CHAPTER 5

THE EVOLUTION OF ALGEBRA, II

Hamilton and Quaternions

IN 1833 HAMILTON PRESENTED a paper before the Irish Academy in which he introduced a formal algebra of real number pairs whose rules of combination are precisely those for complex numbers. The important rule for multiplication of these pairs, corresponding to the rule

$$(a + b\mathrm{i})(c + d\mathrm{i}) = (ac - bd) + (ad + bc)\mathrm{i}$$

is

$$(a,b)(c,d) = (ac - bd\,,\, ad + bc),$$

which he interpreted as an operation involving rotation. Hamilton's paper provided the definitive formulation of complex numbers as pairs of real numbers.

Hamilton grasped that his ordered pairs could be thought of as directed entities — *vectors*—in the plane, and he naturally sought to extend this conception to three dimensions by moving from number pairs $(a,b) = a + b\mathrm{i}$ to number triples $(a,b,c) = a + b\mathbf{i} + c\mathbf{j}$. How should the operations of addition and multiplication on these entities then be defined? The operation of addition presented no problem—indeed, the rule is obvious. For several years, however, he was unable to see how to define multiplication for triples. It is part of mathematical legend that, in October 1843, as he was walking with his wife along the Royal Canal near Dublin, in a flash of inspiration he saw that the difficulty would vanish if triples were to be replaced by *quadruples* $(a,b,c,d) = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ and *the commutative law of multiplication jettisoned.* He had already discerned that for number quadruples one should take $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1$; now in addition he realized that one should not only take $\mathbf{ij} = \mathbf{k}$, but also $\mathbf{ji} = -\mathbf{k}$ and similarly $\mathbf{jk} = \mathbf{i} = -\mathbf{kj}$ and $\mathbf{ki} = \mathbf{j} = -\mathbf{ik}$. The remaining laws of operation—associativity, distributivity—were then to be those of ordinary algebra. By the revolutionary act[1] of abandoning the

[1] It is worth noting in this connection that Hamilton's mathematical work was strongly influenced by his philosophical views, which were derived in the main from Kant. Kant had maintained that space

commutative law of multiplication, Hamilton had in effect created a self-consistent new algebra—the *quaternion algebra*. It is said that, so impressed was he by this discovery, with a knife he inscribed the fundamental formula for quaternions $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk}$ on a stone of Brougham Bridge.

Accordingly a *quaternion* is an expression of the form

$$a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = \mathbf{u},$$

where a, b, c, d are real numbers. The terms $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are called *units*. The real number a is called the *scalar part*, and the remainder $b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ the *vector part*, of $\mathbf{u}$. The three coefficients of the vector part of $\mathbf{u}$ may be thought of as rectangular Cartesian coordinates of a point and the units $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ as unit vectors directed along the three axes. Two quaternions $\mathbf{u} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ and $\mathbf{u}' = a' + b'\mathbf{i} + c'\mathbf{j} + d'\mathbf{k}$ are *equal* if $a = a'$, $b = b'$, $c = c'$, and $d = d'$. Quaternions are *added* by the rule

$$\mathbf{u} + \mathbf{u}' = (a + a') + (b + b')\mathbf{i} + (c + c')\mathbf{j} + (d + d')\mathbf{k}.$$

The *zero* quaternion is defined as

$$\mathbf{0} = 0 + 0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k};$$

clearly $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for any quaternion $\mathbf{u}$. To each real number a there corresponds a quaternion, namely

$$\mathbf{a} = a + 0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}.$$

Products of quaternions are calculated using the familiar algebraic rules of manipulation, except that in forming products of the units $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ the following rules—mentioned above—are to be observed:

$$\mathbf{jk} = \mathbf{i},\ \mathbf{kj} = -\mathbf{i},\ \mathbf{ki} = \mathbf{j},\ \mathbf{ik} = -\mathbf{j},\ \mathbf{ij} = \mathbf{k},\ \mathbf{ji} = -\mathbf{k}$$
$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -\mathbf{1}$$

Thus

$$(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k})(a' + b'\mathbf{i} + c'\mathbf{j} + d'\mathbf{k}) = A + B\mathbf{i} + C\mathbf{j} + D\mathbf{k},$$

where

$$A = aa' - bb' - cc' - dd'$$

---

and time were the two essential forms of sensuous intuition, and Hamilton went so far as to proclaim that, just as geometry is the science of pure space, so algebra must be the science of pure time.

$$B = ab' + ba' + cd' - dc'$$
$$C = ac' + ca' + db' - bd'$$
$$D = ad' + da' + bc' - cb'.$$

So, for instance,

$$(\mathbf{1} + \mathbf{i} + \mathbf{j} + \mathbf{k})(\mathbf{1} - \mathbf{i} + \mathbf{j} + \mathbf{k}) = 4\mathbf{k},$$

while

$$(\mathbf{1} - \mathbf{i} + \mathbf{j} + \mathbf{k})(\mathbf{1} + \mathbf{i} + \mathbf{j} + \mathbf{k}) = 4\mathbf{j}.$$

That is, *multiplication of quaternions is not commutative*. On the other hand, it may be verified from the above rules for multiplying units that the product of units is associative, and it follows from this that *multiplication of quaternions is associative*.

Quaternions form what is known as a *division algebra*, in that, like the real and complex numbers, division by nonzero elements can be effected. We first determine the *reciprocal* or *inverse* $\mathbf{u}^{-1}$ of a nonzero quaternion $\mathbf{u} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$. This will have to satisfy

$$\mathbf{u}\mathbf{u}^{-1} = \mathbf{u}^{-1}\mathbf{u} = \mathbf{1} \qquad (1)$$

We note that if we define

$$\mathbf{u}^{-1} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k},$$

then

$$\mathbf{u}\mathbf{u}^{-1} = \mathbf{u}^{-1}\mathbf{u} = a^2 + b^2 + c^2 + d^2. \qquad (2)$$

The quantity on the right side is called the *norm* of $\mathbf{u}$ and is written $\|\mathbf{u}\|$. It follows then from (2) that if we define

$$\mathbf{u}^{-1} = a/\|\mathbf{u}\| - (b/\|\mathbf{u}\|)\mathbf{i} - (c/\|\mathbf{u}\|)\mathbf{j} - (d/\|\mathbf{u}\|)\mathbf{k},$$

then (1) will be satisfied.

Now let $\mathbf{v}$ be any quaternion. Since multiplication is not commutative, there will in general be two quotients of $\mathbf{v}$ by $\mathbf{u}$, namely $\mathbf{r}$ for which $\mathbf{ru} = \mathbf{v}$ and $\mathbf{s}$ for which $\mathbf{us} = \mathbf{v}$. To find these, we multiply $\mathbf{ru} = \mathbf{v}$ by $\mathbf{u}^{-1}$ on both sides on the right to get $\mathbf{r} = \mathbf{ruu}^{-1} = \mathbf{vu}^{-1}$ and $\mathbf{us} = \mathbf{v}$ on the left to get $\mathbf{s} = \mathbf{u}^{-1}\mathbf{us} = \mathbf{u}^{-1}\mathbf{v}$.

Since the quaternions form a division algebra, it makes sense to ask whether, like the complex numbers, the *Fundamental Theorem of Algebra* holds for them. This question was answered in the affirmative in 1944 when Eilenberg and Niven showed that any polynomial equation $\mathbf{a}_0 + \mathbf{a}_1 x + \ldots + \mathbf{a}_n x^n = \mathbf{0}$ with quaternion coefficients has a quaternion root *x*.

Quaternions may be regarded as describing *rotations* and *stretchings* (or *contractions*) in three-dimensional space. Given two vectors $v = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ and $v' = a'\mathbf{i} + b'\mathbf{j} + c'\mathbf{k}$ we seek a quaternion $\mathbf{q} = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ which when "applied" to *v* rotates or stretches it to coincide with *v′* in the sense that $v' = \mathbf{qv}$, that is,

$$(w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k})(a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) = a'\mathbf{i} + b'\mathbf{j} + c'\mathbf{k}.$$

By multiplying the left side of this equality as quaternions and equating coefficients of both sides we get four equations in the unknowns *w, x, y, z* whose values are then uniquely determined. From the laws of the multiplication of the units **i, j, k**, we see that "applying" **i** in this sense has the effect of rotating the system of unit vectors **i, j, k** about **i** so as to bring **j** into coincidence with **k** and **k** with **–j;** and similarly for "applications" of **j, k.**

Grassmann's "Calculus of Extension"

In 1844 the German mathematician *Hermann Grassmann* (1809–1877) published a work—*Die Lineale Ausdehnungslehre* ("The Calculus of Extension")—in which is formulated a symbolic algebra far surpassing Hamilton's quaternions in generality. Indeed, Grassmann's algebra is nothing less than a full *n*-dimensional vector calculus.

Grassmann starts with a set of quantities $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ that he calls *primary units*: these may be thought of geometrically as a system of *n* mutually perpendicular line segments of unit length drawn from a common origin in *n*-dimensional space. It is assumed that these units can be added together, and multiplied by real numbers. Grassmann's basic notion is that of *extensive quantity*[2], which he defines to be an expression of the form

$$a_1\mathbf{e}_1 + \ldots + a_n\mathbf{e}_n = \mathbf{a},$$

where $a_1, \ldots, a_n$ are any real numbers. Addition of extensive quantities is defined by

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1)\mathbf{e}_1 + \ldots + (a_n + b_n)\mathbf{e}_n.$$

[2] Extensive quantities are to be contrasted with *intensive quantities.* Quantities such as mass or volume are extensive in the sense that they are defined over extended regions of space and are therefore additive: thus 2 pounds + 2 pounds = 4 pounds. Vector quantities such as velocity or acceleration, being additive in this sense, also count as extensive quantities. On the other hand, quantities such as temperature or density are intensive in that they are defined at a point and are not additive: thus on mixing two buckets of water each having a uniform temperature of 50 degrees one obtains a quantity of water at a temperature of 50, rather than 100, degrees.

Grassmann introduces a very general concept of *multiplication* of extensive quantities. He first assumes that for each pair $\mathbf{e}_i$, $\mathbf{e}_j$ of units a new unit $[\mathbf{e}_i\mathbf{e}_j]$ called their *product* is given. Then he defines the product of the extensive quantities **a** and **b** to be the extensive quantity

$$[\mathbf{ab}] = a_1b_1[\mathbf{e}_1\mathbf{e}_1] + a_2b_1[\mathbf{e}_2\mathbf{e}_1] + \dots + a_nb_{n-1}[\mathbf{e}_n\mathbf{e}_{n-1}] + a_nb_n[\mathbf{e}_n\mathbf{e}_n].$$

By imposing various conditions on the unit products $[\mathbf{e}_i\mathbf{e}_j]$ Grassmann obtains different sorts of product. For simplicity let us assume that $n = 3$. Then, for example, if we take $[\mathbf{e}_2\mathbf{e}_1] = [\mathbf{e}_1\mathbf{e}_2]$, $[\mathbf{e}_1\mathbf{e}_3] = [\mathbf{e}_3\mathbf{e}_1]$, and $[\mathbf{e}_2\mathbf{e}_3] = [\mathbf{e}_2\mathbf{e}_3]$, the coefficients of $[\mathbf{ab}]$ are

$$a_1b_1,\ a_2b_2,\ a_3b_3,\ a_1b_2 + a_2b_1,\ a_1b_3 + a_3b_1,\ a_2b_3 + a_3b_2,$$

and so in this case the laws governing the product are just those of ordinary multiplication. If we agree to take $[\mathbf{e}_1\mathbf{e}_1] = [\mathbf{e}_2\mathbf{e}_2] = [\mathbf{e}_3\mathbf{e}_3] = 0$ and $[\mathbf{e}_2\mathbf{e}_1] = -[\mathbf{e}_1\mathbf{e}_2]$, $[\mathbf{e}_1\mathbf{e}_3] = -[\mathbf{e}_3\mathbf{e}_1]$, $[\mathbf{e}_2\mathbf{e}_3] = -[\mathbf{e}_3\mathbf{e}_2]$, then $[\mathbf{ab}]$ has the coefficients

$$a_2b_3 - a_3b_2,\quad a_3b_1 - a_1b_3,\quad a_1b_2 - a_2b_1.$$

This Grassmann terms a *combinatory* (also known as *vector*) product. Finally, if we choose $[\mathbf{e}_2\mathbf{e}_1] = [\mathbf{e}_1\mathbf{e}_2] = [\mathbf{e}_1\mathbf{e}_3] = [\mathbf{e}_3\mathbf{e}_1] = [\mathbf{e}_2\mathbf{e}_3] = [\mathbf{e}_3\mathbf{e}_2] = 0$ and $[\mathbf{e}_1\mathbf{e}_1] = [\mathbf{e}_2\mathbf{e}_2] = [\mathbf{e}_3\mathbf{e}_3] = 1$, then $[\mathbf{ab}]$ becomes the numerical quantity

$$a_1b_1 + a_2b_2 + a_3b_3;$$

this Grassmann calls the *inner product* of **a** and **b**.

Grassmann defines the *numerical value*—what we would today call the *norm* or *magnitude*—of an extensive quantity $\mathbf{a} = a_1\mathbf{e}_1 + \dots + a_n\mathbf{e}_n$ to be the real number

$$\|\mathbf{a}\| = (a_1^2 + \dots + a_n^2).$$

In the case $n = 3$, if we take $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ to be 3 mutually perpendicular line segments of unit length, then each extensive quantity $\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$ is a vector in 3-dimensional space with components a_1, a_2, a_3 along $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$. The magnitude $\|\mathbf{a}\|$ is the length of **a**, the inner product of **a** and **b** is the product of the length of **a** with that of the projection of **b** onto **a** (or vice-versa), and the magnitude of the combinatory product $[\mathbf{ab}]$ is the area of the parallelogram with adjacent sides **a** and **b**.

Grassmann also considers the product formed by taking the inner product of an extensive quantity with the combinatory product of two extensive quantities, as in $[\mathbf{ab}]\mathbf{c}$. In the 3-dimensional case this quantity may be interpreted as the volume of a parallelepiped with adjacent sides **a, b, c.**

The novelty and unconventional presentation of Grassmann's work prevented its full significance from being recognized by his contemporaries and for this reason he was led to publish a revised version of the *Ausdehnungslehre* in 1862. Nevertheless, it proved to be an important influence in the development of the three-dimensional vector calculus whose use was to become standard in physics, and later in mathematics also.

Finite-Dimensional Linear Algebras

Like the complex numbers, quaternions can be regarded as constituting an algebra of *hypercomplex numbers* extending the algebra of the real numbers. The nineteenth century saw the creation of many new algebras of hypercomplex numbers which, while different from the quaternion algebra, continued to share many of its properties.

The British algebraist *Arthur Cayley* (1821–1895) (and independently his friend *J. T. Graves*) formulated an eight unit generalization of the quaternions—the *octonion algebra*. His units were $\mathbf{1}, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_7$ with multiplication table

$$\mathbf{e}_i^{\,2} = -\mathbf{1}, \quad \mathbf{e}_i\mathbf{e}_j = -\mathbf{e}_j\mathbf{e}_i \text{ for } i, j = 1, 2, \ldots, 7 \text{ and } i \neq j$$
$$\mathbf{e}_1\mathbf{e}_2 = \mathbf{e}_3, \ \mathbf{e}_1\mathbf{e}_4 = \mathbf{e}_5, \ \mathbf{e}_1\mathbf{e}_6 = \mathbf{e}_7, \ \mathbf{e}_2\mathbf{e}_5 = \mathbf{e}_7, \ \mathbf{e}_2\mathbf{e}_4 = -\mathbf{e}_6, \ \mathbf{e}_3\mathbf{e}_4 = \mathbf{e}_7, \ \mathbf{e}_3\mathbf{e}_5 = \mathbf{e}_6,$$

together with the fourteen equations obtained from the last seven by cyclic permutation of the indices; e.g. $\mathbf{e}_2\mathbf{e}_3 = \mathbf{e}_1$, $\mathbf{e}_3\mathbf{e}_1 = \mathbf{e}_2$. An *octonion* is defined to be an expression of the form

$$u_0 + u_1\mathbf{e}_1 + \ldots + u_7\mathbf{e}_7 = \mathbf{u},$$

where the u_i are real numbers. The *norm* $\|\mathbf{u}\|$ of $\mathbf{u}$ is defined analogously as for quaternions, viz.,

$$\|\mathbf{u}\| = u_0^2 + u_1^2 + \ldots + u_7^2.$$

The product of octonions is obtained termwise as expected:

$$(u_0 + u_1\mathbf{e}_1 + \ldots + u_7\mathbf{e}_7)(v_0 + v_1\mathbf{e}_1 + \ldots + v_7\mathbf{e}_7) = u_0v_0 + \ldots + u_iv_j\,\mathbf{e}_i\mathbf{e}_j + \ldots,$$

where each unit product $\mathbf{e}_i\mathbf{e}_j$ is calculated according to the above table.

Unlike quaternions, multiplication of octonions is *nonassociative*, that is, in general $\mathbf{u}(\mathbf{w}) \neq (\mathbf{uv})\mathbf{w}$ (nor is it commutative). Nevertheless, like the quaternions, the octonions form a *division algebra* (a fact which seems to have escaped Cayley) in the sense that any octonion $\mathbf{u} \neq \mathbf{0}$ has a reciprocal $\mathbf{u}^{-1}$ satisfying $\mathbf{u}^{-1}\mathbf{u} = \mathbf{u}\mathbf{u}^{-1} = \mathbf{1}$. In fact we may take

$$\mathbf{u}^{-1} = (u_0 - u_1\mathbf{e}_1 - \ldots - u_7\mathbf{e}_7)/\|\mathbf{u}\|.$$

The algebra $\mathbb{C}$ of complex numbers, $\mathbb{H}$ of quaternions, and $\mathbb{O}$ of octonions are all examples of what are known as *linear algebras* over the field $\mathbb{R}$ of real numbers—as is the algebra ℝ itself. This means that the elements of each can be added and multiplied by real numbers and with one another in such a way that the usual algebraic laws—with the possible exception of commutativity and associativity of multiplication—are satisfied. Each is of *finite dimension* in the sense that it is generated by a finite number of "units"—1, 2, 4, and 8 for ℝ, ℂ, ℍ and 𝕆 respectively. Each is a *division algebra* in the sense that every nonzero element has a reciprocal. Multiplication in ℝ, ℂ, and ℍ is associative, and in ℝ and ℂ, commutative. Once these facts were recognized, mathematicians took the natural step of asking whether any more algebras of this kind remained to be discovered. Gauss was convinced that an extension of the complex numbers preserving all its properties was impossible. When Hamilton abandoned the search for a three-dimensional algebra in favour of the four-dimensional algebra of quaternions he had no proof that a three-dimensional algebra could not exist. And neither did Grassmann.

In the 1860s the American mathematician *Benjamin Peirce*[3] (1809–1880) and his son *Charles Sanders Peirce* (1839–1914) undertook a general investigation of associative linear algebras. The elder Peirce found 162 such algebras, but the only division algebras among them were ℝ, ℂ, and ℍ. In 1881 the younger Peirce and the German mathematician *Georg Frobenius* (1849–1917) independently proved that these three are the *only* associative finite dimensional linear division algebras over ℝ. The question of whether there are any *non*associative division algebras of this kind other than the octonion algebra 𝕆 was not resolved—in the negative—until 1958, and required the full resources of modern algebraic topology for its resolution.

These algebras have a connection with a problem concerning *sums of squares* considered by *Adolf Hurwitz* (1859–1919) in 1898. Certain identities concerning sums of squares of real (or complex) numbers had been known to mathematicians for some time. Each took the form

$$\left(\sum_1^n x_i^2\right)\left(\sum_1^n y_i^2\right) = \sum_1^n z_i^2 \qquad (1)$$

where the z_i have the form

$$z_i = \sum_{j,k=1}^n a_{ijk} x_i y_k ,$$

with a_{ijk} real numbers. The known ones were identities for $n = 1, 2, 4, 8$. The first of these is the trivial one:

$$x_1^2 y_1^2 = (x_1 y_1)^2 .$$

The next two are already nontrivial, namely

---

[3] It was the elder Peirce who, in 1870, formulated the well-known definition: *Mathematics is the science that draws necessary conclusions.*

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 - x_2y_2)^2 + (x_1y_2 + x_2y_1)^2$$

$$\begin{aligned}(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) &= (x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4)^2 \\ &+ (x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3)^2 \\ &+ (x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2)^2 \\ &+ (x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1)^2.\end{aligned}$$

The corresponding identity for $n = 8$ is somewhat tedious to write down. It is not known who first discovered the above identity for $n = 2$. The one for $n = 4$ (which plays a central role in the proof of Lagrange's theorem—mentioned in Chapter 3—that every number is the sum of four squares) seems to have been found by Euler and the one for $n = 8$ by C.F. Degen in 1822.

Each of these identities can be deduced from the properties of multiplication in $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$. For instance, when $n = 2$,

$$\begin{aligned}(x_1^2 + x_2^2)(y_1^2 + y_2^2) &= (x_1 + \mathrm{i}x_2)(x_1 - \mathrm{i}x_2)(y_1 + \mathrm{i}y_2)(y_1 - \mathrm{i}y_2) \\ &= (x_1 + \mathrm{i}x_2)(y_1 + \mathrm{i}y_2)(x_1 - \mathrm{i}x_2)(y_1 - \mathrm{i}y_2) \\ &= [x_1y_1 - x_2y_2 + \mathrm{i}(x_1y_2 + x_2y_1)][x_1y_1 - x_2y_2 - \mathrm{i}(x_1y_2 + x_2y_1)] \\ &= (x_1y_1 - x_2y_2)^2 + (x_1y_2 - x_2y_1)^2.\end{aligned}$$

Matrices

One of the most fruitful algebraic concepts to have emerged in the nineteenth century is that of a *matrix* which was originally introduced by Cayley in 1855 as a convenient notation for linear transformations. A pair of equations of the form

$$x' = ax + by$$
$$y' = cx + dy$$

is said to specify a *linear transformation* of x, y into x', y'. Cayley introduced the array of numbers

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

—a 2×2 *matrix*— to represent this transformation. In general, the $m \times n$ matrix

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

represents the linear transformation of the variables $x_1,\ldots, x_n$ into the variables $x_1',\ldots, x_m'$ given by the equations

$$\begin{aligned} x_1' &= a_{11}x_1 + \ldots + a_{1n}x_n \\ &\vdots \\ x_m' &= a_{m1}x_1 + \ldots + a_{mn}x_n. \end{aligned}$$

An $n \times n$ matrix is said to be *square*.

Cayley defined a number of operations on matrices and showed essentially that, for each n, the $n \times n$ matrices form a linear associative algebra over the real numbers. Thus, for $n = 2$, Cayley defined the *sum* of the two matrices

$$\mathbf{M} = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathbf{M}' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$$

to be the matrix

$$\mathbf{M} + \mathbf{M}' = \begin{pmatrix} a + a' & b + b' \\ c + c' & d + d' \end{pmatrix}$$

and, for any real number α, the *product* of **M** by α to be the matrix

$$\alpha\mathbf{M} = \begin{pmatrix} \alpha a & \alpha b \\ \alpha c & \alpha d \end{pmatrix}$$

To multiply two matrices Cayley analyzed the effect of two successive transformations. Thus if the transformation

$$\begin{aligned} x' &= a_{11}x + a_{12}y \\ y' &= a_{21}x + a_{22}y \end{aligned}$$

is followed by the transformation

$$\begin{aligned} x'' &= b_{11}x' + b_{12}y' \\ y'' &= b_{21}x' + b_{22}y', \end{aligned}$$

then x'', y'' and x, y are connected by the equations

$$x'' = (b_{11}a_{11} + b_{12}a_{21})x + (b_{11}a_{12} + b_{12}a_{22})y$$
$$y'' = (b_{21}a_{11} + b_{22}a_{21})x + (b_{21}a_{12} + b_{22}a_{22})y.$$

Accordingly Cayley defined the product of the matrices to be

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} \\ b_{21}a_{21} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{pmatrix}$$

Multiplication of matrices is associative but not generally commutative. Cayley observed that—according to his definition—an $m \times n$ matrix can be multiplied only by an $n \times p$ matrix. The *zero* and *unit* $n \times n$ matrices are given by

$$\mathbf{0} = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \qquad \mathbf{I} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

These have the property that for every $n \times n$ matrix $\mathbf{M}$,

$$\mathbf{M + 0 = 0 + M = M, \quad 0.M = M.0 = 0, \; M.I = I.M = M.}$$

The *inverse* $\mathbf{M}^{-1}$ —if it exists—of a matrix $\mathbf{M}$ satisfies $\mathbf{MM}^{-1} = \mathbf{M}^{-1}\mathbf{M} = \mathbf{I}.$ Cayley noted that some matrices fail to have inverses in this sense, so that matrices do not in general form a division algebra. In calculating inverses of matrices Cayley made use of *determinants,* devices originally introduced in the eighteenth century in connection with the solution of systems of linear equations and from which the concept of matrix itself may be considered to have originated. The *determinant* det($\mathbf{M}$) of a 2×2 or 3×3 matrix $\mathbf{M}$ is given by the following rules: if

$$\mathbf{M} = \begin{pmatrix} a & b \\ a' & b' \end{pmatrix}$$

then

$$\det(\mathbf{M}) = \begin{vmatrix} a & b \\ a' & b' \end{vmatrix} = ab' - ba';$$

if

$$\mathbf{M} = \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}$$

then

$$\det(\mathbf{M}) = \begin{vmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{vmatrix} = a\begin{vmatrix} b' & c' \\ b'' & c'' \end{vmatrix} - b\begin{vmatrix} a' & c' \\ a'' & c'' \end{vmatrix} + c\begin{vmatrix} a' & b' \\ a'' & b'' \end{vmatrix}$$
$$= ab'c'' + bc'a'' + ca'b'' - (ac'b'' + ba'c'' + cb''a').$$

The significance of the determinant in connection with linear equations arises from the fact that a system of equations such as

$$a + bx + cy = 0$$
$$a' + b'x + c'y = 0$$
$$a'' + b''x + c''y = 0$$

has a solution if and only if

$$\begin{vmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{vmatrix} = 0$$

The *minor* corresponding to an entry m in a square matrix **M** is the determinant of the matrix obtained from **M** by deleting the row and column containing m. Using this idea we can determine the inverse of the 3×3 matrix **M** above: it is given by

$$\mathbf{M}^{-1} = \frac{1}{\det(\mathbf{M})}\begin{pmatrix} A & -A' & A'' \\ -B & B' & -B'' \\ C & -C' & C'' \end{pmatrix}$$

where A is the minor of a, A' that of a', etc. Notice that $\mathbf{M}^{-1}$ exists only when det $(\mathbf{M}) \neq 0$. When det$(\mathbf{M}) = 0$, Cayley called **M** *indeterminate* (the modern term is *singular*).

So we see that determinants are two-dimensional devices for representing *individual numbers,* and matrices may be considered as having arisen from determinants through a literal reading of their two-dimensional character. Matrices may accordingly be regarded as *two-dimensional numbers.*

An important result concerning matrices announced by Cayley (but not proved fully by him) is now known as the *Cayley-Hamilton theorem* This involves the so-called *characteristic equation* of a matrix, which for a matrix **M** is defined to be

$$\det(\mathbf{M} - x\mathbf{I}) = 0.$$

Accordingly if

$$\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

the characteristic equation is

$$x^2 - (a + d)x + ad - bc = 0.$$

The Cayley-Hamilton theorem is the assertion that, if **M** is substituted for x in its characteristic equation, the resulting matrix is the zero matrix. Hamilton had proved the special case for 3×3 matrices; the first complete proof was given by Frobenius in 1878.

The roots of the characteristic equation of a matrix **M** turn out to be of particular importance. Consider, for example, the 3×3 matrix

$$\mathbf{M} = \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}.$$

M represents the linear transformation given by the equations

$$\begin{aligned} ax + by + cz &= u \\ a'x + b'y + c'z &= v \\ a''x + b''y + c''z &= w \end{aligned}$$

which may in turn be written

$$\begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} u \\ v \\ w \end{pmatrix} \qquad (1)$$

Here the *column matrices*

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad \mathbf{u} = \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

may be thought of as 3-dimensional vectors with components x, y, z and u, v, w, respectively. Equation (1) may then be written

$$\mathbf{Mx} = \mathbf{u}, \qquad (2)$$

which expresses the idea that **M** *transforms* the vector **x** into the vector **u**. Thus a 3×3 matrix can be thought of as a *linear operator* on the space $\mathbf{R}^3$ of 3-dimensional vectors, and in general, an $n \times n$ matrix as a linear operator on the space $\mathbf{R}^n$ of n-dimensional vectors.

Ordinarily, the vectors **x** and **Mx** do not lie in the same line—are not *collinear*— but an important fact in applications is that, for certain vectors $\mathbf{x} \neq \mathbf{0}$, **Mx** and **x** are collinear, that is, there exists a real number λ for which

$$\mathbf{Mx} = \lambda\mathbf{x}. \qquad (3)$$

The *characteristic value* or *eigenvalue problem* is the following : given an $n \times n$ matrix **M**, for which vectors $\mathbf{x} \neq \mathbf{0}$ and for what numbers λ does (3) hold? Numbers λ satisfying (3) for some vector **x** are called *characteristic values* or *eigenvalues* (from German *eigen*, "own") of **M.**

There is an important connection between the eigenvalues of a matrix and its characteristic equation which is revealed when equation (3) is rewritten in the form

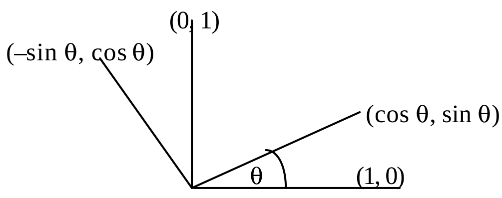
$$(\mathbf{M} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}.$$

This equation represents a system of n homogeneous[4] first-degree equations in n unknowns, and it is a basic fact about such a system that it has a nonzero solution precisely when

$$\det(\mathbf{M} - \lambda\mathbf{I}) = 0.$$

It follows that the eigenvalues of **M** are precisely the roots of its characteristic equation.

Many natural *geometric* operations in the plane or in space give rise to linear transformations, which in turn may be conveniently represented by their associated matrices. In the plane, for instance, a counterclockwise rotation about the origin through an angle θ:



[4] A system of first-degree equations in the unknowns $x, y, z, \ldots$ is said to be *homogeneous* if each is of the form $ax + by + cz + \ldots = 0$.

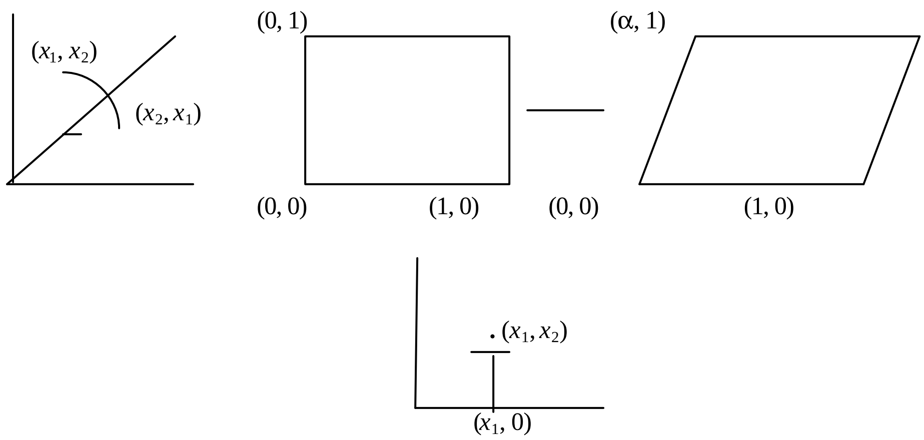
is represented by the matrix

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

so that, in particular, a 90° rotation is represented by

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

Reflection in a 45° line, "shear," and projection onto the horizontal axis:



are linear transformations represented by the matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Matrices play an important role in representing algebras of hypercomplex numbers. Consider, for instance, the algebra $\mathbb{C}$ of complex numbers. If with each complex number $a + b$i we associate the 2×2 matrix

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

then it is easily verified that the sum and product of complex numbers corresponds exactly to the sum and product of the associated matrices. In particular the matrix associated with i itself is

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

precisely the matrix corresponding to a 90° counterclockwise rotation.

Quaternions may be represented by *4 ´ 4 matrices*. In this case, with the quaternion $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ is associated the matrix

$$\begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix},$$

so that the matrices associated with **i, j, k** are, respectively,

$$\mathbf{i^*} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \mathbf{j^*} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \quad \mathbf{k^*} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

If we now identify **1, i, j, k** with the "column vectors"

$$\underline{\mathbf{1}} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \underline{\mathbf{i}} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \underline{\mathbf{j}} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \underline{\mathbf{k}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

then it is readily checked by matrix multiplication that

$$\mathbf{j^*}\underline{\mathbf{k}} = \underline{\mathbf{i}}, \quad \mathbf{k^*}\underline{\mathbf{j}} = -\underline{\mathbf{i}}, \quad \mathbf{k^*}\underline{\mathbf{i}} = \underline{\mathbf{j}}, \quad \mathbf{i^*}\underline{\mathbf{k}} = -\underline{\mathbf{j}}, \quad \mathbf{i^*j} = \underline{\mathbf{k}}, \quad \mathbf{j^*}\underline{\mathbf{i}} = -\underline{\mathbf{k}},$$

$$\mathbf{i^*}\underline{\mathbf{i}} = \mathbf{j^*j} = \mathbf{k^*}\underline{\mathbf{k}} = -\mathbf{1}.$$

These are just the multiplication tables for the quaternion units written in terms of the matrices **i*, j*, k*** and the vectors $\underline{\mathbf{1}}, \underline{\mathbf{i}}, \underline{\mathbf{j}}, \underline{\mathbf{k}}$.

It can be shown that *every* linear associative algebra over **R** can be "represented" as an algebra of matrices in this way. To put it succinctly, *hypercomplex numbers are representable as two-dimensional numbers.*

Quaternions can also be represented by matrices built from *complex* numbers. In fact the algebra of quaternions is precisely the same as the algebra generated by the following special matrices with (some) imaginary entries:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad s_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \quad s_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad s_3 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$$

where, as usual, $i = \sqrt{-1}$. One easily verifies that under matrix multiplication **I,** s_1**,** s_2**,** s_3**, I** satisfy precisely the same rules as do the quaternion units **1, i, j, k.** In this representation the general quaternion $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ is correlated with the comp lex matrix

$$\begin{pmatrix} a - id & c + ib \\ -c + ib & a + id \end{pmatrix}$$

These 2×2 complex matrices (and the corresponding quaternions) play an important role in modern quantum physics. There they are regarded as embodying rotations of two dimensional complex vectors known as *spinors,* which describe the state of an electron—a particle with *spin*, that is, intrinsic angular momentum. The matrices s_1, s_2 and s_3, multiplied by –i, are the *Pauli spin matrices.* Even the visionary Hamilton, with all his enthusiasm for quaternions, could not have foreseen this application of his invention!

Lie Algebras

While Cayley's algebra of octonions was the first nonassociative linear algebra to be discovered, it is really not much more than a mathematical curiosity. The most significant type of nonassociative algebra—now known as a *Lie algebra*—was introduced in 1876 by the Norwegian mathematician *Sophus Lie* (1842–1899). These algebras were the distillation of Lie's study of the structure of groups of spatial transformations, which in turn involved the idea of an *infinitesimal transformation* in space, that is, one which moves any point in space an infinitesimal distance. Infinitesimal transformations may be added and multiplied by real numbers in a natural way. Given two infinitesimal transformations *X, Y,* we may also consider their *composite XY,* that is, the transformation that results by applying first *Y* and then *X*. Now *XY* and *YX* are in general not the same, nor indeed are they infinitesimal transformations. However, the transformation

$$XY - YX$$

is infinitesimal: it is called the *commutator* or *Lie product* of *X* and *Y* and is written $[X, Y]$. If we regard this Lie product as a multiplication operation on infinitesimal transformations, then these form a *nonassociative* linear algebra over **R**. Moreover, the Lie product satisfies the laws

(1) $[X, X] = 0$
(2) $[X, Y] = -[Y, X]$
(3) $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0.$

Equation (3), known as the *Jacobi identity*, had already arisen in connection with the researches of the German mathematician *C. J. G. Jacobi* (1804–1851) into the solution of partial differential equations. Equations (1) and (2) are immediate consequences of the definition of $[X, Y]$, and 3) follows by noting that, if in

$$[[X, Y], Z] = XYZ - YXZ - ZXY + ZXY$$

we permute X, Y and Z cyclically, and add all the terms so obtained, we obtain zero.

A linear algebra satisfying (1), (2) and (3) is called a *Lie algebra*. A noteworthy example of a Lie algebra is the algebra of vectors in 3-dimensional space with multiplication given by Grassmann's combinatory or vector product. Any linear associative algebra gives rise to a corresponding Lie algebra whose Lie product is determined by

$$[x, y] = xy - yx.$$

So in particular every algebra of $n \times n$ matrices can be construed as a Lie algebra. Lie himself initially believed he had shown that every Lie algebra can be represented as a Lie algebra of matrices in this sense, but he quickly came to recognize that his proof was incomplete. This fact was established conclusively only in 1935 by the Russian mathematician *I. Ado*.