Minimum upper bound of the Cayley transform of an orthogonal matrix multiplied by signed permutation matrices

Takuma Mitamura[a], Akira Tanaka[b]

[a]Research Institute of Systems Planning Inc, Nihonkaikan, 18-6, Sakuragaoka, Shibuya, 150-0031, Tokyo, Japan
[b]Faculty of Information Science and Technology, Hokkaido University, N14W9, Kita-ku, Sapporo, 060-0814, Hokkaido, Japan

Abstract

O'Dorney has proved that for any orthogonal matrix U, there exists a signature matrix D (a diagonal matrix with diagonal entries ± 1) by which $|s_{ij}| \leq 1$, where $S = (s_{ij}) = Q(UD)$ is the Cayley transform of UD. In this paper, we prove that the upper bound can be reduced to $\sqrt{2} - 1$ by multiplying by a certain signed permutation matrix, instead of a signature matrix.

Keywords: skew-symmetric matrix, Cayley transform, orthogonal matrix, signed permutation matrix
2000 MSC: 15A06, 20G05

1. Introduction

Throughout this paper, we consider $n \times n$ real matrices with an arbitrarily fixed $n \geq 2$. The Cayley transform of a matrix A, satisfying $\det(I + A) \neq 0$ (in other words, A does not have eigenvalues -1) is defined as

$$Q(A) = (I - A)(I + A)^{-1}, \tag{1}$$

where I denotes the identity matrix. Let $\mathrm{Skew}(n)$, $\mathrm{O}(n)$, and $\mathrm{T}(n)$ denote the sets of skew-symmetric matrices (i.e. $A^T = -A$), orthogonal matrices (i.e. $A^T A = AA^T = I$), and square matrices without eigenvalues -1, respectively. The following are fundamental properties of the Cayley transform:

- $\forall A \in \mathrm{T}(n),\ Q(Q(A)) = A$,

- $\forall S \in \mathrm{Skew}(n),\ Q(S) \in \mathrm{O}(n) \cap \mathrm{T}(n)$,
- $\forall U \in \mathrm{O}(n) \cap \mathrm{T}(n),\ Q(U) \in \mathrm{Skew}(n)$.

Note that $\mathrm{Skew}(n) \subset \mathrm{T}(n)$, since eigenvalues of skew-symmetric matrices are purely imaginary numbers. Also note that $\mathrm{O}(n) \cap \mathrm{T}(n)$ is a proper subset of $\mathrm{SO}(n)$, the set of special orthogonal matrices (i.e. $U \in \mathrm{O}(n)$ satisfying $\det(U) = 1$), since a special orthogonal matrix can have an even number of eigenvalues -1, and then $\mathrm{SO}(n) \not\subset \mathrm{T}(n)$. Obviously, when an orthogonal matrix $U \in \mathrm{O}(n)$ has eigenvalue(s) -1, there does not exist a skew-symmetric matrix $S \in \mathrm{Skew}(n)$ satisfying $S = Q(U)$ since the Cayley transform is not defined for such an orthogonal matrix.

Let $\mathrm{S}(n)$ denote the set of signature matrices, defined as $\mathrm{S}(n) = \{D \mid D = (\delta_{ij} d_i),\ d_i \in \{-1, 1\}\}$, where δ_{ij} stands for the Kronecker delta. Kahan [1] has proved that for any matrix A, there exists a signature matrix $D \in \mathrm{S}(n)$ such that $\det(I + AD) \neq 0$, which immediately leads the following theorem:

Theorem 1 ([1]).

$$\forall U \in \mathrm{O}(n),\ \exists D \in \mathrm{S}(n), S \in \mathrm{Skew}(n) \quad \text{s.t.} \quad S = Q(UD). \tag{2}$$

This theorem connects $\mathrm{Skew}(n)$ and the whole $\mathrm{O}(n)$, instead of its subset $\mathrm{O}(n) \cap \mathrm{T}(n)$. Kahan [1] also discussed the magnitude of entries of $S = Q(UD)$, and raised a problem, "How small can we make $|s_{ij}|$ with $S = (s_{ij}) = Q(UD)$?" O'Dorney [2] has resolved this open problem by the following theorem:

Theorem 2 ([2]).

$$\begin{aligned} &\forall U \in \mathrm{O}(n),\ \exists D \in \mathrm{S}(n), S \in \mathrm{Skew}(n) \\ &\qquad \text{s.t.} \quad S = Q(UD) \quad \text{with} \quad |s_{ij}| \leq 1. \end{aligned} \tag{3}$$

As is well known, multiplying an orthogonal matrix $U \in \mathrm{O}(n)$ by a signature matrix $D \in \mathrm{S}(n)$ may flip the sign of the determinant (i.e. $\det(UD) = -\det(U)$ with a certain $D \in \mathrm{S}(n)$). Similarly, a permutation matrix may also flip the sign of the determinant (i.e. $\det(UP) = -\det(U)$ with a certain

$P \in \mathrm{P}(n)$, where $\mathrm{P}(n)$ denotes the set of permutation matrices). On the basis of this similarity, we discussed the possibility to eliminate eigenvalues -1 from an orthogonal matrix $U \in \mathrm{O}(n)$ by multiplying by a permutation matrix $P \in \mathrm{P}(n)$, and proved the following theorem [3]:

Theorem 3 ([3]).

$$\forall U \in \mathrm{O}(n),\ \exists P \in (\mathrm{P}(n) \cup -\mathrm{P}(n)), S \in \mathrm{Skew}(n) \quad \text{s.t.} \quad S = Q(UP), \tag{4}$$

where $-\mathrm{P}(n) = \{-P \mid P \in \mathrm{P}(n)\}$.

Let $\mathrm{SP}(n)$ denote the set of signed permutation matrices, defined as $\mathrm{SP}(n) = \{PD \mid P \in \mathrm{P}(n), D \in \mathrm{S}(n)\}$, which trivially includes $\mathrm{S}(n)$, $\mathrm{P}(n)$ and $-\mathrm{P}(n)$. Then, the following corollary follows immediately from either Theorem 1 or Theorem 3, which is pointed out in [3].

Corollary 1 ([3]).

$$\forall U \in \mathrm{O}(n),\ \exists M \in \mathrm{SP}(n), S \in \mathrm{Skew}(n) \quad \text{s.t.} \quad S = Q(UM). \tag{5}$$

Here, a natural question will arise, that is, "Can the upper bound in Theorem 2 be reduced or not by replacing $\mathrm{S}(n)$ with $\mathrm{SP}(n)$?" This problem is now positively resolved in this paper by the following theorem, which was a conjecture obtained through numerical experiments in our previous work [3].

Theorem 4.

$$\begin{aligned} &\forall U \in \mathrm{O}(n),\ \exists M \in \mathrm{SP}(n), S \in \mathrm{Skew}(n) \\ &\text{s.t.} \quad S = Q(UM) \quad \text{with} \quad |s_{ij}| \le \sqrt{2} - 1. \end{aligned} \tag{6}$$

2. Preliminaries

Let $J \subset \{1, \ldots, n\}$ be a certain index set and let J^C be its complement defined as $J^C = \{1, \ldots, n\} \setminus J$. Firstly, we define some matrices as follows:

$$
\begin{aligned}
I^J &= \textstyle\sum_{k \in J^C} \boldsymbol{e}_k \boldsymbol{e}_k^T, \\
F^{(i,j)} &= \boldsymbol{e}_i \boldsymbol{e}_j^T - \boldsymbol{e}_j \boldsymbol{e}_i^T, \\
C^{(i,j)} &= I^{\{i,j\}} + F^{(i,j)},
\end{aligned}
$$

where $\boldsymbol{e}_i$ stands for the n-dimensional unit vector whose i-th entry is unity. Note that $C^{(i,j)} \in \mathrm{SP}(n)$.

Let $U \in \mathrm{O}(n) \cap \mathrm{T}(n)$ be an arbitrary orthogonal matrix without eigenvalues -1. The Cayley transform Eq.(1) of U can be also represented as

$$
\begin{aligned}
S &= (I - U)(I + U)^{-1} \\
&= (2I - (I + U))(I + U)^{-1} \\
&= 2(I + U)^{-1} - I \in \mathrm{Skew}(n).
\end{aligned} \tag{7}
$$

Let $B = (b_{ij}) = (I+U)^{-1}$. Then $B = \frac{1}{2}(S+I)$. Thus, we have $b_{ij} = -b_{ji}$, $s_{ij} = 2b_{ij}$ with $i \neq j$, and $b_{ii} = 1/2$. Also, $(\det(I + U))B = \widetilde{(I + U)}$, where $\widetilde{(I + U)}$ stands for the adjugate matrix (transposed cofactor matrix) of $I+U$. Note that the cofactor of the (j, i)-entry of $I + U$, denoted as $\mathrm{Cof}(I + U)_{ji}$, is represented as

$$
\begin{aligned}
\mathrm{Cof}(I + U)_{ji} &= (-1)^{j+i} \mathrm{Minor}(I + U)_{ji} \\
&= \det(I^{\{j\}}(I + U)I^{\{i\}} + \boldsymbol{e}_j \boldsymbol{e}_i^T) \\
&= \det((I + U)I^{\{i\}} + \boldsymbol{e}_j \boldsymbol{e}_i^T),
\end{aligned}
$$

where $\mathrm{Minor}(I + U)_{ji}$ stands for the first minor of the (j, i)-entry of $I + U$. Thus, we have

$$
\det(I + U)b_{ij} = \det((I + U)I^{\{i\}} + \boldsymbol{e}_j \boldsymbol{e}_i^T). \tag{8}
$$

For an $n \times n$ square matrix A, A_{J^C} denotes the matrix which is obtained by eliminating the k-th column and the k-th row from A for each $k \in J$. Hereafter, we use the notation $U + I$ instead of $I + U$ for readability.

Lemma 1. *Let $U \in \mathrm{O}(n) \cap \mathrm{T}(n)$ and let $S = (s_{ij}) = Q(U)$. Then for $i \neq j$,*

$$s_{ij}^2 + 1 = 4\frac{\det((U+I)_{\{i,j\}^C})}{\det(U+I)}. \tag{9}$$

Proof Due to the multi-linearity of determinants,

$$\begin{aligned}
&\det(U + I + F^{(i,j)}) \\
&= \det(U+I) + \det((U+I)I^{\{i\}} - \boldsymbol{e}_j\boldsymbol{e}_i^T) + \det((U+I)I^{\{j\}} + \boldsymbol{e}_i\boldsymbol{e}_j^T) \\
&\quad + \det((U+I)I^{\{i,j\}} + F^{(i,j)}) \\
&= \det(U+I) - \det((U+I)I^{\{i\}} + \boldsymbol{e}_j\boldsymbol{e}_i^T) + \det((U+I)I^{\{j\}} + \boldsymbol{e}_i\boldsymbol{e}_j^T) \\
&\quad - \det((U+I)I^{\{i,j\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T - \boldsymbol{e}_j\boldsymbol{e}_j^T) \\
&= \det(U+I) - \det((U+I)I^{\{i\}} + \boldsymbol{e}_j\boldsymbol{e}_i^T) + \det((U+I)I^{\{j\}} + \boldsymbol{e}_i\boldsymbol{e}_j^T) \\
&\quad + \det((U+I)I^{\{i,j\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T + \boldsymbol{e}_j\boldsymbol{e}_j^T) \\
&= \det(U+I) - \det(U+I)b_{ij} + \det(U+I)b_{ji} + \det((U+I)_{\{i,j\}^C})
\end{aligned}$$

which yields

$$\frac{\det(U+I+F^{(i,j)})}{\det(U+I)} = 1 + (b_{ji} - b_{ij}) + \frac{\det((U+I)_{\{i,j\}^C})}{\det(U+I)}. \tag{10}$$

On the other hand,

$$\begin{aligned}
\frac{\det(U+I+F^{(i,j)})}{\det(U+I)} &= \det((U+I+F^{(i,j)})(U+I)^{-1}) \\
&= \det(I + F^{(i,j)}(U+I)^{-1}) \\
&= \det(I + F^{(i,j)}B) \\
&= (b_{ji}+1)(-b_{ij}+1) + b_{ii}b_{jj} \\
&= b_{ii}b_{jj} - b_{ji}b_{ij} + b_{ji} - b_{ij} + 1.
\end{aligned} \tag{11}$$

From Eqs.(10) and (11), we have

$$\frac{\det((U+I)_{\{i,j\}^C})}{\det(U+I)} = b_{ii}b_{jj} - b_{ji}b_{ij} = \frac{1}{4} + \frac{1}{4}s_{ij}^2, \tag{12}$$

which implies

$$s_{ij}^2 + 1 = 4\frac{\det((U+I)_{\{i,j\}^C})}{\det(U+I)}. \tag{13}$$

□

Lemma 2. *Let $U \in \mathrm{O}(n) \cap \mathrm{T}(n)$. Then for $i \neq j$,*

$$\det(U + I^{\{i,j\}}) = \det((U+I)_{\{i,j\}^C}). \tag{14}$$

Proof Due to the multi-linearity of determinants, we have

$$\begin{aligned}
&\det(U+I)\\
&= \det(U+I^{\{i,j\}}) + \det((U+I^{\{j\}})I^{\{i\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T)\\
&\quad + \det((U+I^{\{i\}})I^{\{j\}} + \boldsymbol{e}_j\boldsymbol{e}_j^T) + \det((U+I)I^{\{i,j\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T + \boldsymbol{e}_j\boldsymbol{e}_j^T)\\
&= \det(U+I^{\{i,j\}}) + \det((U+I^{\{j\}})I^{\{i\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T)\\
&\quad + \det((U+I^{\{i\}})I^{\{j\}} + \boldsymbol{e}_j\boldsymbol{e}_j^T) + \det((U+I)_{\{i,j\}^C}).
\end{aligned} \tag{15}$$

Since the cofactor of the (i,i)-entry of $(U+I)$ is

$$(\widetilde{U+I})_{ii} = \det(U+I)b_{ii} = \frac{\det(U+I)}{2}, \tag{16}$$

for any $i \in \{1,\ldots,n\}$, and

$$\begin{aligned}
(\widetilde{U+I})_{ii} &= \det((U+I)I^{\{i\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T)\\
&= \det((U+I)I^{\{i,j\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T + \boldsymbol{e}_j\boldsymbol{e}_j^T)\\
&\quad + \det((U+I^{\{j\}})I^{\{i\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T)\\
&= \det((U+I)_{\{i,j\}^C}) + \det((U+I^{\{j\}})I^{\{i\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T)
\end{aligned} \tag{17}$$

for any $i \in \{1,\ldots,n\}$, we have

$$\begin{aligned}
\det(U+I) &= (\widetilde{U+I})_{ii} + (\widetilde{U+I})_{jj}\\
&= 2\det((U+I)_{\{i,j\}^C}) + \det((U+I^{\{j\}})I^{\{i\}} + \boldsymbol{e}_i\boldsymbol{e}_i^T)\\
&\quad + \det((U+I^{\{i\}})I^{\{j\}} + \boldsymbol{e}_j\boldsymbol{e}_j^T).
\end{aligned} \tag{18}$$

Therefore, from Eqs.(15) and (18), Eq.(14) immediately follows. □

Lemma 3.

$$\forall U \in \mathrm{O}(n) \cap \mathrm{T}(n), \quad \det(U + I) > 0. \tag{19}$$

Proof An eigenvalue of U, denoted as λ, satisfies $|\lambda| = 1$, and its complex conjugate $\overline{\lambda}$ is also an eigenvalue of U. Thus, without loss of generality, we can assume that U has n_p eigenvalues 1 and complex (non-real) eigenvalues $\lambda_1, \ldots, \lambda_{n_c}, \overline{\lambda_1}, \ldots, \overline{\lambda_{n_c}}$, with $n_p + 2n_c = n,\ n_p \geq 0,\ n_c \geq 0$. Then, it is trivial that

$$\begin{aligned}\det(U + I) &= (1+1)^{n_p} \Pi_{k=1}^{n_c} (1 + \lambda_k)(1 + \overline{\lambda_k}) \\ &= 2^{n_p} \Pi_{k=1}^{n_c} (1 + 2\mathrm{Re}(\lambda_k) + |\lambda_k|^2) \\ &= 2^{n_p + n_c} \Pi_{k=1}^{n_c} (1 + \mathrm{Re}(\lambda_k)) > 0, \end{aligned} \tag{20}$$

since $-1 < \mathrm{Re}(\lambda_k) < 1$. □

3. Proof of Theorem 4

Firstly, we give the following theorem, in which we identify the signed permutation matrix that attains the minimum upper bound in Theorem 4.

Theorem 5. *Let $U \in \mathrm{O}(n)$ be an arbitrary orthogonal matrix and let*

$$M_0 = \mathrm{argmax}_{M \in \mathrm{SP}(n)} \det(UM + I). \tag{21}$$

Then $S = (s_{ij}) = Q(UM_0)$ is defined, and

$$|s_{ij}| \leq \sqrt{2} - 1. \tag{22}$$

Proof Corollary 1 guarantees that there exists $M \in \mathrm{SP}(n)$ such that $\det(UM + I) \neq 0$, and then $\det(UM + I) > 0$ by Lemma 3. Therefore, we conclude that $\det(UM_0 + I) \neq 0$, which implies that the Cayley transform $S = Q(UM_0)$ is defined.

Let $U_0 = UM_0$. Then

$$\det(U_0 + I) \geq \det(U_0 M + I) \tag{23}$$

for any $M \in \mathrm{SP}(n)$ by assumption. Since $C^{(i,j)} \in \mathrm{SP}(n)$, $\det(C^{(i,j)}) = 1$, and $(C^{(i,j)})^T = C^{(j,i)}$,

$$\begin{aligned}
\det(U_0 + I) &\geq \det(U_0 C^{(i,j)} + I) \\
&= \det(U_0 + C^{(j,i)}) \\
&= \det(U_0 + I^{\{i,j\}}) + \det((U_0 + I)I^{\{i,j\}} - \boldsymbol{e}_i \boldsymbol{e}_j^T + \boldsymbol{e}_j \boldsymbol{e}_i^T) \\
&\quad + \det((U_0 + I)I^{\{i\}} + \boldsymbol{e}_j \boldsymbol{e}_i^T) + \det((U_0 + I)I^{\{j\}} - \boldsymbol{e}_i \boldsymbol{e}_j^T) \\
&= \det(U_0 + I^{\{i,j\}}) + \det((U_0 + I)_{\{i,j\}^C}) \\
&\quad + \det(U_0 + I)b_{ij} - \det(U_0 + I)b_{ji} \\
&= 2\det((U_0 + I)_{\{i,j\}^C}) + 2\det(U_0 + I)b_{ij} \quad \text{(by Lemma 2)}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\det(U_0 + I) &\geq \det(U_0 C^{(j,i)} + I) \\
&= \det(U_0 + C^{(i,j)}) \\
&= 2\det((U_0 + I)_{\{i,j\}^C}) - 2\det(U_0 + I)b_{ij}.
\end{aligned}$$

Thus, we conclude that

$$\begin{aligned}
\det(U_0 + I) &\geq 2\det((U_0 + I)_{\{i,j\}^C}) + 2\det(U_0 + I)|b_{ij}| \\
&= 2\det((U_0 + I)_{\{i,j\}^C}) + \det(U_0 + I)|s_{ij}|, \qquad (24)
\end{aligned}$$

which implies

$$1 \geq 2\frac{\det((U_0 + I)_{\{i,j\}^C})}{\det(U_0 + I)} + |s_{ij}|, \qquad (25)$$

and then

$$2(1 - |s_{ij}|) \geq 4\frac{\det((U_0 + I)_{\{i,j\}^C})}{\det(U_0 + I)}. \qquad (26)$$

Thus, from Lemma 1 and Eq.(26),

$$\begin{aligned}
|s_{ij}|^2 + 1 &\leq 2(1 - |s_{ij}|) \\
(|s_{ij}| + 1)^2 &\leq 2 \\
|s_{ij}| &\leq \sqrt{2} - 1,
\end{aligned}$$

which concludes the proof. □

Accordingly, Theorem 4 has been proven since Theorem 4 is a trivial corollary of Theorem 5.

4. A remark on the tightness of the upper bound

In [3], we gave a proof of Theorem 4 only for $n = 2$, together with the fact that there exists $U \in \mathrm{O}(2)$ by which

$$\min_{M \in \mathrm{SP}(2)} \max_{i,j} |s_{ij}^{(M)}| = \sqrt{2} - 1, \tag{27}$$

where $S^{(M)} = (s_{ij}^{(M)}) = Q(UM)$. This implies that the upper bound $\sqrt{2} - 1$ is the tightest one. The matrix

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \in \mathrm{O}(2)$$

is such an example. The following is an excerpt from the proof in [3] on this issue.

SP(2) consists of the following 8 matrices:

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ M_2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \ M_3 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \ M_4 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

$$M_5 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \ M_6 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \ M_7 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ M_8 = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Since $\det(UM_k + I) \neq 0$ for $k \in \{1, 2, 3, 4\}$, $Q(UM_k)$ is defined for $k \in \{1, 2, 3, 4\}$, while $Q(UM_k)$ is not defined for $k \in \{5, 6, 7, 8\}$ due to $\det(UM_k + I) = 0$ for $k \in \{5, 6, 7, 8\}$. The Cayley transforms of UM_k with $k \in \{1, 2, 3, 4\}$ are given as follows:

$$S_1 = (s_{ij}^{(1)}) = Q(UM_1) = \begin{bmatrix} 0 & \sqrt{2} - 1 \\ -(\sqrt{2} - 1) & 0 \end{bmatrix},$$

$$S_2 = (s_{ij}^{(2)}) = Q(UM_2) = \begin{bmatrix} 0 & -(\sqrt{2} + 1) \\ \sqrt{2} + 1 & 0 \end{bmatrix},$$

$$S_3 = (s_{ij}^{(3)}) = Q(UM_3) = \begin{bmatrix} 0 & \sqrt{2} + 1 \\ -(\sqrt{2} + 1) & 0 \end{bmatrix},$$

$$S_4 = (s_{ij}^{(4)}) = Q(UM_4) = \begin{bmatrix} 0 & -(\sqrt{2} - 1) \\ \sqrt{2} - 1 & 0 \end{bmatrix}.$$

Therefore,

$$\min_{k \in \{1,2,3,4\}} \max_{i,j} |s_{ij}^{(k)}| = \sqrt{2} - 1$$

for this U.

Declaration of competing interest

We have no conflict of interest to declare.

Data availability

No data was used for the research described in the article.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments. This work was partially supported by JSPS KAKENHI Grant Number JP20H04238.

References

[1] W. Kahan, Is there a small skew Cayley transform with zero diagonal?, Linear Algebra and its Applications 417 (2–3) (2006) 335–341.

[2] E. O'Dorney, Minimizing the Cayley transform of an orthogonal matrix by multiplying by signature matrices, Linear Algebra and its Applications 448 (2014) 97–103.

[3] T. Mitamura, A. Tanaka, Properties of the set of orthogonal matrices obtained as the Cayley transform of skew-symmetric matrices, IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences J105-A (11), japanese only; DOI: 10.14923/transfunj.2022JAP1007 (2022).