Operators
and
Matrices
Volume 16, Number 3 (2022), 775–792
doi:10.7153/oam-2022-16-55

SINGULAR VALUE STRUCTURE OF REAL MATRICES WHICH CAN BE EXPRESSED AS A LINEAR COMBINATION OF TWO ORTHOGONAL MATRICES

ZHENG LI*, TIE ZHANG AND CHANG-JUN LI

(*Communicated by E. Poon*)

Abstract. Can every real matrix be expressed as a linear combination of a certain number of orthogonal matrices? What is the smallest number of these orthogonal matrices? These elegant and interesting questions were initially raised by Zhan [7]. Soon afterwards, Li and Poon [4] proved that k_{min}, the smallest number of these orthogonal matrices, is not greater than 4. These classic results inspire us to further explore the improvement or supplement of this theory.

We investigate some fundamental properties of $\mathscr{A}_n(k)$, which is the set of all $n\times n$ real matrices that can be expressed as a linear combination of k orthogonal matrices. Furthermore, we characterize the singular value structure of matrices in set $\mathscr{A}_n(2)$ and the block structure of related orthogonal matrices. We obtain an equivalent condition and some sufficient or necessary conditions of $A\in\mathscr{A}_n(2)$. Based on these results, we demonstrate the existence of matrices that are not in set $\mathscr{A}_n(2)$, and prove that $k_{min}>2$ (for $n\geqslant 3$).

1. Introduction

Additive decomposition (or called matrix splitting) of a matrix A is to decompose A into the sum of some special types of matrices. This technique is fundamental and usually useful in the analysis of matrix computation or other applications of matrix theory.

Let $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ be the spaces of $n\times n$ real matrices and complex matrices, respectively. Wu [6] and Zhan [7] have pointed out that every matrix in $M_n(\mathbb{C})$ is a linear combination of two unitary matrices. However, the situation in $M_n(\mathbb{R})$ is different. Zhan [7, Observation 7] initially proposed the following result.

THEOREM 1.1. [7] *Every matrix in $M_n(\mathbb{R})$ is a linear combination of n real orthogonal matrices.*

At the same time, an elegant and interesting question was raised as follows.

Mathematics subject classification (2020): 15A16, 15A18.
Keywords and phrases: Orthogonal matrices, linear combination, singular value, additive decomposition, cone.
* Corresponding author.

Paper OaM-16-55

QUESTION 1.1. [7, Question 3] Does there exist a fixed positive integer k, independent of the matrix order n, such that every matrix in $M_n(\mathbb{R})$ is a real linear combination of at most k orthogonal matrices?

Soon after that, Li and Poon [4, Proposition 1] gave a sharper result, which has been included in [8, Page 12, Exercise 13], as follows.

THEOREM 1.2. [4, 8] *Every matrix* $A \in M_n(\mathbb{R})$ *is a linear combination of four orthogonal matrices.*

Let k_{min} denote the minimum value of k mentioned in Question 1.1, then Theorem 1.2 essentially reveals that $k_{min} \leqslant 4$. Naturally, the following question arose.

QUESTION 1.2. [9, page 232, Question 14] Is the number 4 of the terms in the above expression least possible?

So far as we know, Question 1.2 is still an open problem.

The remaining sections of this paper are organized as follows. In Section 2 we introduce some terms and basic lemmas. In Section 3, we investigate some fundamental properties of $\mathscr{A}_n(k)$, the set of all $n \times n$ real matrices that can be expressed as a linear combination of k orthogonal matrices. In Section 4, we study set $\mathscr{A}_n(2)$. Specifically, in Lemma 4.1 and Lemma 4.3, via some elementary analysis, we characterize the singular value structures of nonsingular diagonal matrices and singular diagonal matrices in set $\mathscr{A}_n(2)$ and the block structure of related orthogonal matrices. Based on these results, in Proposition 4.4 we obtain an equivalent condition of $A \in \mathscr{A}_n(2)$. Furthermore, we give some sufficient or necessary conditions as the specific criteria to identify whether $A \in \mathscr{A}_n(2)$ or not. Subsequently, we reveal that set $\mathscr{A}_n(2)$ is not closed with respect to matrix addition, and demonstrate the existence of matrices that are not in $\mathscr{A}_n(2)$, which essentially prove that $k_{min} > 2$ (for $n \geqslant 3$). In Section 5, we briefly summarize the work of this paper.

2. Preliminaries

Let $m,\ n \in \mathbb{N}$. The set of all $m \times n$ real matrices is denoted by $M_{m,\,n}(\mathbb{R})$, or abbreviated as $M_{m,\,n}$ without ambiguity. In particular, the set $M_{n,\,n}$ is abbreviated as M_n, and $\mathbb{R}^n := M_{n,\,1}$.

DEFINITION 2.1. Let $A \in M_n$ (or $M_{m,\,n}$, which is not covered in this article).

- The multiplicity of a singular value α of A is denoted by $\mu[\alpha]$. The singular value α is called a *simple singular value* if $\mu[\alpha] = 1$, or a *multiple singular value* if $\mu[\alpha] \geqslant 2$. The singular value α is called an *odd singular value* when it has an odd multiplicity, or an *even singular value* when it has an even multiplicity.
- The set of all singular values of A is denoted by $\sigma(A)$. The maximum singular value of A is denoted by $\sigma_{max}(A)$. The sets of all simple singular values, multiple

singular values, odd singular values, and even singular values of A are denoted by $\sigma_s(A)$, $\sigma_m(A)$, $\sigma_o(A)$ and $\sigma_e(A)$ respectively.

- For any set B with finite elements, its cardinality is denoted by $Card(B)$.

For example, let $A = diag(1, 1, 1, 2, 3, 3, 4, 5) \in M_8$, then $\sigma_{max}(A) = 5$, and $\sigma(A) = \{1, 2, 3, 4, 5\}$, $Card(\sigma(A)) = 5$; $\sigma_s(A) = \{2, 4, 5\}$, $Card(\sigma_s(A)) = 3$; $\sigma_m(A) = \{1, 3\}$, $Card(\sigma_m(A)) = 2$; $\sigma_o(A) = \{1, 2, 4, 5\}$, $Card(\sigma_o(A)) = 4$; $\sigma_e(A) = \{3\}$, and $Card(\sigma_e(A)) = 1$.

Let $\mathscr{O}_n$ denote the set of all $n \times n$ orthogonal matrices (in particular, $\mathscr{O}_1 = \{1, -1\}$), then a well-known result is as follows.

LEMMA 2.1. (Unitary equivalence of the singular values) [3, 8] *Let $A \in M_n$. Then*

$$\sigma(UAV) = \sigma(A), \quad \forall\, U, V \in \mathscr{O}_n.$$

Let the symbol O always represent zero matrices (which may have different sizes in different occasions), and $\|\cdot\|$ the Euclidean norm of vectors or matrices. We have the following lemma.

LEMMA 2.2. *Let*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \in \mathscr{O}_{n_1+n_2},$$

where $P_{11} \in M_{n_1}$, and $P_{22} \in M_{n_2}$. Then the following four conclusions are equivalent.

(i) $P_{12} = O$. (ii) $P_{21} = O$. (iii) $P_{11} \in \mathscr{O}_{n_1}$. (iv) $P_{22} \in \mathscr{O}_{n_2}$.

Proof. The fact that each row vector or column vector of matrix P is normalized yields

$$\|P_{11}\|^2 + \|P_{12}\|^2 = n_1 = \|P_{11}\|^2 + \|P_{21}\|^2, \tag{2.1}$$

and

$$\|P_{22}\|^2 + \|P_{12}\|^2 = n_2 = \|P_{22}\|^2 + \|P_{21}\|^2, \tag{2.2}$$

which show that conclusions (i) and (ii) are equivalent, and they are equivalent to

$$P = \begin{pmatrix} P_{11} & O \\ O & P_{22} \end{pmatrix}. \tag{2.3}$$

Equality (2.3) and the orthogonality of P immediately lead to conclusions (iii) and (iv).

Conversely, if condition (iii) holds, then $\|P_{11}\|^2 = n_1$, and (2.3) can be deduced from (2.1). Similarly, condition (iv) yields $\|P_{22}\|^2 = n_2$, while (2.3) can also be derived from (2.2). □

3. Some fundamental properties of set $\mathscr{A}_n(k)$

We use the notation $X \subset Y$ to indicate that X is a subset of Y, including the case $X = Y$. Let I_t represent the $t \times t$ identity matrix (in particular, $I_1 = 1$). Based on some basic concepts and theoretical results such as Theorem 1.1 and Theorem 1.2, we summarize some fundamental properties of set $\mathscr{A}_n(k)$ as follows.

PROPOSITION 3.1. *For any given $n, k \in \mathbb{N}$, the set $\mathscr{A}_n(k)$ has the following properties.*

1. $O \in \mathscr{A}_n(k)$.
2. $\mathscr{O}_n \subset \mathscr{A}_n(k) \subset M_n$. *In particular,* $I_n \in \mathscr{A}_n(k) \subset M_n$.
3. $\mathscr{A}_n(k_1) \subset \mathscr{A}_n(k_2)$, *for* $k_1, k_2 \in \mathbb{N}$ *and* $k_1 < k_2$.
4. *If* $k \geqslant 4$, *then* $\mathscr{A}_n(k) = \mathscr{A}_n(k+k_1) = M_n$, $\forall\, k_1 \in \mathbb{N}$.
5. *If* $c \in \mathbb{R}$ *and* $A \in \mathscr{A}_n(k)$, *then* $cA \in \mathscr{A}_n(k)$.
6. $A \in \mathscr{A}_n(k)$ *if and only if* $UAV \in \mathscr{A}_n(k)$, $\forall\, U, V \in \mathscr{O}_n$.
7. *Let* $k_1, k_2, n_1, n_2 \in \mathbb{N}$. *If* $A \in \mathscr{A}_{n_1}(k_1)$ *and* $B \in \mathscr{A}_{n_2}(k_2)$, *then*
$$A \otimes B \in \mathscr{A}_{n_1 n_2}(min\{k_1 k_2, 4\}),$$
where $\otimes$ *denotes the Kronecker product (or called tensor product) of matrices.*

Proof. Properties 1-5 can be directly derived from Theorem 1.2 and the definition of $\mathscr{A}_n(k)$. The details are omitted. For any $U, V \in \mathscr{O}_n$, we have

$$\begin{aligned} A \in \mathscr{A}_n(k) &\Longleftrightarrow A = \sum_{i=1}^{k} c_i Q_i, \ \text{ where } c_i \in \mathbb{R},\, Q_i \in \mathscr{O}_n,\, i = 1, \dots, k. \\ &\Longleftrightarrow UAV = \sum_{i=1}^{k} c_i U Q_i V, \ \text{ where } c_i \in \mathbb{R},\, UQ_iV \in \mathscr{O}_n,\, i = 1, \dots, k. \\ &\Longleftrightarrow UAV \in \mathscr{A}_n(k), \end{aligned}$$

which leads to Property 6.

If $A \in \mathscr{A}_{n_1}(k_1)$ and $B \in \mathscr{A}_{n_2}(k_2)$, then $A = \sum_{i=1}^{k_1} c_i P_i$, where $c_i \in \mathbb{R}$, $P_i \in \mathscr{O}_{n_1}$, $i = 1, \dots, k_1$, and $B = \sum_{j=1}^{k_2} t_j Q_j$, where $t_j \in \mathbb{R}$, $Q_j \in \mathscr{O}_{n_2}$, $j = 1, \dots, k_2$. Thus we have

$$A \otimes B = \left(\sum_{i=1}^{k_1} c_i P_i\right) \otimes \left(\sum_{j=1}^{k_2} t_j Q_j\right) = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} c_i t_j P_i \otimes Q_j,$$

where $P_i \otimes Q_j \in \mathscr{O}_{n_1 n_2}$, $i = 1, \dots, k_1$, $j = 1, \dots, k_2$, (see [2, 8]) which indicates $A \otimes B \in \mathscr{A}_{n_1 n_2}(k_1 k_2)$. Combining this result with Property 4, we obtain Property 7. $\square$

Property 5 of Proposition 3.1 shows that set $\mathscr{A}_n(k)$ is a cone (see [5] for the definition) in M_n.

Throughout this paper, let the symbol T express the transpose of vectors or matrices, and symbol $|\cdot|$ the absolute value function. Some further properties of $\mathscr{A}_n(k)$ are presented as follows.

PROPOSITION 3.2. *Let $n \in \mathbb{N}$ and $A \in M_n$. Then $A \in \mathscr{A}_n(1)$ if and only if*

$$Card(\sigma(A)) = 1.$$

Proof. "$\Rightarrow$". Assume $A \in \mathscr{A}_n(1)$, and $A = cQ$ with $c \in \mathbb{R}$ and $Q \in \mathscr{O}_n$. It is evident that

$$A^T A = c^2 Q^T Q = c^2 I,$$

which implies that A contains a unique singular value $|c|$. Thus the necessity is proved.

"$\Leftarrow$". Let $Card(\sigma(A)) = 1$. Applying singular value decomposition (SVD) (see [1, 3]), we can choose matrices $U,\ V \in \mathscr{O}_n$ such that

$$D = UAV = diag(\sigma_1,\ \sigma_2,\ \ldots,\ \sigma_n),$$

where $\sigma_j \in \sigma(D)$, $j = 1, \ldots, n$. Lemma 2.1 and the assumption "$Card(\sigma(A)) = 1$" yield $\sigma(A) = \sigma(D)$ and $\sigma_1 = \sigma_2 = \ldots = \sigma_n = c$, which means $UAV = cI$. From Properties 2, 5 and 6 of Proposition 3.1, it is concluded that $A \in \mathscr{A}_n(1)$. Hence the sufficiency of this proposition is proved. $\square$

REMARK 3.1. The proof of Proposition 3.2 takes advantage of SVD. This technique has been applied in many references, such as [7, 4], and enlighten us that it is usually sufficient to focus only on diagonal matrices.

Proposition 3.2 gives a clear characterization of set $\mathscr{A}_n(1)$. Some characterizations of $\mathscr{A}_n(2)$ will be presented in the next section. To begin with, let $rank(\cdot)$ denote the rank of the matrix, and for any two vectors $x = (x_1, \ldots, x_n)^T$, $y = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$, define their inner product as

$$\langle x,\ y\rangle = \sum_{i=1}^{n} x_i y_i, \tag{3.1}$$

and we prepare the following useful results.

PROPOSITION 3.3. *Suppose $k \in \mathbb{N}$, $k \geqslant 2$, and $A \in \mathscr{A}_n(k)$ with $A = \sum_{i=1}^{k} c_i Q_i$, where $Q_i \in \mathscr{O}_n$, $c_i \in \mathbb{R}$, $i = 1, \ldots, k$. Let*

$$\lambda(c_1, \ldots,\ c_k) = \sum_{i=1}^{k} c_i^2 - 2 \sum_{\substack{i,\ j=1 \\ i<j}}^{k} |c_i c_j|. \tag{3.2}$$

If $\lambda(c_1, \ldots,\ c_k) > 0$, then $rank(A) = n$.

Proof. For arbitrary nonzero $x \in \mathbb{R}^n$, we have

$$\|Ax\|^2 = \left\langle \left(\sum_{i=1}^{k} c_i Q_i\right) x,\ \left(\sum_{j=1}^{k} c_j Q_j\right) x \right\rangle$$
$$= \left(\sum_{i=1}^{k} c_i^2\right) \|x\|^2 + \sum_{\substack{i,\, j=1 \\ i<j}}^{k} 2 c_i c_j \left\langle Q_i x,\ Q_j x \right\rangle. \tag{3.3}$$

Cauchy-Schwartz inequality and the orthogonal invariance of the Euclidean-norm yield

$$|\left\langle Q_i x,\ Q_j x \right\rangle| \leqslant \|Q_i x\| \|Q_j x\| = \|x\|^2.$$

Hence (3.3) leads to the fact that

$$x^T A^T A x = \|Ax\|^2 \geqslant \lambda(c_1, \ldots, c_k) \|x\|^2, \tag{3.4}$$

where $\lambda(c_1, \ldots, c_k)$ is defined by (3.2). If $\lambda(c_1, \ldots, c_k) > 0$, then from (3.4) it follows that $x^T A^T A x > 0$, which indicates that matrix $A^T A$ is positive definite, and implies $rank(A) = n$. $\square$

COROLLARY 3.4. *Suppose $A \in \mathscr{A}_n(2)$, and $A = c_1 Q_1 + c_2 Q_2$, where $Q_1, Q_2 \in \mathscr{O}_n$, and $c_1, c_2 \in \mathbb{R}$. If $|c_1| \neq |c_2|$, then $rank(A) = n$.*

Proof. Consider the case $k = 2$ of Proposition 3.3. Now

$$\lambda(c_1, c_2) = c_1^2 + c_2^2 - 2|c_1||c_2| = (|c_1| - |c_2|)^2.$$

Therefore, $\lambda(c_1, c_2) > 0$ if (and only if) $|c_1| \neq |c_2|$. Thus the proof is completed. $\square$

REMARK 3.2. An equivalent statement of the conclusion of Corollary 3.4 is:

$$\text{If } rank(A) < n, \text{ then } |c_1| = |c_2|.$$

In the previous discussion on $A = \sum_{i=1}^{k} c_i Q_i$, where $Q_i \in \mathscr{O}_n$, we often assume that the coefficients $c_i \in \mathbb{R}$, for $i = 1, \ldots, k$. However, noticing the fact that $Q \in \mathscr{O}_n$ if and only if $-Q \in \mathscr{O}_n$, and let

$$\mathbb{R}_0^+ := \{x \in \mathbb{R} |\, x \geqslant 0\}, \text{ and } \mathbb{R}^+ := \{x \in \mathbb{R} |\, x > 0\}.$$

we shall only investigate the case $c_i \in \mathbb{R}_0^+$ (sometimes $c_i \in \mathbb{R}^+$), for $i = 1, \ldots, k$, in the upcoming discussion in Section 4.

4. Singular value structure of matrices in set $\mathscr{A}_n(2)$ and block structure of related orthogonal matrices

LEMMA 4.1. *For a nonsingular diagonal matrix $D \in M_n$, let $n_s = Card(\sigma_s(D))$, $n_m = Card(\sigma_m(D))$, $\sigma_s(D) = \{d_1, \dots, d_{n_s}\} \subset \mathbb{R}^+$, $\sigma_m(D) = \{\tilde{d}_{n_s+1}, \dots, \tilde{d}_{n_s+n_m}\} \subset \mathbb{R}^+$, and $D = diag(d_1, \dots, d_n)$ with the block structure as*

$$D = diag\left(\tilde{d}_1, \dots, \tilde{d}_{n_s}, \tilde{d}_{n_s+1} I_{\mu[\tilde{d}_{n_s+1}]}, \dots, \tilde{d}_{n_s+n_m} I_{\mu[\tilde{d}_{n_s+n_m}]}\right), \tag{4.1}$$

where $\tilde{d}_j = d_j$ for $j = 1, \dots, n_s$, and $\mu[\tilde{d}_l] \geqslant 2$ for $l = n_s+1, \dots, n_s+n_m$. If

$$D = c_1 P + c_2 Q, \text{ where } c_1, c_2 \in \mathbb{R}^+, \text{ and } P = (p_{ij}), Q = (q_{ij}) \in \mathscr{O}_n, \tag{4.2}$$

then the following conclusions are true.

(i)

$$p_{jj} = \frac{d_j^2 + c_1^2 - c_2^2}{2 d_j c_1}, \quad q_{jj} = \frac{d_j^2 + c_2^2 - c_1^2}{2 d_j c_2}, \quad \text{for } j = 1, 2, \dots, n; \tag{4.3}$$

$$p_{ij} = \frac{-c_2}{c_1} q_{ij}, \quad p_{ij} = -p_{ji}, \quad q_{ij} = -q_{ji}, \quad \text{for } i, j = 1, \dots, n, \text{ and } i \neq j. \tag{4.4}$$

(ii) Matrices P and Q have the block structures as

$$\begin{cases} P = diag(\tilde{P}_1 \dots, \tilde{P}_{n_s}, \tilde{P}_{n_s+1}, \dots, \tilde{P}_{n_s+n_m}), \\ Q = diag(\tilde{Q}_1 \dots, \tilde{Q}_{n_s}, \tilde{Q}_{n_s+1}, \dots, \tilde{Q}_{n_s+n_m}), \end{cases} \tag{4.5}$$

where $\tilde{P}_j, \tilde{Q}_j \in \mathscr{O}_1$ for $j = 1, \dots, n_s$, and $\tilde{P}_l, \tilde{Q}_l \in \mathscr{O}_{\mu[\tilde{d}_l]}$ with $\mu[\tilde{d}_l] \geqslant 2$ for $l = n_s+1, \dots, n_s+n_m$.

(iii)

$$Card(\sigma_s(D)) \leqslant 2.$$

Proof. The assumption in (4.2) immediately yields

$$c_1 p_{ij} = -c_2 q_{ij}, \quad i, j = 1, 2, \dots, n, \text{ and } i \neq j, \tag{4.6}$$

$$c_1 p_{jj} + c_2 q_{jj} = d_j, \quad j = 1, 2, \dots, n. \tag{4.7}$$

Define

$$\hat{p}_j = (p_{1,j}, \dots, p_{j-1,j}, p_{j+1,j}, \dots, p_{n,j})^T \in \mathbb{R}^{n-1},$$
$$\hat{q}_j = (q_{1,j}, \dots, q_{j-1,j}, q_{j+1,j}, \dots, q_{n,j})^T \in \mathbb{R}^{n-1},$$

then (4.6) is equivalent to

$$c_1 \hat{p}_j = -c_2 \hat{q}_j, \quad j = 1, \dots, n. \tag{4.8}$$

Since all columns of P and Q are normalized, it follows that

$$\|\hat{p}_j\|^2 = 1 - p_{jj}^2, \quad \|\hat{q}_j\|^2 = 1 - q_{jj}^2. \tag{4.9}$$

Substituting (4.8) into (4.9), we have

$$c_1^2(1-p_{jj}^2) = c_2^2(1-q_{jj}^2). \tag{4.10}$$

Then substitute (4.7) into (4.10) and after some reduction, we obtain (4.3).

Moreover, for any matrix W, let $W(i\,,:)$ denote the i-th row vector, and $W(:\,,\,i)$ the i-th column vector. Then from the orthogonality among the respective column vectors of matrices P and Q, it follows that for any $i \neq j$, we have

$$0 = \langle P(:\,,\,i)\,,\,P(:\,,\,j)\rangle = \sum_{\substack{k=1\\k\neq i,j}}^{n} p_{ki}p_{kj} + p_{ii}p_{ij} + p_{ji}p_{jj}, \tag{4.11}$$

and

$$0 = \langle Q(:\,,\,i)\,,\,Q(:\,,\,j)\rangle = \sum_{\substack{k=1\\k\neq i,j}}^{n} q_{ki}q_{kj} + q_{ii}q_{ij} + q_{ji}q_{jj}. \tag{4.12}$$

From (4.3), (4.6) and via some deduction, equality (4.12) can be rewritten as

$$0 = c_1 \sum_{\substack{k=1\\k\neq i,j}}^{n} p_{ki}p_{kj} - p_{ij}\frac{d_i^2+c_2^2-c_1^2}{2d_i} - p_{ji}\frac{d_j^2+c_2^2-c_1^2}{2d_j}. \tag{4.13}$$

Meanwhile, after substituting (4.3) into (4.11) and multiplying both sides of equation (4.11) by c_1, we get

$$0 = c_1 \sum_{\substack{k=1\\k\neq i,j}}^{n} p_{ki}p_{kj} + p_{ij}\frac{d_i^2+c_1^2-c_2^2}{2d_i} + p_{ji}\frac{d_j^2+c_1^2-c_2^2}{2d_j}. \tag{4.14}$$

Subtract both sides of equation (4.14) from that of equation (4.13) we obtain

$$p_{ij}d_i + p_{ji}d_j = 0. \tag{4.15}$$

In addition, from the orthogonality among the respective row vectors of matrices P and Q, for any $i \neq j$, we have

$$0 = \langle P(i,\,:)\,,\,P(j,\,:)\rangle = \sum_{\substack{k=1\\k\neq i,j}}^{n} p_{ik}p_{jk} + p_{ii}p_{ji} + p_{ij}p_{jj}, \tag{4.16}$$

$$0 = \langle Q(i,\,:)\,,\,Q(j,\,:)\rangle = \sum_{\substack{k=1\\k\neq i,j}}^{n} q_{ik}q_{jk} + q_{ii}q_{ji} + q_{ij}q_{jj}, \tag{4.17}$$

Similar to the discussion from (4.11) to (4.15), we can also obtain

$$p_{ij}d_j + p_{ji}d_i = 0. \tag{4.18}$$

For every coordinate $(i,\ j)$ with $i \neq j$, the solutions of equalities (4.15) and (4.18) have two cases, i.e.,

$$p_{ij} = -p_{ji} \text{ (may be 0 or not)}, \text{ (symmetrically, } q_{ij} = -q_{ji}), \quad \text{if } d_i = d_j; \tag{4.19}$$

or

$$p_{ij} = p_{ji} = 0, \quad \text{(symmetrically, } q_{ij} = q_{ji} = 0), \quad \text{if } d_i \neq d_j. \tag{4.20}$$

Therefore, conclusion (i) is true for any case of (4.19) or (4.20).

For any given $j \in \{1,\ldots,\ n_s\}$, since $d_j = \tilde{d}_j \in \sigma_s(D)$, then $d_j \neq d_i$ for $i = 1,\ldots,\ n$, and $i \neq j$. By (4.20) this means

$$p_{ij} = p_{ji} = 0, \quad \text{for } j = 1,\ldots,\ n_s,\ i = 1,\ldots,\ n, \text{ and } i \neq j, \tag{4.21}$$

and the orthogonality of matrix P leads to

$$|p_{jj}| = 1, \quad \text{(symmetrically, } |q_{jj}| = 1), \quad \text{for } j = 1,\ldots,\ n_s. \tag{4.22}$$

On the other hand, for any given $l \in \{n_s+1,\ldots,\ n_s+n_m\}$, there is a unique value $\tilde{d}_l \in \sigma_m(D)$ and some d_j ($j \in \{n_s+1,\ldots,\ n\}$) with multiplicity $\mu[\tilde{d}_l]$ such that

$$d_j = \tilde{d}_l, \quad \text{for } j = \sum_{r=1}^{l-1}\mu[\tilde{d}_r]+1,\ \sum_{r=1}^{l-1}\mu[\tilde{d}_r]+2,\ldots,\ \sum_{r=1}^{l}\mu[\tilde{d}_r].$$

(We note that $\sum_{r=1}^{l}\mu[\tilde{d}_r] - \sum_{r=1}^{l-1}\mu[\tilde{d}_r] = \mu[\tilde{d}_l]$. In particular, $\sum_{r=1}^{l-1}\mu[\tilde{d}_r] = 0$ if $l = 1$, which means $\sigma_s(D) = \varnothing$ in this case, where the symbol $\varnothing$ always denotes the empty set throughout this paper.) Thus (4.19) and (4.20) force the entries of columns n_s+1 to n of matrices P and Q to satisfy the following conditions:

$$\begin{cases} \left.\begin{aligned} p_{ij} &= -p_{ji}, \\ q_{ij} &= -q_{ji}, \end{aligned}\right\} & i,\ j = \sum\limits_{r=1}^{l-1}\mu[\tilde{d}_r]+1,\ \sum\limits_{r=1}^{l-1}\mu[\tilde{d}_r]+2,\ldots,\ \sum\limits_{r=1}^{l}\mu[\tilde{d}_r], \\ \left.\begin{aligned} p_{ij} &= p_{ji} = 0, \\ q_{ij} &= q_{ji} = 0, \end{aligned}\right\} & \text{else}, \end{cases} \tag{4.23}$$

for every $l = n_s+1,\ n_s+2,\ldots,\ n_s+n_m$. Therefore, combining (4.21)–(4.23) with Lemma 2.2, we see that matrices P and Q have the structures as (4.5), and obtain the conclusion (ii).

For $d_j \in \sigma_s(D)$, since (4.22) holds, then equality (4.3) yields

$$d_j^2 + c_1^2 - c_2^2 = \pm 2d_jc_1, \quad d_j^2 + c_2^2 - c_1^2 = \pm 2d_jc_2, \tag{4.24}$$

or equivalently

$$(d_j \pm c_1)^2 = c_2^2, \quad (d_j \pm c_2)^2 = c_1^2. \tag{4.25}$$

Because $d_j,\ c_1,\ c_2$ are all assumed to be positive numbers, there are at most two possibilities for the solutions of all the $d_j \in \sigma_s(D)$, i.e., $d_j = c_1 + c_2$, or $d_j = |c_1 - c_2|$. This implies that matrix D contains at most two simple singular values. Hence we prove the conclusion (iii). $\square$

LEMMA 4.2. *Under the assumptions of Lemma 4.1, if $c_1 = c_2$, then the following conclusions are true.*

(i) Matrix P has the same structure as (4.5) with each block $\tilde{P}_j$ satisfying

$$\tilde{P}_j = \frac{d_j}{2c} I_{\mu[\tilde{d}_j]} + X_j, \text{ with } X_j^T = -X_j,\ j = 1, \ldots, n_s + n_m, \tag{4.26}$$

and matrix Q satisfies $Q = P^T$.

(ii) $Card(\sigma_s(D)) \leqslant 1$.

Proof. The proof of Lemma 4.1 can be roughly copied here. Now that $c_1 = c_2 = c$, from (4.3) we immediately get the conclusion (i).

If $d_j \in \sigma_s(D)$, since $c\ (= c_1 = c_2)$, $d_j \in \mathbb{R}^+$, then equation (4.25) has at most a positive solution $d_j = 2c$. Thus conclusion (ii) is proved. □

Based on Lemma 4.2, we prepare a similar result for singular matrices.

LEMMA 4.3. *Suppose τ, $n \in \mathbb{N}$ satisfying $1 \leqslant \tau \leqslant n-1$, and a singular diagonal matrix $D \in M_n$ with the form $D = \begin{pmatrix} D_\tau & O \\ O & O \end{pmatrix}$, where $D_\tau \in M_\tau$ is nonsingular diagonal. Let $n_s = Card(\sigma_s(D_\tau))$, $n_m = Card(\sigma_m(D_\tau))$, $\sigma_s(D_\tau) = \{d_1, \ldots, d_{n_s}\} \subset \mathbb{R}^+$, $\sigma_m(D_\tau) = \{\tilde{d}_{n_s+1}, \ldots, \tilde{d}_{n_s+n_m}\} \subset \mathbb{R}^+$, and $D_\tau = diag(d_1, \ldots, d_\tau)$ with the block structure as*

$$D_\tau = diag\left(\tilde{d}_1, \ldots, \tilde{d}_{n_s}, \tilde{d}_{n_s+1} I_{\mu[\tilde{d}_{n_s+1}]}, \ldots, \tilde{d}_{n_s+n_m} I_{\mu[\tilde{d}_{n_s+n_m}]}\right),$$

where $\tilde{d}_j = d_j$ for $j = 1, \ldots, n_s$, and $\mu[\tilde{d}_l] \geqslant 2$ for $l = n_s+1, \ldots, n_s+n_m$. If

$$D = c_1 P + c_2 Q, \text{ where } c_1, c_2 \in \mathbb{R}^+, \text{ and } P, Q \in \mathscr{O}_n, \tag{4.27}$$

then the following are true.

(i) $c_1 = c_2$, and $D_\tau \in \mathscr{A}_\tau(2)$.

(ii) $Card(\sigma_s(D)) \leqslant 2$, and matrices P and Q have the structures as

$$\begin{cases} P = diag\left(\tilde{P}_1, \ldots, \tilde{P}_{n_s}, \tilde{P}_{n_s+1}, \ldots, \tilde{P}_{n_s+n_m}, \tilde{P}_{n_s+n_m+1}\right), \\ Q = diag\left(\tilde{Q}_1, \ldots, \tilde{Q}_{n_s}, \tilde{Q}_{n_s+1}, \ldots, \tilde{Q}_{n_s+n_m}, \tilde{Q}_{n_s+n_m+1}\right), \end{cases} \tag{4.28}$$

where $\tilde{P}_j$, $\tilde{Q}_j \in \mathscr{O}_{\mu[\tilde{d}_j]}$ with $\tilde{Q}_j = \tilde{P}_j^T$, and $\tilde{P}_j^T$ satisfy (4.26) for $j = 1, \ldots, n_s + n_m$, and

$$\tilde{P}_{n_s+n_m+1} = -\tilde{Q}_{n_s+n_m+1}, \text{ where } \tilde{P}_{n_s+n_m+1}, \tilde{Q}_{n_s+n_m+1} \in \mathscr{O}_{n-\tau}. \tag{4.29}$$

Proof. Since $rank(D) < n$, from Corollary 3.4 it follows that

$$c_1 = c_2 = c \in \mathbb{R}^+. \tag{4.30}$$

Matrices P and Q can be written as

$$P = (p_1, \ldots, p_\tau, P_{n-\tau}), \text{ and } Q = (q_1, \ldots, q_\tau, Q_{n-\tau}),$$

where $p_j,\ q_j,\ j=1,\dots,\ \tau$, are the first τ columns of matrices P and Q, respectively, while blocks $P_{n-\tau}$ and $Q_{n-\tau}$ are composed of the last $n-\tau$ columns of matrices P and Q, respectively. Let e_j denote the j-th canonical (column) vector. From (4.27) and (4.30) it follows that

$$p_j+q_j=(d_j/c)e_j,\quad \text{for } j=1,\ 2,\dots,\tau; \tag{4.31}$$

$$P_{n-\tau}=-Q_{n-\tau}. \tag{4.32}$$

Combining (4.32) with (4.31), and from the orthogonality of matrices P and Q we get

$$P^T_{n-\tau}e_j=\frac{c}{d_j}\left(P^T_{n-\tau}p_j-Q^T_{n-\tau}q_j\right)=\frac{c}{d_j}(\mathbf{0}-\mathbf{0})=\mathbf{0}\in\mathbb{R}^{n-\tau},\ \text{ for } j=1,\dots,\ \tau, \tag{4.33}$$

where $\mathbf{0}$ denotes the zero column vector. Equation (4.33) means the block $P_{n-\tau}$ has the form $P_{n-\tau}=\begin{pmatrix} O\\ \bar{P}_{n-\tau}\end{pmatrix}$, where $\bar{P}_{n-\tau}\in M_{n-\tau}$. Consequently $Q_{n-\tau}=\begin{pmatrix} O\\ \bar{Q}_{n-\tau}\end{pmatrix}$, where $\bar{Q}_{n-\tau}\in M_{n-\tau}$. Thus according to Lemma 2.2, matrices P and Q have the forms as

$$P=\begin{pmatrix}\bar{P}_\tau & O\\ O & \bar{P}_{n-\tau}\end{pmatrix},\quad Q=\begin{pmatrix}\bar{Q}_\tau & O\\ O & \bar{Q}_{n-\tau}\end{pmatrix},$$

where $\bar{P}_\tau,\ \bar{Q}_\tau\in\mathscr{O}_\tau$ and $\bar{P}_{n-\tau},\ \bar{Q}_{n-\tau}\in\mathscr{O}_{n-\tau}$. Therefore, (4.27) and (4.30) lead to equations

$$D_\tau=c\bar{P}_\tau+c\bar{Q}_\tau,\ \text{ where } \bar{P}_\tau,\ \bar{Q}_\tau\in\mathscr{O}_\tau, \tag{4.34}$$

$$\bar{P}_{n-\tau}=-\bar{Q}_{n-\tau},\ \text{ where } \bar{P}_{n-\tau},\ \bar{Q}_{n-\tau}\in\mathscr{O}_{n-\tau}. \tag{4.35}$$

Equation (4.34) shows that $D_\tau\in\mathscr{A}_\tau(2)$. Thus conclusion (i) is proved.

Meanwhile, equation (4.34) and Lemma 4.2 yield $n_s\equiv Card(\sigma_s(D_\tau))\leqslant 1$. Therefore, including singular value 0, the number of simple singular values of matrix D should not exceed 2, i.e.,

$$Card(\sigma_s(D))\leqslant 2.$$

Similar to the discussion of Lemma 4.2, blocks $\bar{P}_\tau$ and $\bar{Q}_\tau$ have the same structures as described in conclusion (i) of Lemma 4.2. The result (4.29) can be obtained directly from (4.35). □

Summarizing above discussion, we obtain an equivalent condition of $A\in\mathscr{A}_n(2)$.

PROPOSITION 4.4. *Suppose $n\in\mathbb{N}$, $n\geqslant 3$, and $A\in M_n$. Let $n_s=Card(\sigma_s(A))$, $n_m=Card(\sigma_m(A))$, $\sigma_s(A)=\{\sigma_1,\dots,\sigma_{n_s}\}$, and $\sigma_m(A)=\{\sigma_{n_s+1},\dots,\sigma_{n_s+n_m}\}$. Then $A\in\mathscr{A}_n(2)$ if and only if the following conditions hold.*

(i) $Card(\sigma_s(A))\leqslant 2$;

(ii) There exist $c_1,\ c_2\in\mathbb{R}^+_0$, $P_j,\ Q_j\in\bigcup_{t=1}^{n_s+n_m}\mathscr{O}_{\mu[\sigma_t]}$, $j=1,\dots,\ n_s+n_m$, such that the coefficients c_1 and c_2 are common to all the following equations:

$$\sigma_j I_{\mu[\sigma_j]}=c_1P_j+c_2Q_j,\ \textit{ where } P_j,\ Q_j\in\mathscr{O}_{\mu[\sigma_j]},\ j=1,\dots,\ n_s+n_m. \tag{4.36}$$

Moreover, equalities (4.3) and (4.4) hold if $rank(A) = n$*, or*

$$\begin{cases} P_j = Q_j^T, & \text{with } P_j \text{ satisfying (4.26)}, \quad \text{for } \sigma_j \neq 0, \\ P_j = -Q_j, & \text{for } \sigma_j = 0, \\ c_1 = c_2, & \end{cases} \tag{4.37}$$

holds if $rank(A) < n$.

Proof. "$\Longleftarrow$". If (4.36) holds, applying SVD we choose $U,\ V \in \mathscr{O}_n$ such that

$$D = UAV = diag\left(\sigma_1, \dots,\ \sigma_{n_s},\ \sigma_{n_s+1} I_{\mu[\sigma_{n_s+1}]}, \dots,\ \sigma_{n_s+n_m} I_{\mu[\sigma_{n_s+n_m}]}\right). \tag{4.38}$$

Let $P = diag(P_1, \dots,\ P_{n_s+n_m})$, and $Q = diag(Q_1, \dots,\ Q_{n_s+n_m})$, where $P_j,\ Q_j \in \mathscr{O}_1$ for $j = 1, \dots,\ n_s$, and $P_j,\ Q_j \in \mathscr{O}_{\mu[\sigma_j]}$ for $j = n_s+1, \dots,\ n_s+n_m$, then obviously $P,\ Q \in \mathscr{O}_n$ and $D = c_1 P + c_2 Q$. Thus from Property 6 of Proposition 3.1 it follows that $A \in \mathscr{A}_n(2)$, and the sufficiency is proved.

"$\Longrightarrow$". If $A \in \mathscr{A}_n(2)$, then there exist $P,\ Q \in \mathscr{O}_n$ such that $A = c_1 P + c_2 Q$. Using SVD we have $D = UAV$ with the structure as (4.38). From Property 6 of Proposition 3.1 we see $D \in \mathscr{A}_n(2)$. Therefore, according to the conclusion (ii) of Lemma 4.1 or Lemma 4.3, matrices P and Q have the structures as in (4.5) (if A is non-singular) or (4.28) (if A is singular). Thus (4.36) holds. Conclusion (i) of this proposition and the properties of P_j and Q_j described in (4.37) can also be obtained directly from the results of Lemma 4.1 or Lemma 4.3. □

Some potentially useful criteria are stated as follows.

COROLLARY 4.5. *Let* $A \in M_n$. *If* $Card(\sigma(A)) \leqslant 2$, *then* $A \in \mathscr{A}_n(2)$.

Proof. The case $Card(\sigma(A)) = 1$ has been proved in Proposition 3.2. Next, we suppose $\sigma(A) = \{\sigma_1,\ \sigma_2\}$ where $\sigma_1,\ \sigma_2 \in \mathbb{R}_0^+$ and $\sigma_1 < \sigma_2$. For equations

$$\begin{cases} \sigma_1 I_{\mu[\sigma_1]} = c_1 P_1 + c_2 Q_1, & \text{where } P_1,\ Q_1 \in \mathscr{O}_{\mu[\sigma_1]}, \\ \sigma_2 I_{\mu[\sigma_2]} = c_1 P_2 + c_2 Q_2, & \text{where } P_2,\ Q_2 \in \mathscr{O}_{\mu[\sigma_2]}, \end{cases} \tag{4.39}$$

we can always choose

$$P_1 = Q_1 = I_{\mu[\sigma_1]},\ P_2 = I_{\mu[\sigma_2]},\ Q_2 = -I_{\mu[\sigma_2]},\ c_1 = (\sigma_1 + \sigma_2)/2,\ c_2 = (\sigma_1 - \sigma_2)/2,$$

such that the equations in (4.39) hold. Thus from Proposition 4.4 we complete this proof. □

COROLLARY 4.6. *Let* $n \in \mathbb{N}$ *and* $n \geqslant 3$. *If* $A \in \mathscr{A}_n(2)$ *and* $A = c_1 P + c_2 Q$, *where* $P,\ Q \in \mathscr{O}_n$, *and* $c_1,\ c_2 \in \mathbb{R}_0^+$, *then*

(i)

$$Card(\sigma(A)) \leqslant \begin{cases} n/2 + 1, & \text{if } n \text{ is even}, \\ (n+1)/2, & \text{if } n \text{ is odd}. \end{cases} \tag{4.40}$$

(ii)

$$c_1 + c_2 \geqslant \sigma_{max}(A). \tag{4.41}$$

In particular, if $c = c_1 = c_2$, *then*

$$c \geqslant \sigma_{max}(A)/2. \tag{4.42}$$

Proof. Let $n_s = Card(\sigma_s(A))$ and $n_m = Card(\sigma_m(A))$. It is evident that the largest $Card(\sigma(D))$ occurs when $\mu[\sigma_j] = 2$ for all $\sigma_j \in \sigma_m(A)$, and $n_s = 2$ if n is even, or $n_s = 1$ if n is odd. Thus we have

$$\begin{cases} 2n_m + 2 = n, & \text{if } n \text{ is even}, \\ 2n_m + 1 = n, & \text{if } n \text{ is odd}, \end{cases}$$

and obtain (4.40). Let $\|\cdot\|_2$ denote the spectral norm of matrices. Under the assumption of this corollary, it follows that

$$\sigma_{max}(A) = \|A\|_2 \leqslant c_1\|P\|_2 + c_2\|Q\|_2 = c_1 + c_2,$$

which yields (4.41). Inequality (4.42) consequently holds. □

Next we will show that the upper bounds described in (4.40) are sometimes attainable.

LEMMA 4.7. *If* $r \in \mathbb{N}$ *is an even number, then for identity matrix* $I \in M_r$ *and any given* $\alpha \in [0,\ 1] \subset \mathbb{R}$, *there exist* $P,\ Q \in \mathscr{O}_r$ *such that*

$$\alpha I = \frac{1}{2}P + \frac{1}{2}Q. \tag{4.43}$$

Proof. Define

$$H(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in \mathscr{O}_2, \text{ and } H(\theta)^T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathscr{O}_2, \tag{4.44}$$

with $\theta = \arccos(\alpha)$. Let

$$P = diag(\underbrace{H(\theta), \ldots, H(\theta)}_{r/2}), \quad Q = P^T, \tag{4.45}$$

then we obtain (4.43). □

COROLLARY 4.8. *Let* $n \in \mathbb{N}$, $n \geqslant 3$ *and* $A \in M_n$. *If* $\sigma_o(A) \subset \{0,\ \sigma_{max}(A)\}$, *then* $A \in \mathscr{A}_n(2)$.

REMARK 4.1. Recall that $\sigma_o(A)$ is a superset of $\sigma_s(A)$, while $\sigma_e(A) \subset \sigma_m(A)$.

Proof of Corollary 4.8. The case of $A = O$ is trivial. Next, the discussion will be carried out under the assumption of $A \neq O$.

For the case

$$\sigma_o(A) = \{0,\ \sigma_{max}(A)\}, \tag{4.46}$$

Since $\sigma_{max}(A/\|A\|_2) = 1$, then by SVD we can choose $U, V \in \mathscr{O}_n$ such that

$$D = U(A/\|A\|_2)V = diag(d_1,\ d_2,\ \tilde{d}_3 I_2, \dots,\ \tilde{d}_{n/2+1} I_2), \tag{4.47}$$

where $d_1 = 0$, $d_2 = 1 \in \sigma_o(A/\|A\|_2)$ and $\tilde{d}_j I_2 \in M_2$ with $\tilde{d}_j \in [0,\ 1]$ for $j = 3, \dots, n/2+1$ (There may be some $\tilde{d}_j$ such that $\tilde{d}_j = d_1$ or $\tilde{d}_j = d_2$.) We can set the orthogonal matrices P and Q as

$$\begin{cases} P = diag\left(1,\ 1,\ H(\theta_3),\ \dots,\ H(\theta_{n/2+1})\right), \\ Q = diag\left(-1,\ 1,\ H(\theta_3)^T, \dots,\ H(\theta_{n/2+1})^T\right), \end{cases} \tag{4.48}$$

with $\theta_j = \arccos(\tilde{d}_j)$, $j = 3, \dots, n/2+1$, and each $H(\theta_j)$ defined as in (4.44). Then according to Lemma 4.7 we have

$$D = \frac{1}{2}P + \frac{1}{2}Q \in \mathscr{A}_n(2). \tag{4.49}$$

For the case $\sigma_o(A) = \{0\}$, after a SVD similar to the above discussion, we can reset the orthogonal matrices P and Q in (4.48) to

$$\begin{cases} P = diag\left(1,\ H(\theta_2), \dots,\ H(\theta_{(n+1)/2})\right), \\ Q = diag\left(-1,\ H(\theta_2)^T, \dots,\ H(\theta_{(n+1)/2})^T\right), \end{cases}$$

with $\theta_j = \arccos(\tilde{d}_j)$, $j = 2, \dots, (n+1)/2$.

For the case $\sigma_o(A) = \{\sigma_{max}(A)\}$, similarly we can reset the orthogonal matrices P and Q in (4.48) to

$$\begin{cases} P = diag\left(1,\ H(\theta_2), \dots,\ H(\theta_{(n+1)/2})\right), \\ Q = diag\left(1,\ H(\theta_2)^T, \dots,\ H(\theta_{(n+1)/2})^T\right), \end{cases}$$

with $\theta_j = \arccos(\tilde{d}_j)$, $j = 2, \dots, (n+1)/2$.

And for the case $\sigma_o(A) = \varnothing$, we can reset the orthogonal matrices P and Q in (4.48) to

$$\begin{cases} P = diag\left(H(\theta_1), \dots,\ H(\theta_{n/2})\right), \\ Q = diag\left(H(\theta_1)^T, \dots,\ H(\theta_{n/2})^T\right), \end{cases}$$

with $\theta_j = \arccos(\tilde{d}_j)$, $j = 1, \dots, n/2$.

Obviously, equation (4.49) always holds for all the above cases. Hence from Property 6 of Proposition 3.1, it follows that $A \in \mathscr{A}_n(2)$ under the assumption of this corollary. □

However, the following result shows that conclusion (i) of Corollary 4.6 is only necessary rather than sufficient in general.

COROLLARY 4.9. *Let $n \in \mathbb{N}$, $n \geqslant 7$, and $A \in M_n$. If $rank(A) < n$, and there exist at least two distinct singular values σ_1, σ_2 of A satisfying $\sigma_1, \sigma_2 \in \mathbb{R}^+$ with $\mu[\sigma_1] = \mu[\sigma_2] = 3$, then $A \notin \mathscr{A}_n(2)$.*

Proof. Suppose $A \in \mathscr{A}_n(2)$. Since $rank(A) < n$, then according to equalities (4.37) and (4.26), we directly suppose there exist P_1, $P_2 \in \mathscr{O}_3$ and $c \in \mathbb{R}^+$ such that

$$\sigma_1 I_3 = cP_1 + cP_1^T, \text{ where } P_1 = \frac{\sigma_1}{2c} I_3 + X_1, \text{ and } X_1^T = -X_1, \tag{4.50}$$

and

$$\sigma_2 I_3 = cP_2 + cP_2^T, \text{ where } P_2 = \frac{\sigma_2}{2c} I_3 + X_2, \text{ and } X_2^T = -X_2. \tag{4.51}$$

From (4.50) we have

$$I = P_1^T P_1 = \frac{\sigma_1^2}{4c^2} I + \frac{\sigma_1}{2c}(X_1^T + X_1) + X_1^T X_1 = \frac{\sigma_1^2}{4c^2} I - X_1^2. \tag{4.52}$$

Let $P_1 = (p_{ij}) \in M_3$, and

$$X_1 = \begin{pmatrix} 0 & p_{12} & p_{13} \\ -p_{12} & 0 & p_{23} \\ -p_{13} & -p_{23} & 0 \end{pmatrix},$$

then (4.52) indicates that the entries at the non-diagonal part of X_1^2 are all 0, which yields

$$p_{13}p_{23} = p_{12}p_{23} = p_{12}p_{13} = 0. \tag{4.53}$$

Equalities in (4.53) imply that at least two of the three entries p_{12}, p_{13} and p_{23} should be 0, which leads to that at least one of the diagonal entries of P_1 should be ± 1 from the orthogonality of P_1. Since $\sigma_1, c \in \mathbb{R}^+$, then the diagonal entries of P_1 cannot be negative. So we obtain that $p_{11} = p_{22} = p_{33} = 1$ since all diagonal entries of P_1 are equal to a constant $\sigma_1/(2c)$ (This also means that $P_1 = I_3$), and immediately get

$$c = \sigma_1/2. \tag{4.54}$$

However, from (4.51) and via a similar analysis on another matrix P_2, we can also get

$$c = \sigma_2/2, \tag{4.55}$$

which contradicts (4.54) since $\sigma_1 \neq \sigma_2$. Thus equations (4.50) and (4.51) cannot synchronously hold. Hence from Proposition 4.4 we assert that the assumption "$A \in \mathscr{A}_2$" is not true. □

The above theoretical results can be applied to some specific procedures.

PROCEDURE 4.1. Given a matrix $A \in M_n$, the following procedure identifies whether $A \in \mathscr{A}_n(2)$.

1. If $Card(\sigma(A)) \leqslant 2$, then $A \in \mathscr{A}_n(2)$, goto End;

2. Else if $Card(\sigma_s(A)) > 2$, then $A \notin \mathscr{A}_n(2)$, goto End;

3. Else if $\sigma_o(A) \subset \{0,\ \sigma_{max}(A)\}$, then $A \in \mathscr{A}_n(2)$, goto End;

4. Else if $rank(A) < n$ and there exist $\sigma_1,\ \sigma_2 \in \sigma(A)$ such that $\mu[\sigma_1] = \mu[\sigma_2] = 3$, then $A \notin \mathscr{A}_n(2)$, goto End;

5. Else, apply Proposition 4.4 or other criteria to identify whether $A \in \mathscr{A}_n(2)$.

6. End.

According to Procedure 4.1, we give some concrete examples as follows.

EXAMPLE 4.1. According to Corollary 4.5,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathscr{A}_4(2), \quad B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathscr{A}_4(2).$$

However, from the conclusion (i) of Proposition 4.4 we see that

$$A + B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \notin \mathscr{A}_4(2)$$

since $Card(\sigma_s(A+B)) = 4 > 2$.

EXAMPLE 4.2. Corollary 4.6 or the conclusion (i) of Proposition 4.4 tells us that

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \notin \mathscr{A}_3(2), \quad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \notin \mathscr{A}_3(2),$$

but Corollary 4.5 and Corollary 4.8 indicate that

$$C + D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 6 \end{pmatrix} \in \mathscr{A}_3(2), \quad F = \begin{pmatrix} I_2 & 0 & 0 \\ 0 & 2I_2 & 0 \\ 0 & 0 & 3I_2 \end{pmatrix} \in \mathscr{A}_6(2).$$

EXAMPLE 4.3. According to Corollary 4.5 and Corollary 4.9, we see that

$$G = \begin{pmatrix} I_3 & 0 \\ 0 & 2I_3 \end{pmatrix} \in \mathscr{A}_6(2), \quad \textit{but} \quad T = \begin{pmatrix} I_3 & 0 & 0 \\ 0 & 2I_3 & 0 \\ 0 & 0 & 0 \end{pmatrix} \notin \mathscr{A}_7(2).$$

All of above examples reveal the complicacy of $\mathscr{A}_n(2)$. Example 4.1 specially shows that generally the set $\mathscr{A}_n(2)$ $(n \geqslant 3)$ is not closed with respect to matrix addition. These examples also demonstrate that $\mathscr{A}_n(2) \subsetneq M_n$, i.e., there exist matrices in M_n but not in $\mathscr{A}_n(2)$ (for $n \geqslant 3$).

Based on Theorem 1.2 and the results discussed above, we conclude the following assertion.

PROPOSITION 4.10. *Let n be the order index of space M_n, and k_{min} the smallest number of which every matrix in M_n can be expressed as a linear combination of orthogonal matrices. If $n \geqslant 3$, then $2 < k_{min} \leqslant 4$.*

5. Conclusion

In this paper, via some elementary analysis we obtain an equivalent condition of $A \in \mathscr{A}_n(2)$ (Proposition 4.4) and some sufficient or necessary conditions (Corollaries 4.5, 4.6, 4.8 and 4.9), which can be used as specific criteria to judge whether $A \in \mathscr{A}_n(2)$. Based on these results, we reveal that cone $\mathscr{A}_n(2)$ is not closed with respect to matrix addition, and demonstrate the existence of matrices that are not in $\mathscr{A}_n(2)$, which indicates that $k_{min} > 2$ (for $n \geqslant 3$).

We have to realize that the work in this article is only a preliminary research on set $\mathscr{A}_n(2)$. The criteria in step 5 of Procedure 4.1 need to be further improved. More exact and comprehensive characterizations of $\mathscr{A}_n(2)$ are expected.

The last outstanding question "Whether $k_{min} = 3$ (for $n \geqslant 4$)?" seems more challenging and need further study. Some results of this article may be helpful for future research.

Acknowledgements. The authors would like to pay tribute to Professor Zhan, Professor Li and Professor Poon for their pioneering work in literature [7, 4] and enlightening guidance, which greatly inspired the idea of this article. The authors are grateful to the anonymous referees for their careful review, constructive comments and detailed suggestions, which effectively improved the quality of the manuscript. The second author Tie Zhang was supported by the State Key Laboratory of Synthetical Automation for Process Industries Fundamental Research Funds, China (No. 2013ZCX02), and the third author Chang-Jun Li was supported by the National Natural Science Foundation of China (Nos. 61575090, 61775169).

REFERENCES

[1] G. H. GOLUB AND C. F. VAN LOAN, *Matrix computations*, Posts & Telecom Press, 3rd edition, Beijing, 2011.

[2] R. A. HORN AND C. R. JOHNSON, *Topics in matrix analysis*, Posts & Telecom Press, Beijing, 2011.

[3] R. A. HORN AND C. R. JOHNSON, *Matrix analysis*, Posts & Telecom Press, 2nd edition, Beijing, 2015.

[4] C.-K. LI AND E. POON, *Additive decomposition of real matrices*, Linear and Multilinear Algebra **50**, 4 (2002), 321–326.

[5] R. T. ROCKAFELLAR, *Convex analysis*, Princeton University Press, Princeton, New Jersey, 1997.

[6] P. Y. WU, *Additive combinations of special operators*, Banach Center Publications **30**, 1 (1994), 337–361.
[7] X. ZHAN, *Span of the orthogonal orbit of real matrices*, Linear and Multilinear Algebra **49**, 4 (2001), 337–346.
[8] X. ZHAN, *Matrix theory*, (in Chinese), Higher Education Press, Beijing, 2008.
[9] X. ZHAN, *Matrix theory*, American Mathematical Society, Graduate Studies in Mathematics, vol. 147, Providence, Rhode Island, 2013.

(Received January 11, 2022)

Zheng Li
Department of Mathematics
Northeastern University
Shenyang 110004, P. R. China
e-mail: neu_lizheng@hotmail.com

Tie Zhang
Department of Mathematics
Northeastern University
Shenyang 110004, P. R. China
e-mail: ztmath@163.com

Chang-Jun Li
School of Computer and Software Engineering
University of Science and Technology Liaoning
Anshan, 114051, P. R. China
e-mail: cjliustl@sina.com