*

$$\|A - \mathscr{N}_7\|_F^2 \leqslant (1 - \frac{9393}{64921})\mathrm{dep}^2(A). \tag{8.15}$$

Presumably, it is possible to continue in the same way and prove sharper bounds for $n = 8, 9, \ldots$. However, the LP problems get more and more messy and the results in the cases n=5, 6 and 7 are only slightly sharper than (5.1). There does not seem to be much point in solving even messier LP problems to get presumably small improvements. Therefore we stop here.

9. Summary

We have proved that Láslzló's conjecture (1.2) is correct for all $n \leqslant 8$ and all even n. However, for the dimensions $n = 3, 5, 6, 7$ we have proved the sharper bounds (2.1), (7.1), (8.1), (8.15). It can be mentioned that the author has also tried to find a sharper bound in the case $n = 4$ without success, however in the case $n = 8$ the author has found a sharper bound which is omitted in this paper. Presumably, it is also possible to use the ideas in the proofs to find sharper bounds for other dimensions. We have also proved the general bound (5.1) which is close to (1.2) when n is large.

Acknowledgements. The author would like to thank Professor Lajos László, Professor Bo Kågström and the referee for comments on the first drafts of this paper.

REFERENCES

[1] L. Elsner and Kh. D. Ikramov, *Towards a proof of László's conjecture*, private communication with L. Elsner (hand-written letter from 1996).
[2] P. E. Gill, W. Murray and M. H. Wright, *Numerical Linear Algebra and Optimization* Vol 1, Addison-Wesley, 1990.
[3] Kh. D. Ikramov, *On normal completions of triangular matrices*, Doklady Akademii Nauk, 351 (1996), pp. 1–2 (in Russian).
[4] Kh. D. Ikramov, *On normal dilations of triangular matrices*, Mathematical notes, 60, N6(1996), pp. 861–872 (in Russian).
[5] L. László, *Upper bounds for the best normal approximation*, *Mathematica Pannonica* **9**, 1 (1998), pp. 121–129.

(Received December 12, 1997)

Anders Barrlund
Department of Computing Science
University of Umeå
S-90187 Umeå, Sweden.
e-mail: abrr@cs.umu.se