Linear Algebra, Vector Analysis, and the Beginnings of Functional Analysis

© S. A. Fulling 2003

Portions appeared previously in *Linearity and the Mathematics of Several Variables*, by S. A. Fulling, World Scientific, Singapore, 2000.

Linear algebra is a vast, multilevel subject, ranging from the solution of two linear equations in two unknowns to the abstract study of vector spaces of infinitely many dimensions and their practical realizations as spaces of functions, which provide the arenas for differential and integral equations.

Ninth-graders learn how to solve linear equations by "addition and subtraction": Suppose

$$x - 2y = 0,$$
$$3x + 4y = 5.$$

Subtract 3 times the first equation from the second:

$$10y = 5.$$

Divide by 10 to get half of the answer:

$$y = \tfrac{1}{2}\,.$$

Add twice this result to the other equation to get the other answer:

$$x = 1.$$

The first step in a college linear algebra course is to learn how to mechanize and systematize this process by pulling out the meaningful numbers of the problem into a table called a *matrix*, and then performing the process of *row reduction* or *Gaussian elimination*:

$$\begin{pmatrix} 1 & -2 & 0 \\ 3 & 4 & 5 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 0 \\ 0 & 10 & 5 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & \frac{1}{2} \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \end{pmatrix}.$$

The final matrix represents the system of equations

$$x = 1,$$
$$y = \tfrac{1}{2}\,,$$

which is logically equivalent to the original system but is already "solved".

Row reduction (not just for two equations and unknowns) is demonstrated in a Chinese manuscript from roughly 200 B.C.E. (Actually, the Chinese did *column* reduction, since they wrote the equations *down* the page like any other Chinese sentence.) As usual in ancient cultures, the method was presented by an example, without explanation of whether and how it was guaranteed to work in a general case. It is striking that the full development of the general method in Western mathematics not only required 2000 more years, but has the name of one of the most famous mathematicians of all time, C. F. Gauss, attached to it.

The transition from addition-and-subtraction to row reduction demonstrates two of the familiar themes of mathematical progress:

1. the codification of an *algorithm* stating the strategy to be followed in solving a general class of problems, as opposed to hit-or-miss progress on particular problems. (Every algebra teacher knows that students are likely to add-and-subtract in circles until given some guidance on how to proceed systematically. "Make the lower left corner of the matrix all zeros" summarizes the guidance.)
2. the *abstraction* of new kinds of mathematical objects. The two unknowns go together to make a single, two-component thing, a *vector* $\binom{x}{y}$, and the two equations are combined into a single *linear function* or *operator* that maps vectors into vectors. This function is represented by a matrix:

$$\begin{pmatrix} 1 & -2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}.$$

   (The matrix in the algorithm is formed by sticking together the matrix of the function and the vector that appears as the value of the function at the unknown point.)

Since around 1950 linear algebra has been a standard and rather static element of the college mathematics curriculum. Nevertheless, linear algebra as we know it today is a rather recent consensus. Its serious development began around 1800; it reached its present form in the 1920s; and *only then* did it attain its central, fundamental place in pure and applied mathematics.

The concept of a vector emerged out of two distinct areas of mathematics, geometry and algebra. "Algebra" in the nineteenth century meant largely the study of the transformation (with luck, simplification) of equations by various changes of variables, but also the study of "systems" of "quantities" satisfying various rules for multiplication, etc. Near the beginning of the nineteenth century, several people[1] independently recognized that complex numbers could be represented by line segments (characterized by length and direction) in a plane, and vice versa. It was natural to ask whether three-dimensional space also had such an algebraic structure.

The Irish mathematician and theoretical physicist William Rowan Hamilton (1895–1865) devoted the last two-thirds of his career[2] to this question — first to finding an answer, then to developing and promoting it. Progress was hampered by the fact that nineteenth-century mathematicians were trying to think of vectors as a new kind of "number", much as

[1] Notably Caspar Wessel (1745–1818), a Norwegian surveyor who published no other mathematical research.

[2] The first third was devoted to landmark work on optics and particle mechanics. Hamilton showed that, mathematically, high-frequency waves behave like ensembles of particles moving along a family of roughly parallel trajectories. This explained why calculations can be done successfully with "light rays" although experiments show that light is actually a wave phenomenon. Conversely, a problem in classical mechanics could be converted into a wavelike partial differential equation (Hamilton–Jacobi equation), which became, a century later, the foundation of our understanding of how classical particle-like behavior of electrons, etc., emerges from quantum mechanics, which is based on a wave equation (Schrödinger equation).

the negative integers had grown out of the positive ones and had in turn been extended to the rational numbers and then to the complex numbers. (Logically, the real numbers should come between the rationals and the complexes, but historically a solid understanding of the reals did not come until later in the century.) It was expected, therefore, that two vectors could be multiplied or divided to yield a third vector. Hamilton discovered that multiplication of three-dimensional vectors was possible, at the cost of abandoning the commutative law of multiplication; thereby he almost invented the modern vector cross product,

$$\mathbf{v} \times \mathbf{w} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{pmatrix}, \tag{1}$$

which is central to many formulas of physics. However, he also maintained the associative law of multiplication and the possibility of division (both false for the vector cross product) at the cost of introducing a fourth dimension. He called his new hypercomplex numbers *quaternions*, and he and his followers proceeded to reformulate physics in terms of them. Quaternions did indeed provide a way of thinking about three-dimensional geometry and physics that was independent of any particular choice of coordinate axes; velocities and forces could be comprehended as single objects, rather than triples of numbers.

Unfortunately, Hamilton had been somewhat misled by a numerical coincidence special to three dimensions. To explain it, let's look back at *rotations* in two-dimensional space. A rotation in two dimensions through an angle θ is represented by the matrix

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

If we differentiate (each element of) R with respect to θ and then set $\theta = 0$, we get

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

This matrix (let's call it L) is *antisymmetric*; that is, it changes sign when rows are interchanged with columns — in index notation, $L_{kj} = -L_{jk}$. Now consider a rotation in n-dimensional space, again through a variable angle θ. The geometrical property of being a rotation can be expressed by the equation

$$R(\theta)^{\mathrm{t}} R(\theta) = I$$

where I is the n-dimensional *unit matrix* and R^{t} is the *transpose*[3] of R. The product of two matrices (defined implicitly in the footnote above) satisfies the ordinary product rule

[3] In index notation, $R^{\mathrm{t}}_{jk} = R_{kj}$ and the rotation equation reads

$$\sum_{i=1}^{n} R_{ij} R_{ik} = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

under differentiation, so one has

$$\frac{d}{d\theta}\left[R(\theta)^{\mathrm{t}}\right]R(\theta)+R(\theta)^{\mathrm{t}}R'(\theta)=0.$$

In particular, when $\theta=0$,

$$R'(0)^{\mathrm{t}}=\frac{d}{d\theta}\left[R(\theta)^{\mathrm{t}}\right]\Big|_0=-R'(0),$$

so $R'(0)$ is an antisymmetric matrix. The big difference from dimension 2 is that there L is the only possible antisymmetric matrix, up to a numerical factor, but in dimension n the antisymmetric matrices form a space of dimension $\frac{1}{2}n(n-1)$: The matrix elements $R'(0)_{jj}$ along the diagonal must be 0, and the elements above the diagonal are determined by those below, but there are still $\frac{1}{2}n(n-1)$ elements below the diagonal ($R'(0)_{jk}$ with $j>k$) that can be chosen completely freely. For example, an antisymmetric matrix in dimension 3 has the form

$$R'(0)=\begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}. \tag{2}$$

The amazing coincidence is that 3 is the only solution (except 0) of the equation

$$\tfrac{1}{2}n(n-1)=n.$$

That is, for the physical space we live in, the dimension of the space of antisymmetric matrices (and hence the number of independent types of rotations[4]) is the same as the dimension of the original space. Furthermore, comparing equations (2) and (1) one can see that

$$R'(0)\mathbf{w}=\mathbf{v}\times\mathbf{w}.$$

That is, in three dimensions (and only there) an antisymmetric matrix (or rotation type[5]) can be represented by a vector, and the deep meaning of the vector cross product (1) is that it expresses the action of an antisymmetric matrix upon other vectors. Every physics formula containing a cross product contains at least one antisymmetric matrix, disguised as a vector!

In the mid-1800s the foregoing was far from understood. From a modern point of view, in the quaternion formalism the vectors and the rotation operators that act upon them are jumbled together. In fact, Hamilton consistently made an error of a factor of 2 in rotation angles as a result of trying to force the vectors and the rotations to be the same

4 For experts: the dimension of the rotation group, regarded as a manifold.

5 For the experts: a one-parameter subgroup of rotations.

thing. The relation between vectors and rotations was more accurately understood in that era by a less famous mathematician, Olinde Rodrigues (1795(?)–1850(?))[6].

Simultaneously, a different algebraic approach to geometry was developed by Hermann Grassmann (1809–1877), a German schoolteacher. His theory, called *Ausdehnungslehre* ("theory of extension"), had the great advantage of applying in all dimensions, not just 3. He was largely concerned with the relation between p-dimensional subspaces and antisymmetric algebraic combinations of p vectors. Grassmann's best ideas were almost a century ahead of their time, and they were buried in other material that turned out to be irrelevant or misguided. His books were considered unreadable, even by other Germans, so for a long time he was essentially ignored.

The era of Hamilton and Grassmann, occupying the middle of the nineteenth century, was the Mesozoic Era of vector theory. The last two decades of that century saw the emergence of vectors and vector spaces in their modern forms.

On the one hand, J. Willard Gibbs (American physicist, 1839–1903)[7] and Oliver Heaviside (British engineer, 1850–1925)[8] distilled from the quaternion theory its useful essence, the three-dimensional vectors we see in physics and engineering textbooks today, with their dot and cross products and nabla[9] operators. This quickly became the standard language of physics and applied mathematics.

On the other hand, several mathematicians in Italy, especially Giuseppe Peano (1858–

---

[6] According to the article of Altmann (1989), many basic facts of Rodrigues's life are obscure. He was a French Jew of Spanish or Portuguese ancestry. In his 1816 doctoral dissertation he gave a formula for the nth-degree Legendre polynomial (a key ingredient in the solution of Laplace's equation and other partial differential equations in spherical coordinates), and it is for that for which he is best known. At that time it was impossible for Jews to get good academic positions in France, so Rodrigues became a banker, like most of his family. He was active in the (pre-Marx) socialist movement, but remained sufficiently acceptable to the establishment as to be "influential in the development of the French railways" (Altmann). Somehow he managed in the midst of this career to publish his fundamental but neglected paper on the rotation group in 1840.

[7] Gibbs was the first important American theoretical physicist. He spent almost all of his career at Yale University — indeed, almost all his life, since his father was a professor there. His most important physics research was in the foundations of thermodynamics and statistical mechanics. Of interest here is that he established the notation used in vector calculus to this day, especially in applied work. His third most famous contribution, the "Gibbs phenomenon", is a side remark that occurred while he was trying to mediate a dispute in the Letters column of *Nature* about the convergence of Fourier series: He discovered that the partial sums of a Fourier series "overshoot" near a discontinuity in the function they are approximating.

[8] Heaviside, a major figure in the development of electromagnetic theory, was apparently the first to write Maxwell's equations for the electromagnetic field in a vectorial form, instead of as separate equations for each component. He was the prototype of the engineer or physicist with impatience for the mathematicians' demands for rigorous proofs. He developed an influential but controversial method for solving differential equations, called "operational calculus", which was eventually turned into respectable mathematics by the work of T. J. Bromwich (1875–1929) on Laplace transforms and complex integrals and of Laurent Schwartz (1915–2002) on "distributions" as linear functionals.

[9] Nineteenth-century British mathematical physicists named the symbol ∇ used to construct the differential operators of gradient, divergence, curl, and Laplacian (for example, $\nabla \cdot \mathbf{A} = \frac{\partial A_1}{\partial x} + \frac{\partial A_2}{\partial y} + \frac{\partial A_3}{\partial z}$) after a certain kind of Assyrian harp, the nabla.

1932)[10], were separating the wheat from the chaff in Grassmann's work. The result was the modern definition of a vector space. However, this development did not attract much attention at the time. A vector space is a set of mathematical objects that can be added to each other and multiplied by numbers (not necessarily multiplied by each other). The addition and multiplication operators are required to satisfy 8 basic properties that generalize the associative, distributive, etc. laws of elementary algebra.

The triumph of the abstract vector concept did not happen until the study of *infinite-dimensional* vectors made it necessary. A number of mathematicians, notably Stefan Banach (1892–1945), independently reinvented the concept of a vector space (equipped with an inner product or a norm, in order to define distances and convergence unambiguously) around 1920 in order to deal with infinite-dimensional problems, in which the vectors were functions or infinite sequences. (Banach was Polish, and Poland has remained a center of research on functional analysis to this day.) It is striking that this development roughly coincided with the rise of quantum mechanics in physics. Quantum theory would be almost impossible without the idea of an infinite-dimensional vector space, but it was not quantum theory that sparked the creation of the mathematical theory. Functional analysis, or infinite-dimensional linear algebra, developed just a few years earlier, largely in response to purely classical physical problems involving integral and differential equations.

"Banach spaces" and more general kinds of infinite-dimensional vector spaces are the subject of intense pure mathematical research still today. It is worth noting, however, that the title of Banach's own book on the subject is *Théorie des Opérations Linéaires* (and not " ... *des Espaces* ... "). It reminds us that we started with two equations to be solved for two unknowns, and the vector space $\mathbf{R}^2$ came in as a conceptual tool for dealing with that problem. That is, the important thing was the matrix

$$M = \begin{pmatrix} 1 & -2 \\ 3 & 4 \end{pmatrix},$$

or rather the linear function that it represents, and the vector space was just a home for that function to act in. Similarly, functional analysis is rooted in applications that involve linear operations on functions. In analogy to our 2×2 algebraic system, consider the ordinary differential equation

$$\frac{d^2y}{dx^2} + 4y = f(x) \quad (\text{for } 0 < x < 1),$$

to be solved for $y(x)$ subject to the boundary conditions

$$y(0) = 0 = y(1);$$

[10] Peano is best known for framing 5 postulates that give a complete abstract characterization of the natural numbers (nonnegative integers). His second most famous discovery is a continuous curve that completely fills a two-dimensional region, proving that the difference between one and two dimensions is more subtle than it appears. In his later years he displayed a grandiose vision reminiscent of Leibniz (and very different from most 20th-century scientists, who concentrated on technical excellence in narrowly circumscribed areas or problems): He devoted himself to the construction of a universal language, based on Latin stripped of its grammar, and a unified formulation of all mathematical knowledge.

f is understood to be given but arbitrary. The modern way of understanding the problem is this:

$$M = \frac{d^2}{dx^2} + 4$$

is a linear operator acting on certain functions (y) and thereby producing other functions (f). To study this operation properly one must be quite careful and technical about the vector spaces where these functions lie. We can take the domain (space of allowed ys) to be all the real-valued functions on the interval $(0, 1)$ that are twice differentiable and vanish at the endpoints. (The first condition seems to be necessary for My to be defined; the second condition conveniently builds the boundary conditions into the definition of the space and assures that the solution will be unique.) Then we could take the codomain (space of allowed fs) to be all the continuous real-valued functions on $(0, 1)$.[11] Now one can show that M has an inverse: The solution of the differential equation can be written

$$y = Gf,$$

where G is another linear function, mapping the codomain space into the domain. In this case one can actually find an integral formula

$$y(x) = \int_0^1 G(x, z) f(z)\, dz,$$

$$G(x, z) = \frac{\sin(2 \min(x, z)) \sin(2(1 - \max(x, z))}{-2 \sin 2},$$

so the analogy between G and a finite-dimensional matrix is really quite close.

In more recent years both Grassmann and Hamilton have been partially vindicated. Grassmann's geometrical ideas evolved into the modern theory of *differential forms*. Hamilton's strange four-dimensional space, with a few sign changes, turned out to be related to the space-time of special relativity, in a formalism well adapted to the study of rotations acting on the fields and wave functions that represent electrons and other half-integral-spin particles (*fermions*). At the same time that Gibbs was cleaning up the vector half of the quaternion formalism, mathematicians such as Felix Klein (1849–1925) and Sophus Lie (1842–1899) were, in effect, rescuing its rotational half, by creating the modern theory of the groups of rotations in n-dimensional space, and the more general theory of Lie groups and Lie algebras. In short, the *Pauli matrices* of quantum mechanics,

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

are Hamilton's unit quaternions in disguise — or, better, in their true form with their nineteenth-century disguise removed. The quaternions in their original form retain some

[11] In fact, it turns out to be better to include certain functions that are not continuous, because they are limits of sequences of functions that are. This is analogous to enlarging the rational numbers to the real numbers. For experts: The codomain is the Hilbert space $L^2(0, 1)$ and the domain is the Sobolev space $H_0^2(0, 1)$.

followers among applied researchers whose work heavily involves rotations (see Kuipers 1999). They have always been of interest to pure mathematicians as a very special kind of algebraic system.

Bibliography

S. L. Altmann, Hamilton, Rodrigues, and the quaternion scandal, *Math. Mag.* **62**, 291–308 (1989).

S. Banach, *Théorie des Opérations Linéaires*, 1932, reprinted Chelsea Publ. Co., 1978.

C. C. Cowen, When did linear algebra enter the curriculum?, *Focus* (MAA newsletter) **17** No. 4, 6–7 (1997).

M. J. Crowe, *A History of Vector Analysis*, University of Notre Dame Press, 1967, reprinted Dover Publications, 1994.

J.-L. Dorier, A general outline of the genesis of vector space theory, *Historia Mathematica* **22**, 227–261 (1995).

J. B. Fraleigh and R. A. Beauregard, *Linear Algebra*, Addison–Wesley, 1987, with historical notes by Victor Katz.

V. J. Katz, The history of Stokes' theorem, *Math. Mag.* **52**, 146–156 (1979).

J. B. Kuipers, *Quaternions and Rotation Sequences, A Primer with Applications to Orbits, Aerospace, and Virtual Reality*, Princeton University Press, 1999.

A. F. Monna, *Functional Analysis in Historical Perspective*, Halsted Press, 1973.

G. H. Moore, The axiomatization of linear algebra: 1875–1940, *Hist. Math.* **22**, 262–305 (1995).

F. Swetz *et al.*, eds., *Learn from the Masters!*, Mathematical Association of America, 1995, especially the articles by V. J. Katz, O. B. Bekken, and K. Reich.

`http://www-groups.dcs.st-and.ac.uk/~history/`, The MacTutor History of Mathematics Archive, especially the pages on "Matrices and Determinants" and "Abstract Linear Spaces".