Matrices and Quaternions

Gary D. Simpson

gsim100887@aol.com

rev 00 Mar 01, 2014

Summary

A possibly novel mathematical structure is presented. The structure is a matrix whose elements are quaternions. The structure is distinct from a tensor.

Preface

A knowledge of quaternions, tensors, and linear algebra is required.

Discussion

The objective of this text is to elaborate upon a mathematical structure that was first presented by Wolfgang Pauli[1]. That structure was a matrix whose elements were 0, ±**i**, or ±1. These matrices are today referred to as Pauli matrices. They were later applied and further developed by Paul Dirac[2] into matrices that describe electron spin. Hence, those matrices are now referred to as Dirac Spinors. Dirac created those matrices to satisfy the equations ab = -ba and $a^2 = -b^2$. The author will argue that these matrices are actually members of a more general matrix structure. The author believes that this structure is distinct from a tensor.

Before beginning the discussion of the topic, it is worthwhile to briefly describe quaternion multiplication and division.

Quaternion Multiplication:

Consider the details of a generic multiplication **AB** = **C**.

$$\left(a_0 + a_i\mathbf{i} + a_j\mathbf{j} + a_k\mathbf{k}\right)\left(b_0 + b_i\mathbf{i} + b_j\mathbf{j} + b_k\mathbf{k}\right) = c_0 + c_i\mathbf{i} + c_j\mathbf{j} + c_k\mathbf{k}$$

This produces the following four equations for the scalar and individual vector components respectively:

$$scalar\colon a_0b_0 - a_ib_i - a_jb_j - a_kb_k = c_0$$

$$\mathbf{i}\colon a_ib_0 + a_0b_i - a_kb_j + a_jb_k = c_i$$

$$\mathbf{j}\colon a_jb_0 + a_kb_i + a_0b_j - a_ib_k = c_j$$

$$\mathbf{k}\colon a_kb_0 - a_jb_i + a_ib_j + a_0b_k = c_k$$

This can be presented in matrix form as follows:

$$\begin{bmatrix} a_0 & -a_i & -a_j & -a_k \\ a_i & a_0 & -a_k & a_j \\ a_j & a_k & a_0 & -a_i \\ a_k & -a_j & a_i & a_0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_i \\ b_j \\ b_k \end{bmatrix} = \begin{bmatrix} c_0 \\ c_i \\ c_j \\ c_k \end{bmatrix}$$

The matrix form can be used to solve for the coefficients of **B** if **A** and **C** are known. If **A** and **B** are known then the coefficients of **C** can be determined directly.

Quaternion Division:

The method described below for matrix reduction requires the user to perform quaternion division to determine a ratio between two quaternions. It might not be immediately obvious how to perform this task. Fortunately, it is fairly straightforward. The problem is set up as follows:

$$\mathbf{A} = \frac{\mathbf{C}}{\mathbf{B}}$$

$$\mathbf{AB} = \mathbf{C}$$

This produces four simultaneous linear equations, with four unknown coefficients, as described above. This system is then solved to determine the four coefficients of **A** (rather than **B** as was done previously).

General:

Consider an arbitrary matrix [**A**] whose elements are all quaternions. The bold capital letter **A** is used to designate a quaternion and the [brackets] are used to designate a matrix. In principle, the matrix can be of any size. The matrix must be square if it is to be multiplied by another matrix as described above in the first paragraph. The author is primarily interested in matrices that can be used to manipulate quaternions. Therefore, the author will only consider matrices that are 4 x 4.

$$[\mathbf{A}] = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix}$$

$$\mathbf{A}_{m,n} = (a_0)_{m,n} + (a_i)_{m,n}\mathbf{i} + \left(a_j\right)_{m,n}\mathbf{j} + (a_k)_{m,n}\mathbf{k} = (a_0)_{m,n} + (\boldsymbol{a})_{m,n}$$

The variables "m" and "n" are used as indices respectively for the row and column of [**A**].

Next, define matrices [**B**], [**C**], and [**D**] as follows by multiplying arbitrary matrix [**A**] by a unit vector:

$$[\mathbf{B}] = [\mathbf{A}]\mathbf{i}\ ;\ [\mathbf{C}] = [\mathbf{A}]\mathbf{j}\ ;\ [\mathbf{D}] = [\mathbf{A}]\mathbf{k}$$

Next, take the square of each matrix.

$$[\mathbf{B}]^2 = -[\mathbf{A}]^2\ ;\ [\mathbf{C}]^2 = -[\mathbf{A}]^2\ ;\ [\mathbf{D}]^2 = -[\mathbf{A}]^2 : therefore\ [\mathbf{B}]^2 = [\mathbf{C}]^2 = [\mathbf{D}]^2 = -[\mathbf{A}]^2$$

Next, perform the matrix multiplications.

$$[\mathbf{B}][\mathbf{C}] = [\mathbf{A}][\mathbf{A}]\mathbf{ij} = [\mathbf{A}]^2\mathbf{k};\ [\mathbf{C}][\mathbf{B}] = [\mathbf{A}][\mathbf{A}]\mathbf{ji} = -[\mathbf{A}]^2\mathbf{k} : therefore\ [\mathbf{B}][\mathbf{C}] = -[\mathbf{C}][\mathbf{B}]$$

$$[\mathbf{B}][\mathbf{D}] = [\mathbf{A}][\mathbf{A}]\mathbf{ik} = -[\mathbf{A}]^2\mathbf{j};\ [\mathbf{D}][\mathbf{B}] = [\mathbf{A}][\mathbf{A}]\mathbf{ki} = [\mathbf{A}]^2\mathbf{j} : therefore\ [\mathbf{B}][\mathbf{D}] = -[\mathbf{D}][\mathbf{B}]$$

$$[\mathbf{C}][\mathbf{D}] = [\mathbf{A}][\mathbf{A}]\mathbf{jk} = [\mathbf{A}]^2\mathbf{i};\ [\mathbf{D}][\mathbf{C}] = [\mathbf{A}][\mathbf{A}]\mathbf{kj} = -[\mathbf{A}]^2\mathbf{i} : therefore\ [\mathbf{C}][\mathbf{D}] = -[\mathbf{D}][\mathbf{C}]$$

These are precisely the requirements that Dirac sought. Any matrix [**A**] will satisfy the above.

There is an additional identity for this method:

$$[\mathbf{B}][\mathbf{C}][\mathbf{D}] = [\mathbf{A}]^3\mathbf{ijk} = -[\mathbf{A}]^3$$

The sequence of multiplication for this identity may be altered to change the sign between (+) and (-).

The reader may wish to envision [**A**] as a cube of dimensions 4 x 4 x 4. The front layer contains the scalar values. The next layer contains the **i** values. The third layer contains the **j** values, and the last layer contains the **k** values.

The various rules, identities, and methods of linear algebra and matrices should be applicable to this structure. For example, there must be an identity matrix [**I**] that satisfies [**A**] [**I**] = [**A**]. There should also be an inverse matrix $[\mathbf{A}]^{-1}$ that satisfies [**A**] $[\mathbf{A}]^{-1}$ = [**I**]. It should also be possible to formulate various physical problems in terms of matrix operations. For example, the simple algebraic expression y = mx + b should have a matrix counterpart such as [**Y**] = [**M**][**X**] + [**B**].

$$\begin{bmatrix}\mathbf{Y}_1\\\mathbf{Y}_2\\\mathbf{Y}_3\\\mathbf{Y}_4\end{bmatrix}=\begin{bmatrix}\mathbf{M}_{1,1}&\mathbf{M}_{1,2}&\mathbf{M}_{1,3}&\mathbf{M}_{1,4}\\\mathbf{M}_{2,1}&\mathbf{M}_{2,2}&\mathbf{M}_{2,3}&\mathbf{M}_{2,4}\\\mathbf{M}_{3,1}&\mathbf{M}_{3,2}&\mathbf{M}_{3,3}&\mathbf{M}_{3,4}\\\mathbf{M}_{4,1}&\mathbf{M}_{4,2}&\mathbf{M}_{4,3}&\mathbf{M}_{4,4}\end{bmatrix}\begin{bmatrix}\mathbf{X}_1\\\mathbf{X}_2\\\mathbf{X}_3\\\mathbf{X}_4\end{bmatrix}+\begin{bmatrix}\mathbf{B}_1\\\mathbf{B}_2\\\mathbf{B}_3\\\mathbf{B}_4\end{bmatrix}$$

The solution set [**X**] would be determined by [**X**] = $[\mathbf{M}]^{-1}$([**Y**] - [**B**]).

Tensor Comparison:

The author has reviewed portions of Flügge[3], Bishop and Goldberg[4], and Simmonds[5]. Based upon their texts, the tensor structure that would most closely resemble the structure presented here is a 3rd order tensor with 4 dimensions. It would have 64 elements (4^3) just as the structure presented here has 64 elements. The difference between these two structures is that the elements of a tensor are all scalars, whereas the structure presented here includes the unit vectors. Both structures can operate upon vectors or quaternions.

Multiplication:

Now consider the multiplication of two arbitrary, square matrices [**A**] and [**B**] such that [**A**][**B**] = [**C**].

$$\begin{bmatrix}\mathbf{C}_{1,1}&\mathbf{C}_{1,2}&\mathbf{C}_{1,3}&\mathbf{C}_{1,4}\\\mathbf{C}_{2,1}&\mathbf{C}_{2,2}&\mathbf{C}_{2,3}&\mathbf{C}_{2,4}\\\mathbf{C}_{3,1}&\mathbf{C}_{3,2}&\mathbf{C}_{3,3}&\mathbf{C}_{3,4}\\\mathbf{C}_{4,1}&\mathbf{C}_{4,2}&\mathbf{C}_{4,3}&\mathbf{C}_{4,4}\end{bmatrix}=\begin{bmatrix}\mathbf{A}_{1,1}&\mathbf{A}_{1,2}&\mathbf{A}_{1,3}&\mathbf{A}_{1,4}\\\mathbf{A}_{2,1}&\mathbf{A}_{2,2}&\mathbf{A}_{2,3}&\mathbf{A}_{2,4}\\\mathbf{A}_{3,1}&\mathbf{A}_{3,2}&\mathbf{A}_{3,3}&\mathbf{A}_{3,4}\\\mathbf{A}_{4,1}&\mathbf{A}_{4,2}&\mathbf{A}_{4,3}&\mathbf{A}_{4,4}\end{bmatrix}\begin{bmatrix}\mathbf{B}_{1,1}&\mathbf{B}_{1,2}&\mathbf{B}_{1,3}&\mathbf{B}_{1,4}\\\mathbf{B}_{2,1}&\mathbf{B}_{2,2}&\mathbf{B}_{2,3}&\mathbf{B}_{2,4}\\\mathbf{B}_{3,1}&\mathbf{B}_{3,2}&\mathbf{B}_{3,3}&\mathbf{B}_{3,4}\\\mathbf{B}_{4,1}&\mathbf{B}_{4,2}&\mathbf{B}_{4,3}&\mathbf{B}_{4,4}\end{bmatrix}$$

The 1st term produced by this multiplication is

$$\mathbf{C}_{1,1}=\ \mathbf{A}_{1,1}\mathbf{B}_{1,1}+\ \mathbf{A}_{1,2}\mathbf{B}_{2,1}+\ \mathbf{A}_{1,3}\mathbf{B}_{3,1}+\ \mathbf{A}_{1,4}\mathbf{B}_{4,1}$$

This equation appears to be fairly simple, but it is, in fact, quite difficult. Each of the two-term multiplications (example, $\mathbf{A}_{1,1}\mathbf{B}_{1,1}$) produces 16 terms (example, $(a_0)_{1,1}(b_0)_{1,1}$). Therefore, $\mathbf{C}_{1,1}$ is actually the sum of 64 terms. The matrix [**C**] contains 16 of these elements. Therefore, there are a total of 1024 terms needed to produce [**C**]. This is 2^{10} (i.e., 2 raised to the power of 10) separate terms.

In general, the elements of [**C**] can be represented as follows:

$$\mathbf{C}_{m,n} = \sum_{i=1}^{i=4} \mathbf{A}_{m,i}\mathbf{B}_{i,n}$$

In this summation, it should be understood that the scalar terms will be combined. The **i** terms will be combined, as will the **j** terms and the **k** terms. Therefore, there are actually four summations. The index "i" used in the summation should not be confused with the vector **i**.

Next, consider the multiplication [**B**][**A**] = [**D**] such that the order of multiplication has been swapped.

$$\begin{bmatrix} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} & \mathbf{D}_{1,3} & \mathbf{D}_{1,4} \\ \mathbf{D}_{2,1} & \mathbf{D}_{2,2} & \mathbf{D}_{2,3} & \mathbf{D}_{2,4} \\ \mathbf{D}_{3,1} & \mathbf{D}_{3,2} & \mathbf{D}_{3,3} & \mathbf{D}_{3,4} \\ \mathbf{D}_{4,1} & \mathbf{D}_{4,2} & \mathbf{D}_{4,3} & \mathbf{D}_{4,4} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \mathbf{B}_{1,3} & \mathbf{B}_{1,4} \\ \mathbf{B}_{2,1} & \mathbf{B}_{2,2} & \mathbf{B}_{2,3} & \mathbf{B}_{2,4} \\ \mathbf{B}_{3,1} & \mathbf{B}_{3,2} & \mathbf{B}_{3,3} & \mathbf{B}_{3,4} \\ \mathbf{B}_{4,1} & \mathbf{B}_{4,2} & \mathbf{B}_{4,3} & \mathbf{B}_{4,4} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix}$$

$$\mathbf{D}_{m,n} = \sum_{i=1}^{i=4} \mathbf{B}_{m,i}\mathbf{A}_{i,n}$$

Inversion:

Next, let us consider the inverse of matrix [**A**] such that [**A**][**A**]$^{-1}$ = [**I**]. To avoid confusion, let us define the inverse matrix such that [**A**]$^{-1}$ = [**B**]. Therefore, [**A**][**B**] = [**I**]. For purposes of this discussion, the matrix [**A**] is assumed to be known, and the objective is to determine the matrix [**B**].

$$\begin{bmatrix} \mathbf{I}_{1,1} & \mathbf{I}_{1,2} & \mathbf{I}_{1,3} & \mathbf{I}_{1,4} \\ \mathbf{I}_{2,1} & \mathbf{I}_{2,2} & \mathbf{I}_{2,3} & \mathbf{I}_{2,4} \\ \mathbf{I}_{3,1} & \mathbf{I}_{3,2} & \mathbf{I}_{3,3} & \mathbf{I}_{3,4} \\ \mathbf{I}_{4,1} & \mathbf{I}_{4,2} & \mathbf{I}_{4,3} & \mathbf{I}_{4,4} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \mathbf{B}_{1,3} & \mathbf{B}_{1,4} \\ \mathbf{B}_{2,1} & \mathbf{B}_{2,2} & \mathbf{B}_{2,3} & \mathbf{B}_{2,4} \\ \mathbf{B}_{3,1} & \mathbf{B}_{3,2} & \mathbf{B}_{3,3} & \mathbf{B}_{3,4} \\ \mathbf{B}_{4,1} & \mathbf{B}_{4,2} & \mathbf{B}_{4,3} & \mathbf{B}_{4,4} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \mathbf{B}_{1,3} & \mathbf{B}_{1,4} \\ \mathbf{B}_{2,1} & \mathbf{B}_{2,2} & \mathbf{B}_{2,3} & \mathbf{B}_{2,4} \\ \mathbf{B}_{3,1} & \mathbf{B}_{3,2} & \mathbf{B}_{3,3} & \mathbf{B}_{3,4} \\ \mathbf{B}_{4,1} & \mathbf{B}_{4,2} & \mathbf{B}_{4,3} & \mathbf{B}_{4,4} \end{bmatrix}$$

$$\mathbf{I}_{m,n} = \sum_{i=1}^{i=4} \mathbf{A}_{m,i}\mathbf{B}_{i,n}$$

For m = n, this becomes

$$\mathbf{I}_{m,n=m} = \sum_{i=1}^{i=4} \mathbf{A}_{m,i}\mathbf{B}_{i,n=m} = 1$$

For m ≠ n, this becomes

$$\mathbf{I}_{m,n\neq m} = \sum_{i=1}^{i=4} \mathbf{A}_{m,i}\mathbf{B}_{i,n\neq m} = 0$$

Viewed as a complete system, the matrix [**A**] is completely defined, and the matrix [**I**] is completely defined. The matrix [**B**] is completely **unknown**. Therefore, there are 64 unknowns $(b_0, b_i, b_j, b_k)_{m,n}$ that must be determined to specify [**B**]. Each of the $\mathbf{A}_{m,i}\mathbf{B}_{i,n}$ multiplications produces 4 independent, linear equations. There are 16 of these $\mathbf{A}_{m,i}\mathbf{B}_{i,n}$ multiplications. Therefore, there are 64 independent, linear equations. The entire system is therefore described by 64 independent, linear equations with 64 unknowns. The system is not homogeneous because [**I**] ≠ **0**. Therefore, there should be a unique solution.

The next step is to set the problem up such that it can be solved by a standard method such as Gaussian Elimination. Consider an arbitrary **AB** multiplication as follows:

$$(a_0 + a_i\mathbf{i} + a_j\mathbf{j} + a_k\mathbf{k})(b_0 + b_i\mathbf{i} + b_j\mathbf{j} + b_k\mathbf{k}) = c_0 + c_i\mathbf{i} + c_j\mathbf{j} + c_k\mathbf{k}$$

This produces the following four equations:

$$a_0b_0 - a_ib_i - a_jb_j - a_kb_k = c_0$$

$$a_ib_0 + a_0b_i - a_kb_j + a_jb_k = c_i$$

$$a_jb_0 + a_kb_i + a_0b_j - a_ib_k = c_j$$

$$a_kb_0 - a_jb_i + a_ib_j + a_0b_k = c_k$$

This is then represented in matrix form as:

$$\begin{bmatrix} a_0 & -a_i & -a_j & -a_k \\ a_i & a_0 & -a_k & a_j \\ a_j & a_k & a_0 & -a_i \\ a_k & -a_j & a_i & a_0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_i \\ b_j \\ b_k \end{bmatrix} = \begin{bmatrix} c_0 \\ c_i \\ c_j \\ c_k \end{bmatrix}$$

The next decision to make is what order in which to perform the **AB** multiplications so as to give the resulting coefficient matrix a desirable structure. If the 1st column of [**B**] is multiplied by each of the four rows of [**A**] then all of the information associated with $\mathbf{B}_{i,1}$ (i.e., the 1st column of [**B**]) will be included in the first 16 equations. If this is then repeated for the next three columns of **B**, the resulting matrix system will look something like the following:

$$\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix}$$

All of the values in this representation are scalars. The coefficient matrix is sparse with the non-zero terms clustered around the diagonal. Each letter "a" represents a 16 x 16 group of coefficients. The coefficient values "a" actually repeat 4 times, since each column of [**B**] is multiplied by the same set of values from each row of [**A**]. The "b" values are grouped by column. Each "b" represents the 16 unknown values that are associated with the four $\mathbf{B}_{m,n}$ for that set of row-column multiplications. The values for "I" are mostly "zeros" with there being a single "1" value in each of the four groups. The value "1" is in a different position in each "I". This system can be solved using Gaussian Elimination.

For diagonal matrices, the inversion problem is fairly simple because a quaternion multiplied by its conjugate produces a scalar term only. Therefore, it is simply a question of scaling the conjugate term properly to produce a value of 1 along the diagonal of [**I**].

Here are a few simple examples:

$$\begin{bmatrix} (1+\mathbf{i}) & 0 & \begin{matrix} 0 & 0 \end{matrix} \\ 0 & (1-\mathbf{i}) & \begin{matrix} 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & (\mathbf{j}+\mathbf{k}) & 0 \\ \begin{matrix} 0 & 0 \end{matrix} & 0 & (\mathbf{j}-\mathbf{k}) \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2}(1-\mathbf{i}) & 0 & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ 0 & \frac{1}{2}(1+\mathbf{i}) & \\ \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \frac{1}{2}(-\mathbf{j}-\mathbf{k}) & 0 \\ & 0 & \frac{1}{2}(-\mathbf{j}+\mathbf{k}) \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathbf{i} & 0 & 0 \\ 0 & 0 & \mathbf{j} & 0 \\ 0 & 0 & 0 & \mathbf{k} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -\mathbf{i} & 0 & 0 \\ 0 & 0 & -\mathbf{j} & 0 \\ 0 & 0 & 0 & -\mathbf{k} \end{bmatrix}$$

Matrix Reduction:

In general, it is possible to convert a generic matrix [**A**] into forms such as upper triangular, lower triangular, and diagonal. This process requires the use of quaternion division and quaternion multiplication rather than their scalar counterparts. Suppose that we wish to convert matrix [**A**] into diagonal matrix [**D**]. Begin with matrix [**A**] as follows:

$$[\mathbf{A}] = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix}$$

Define a quaternion ratio $\mathbf{R}_{2,1}$ such that $\mathbf{R}_{2,1} = -\mathbf{A}_{2,1}/\mathbf{A}_{1,1}$ for $\mathbf{A}_{1,1} \neq 0$. This **R** will be used to eliminate $\mathbf{A}_{2,1}$ from row 2 of matrix [**A**]. There will also be **R** values for row 3 and row 4 of [**A**]. The $\mathbf{A}_{2,1}$ term is eliminated by multiplying the entire 1st row by $\mathbf{R}_{2,1}$ and adding the result to the 2nd row to produce a new 2nd row.

$$\mathbf{A'}_{2,n} = \mathbf{A}_{1,n}\mathbf{R}_{2,1} + \mathbf{A}_{2,n} \; for \; n = 1, 2, 3, 4$$

This is repeated for row 3 and 4 as follows:

$$\mathbf{A'}_{3,n} = \mathbf{A}_{1,n}\mathbf{R}_{3,1} + \mathbf{A}_{3,n} \; for\ n = 1, 2, 3, 4$$

$$\mathbf{A'}_{4,n} = \mathbf{A}_{1,n}\mathbf{R}_{4,1} + \mathbf{A}_{4,n} \; for\ n = 1, 2, 3, 4$$

After these operations are completed, the matrix [**A**'] is as follows:

$$[\mathbf{A'}] = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ 0 & \mathbf{A'}_{2,2} & \mathbf{A'}_{2,3} & \mathbf{A'}_{2,4} \\ 0 & \mathbf{A'}_{3,2} & \mathbf{A'}_{3,3} & \mathbf{A'}_{3,4} \\ 0 & \mathbf{A'}_{4,2} & \mathbf{A'}_{4,3} & \mathbf{A'}_{4,4} \end{bmatrix}$$

The next step is to eliminate $\mathbf{A'}_{3,2}$, $\mathbf{A'}_{4,2}$ and $\mathbf{A}_{1,2}$. This is done using $\mathbf{A'}_{2,2}$ in a manner similar to the above. Define a quaternion ratio $\mathbf{R'}_{3,2}$ such that $\mathbf{R'}_{3,2} = -\mathbf{A'}_{3,2}/\mathbf{A'}_{2,2}$ for $\mathbf{A'}_{2,2} \neq 0$. There will also be **R**' values for row 1 and row 4.

$$\mathbf{A''}_{3,n} = \mathbf{A'}_{2,n}\mathbf{R'}_{3,2} + \mathbf{A'}_{3,n} \; for\ n = 2, 3, 4$$

$$\mathbf{A''}_{4,n} = \mathbf{A'}_{2,n}\mathbf{R'}_{4,2} + \mathbf{A'}_{4,n} \; for\ n = 2, 3, 4$$

$$\mathbf{A''}_{1,n} = \mathbf{A'}_{2,n}\mathbf{R'}_{1,2} + \mathbf{A}_{1,n} \; for\ n = 2, 3, 4$$

After these operations are completed, the matrix [**A**''] is as follows:

$$[\mathbf{A''}] = \begin{bmatrix} \mathbf{A}_{1,1} & 0 & \mathbf{A''}_{1,3} & \mathbf{A''}_{1,4} \\ 0 & \mathbf{A'}_{2,2} & \mathbf{A'}_{2,3} & \mathbf{A'}_{2,4} \\ 0 & 0 & \mathbf{A''}_{3,3} & \mathbf{A''}_{3,4} \\ 0 & 0 & \mathbf{A''}_{4,3} & \mathbf{A''}_{4,4} \end{bmatrix}$$

This process is repeated two more times, for a total of four reductions, with the final result being a diagonal matrix [**D**] as follows:

$$[\mathbf{D}] = \begin{bmatrix} \mathbf{A}_{1,1} & 0 & 0 & 0 \\ 0 & \mathbf{A'}_{2,2} & 0 & 0 \\ 0 & 0 & \mathbf{A''}_{3,3} & 0 \\ 0 & 0 & 0 & \mathbf{A'''}_{4,4} \end{bmatrix}$$

This matrix reduction can be represented by multiplication by a transformation matrix [**T**] as follows:

$$[\mathbf{D}] = [\mathbf{T}][\mathbf{A}]$$

and

$$[\mathbf{T}]^{-1}[\mathbf{D}] = [\mathbf{A}]$$

At first consideration, these relations do not appear to be useful since [**D**] must be generated by manual reduction. The matrix $[\mathbf{T}]^{-1}$ is determined using Gauss Elimination on the 64-equation system. The matrix

[**T**] is then determined by inverting $[\mathbf{T}]^{-1}$. The usefulness of [**T**] and $[\mathbf{T}]^{-1}$ will become clear when considering the square root of a matrix.

Upper triangular and lower triangular reductions are produced by the same process except only the rows below or above the diagonal element are reduced.

The Square of a Matrix:

Now let us consider the square of matrix [**A**] such that $[\mathbf{A}]^2 = [\mathbf{B}]$.

$$\begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \mathbf{B}_{1,3} & \mathbf{B}_{1,4} \\ \mathbf{B}_{2,1} & \mathbf{B}_{2,2} & \mathbf{B}_{2,3} & \mathbf{B}_{2,4} \\ \mathbf{B}_{3,1} & \mathbf{B}_{3,2} & \mathbf{B}_{3,3} & \mathbf{B}_{3,4} \\ \mathbf{B}_{4,1} & \mathbf{B}_{4,2} & \mathbf{B}_{4,3} & \mathbf{B}_{4,4} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix}$$

It is simple enough to go from [**A**] to $[\mathbf{A}]^2$. This is a matrix multiplication as described above where

$$\mathbf{B}_{m,n} = \sum_{i=1}^{i=4} \mathbf{A}_{m,i}\mathbf{A}_{i,n}$$

But what of the inverse problem? Is it possible to determine the square root of a matrix? A quick search of the internet indicates that the answer is in the affirmative. The website Wikipedia references several methods of determining the square root of a matrix. The simplest method appears to be diagonalization. Essentially, the matrix [**B**] is converted into diagonal form [**D**] by using transformation matrix [**T**]. The square roots are then determined for each of the diagonal elements. The resulting square root matrices are then transformed back into non-diagonal form using the inverse transformation matrix $[\mathbf{T}]^{-1}$. This process is easier to understand if the matrix [**T**] is thought of as the information needed to convert into diagonal form. Reverting back to non-diagonal form is then accomplished by the information in $[\mathbf{T}]^{-1}$.

Next, let us suppose that the original matrix has been converted into diagonal form [**D**] such that

$$\begin{bmatrix} \mathbf{D}_1 & 0 & 0 & 0 \\ 0 & \mathbf{D}_2 & 0 & 0 \\ 0 & 0 & \mathbf{D}_3 & 0 \\ 0 & 0 & 0 & \mathbf{D}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \end{bmatrix}$$

The problem is to determine a quaternion **A** for each element of the diagonal matrix such that $\mathbf{A}^2 = \mathbf{D}$. The square of quaternion **A** simplifies to the following:

$$\mathbf{A}^2 = a_0^2 - \left(a_i^2 + a_j^2 + a_k^2\right) + 2a_0\left(a_i\mathbf{i} + a_j\mathbf{j} + a_k\mathbf{k}\right)$$

For each diagonal element, there are four equations with four unknown coefficients to determine. Therefore, there should be a solution for each diagonal element. However, the problem is not linear. The complicating factors are that the coefficients are squared for the scalar component, and they are multiplied by twice the scalar coefficient for the vector components.

As an example, let us determine the square root of the following:

$$\begin{bmatrix} \mathbf{D}_1 & 0 & 0 & 0 \\ 0 & \mathbf{D}_2 & 0 & 0 \\ 0 & 0 & \mathbf{D}_3 & 0 \\ 0 & 0 & 0 & \mathbf{D}_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathbf{i} & 0 & 0 \\ 0 & 0 & \mathbf{j} & 0 \\ 0 & 0 & 0 & \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \end{bmatrix}^2$$

The value for $\mathbf{A}_1$ is simply ±1 since there are no vector terms in $\mathbf{D}_1$. The values for $\mathbf{A}_2$, $\mathbf{A}_3$, and $\mathbf{A}_4$ are a little more complicated. Next, consider $\mathbf{A}_2$. Since there is no scalar term (i.e., $a_0^2 - a_i^2 = 0$), it follows that $a_0^2 = a_i^2$, and that $a_0 = \pm a_i$. Since d_i equals 1, it follows that $2a_0a_i = 1$. Therefore, the signs of a_0 and a_i must either both be positive or both be negative. Therefore, $a_0 = a_i = \pm(1/\text{sqrt } 2)$. Since d_j and d_k are each equal to zero, it follows that a_j and a_k are also zero. The $\mathbf{A}_3$ and $\mathbf{A}_4$ elements are solved similarly. Therefore, the square root of matrix [**D**] is as follows:

$$[\mathbf{A}] = \begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \end{bmatrix} = \begin{bmatrix} \pm 1 & 0 & 0 & 0 \\ 0 & \pm\frac{1}{\sqrt{2}}(1+\mathbf{i}) & 0 & 0 \\ 0 & 0 & \pm\frac{1}{\sqrt{2}}(1+\mathbf{j}) & 0 \\ 0 & 0 & 0 & \pm\frac{1}{\sqrt{2}}(1+\mathbf{k}) \end{bmatrix}$$

The next example will illustrate how it is possible for a solution to be not unique.

$$\begin{bmatrix} \mathbf{D}_1 & 0 & 0 & 0 \\ 0 & \mathbf{D}_2 & 0 & 0 \\ 0 & 0 & \mathbf{D}_3 & 0 \\ 0 & 0 & 0 & \mathbf{D}_4 \end{bmatrix} = \begin{bmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \end{bmatrix}^2$$

$$[\mathbf{A}] = \begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \end{bmatrix} = \begin{bmatrix} \pm 1 & 0 & 0 & 0 \\ 0 & \pm\mathbf{i} & 0 & 0 \\ 0 & 0 & \pm\mathbf{j} & 0 \\ 0 & 0 & 0 & \pm\mathbf{k} \end{bmatrix}$$

Other:

In all of the above, the focus is upon multiplication of two square matrices. It is also possible to multiply a square matrix by a column matrix, and it is possible to multiply a square matrix by a single quaternion. These are illustrated below.

$$[\mathbf{C}] = [\mathbf{A}][\mathbf{B}]$$

$$\begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \\ \mathbf{C}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \mathbf{B}_3 \\ \mathbf{B}_4 \end{bmatrix}$$

$$\mathbf{C}_m = \sum_{i=1}^{i=4} \mathbf{A}_{m,i}\mathbf{B}_i$$

$$[\mathbf{C}] = [\mathbf{A}]\mathbf{B}$$

$$\begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \\ \mathbf{C}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \mathbf{A}_{1,4} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \mathbf{A}_{2,4} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \mathbf{A}_{3,4} \\ \mathbf{A}_{4,1} & \mathbf{A}_{4,2} & \mathbf{A}_{4,3} & \mathbf{A}_{4,4} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2\mathbf{i} \\ b_3\mathbf{j} \\ b_4\mathbf{k} \end{bmatrix}$$

$$\mathbf{C}_m = \mathbf{A}_{m,1}b_1 + \mathbf{A}_{m,2}b_2\mathbf{i} + \mathbf{A}_{m,3}b_3\mathbf{j} + \mathbf{A}_{m,4}b_4\mathbf{k}$$

It is somewhat ironic that the notation for multiplication by a single quaternion is the most complex. The previous notations could also be utilized if it is recognized that most of the coefficients are zero.

Acknowledgements

The author thanks Susan Barlow, Karoly Kehrer, and the rest of the WSM user's group on Yahoo.com. The author also thanks viXra.org and Wikipedia.

References

1. Pauli, W. 1973. Pauli Lectures on Physics - Volume 5 - Wave Mechanics, Massachusetts Institute of Technology, USA, pp. 156-159.
2. Dirac, P.A.M. The Principles of Quantum Mechanics - 4th Edition, Snow Ball Publishing, USA, pp. 149-152.
3. Flügge, W. 1972. Tensor Analysis and Continuum Mechanics, Springer-Verlag, Berlin, pp. 1-22.
4. Bishop, R.L., and Goldberg, S.I. 1980. Tensor Analysis on Manifolds, Dover, New York, pp. 59-115.
5. Simmonds, J.G. 1982. A Brief bon Tensor Analysis, Springer, New York, pp. 1-44.