Remarks on the Cayley Representation of Orthogonal Matrices and on Perturbing the Diagonal of a Matrix to Make it Invertible

Jean Gallier

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
jean@cis.upenn.edu

June 20, 2014

Abstract. This note contains two remarks. The first remark concerns the extension of the well-known Cayley representation of rotation matrices by skew-symmetric matrices to rotation matrices admitting -1 as an eigenvalue and then to all orthogonal matrices. We review a method due to Hermann Weyl and another method involving multiplication by a diagonal matrix whose entries are $+1$ or -1. The second remark has to do with ways of flipping the signs of the entries of a diagonal matrix, C, with nonzero diagonal entries, obtaining a new matrix, E, so that $E+A$ is invertible, where A is any given matrix (invertible or not).

1 The Cayley Representation of Orthogonal Matrices

Given any rotation matrix, $R \in \mathbf{SO}(n)$, if R does not admit -1 as an eigenvalue, then there is a unique skew-symmetric matrix, S, ($S^\top = -S$) so that

$$R = (I - S)(I + S)^{-1}.$$

This is a classical result of Cayley [3] (1846) and R is called the *Cayley transform of S*. Among other sources, a proof can be found in Hermann Weyl's beautiful book *The Classical Groups* [7], Chapter II, Section 10, Theorem 2.10.B (page 57).

As we can see, this representation misses rotation matrices admitting the eigenvalue -1, and of course, as $\det((I - S)(I + S)^{-1}) = +1$, it misses improper orthogonal matrices, i.e., those matrices $R \in \mathbf{O}(n)$ with $\det(R) = -1$.

Question 1. Is there a way to extend the Cayley representation to all rotation matrices (matrices in $\mathbf{SO}(n)$)?

Question 2. Is there a way to extend the Cayley representation to all orthogonal matrices (matrices in $\mathbf{O}(n)$)?

Answer: Yes in both cases!

An answer to Question 1 is given in Weyl's book [7], Chapter II, Section 10, Lemma 2.10.D (page 60):

Proposition 1.1 *(Weyl) Every rotation matrix, $R \in \mathbf{SO}(n)$, can be expressed as a product*

$$R = (I - S_1)(I + S_1)^{-1}(I - S_2)(I + S_2)^{-1},$$

where S_1 and S_2 are skew-symmetric matrices.

Thus, if we allow two Cayley representation matrices, we can capture orthogonal matrices having an even number of -1 as eigenvalues. Actually, proposition 1.1 can be sharpened slightly as follows:

Proposition 1.2 *Every rotation matrix, $R \in \mathbf{SO}(n)$, can be expressed as*

$$R = \Big((I - S)(I + S)^{-1}\Big)^2$$

where S is a skew-symmetric matrix.

Proposition 1.2 can be easily proved using the following well-known normal form for orthogonal matrices:

Proposition 1.3 *For every orthogonal matrix, $R \in \mathbf{O}(n)$, there is an orthogonal matrix P and a block diagonal matrix D such that $R = P D P^\top$, where D is of the form*

$$D = \begin{pmatrix} D_1 & & \ldots & \\ & D_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & D_p \end{pmatrix}$$

such that each block D_i is either 1, -1, or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

where $0 < \theta_i < \pi$.

In particular, if R is a rotation matrix ($R \in \mathbf{SO}(n)$), then it has an even number of eigenvalues -1. So, they can be grouped into two-dimensional rotation matrices of the form

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

i.e., we allow $\theta_i = \pi$ and we may assume that D does not contain one-dimensional blocks of the form -1.

A proof of Proposition 1.3 can be found in Gantmacher [5], Chapter IX, Section 13 (page 285), or Berger [2], or Gallier [4], Chapter 11, Section 11.4 (Theorem 11.4.5).

Now, for every two-dimensional rotation matrix

$$T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with $0 < \theta \leq \pi$, observe that

$$T^{\frac{1}{2}} = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}$$

does not admit -1 as an eigenvalue (since $0 < \theta/2 \leq \pi/2$) and $T = \left(T^{\frac{1}{2}}\right)^2$. Thus, if we form the matrix $R^{\frac{1}{2}}$ by replacing each two-dimensional block D_i in the above normal form by $D_i^{\frac{1}{2}}$, we obtain a rotation matrix that does not admit -1 as an eigenvalue, $R = \left(R^{\frac{1}{2}}\right)^2$ and the Cayley transform of $R^{\frac{1}{2}}$ is well defined. Therefore, we have proved Proposition 1.2. $\square$

Next, why is the answer to Question 2 also yes?

This is because

Proposition 1.4 *For any orthogonal matrix, $R \in \mathbf{O}(n)$, there is some diagonal matrix, E, whose entries are $+1$ or -1, and some skew-symmetric matrix, S, so that*

$$R = E(I - S)(I + S)^{-1}.$$

As such matrices E are orthogonal, all matrices $E(I-S)(I+S)^{-1}$ are orthogonal, so we have a Cayley-style representation of all orthogonal matrices.

I am not sure when Proposition 1.4 was discovered and originally published. Since I could not locate this result in Weyl's book [7], I assume that it was not known before 1946, but I did stumble on it as an exercise in Richard Bellman's classic [1], first published in 1960, Chapter 6, Section 4, Exercise 11, page 91-92 (see also, Exercises, 7, 8, 9, and 10).

Why does this work?

Fact E: Because, for every $n \times n$ matrix, A (invertible or not), there some diagonal matrix, E, whose entries are $+1$ or -1, so that $I + EA$ is invertible!

This is Exercise 10 in Bellman [1] (Chapter 6, Section 4, page 91). Using Fact E, it is easy to prove Proposition 1.4.

Proof of Proposition 1.4. Let $R \in \mathbf{O}(n)$ be any orthogonal matrix. By Fact E, we can find a diagonal matrix, E (with diagonal entries ± 1), so that $I + ER$ is invertible. But then, as E is orthogonal, ER is an orthogonal matrix that does not admit the eigenvalue -1 and so, by the Cayley representation theorem, there is a skew-symmetric matrix, S, so that

$$ER = (I - S)(I + S)^{-1}.$$

However, notice that $E^2 = I$, so we get

$$R = E(I - S)(I + S)^{-1},$$

as claimed. $\square$

But Why does Fact E hold?

As we just observed, $E^2 = I$, so by multiplying by E,

$$I + EA \text{ is invertible iff } E + A \text{ is.}$$

Thus, we are naturally led to the following problem: If A is any $n \times n$ matrix, is there a way to perturb the diagonal entries of A, i.e., to add some diagonal matrix, $C = \mathrm{diag}(c_1, \ldots, c_n)$, to A so that $C + A$ becomes invertible?

Indeed this can be done, and we will show in the next section that what matters is not the magnitude of the perturbation but the signs of the entries being added.

2 Perturbing the Diagonal of a Matrix to Make it Invertible

In this section we prove the following result:

Proposition 2.1 *For every $n \times n$ matrix (invertible or not), A, and every any diagonal matrix, $C = \mathrm{diag}(c_1, \ldots, c_n)$, with $c_i \neq 0$ for $i = 1, \ldots, n$, there an assignment of signs, $\epsilon_i = \pm 1$, so that if $E = \mathrm{diag}(\epsilon_1 c_1, \ldots, \epsilon_n c_n)$, then $E + A$ is invertible.*

Proof. Let us evaluate the determinant of $C+A$. We see that $\Delta = \det(C+A)$ is a polynomial of degree n in the variables $c_1, \ldots, c_n$ and that all the monomials of Δ consist of products of distinct variables (i.e., every variable occurring in a monomial has degree 1). In particular, Δ contains the monomial $c_1 \cdots c_n$. In order to prove Proposition 2.1, it will suffice to prove

Proposition 2.2 *Given any polyomial, $P(x_1, \ldots, x_n)$, of degree n (in the indeterminates $x_1, \ldots, x_n$ and over any integral domain of characteristic unequal to 2), if every monomial in P is a product of distinct variables, for every n-tuple $(c_1, \ldots, c_n)$ such that $c_i \neq 0$ for $i = 1, \ldots, n$, then there is an assignment of signs, $\epsilon_i = \pm 1$, so that*

$$P(\epsilon_1 c_1, \ldots, \epsilon_n c_n) \neq 0.$$

Clearly, any assignment of signs given by Proposition 2.2 will make $\det(E + A) \neq 0$, proving Proposition 2.1. $\square$

It remains to prove Proposition 2.2.

Proof of Proposition 2.2. We proceed by induction on n (starting with $n = 1$). For $n = 1$, the polynomial $P(x_1)$ is of the form $P(x_1) = a + bx_1$, with $b \neq 0$ since $\deg(P) = 1$. Obviously, for any $c \neq 0$, either $a + bc \neq 0$ or $a - bc \neq 0$ (otherwise, $2bc = 0$, contradicting $b \neq 0$, $c \neq 0$ and the ring being an integral domain of characteristic $\neq 2$).

Assume the induction hypothesis holds for any $n \geq 1$ and let $P(x_1, \ldots, x_{n+1})$ be a polynomial of degree $n + 1$ satisfying the conditions of Proposition 2.2. Then, P must be of the form

$$P(x_1, \ldots, x_n, x_{n+1}) = Q(x_1, \ldots, x_n) + S(x_1, \ldots, x_n)x_{n+1},$$

where both $Q(x_1, \ldots, x_n)$ and $S(x_1, \ldots, x_n)$ are polynomials in $x_1, \ldots, x_n$ and $S(x_1, \ldots, x_n)$ is of degree n and all monomials in S are products of distinct variables. By the induction hypothesis, we can find $(\epsilon_1, \ldots, \epsilon_n)$, with $\epsilon_i = \pm 1$, so that

$$S(\epsilon_1 c_1, \ldots, \epsilon_n c_n) \neq 0.$$

But now, we are back to the case $n = 1$ with the polynomial

$$Q(\epsilon_1 c_1, \ldots, \epsilon_n c_n) + S(\epsilon_1 c_1, \ldots, \epsilon_n c_n)x_{n+1},$$

and we can find $\epsilon_{n+1} = \pm 1$ so that

$$P(\epsilon_1 c_1, \ldots, \epsilon_n c_n, \epsilon_{n+1} c_{n+1}) = Q(\epsilon_1 c_1, \ldots, \epsilon_n c_n) + S(\epsilon_1 c_1, \ldots, \epsilon_n c_n)\epsilon_{n+1} c_{n+1} \neq 0,$$

establishing the induction hypothesis. □

Note that in Proposition 2.1, the c_i can be made arbitrarily small or large, as long as they are not zero. What matters is the signs that are assigned to the perturbation.

Another nice proof of Fact E is given in a short note by William Kahan [6]. Due to its elegance, we feel compelled to sketch Kahan's proof. This proof uses two facts:

(1) If $A = (A_1, \ldots, A_{n-1}, U)$ and $B = (A_1, \ldots, A_{n-1}, V)$ are two $n \times n$ matrices that differ in their last column, then

$$\det(A + B) = 2^{n-1}(\det(A) + \det(B)).$$

This is because determinants are multilinear (alternating) maps of their columns. Therefore, if $\det(A) = \det(B) = 0$, then $\det(A + B) = 0$. Obviously, this fact also holds whenever A and B differ by just one column (not just the last one).

(2) For every $k = 0, \ldots 2^n - 1$, if we write k in binary as $k = k_n \cdots k_1$, then let E_k be the diagonal matrix whose ith diagonal entry is -1 iff $k_i = 1$, else $+1$ iff $k_i = 0$. For example, $E_0 = I$ and $E_{2^n-1} = -I$. Observe that E_k and E_{k+1} differ by exactly one column. Then, it is easy to see that

$$E_0 + E_1 + \cdots + E_{2^n-1} = 0.$$

The proof proceeds by contradiction. Assume that $\det(I + E_k A) = 0$, for $k = 0, \ldots, 2^n - 1$. The crux of the proof is that

$$\det(I + E_0 A + I + E_1 A + I + E_2 A + \cdots + I + E_{2^n-1} A) = 0.$$

However, as $E_0 + E_1 + \cdots + E_{2^n-1} = 0$, we see that

$$I + E_0 A + I + E_1 A + I + E_2 A + \cdots + I + E_{2^n-1} A = 2^n I,$$

and so,

$$0 = \det(I + E_0 A + I + E_1 A + I + E_2 A + \cdots + I + E_{2^n-1} A) = \det(2^n I) = 2^n \neq 0,$$

a contradiction!

To prove that $\det(I + E_0 A + I + E_1 A + I + E_2 A + \cdots + I + E_{2^n-1} A) = 0$, we observe using fact (2) that,

$$\det(I + E_{2i} A + I + E_{2i+1} A) = \det(I + E_{2i} A) + \det(I + E_{2i+1} A) = 0,$$

for $i = 0, \ldots, 2^{n-1} - 1$; similarly,

$$\det(I + E_{4i}A + I + E_{4i+1}A + I + E_{4i+2}A + I + E_{4i+3}A) = 0,$$

for $i = 0, \ldots, 2^{n-2} - 1$; by induction, we get

$$\det(I + E_0A + I + E_1A + I + E_2A + \cdots + I + E_{2^n-1}A) = 0,$$

which concludes the proof.

Final Questions:

(1) When was Fact E first stated and by whom (similarly for Proposition 1.4)?

(2) Can Proposition 2.1 be generalized to non-diagonal matrices (in an interesting way)?

References

[1] Richard Bellman. *Introduction to Matrix Analysis*. Classics in Applied Mathematics, No. 12. SIAM Publications, first edition, 1995.

[2] Marcel Berger. *Géométrie 1*. Nathan, 1990. English edition: Geometry 1, Universitext, Springer Verlag.

[3] Arthur Cayley. About the algebraic structure of the orthogonal group and the other classical groups in a field of characteristic zero or a prime characteristic. *J. Reine Angew. Math.*, 32, 1846.

[4] Jean H. Gallier. *Geometric Methods and Applications, For Computer Science and Engineering*. TAM, Vol. 38. Springer, first edition, 2000.

[5] F.R. Gantmacher. *The Theory of Matrices, Vol. I*. AMS Chelsea, first edition, 1977.

[6] William Kahan. Is there a small skew Cayley transform with zero diagonal. Technical report, University of California, Berkeley, Mathematics Department, Berkeley, CA 94720-3840, 2004. pdf file available from `http://www.cs.berkeley.edu/~wkahan`.

[7] Hermann Weyl. *The Classical Groups. Their Invariants and Representations*. Princeton Mathematical Series, No. 1. Princeton University Press, second edition, 1946.