2

Preamble

I was unable to devote myself to the learning of this al–jabr and the continued concentration upon it, because of obstacles in the vagaries of Time which hindered me; for we have been deprived of all the people of knowledge save for a group, small in number, with many troubles, whose concern in life is to snatch the opportunity, when Time is asleep, to devote themselves meanwhile to the investigation and perfection of a science; for the majority of people who imitate philosophers confuse the true with the false, and they do nothing but deceive and pretend knowledge, and they do not use what they know of the sciences except for base and material purposes; and if they see a certain person seeking for the right and preferring the truth, doing his best to refute the false and untrue and leaving aside hypocrisy and deceit, they make a fool of him and mock him.

Omar Khayyam,
Risala fi'l–barahin 'ala masa'il al–jabr wa'l–muqabala

2.1 A Historical Enigma

Figure 2.1 shows cuneiform tablet no. 322 in the Plimpton Collection of the Rare Book and Manuscript Library at Columbia University. This compilation of sexagesimal (base 60) numbers[1] is believed to originate from the ancient Mesopotamian city Larsa (Tell Senkereh in modern Iraq) and has been dated to the period 1820–1762 BC. It was discovered in the 1920s and acquired in a market by the antiquities dealer Edgar A. Banks, who then sold it for $10 to George A. Plimpton, a New York publisher and a collector of mathematical artifacts. Plimpton bequeathed his entire collection to Columbia University in 1936, but the significance of the tablet was not fully appreciated until a

[1] Our modern use of *minutes* and *seconds* as measures of time and angle can be traced back to the Mesopotamian sexagesimal number system.

Fig. 2.1. Plimpton 322, the "Pythagorean triples" cuneiform tablet from the ancient city of Larsa in Mesopotamia (∼1820–1762 BC). Reproduced with permission from the Plimpton Collection, Rare Book and Manuscript Library, Columbia University.

thorough transcription and analysis of its contents was published [344] in 1945 by Otto Neugebauer and Abraham Sachs at Brown University.

Of all existing cuneiform mathematical tablets, Plimpton 322 has been the subject of the most intense scholarly research [65, 66, 72, 128, 205, 344, 377, 378, 393]. While its numerical content (and even the correction of calculation and transcription errors therein) is no longer in doubt, the interpretation of its mathematical significance and its "purpose" are still the subject of lively debate and reassessment, some 60 years after its initial decipherment.

The tablet measures approximately $5 \times 3\frac{1}{2}$ inches, but is incomplete — a portion has broken off at the left edge, while parts of the available fragment are damaged and hence illegible. Traces of modern glue have been identified along the broken edge, suggesting that the tablet may have been broken after its modern discovery. The available portion, though incomplete, nevertheless reveals a profound degree of numeracy and algebraic sophistication.

The fragment lists fifteen rows of sexagesimal numbers arranged in four columns, with the last column being simply a counter for the rows. A clearer impression may be gained from the drawing by Eleanor Robson [377] shown in Fig. 2.2. Table 2.1 presents a transcription of Plimpton 322 in modern Indo–Arabic numerals [343], with commas employed to separate the coefficients for successive powers of 60. In the second and third columns, it is assumed that the right–most entries are the coefficients of unity — for example, the quantity 3,31,49 in the fourth row, second column is interpreted as

$$3 \times (60)^2 + 31 \times 60 + 49\,,$$

or 12,709 in familiar decimal notation. However, the quantities in the first column apparently all begin with 1, suggesting a different interpretation with

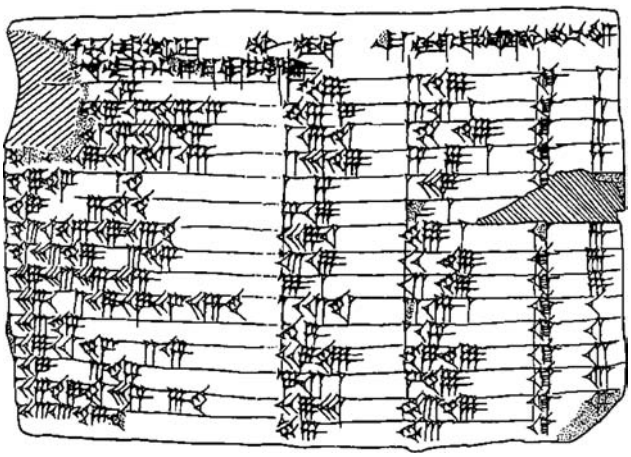

Fig. 2.2. A scale drawing by Eleanor Robson, clarifying the cuneiform sexagesimal numbers tabulated in Plimpton 322 — reproduced with permission from [377].

Table 2.1. Left: the transcription of Plimpton 322 by Neugebauer and Sachs [344], including interpolated missing or corrected values in square brackets. Right: deduced integers p and q that generate the values in the four columns of Plimpton 322.

$f = [\,(p^2+q^2)/2pq\,]^2$	$a = p^2 - q^2$	$c = p^2 + q^2$	#	p	q
[1;59,0,]15	1,59	2,49	1	12	5
[1;56,56,]58,14,50,6,15	56,7	1,20,25	2	1,4	27
[1;55,7,]41,15,33,45	1,16,41	1,50,49	3	1,15	32
[1;]5[3,1]0,29,32,52,16	3,31,49	5,9,1	4	2,5	54
[1;]48,54,1,40	1,5	1,37	5	9	4
[1;]47,6,41,40	5,19	8,1	6	20	9
[1;]43,11,56,28,26,40	38,11	59,1	7	54	25
[1;]41,33,59,3,45	13,19	20,49	8	32	15
[1;]38,33,36,36	8,1	12,49	9	25	12
1;35,10,2,28,27,24,26,40	1,22,41	2,16,1	10	1,21	40
1;33,45	45,0	1,15,0	11	1,0	30
1;29,21,54,2,15	27,59	48,49	12	48	25
[1;]27,0,3,45	2,41	4,49	13	15	8
1;25,48,51,35,6,40	29,31	53,49	14	50	27
[1;]23,13,46,40	56	1,46	15	9	5

the left–most entries as the coefficients of unity.[2] The quantity 1;48,54,1,40 in the fifth row, first column is thus interpreted as

[2] Mesopotamian numbers do not use a "sexagesimal point" to separate whole and fractional parts, and are thus indeterminate by a power of 60 (although this is often resolved by the context). Following Robson [377] we employ semi–colons to denote the putative position of such points.

$$1 + \frac{48}{60} + \frac{54}{(60)^2} + \frac{1}{(60)^3} + \frac{40}{(60)^4} .$$

The parentheses [ ] in Table 2.1 indicate illegible entries that were "restored" by Neugebauer and Sachs, who also corrected several apparent transcription or calculation errors (where the listed values are inconsistent with the overall structure apparent in the tabulation).

From a modern viewpoint, this structure is that the first three columns can be generated from appropriately–selected integers p and q by the expressions

$$f \, = \, \left[\frac{p^2 + q^2}{2pq} \right]^2 , \quad a \, = \, p^2 - q^2 , \quad c \, = \, p^2 + q^2 . \tag{2.1}$$

Deduced values for p and q are appended on the right in Table 2.1, where it can be seen that $1 < q < 60$, $q < p$, and the ratio p/q is steadily decreasing — which also implies that the first–column entries steadily decrease. The column headings in Table 2.1 are repeated here from (2.1) for convenience, and are *not* transcriptions from the original tablet — it must be emphasized that the manipulation of symbolic notations in mathematics was not widely practiced prior to the Renaissance, and was certainly unknown in ancient Mesopotamia.

Neugebauer [343] observed that the values in columns two and three, and also the denominators of the squares of the rational numbers in column one, are intimately connected to a well–known procedure from number theory that generates *Pythagorean triples* of integers (a, b, c) satisfying

$$a^2 + b^2 \, = \, c^2 , \tag{2.2}$$

where a, b, c denote the three sides of a right triangle (see Fig. 2.3). Namely, when p and q range over all pairs of positive integers such that: (i) $q < p$; (ii) p and q are not both odd; and (iii) p and q have no common factor other than 1; then the expressions

$$a \, = \, p^2 - q^2 , \quad b \, = \, 2pq , \quad c \, = \, p^2 + q^2 \tag{2.3}$$

yield *all* primitive integer solutions to (2.2) without repetition (a "primitive" triple is one in which a, b, c have no common factor other than 1 — i.e., we exclude solutions that are merely of the form $(a', b', c') = (ka, kb, kc)$ where (a, b, c) is an integer solution and k is an integer greater than 1).

Mathematicians who have studied Plimpton 322 were tempted to regard it as an exercise in number theory, in which their Mesopotamian predecessors were engaged in computing Pythagorean triples by means of the generating functions (2.3) — or alternatively as a trigonometric table, since the entries in the first column amount to $\sec^2 \theta$ (where θ is the angle between the triangle

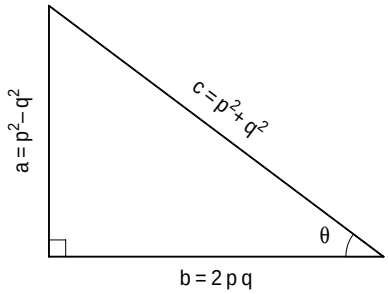

Fig. 2.3. Right triangle with integer sides generated by expressions (2.3).

sides b and c, as in Fig. 2.3) and the resulting θ values decrease in an orderly progression from just under 45° to just over 30°.

However, Robson [377, 378] argues convincingly that such "internalized" mathematical interpretations are unduly colored by the modern perspectives of their authors, and do not adequately take account of the historical, cultural, and linguistic milieu of the tablet's creation. For example, the theory that the Pythagorean generating functions (2.3) were *directly* employed in calculating the column entries contradicts the typical orderly left–to–right calculational progression seen on contemporaneous tablets: one would expect each line to begin explicitly with p and q, and proceed to subsequent derived quantities towards the right. Similarly, a trigonometric reading contradicts the absence of a well–developed notion of angle measure in Mesopotamian mathematics. Robson illustrates this by contrasting the Mesopotamian perspective on the area of a circle with the modern view. The modern formula $A = \pi r^2$ derives from the genesis of a circle by the angular rotation of a vector of length r, the radius. In Mesopotamian thought, however, the circumference C (which we know to be $C = 2\pi r$) is predominant: they expressed the area as $A = C^2/4\pi$ — with, of course, an approximate π value — i.e., they conceived of the circle as the locus of given length C that bounds a symmetric area.

The explication of the *purpose* of Plimpton 322 currently considered most likely [65, 66, 205, 377, 378, 393] is that it represents a "school text" employed to train scribes to perform computations concerned with reciprocal numbers. In Mesopotamian mathematics, the division p/q of two numbers p and q is accomplished by first computing the reciprocal $1/q$ of the denominator, and then multiplying it with the numerator p. The *regular* sexagesimal numbers — i.e., those whose reciprocals have finite sexagesimal expressions — are of particular importance in this regard (such numbers possess factorizations of the form $2^\alpha 3^\beta 5^\gamma$ for positive integers α, β, γ). Lists of regular reciprocal pairs are common among mathematical cuneiform tablets, and presumably served as aides to routine computations.

Now if p/q and q/p are a regular reciprocal pair (i.e., two numbers with finite sexagesimal representations whose product is unity), the Plimpton 322 entries can be readily computed from them using the formulae

$$f = \frac{1}{4}\left(\frac{p}{q} + \frac{q}{p}\right)^2, \quad a = pq\left(\frac{p}{q} - \frac{q}{p}\right), \quad c = pq\left(\frac{p}{q} + \frac{q}{p}\right). \tag{2.4}$$

Such numbers occur in the solution of an equation of the form

$$x = \frac{1}{x} + h \tag{2.5}$$

for a regular reciprocal pair, x and $1/x$, where the former exceeds the latter by some integer amount h. The motivation for this is a "cut–and–paste" geometry problem of the following form: given a rectangle of area $A = 1$ with sides x and $1/x$, the former exceeding the latter by h, we wish to determine x and $1/x$ from this data. We cut off a portion of width $\frac{1}{2}h$ along the side x of the rectangle, and affix it to the top to form an L shape, as shown in Fig. 2.4. The L shape is contained within a square of side $1/x + \frac{1}{2}h$, and its area $A = 1$ must equal the area of this square, minus the area of the smaller shaded square shown in Fig. 2.4, of side $\frac{1}{2}h$. Thus

$$1 = (1/x + \tfrac{1}{2}h)^2 - (\tfrac{1}{2}h)^2,$$

and multiplying both sides by x yields equation (2.5). Now writing $x = p/q$, the quantities $\frac{1}{2}h$ and $1/x + \frac{1}{2}h$ arising in this construction become

$$\frac{1}{2}\left(\frac{p}{q} - \frac{q}{p}\right) \qquad \text{and} \qquad \frac{1}{2}\left(\frac{p}{q} + \frac{q}{p}\right),$$

and if we scale them by $2pq$ to obtain integers, they agree precisely with the quantities a and c in (2.4), while the quantity f represents the (unscaled) area of the square that contains the L shape.

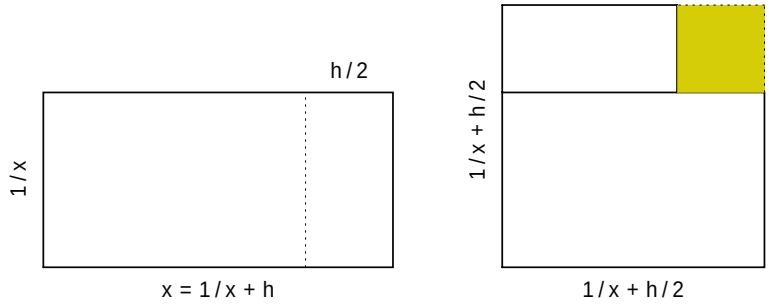

Fig. 2.4. Interpretation of Plimpton 322 in terms of a "cut–and–paste" geometry problem. Left: a rectangle of unit area with reciprocal sides, x and $1/x$, the former exceeding the latter by an integer amount h. Right: cutting off width $\frac{1}{2}h$ and placing it on top produces an L shape within a square of side $1/x + \frac{1}{2}h$. The area of this square minus that of the smaller shaded square, of side $\frac{1}{2}h$, must be equal to 1.

The interpretation of Plimpton 322 as a compilation of "cut–and–paste" geometry exercises involving regular reciprocal pairs is perhaps more mundane (but more credible) than "number theory" or "trigonometry" interpretations. Even as a humble pedagogical tool, however, it suggests at least an *implicit* familiarity with the concept of Pythagorean triples, and imparts respect for the thoroughness of Mesopotamian scribal training. A sense of the dedication and professional pride that Mesopotamian scribes possessed, as the vanguard of human literacy and numeracy, is apparent in the following passage from "In praise of the scribal art," translated [417] by Åke W. Sjöberg:

> *The scribal art is the mother of orators, the father of masters,*
> *The scribal art is delightful, it never satiates you,*
> *The scribal art is not (easily) learned, (but) he who has learned it*
> *need no longer be anxious about it,*
> *Strive to master the scribal art and it will enrich you,*
> *Be industrious in the scribal art*
> *and it will provide you with wealth and abundance,*
> *Do not be careless about the scribal art, do not neglect it ...*

2.2 Theorem of Pythagoras

Pythagoras of Samos (∼580–500 BC) is credited with the famous theorem

$$a^2 + b^2 = c^2 \tag{2.6}$$

that relates the hypotenuse length c of a right triangle to the lengths a, b of the other sides. On account of its simplicity and profundity, and its archetypal role in the emerging concept of *proof*, this mathematical theorem has acquired the unusual distinction of universal recognition. However, modern scholarship — exemplified by the exhaustive treatise of W. Burkert [74] — has demolished the legendary and heroic stature of Pythagoras (concerning his mathematical achievements, at least). According to M. F. Burnyeat [76]:

> *It is hard to let go of Pythagoras. He has meant so much to so many for so long. I can with confidence say to readers of this essay: most of what you believe, or think you know, about Pythagoras is fiction, much of it deliberately contrived.*

The "traditional lore" concerning Pythagoras goes as follows. He is thought to have travelled to Egypt and perhaps Mesopotamia, acquiring scientific and mathematical knowledge there before founding a secretive society called the "Pythagorean school" in Crotone on the south coast of modern Italy — part of *Magna Graecia* in the time of Pythagoras. The Pythagorean school's secretive nature, and the fact that no contemporary biography of Pythagoras survives, have only served to enhance his legendary standing and near–apotheosis. The

followers of Pythagoras supposedly shunned individuality, and believed that the discovery and stewardship of knowledge should be a communal endeavor: it was their custom to credit all discoveries to their leader.

The Pythagorean school was ultimately destroyed in a political upheaval, possibly engendered by external suspicion of their secret and elitist practices. Pythagoras himself fled Crotone but was pursued and killed in Metapontum. The Pythagoreans left no written documents — what we know of their ideas and accomplishments comes from others. It is usually claimed, however, that they were the first intellectual society, pursuing philosophy and mathematics for their own sake,[3] and as a medium for moral advancement. Their putative motto — *All is number* — expresses their faith in the unity of nature's latent mathematical structure, with its diverse manifestations in musical harmony, the planetary motions, and other natural phenomena.[4]

The Pythagoreans pursued a fruitful mixture of algebra and geometry, in which the emphasis was on securing the certainty and universality of results by *rigorous proof*, based upon logical argument, rather than the case–by–case examples that characterized most prior mathematics. Although commonly attributed to Pythagoras, it has not been possible to establish with certainty that he was the first to prove the right–triangle theorem (2.6). The form of the proof is unknown, but is likely to have followed an intuitive geometrical argument, such as that suggested [61] in Fig. 2.5. Four copies of a right–triangle tile are positioned adjacent to each other, so the long side indicates the four compass directions — north, east, south, west. Adding a small square tile (shaded) in the center then yields the square on the hypotenuse. By a simple re–arrangement of these tiles, it is evident that the area of this square equals the areas of the squares on the long and short triangle sides.

The legend that Pythagoras sacrificed a hundred oxen for the Muses, to celebrate his proof of the theorem, is likely apocryphal in view of the strict vegeterianism of the Pythagorean school — motivated by their beliefs in the transmigration of souls and other mystical views. Having established the basic relation (2.6) that governs all right triangles, the Pythagoreans were naturally interested in examples for which it is satisfied by "whole numbers" (a, b, c) — i.e., in *Pythagorean triples of integers*. They were familiar with the simplest triple $(3, 4, 5)$ employed by the Egyptians in the construction of the pyramids, and probably many others transmitted from Mesopotamia or discovered by themselves. But they also devised a procedure to *construct* such triples, by inserting odd numbers m into the expressions

$$a = \tfrac{1}{2}(m^2 - 1)\,, \qquad b = m\,, \qquad c = \tfrac{1}{2}(m^2 + 1)\,.$$

[3] Pythagoras himself supposedly coined the terms *philosophy* for "love of wisdom" and *mathematics* for "that which is learned" to describe the goals of his school.

[4] In medieval times, the *quadrivium* or "four paths" (arithmetic, geometry, music, astronomy) complemented the *trivium* (grammar, dialectic, rhetoric) to form the *seven liberal arts*. Arithmetic was the study of *pure number*; geometry of *number in space*; music of *number in time*; and astronomy of *number in space and time*.

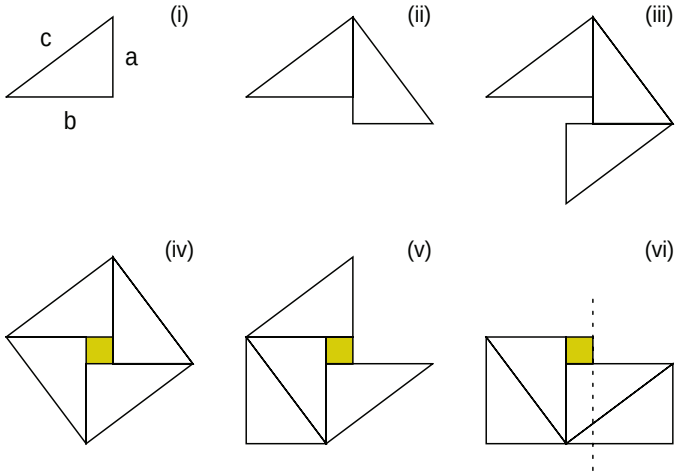

Fig. 2.5. By four–fold replication of the triangle in (i), and addition of the central shaded square of side $b - a$, we obtain the square of area c^2 on the hypotenuse in (iv). This can be re–arranged and divided, as indicated by the dashed line in (vi), into squares of areas a^2 and b^2 — hence, the Pythagorean theorem $a^2 + b^2 = c^2$.

This was subsequently generalized in Euclid's *Elements* — where it is shown that, for integers u and v, the formulae

$$a \,=\, u^2 - v^2\,, \quad b \,=\, 2uv\,, \quad c \,=\, u^2 + v^2 \tag{2.7}$$

yield all Pythagorean triples. If u, v have no common factor ($\gcd(u, v) = 1$), expressions (2.7) define a *primitive* Pythagorean triple in which a, b, c have no common factors. Of course, it is possible to generate other Pythagorean triples by simply multiplying expressions (2.7) by any integer $h > 1$.

But the Pythagorean theorem also proved to be a source of consternation to the Pythagoreans — a severe blow to their belief that *all is number* (where "number" connotes a *whole* number or, at most, a *ratio* of whole numbers). If we choose $a = b = 1$ in (2.6) the resulting value for c, which nowadays we denote by $\sqrt{2}$ and recognize to be irrational, is not a whole number nor a ratio p/q of whole numbers p, q. The Pythagoreans knew this, by one of the first recorded cases of "proof by contradiction" or *reductio ad absurdum*. The argument is as follows: suppose that $\sqrt{2} = p/q$, where p and q are integers with no common factors (and hence not both even). Then

$$p^2 \,=\, 2q^2\,, \tag{2.8}$$

so p^2 is even, and p must also be even, since only the squares of even numbers are even. Thus, $p = 2r$ for some integer r, and substituting into (2.8) gives

$$4r^2 \,=\, 2q^2 \qquad \text{or} \qquad q^2 \,=\, 2r^2\,.$$

So q^2 must be even, and q must also be even. The conclusion that p, q must both be even contradicts the supposition that $\sqrt{2} = p/q$, with p and q not both even, and hence this supposition must be false.

The discovery of "incommensurable" lengths in elementary geometrical configurations incurred a crisis of confidence for the Pythagorean school and subsequent Greek geometers. Their response was to retreat within the safety of intuitive geometrical constructions by straight–edge and compass, a strategy that allowed them to circumvent algebraic confrontations with values that are not exactly expressible as whole–number ratios. As with other mathematical stumbling blocks, the ultimate solution to this impasse was to regard it as an opportunity to define a richer and more general mathematical structure, the continuum of *real numbers*, based on experience in the natural world.

The significance of the Pythagorean theorem, which has been deemed the most fundamental result in all of mathematics, is that it lies at the foundation of *distance measurement*. The use of Cartesian coordinates (x, y) to describe the position of any point $\mathbf{p}$ corresponds to specifying its distances from two orthogonal lines, the *coordinate axes*. The distance

$$d \;=\; \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

between points (x_1, y_1) and (x_2, y_2) is then obtained by applying the theorem to a triangle with horizontal and vertical sides $x_2 - x_1$ and $y_2 - y_1$.

With the advent of calculus, it became possible to precisely define not only the straight–line distance between two points, but also the distance along a curved path, i.e., to *rectify*[5] (compute the arc length of) of curves. Applying the Pythagorean theorem to an infinitesimal segment $\mathrm{d}\xi$ of a differentiable parametric curve $\mathbf{r}(\xi) = (x(\xi), y(\xi))$ allows us to express its arc length as

$$\mathrm{d}s \;=\; \sqrt{x'^2(\xi) + y'^2(\xi)}\;\mathrm{d}\xi\,,$$

and the total arc length S of a finite segment $\xi \in [\,a, b\,]$ is thus given by the integral

$$S \;=\; \int_a^b \sqrt{x'^2(\xi) + y'^2(\xi)}\;\mathrm{d}\xi\,.$$

Under what circumstances can we consider this integral exactly computable? To obtain a closed–form reduction of the integral, the integrand must admit an indefinite integral — or "anti–derivative" — expressible in terms of known analytic functions, i.e., we must be able to identify a function $s(\xi)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\, s(\xi) \;=\; \sqrt{x'^2(\xi) + y'^2(\xi)}\,.$$

[5] The term *rectification* connotes the "straightening out" a curve, as though it were a piece of string, so it can be compared with straight lines of known length.

It is instructive to consider a sequence of progressively more difficult cases:

- if $x(\xi)$, $y(\xi)$ are linear polynomials — i.e., $\mathbf{r}(\xi)$ is a straight line — then $\sqrt{x'^2(\xi) + y'^2(\xi)}$ is a constant, and $s(\xi)$ is linear in ξ;
- if $x(\xi) = r\cos\xi$, $y(\xi) = r\sin\xi$ — i.e., $\mathbf{r}(\xi)$ is a circle of radius r — then $\sqrt{x'^2(\xi) + y'^2(\xi)} = r$ and $s(\xi)$ is again linear in the angular variable ξ;
- if $x(\xi)$, $y(\xi)$ are quadratic, $\mathbf{r}(\xi)$ defines a parabola, and $\sqrt{x'^2(\xi) + y'^2(\xi)}$ is the square root of a quadratic in ξ — a closed–form expression for $s(\xi)$ involving a logarithmic terms is possible;
- when $x(\xi)$, $y(\xi)$ are cubic, $\sqrt{x'^2(\xi) + y'^2(\xi)}$ is the square root of a quartic in ξ, and $s(\xi)$ can be expressed in terms of *incomplete elliptic integrals* — the same is true for the ellipse and hyperbola.

For higher degree curves, the arc length integral $s(\xi)$ does not, in general, admit a closed–form expression. Even in the cases where such an expression is possible, but involves transcendental functions, its cumbersome nature may compromise its practical value.[6] However, the qualification *in general* suggests a possible means to ameliorate this problem: if the argument $x'^2(\xi) + y'^2(\xi)$ of the square root happens to be the *exact square* of some polynomial $\sigma(\xi)$ — i.e., $x'(\xi)$, $y'(\xi)$, $\sigma(\xi)$ constitute a *Pythagorean triple of polynomials* satisfying

$$x'^2(\xi) \, + \, y'^2(\xi) \, \equiv \, \sigma^2(\xi)$$

— then $s(\xi)$ is just the indefinite integral of the polynomial $\sigma(\xi)$, and is thus itself a polynomial (of degree one higher). To make this a viable scheme, we cannot depend on the Pythagorean nature of the triple $x'(\xi)$, $y'(\xi)$, $\sigma(\xi)$ to arise serendipitously — rather, we must ensure that we explicitly incorporate this structure into the polynomials $x'(\xi)$, $y'(\xi)$ that represent the hodograph (derivative) components of a planar curve $\mathbf{r}(\xi) = (x(\xi), y(\xi))$.

Like the integers, polynomials with coefficients in any given field (e.g., the rational, real, or complex numbers) constitute a *unique factorization domain* (UFD). A UFD is, essentially, a set closed under addition or subtraction and (commutative) multiplication, whose members admit unique decompositions into products of *prime* or "irreducible" factors. In the case of integers, these factors are of course the prime numbers. In the case of degree–n polynomials, they are polynomials of degree $\leq n$ with coefficients in the prescribed field that admit no further reduction into products of lower–degree factors with coefficients in that field (we first factor out the highest–order coefficient, to obtain a *monic* polynomial whose irreducible factors are also monic).

Euclid's characterization (2.7) of Pythagorean triples of *integers* may be generalized [292] to the members of *any* unique factorization domain. Thus, three polynomials $a(t)$, $b(t)$, $c(t)$ with coefficients in the field of real numbers and no non–constant common factors will satisfy the Pythagorean condition

$$a^2(t) \, + \, b^2(t) \, \equiv \, c^2(t)$$

[6] See §16.2 for a historical perspective on the curve rectification problem.

if and only if they can be written in terms of two real polynomials $u(t)$, $v(t)$ in the form

$$a(t) \,=\, u^2(t) - v^2(t)\,, \quad b(t) \,=\, 2\,u(t)v(t)\,, \quad c(t) \,=\, u^2(t) + v^2(t)\,.$$

Note that, for polynomials with real coefficients, the roles of $a(t)$ and $b(t)$ are essentially interchangeable since we can obtain the same triple from

$$a(t) \,=\, 2\,\tilde{u}(t)\tilde{v}(t)\,, \quad b(t) \,=\, \tilde{u}^2(t) - \tilde{v}^2(t)\,, \quad c(t) \,=\, \tilde{u}^2(t) + \tilde{v}^2(t)\,,$$

where $\tilde{u}(t) = [\,u(t) + v(t)\,]/\sqrt{2}$ and $\tilde{v}(t) = [\,u(t) - v(t)\,]/\sqrt{2}$. By considering curves defined by hodographs (derivatives) defined in terms of relatively prime polynomials $u(t)$, $v(t)$ in the form

$$x'(t) \,=\, u^2(t) - v^2(t)\,, \qquad y'(t) \,=\, 2u(t)v(t)$$

we resolve the difficulty of rectification. For such *Pythagorean–hodograph* (PH) *curves*, the arc length can be exactly computed through just a few arithmetic operations on the curve coefficients, and we shall find that they possess many other interesting and useful attributes. For space curves, the three hodograph components $x'(t)$, $y'(t)$, $z'(t)$ must be specified in terms of *four* polynomials $u(t)$, $v(t)$, $p(t)$, $q(t)$ in order to satisfy a Pythagorean condition.

To facilitate their construction and analysis, it is advantageous to employ PH curve formulations based on appropriate algebras — the complex numbers and quaternions for planar and spatial PH curves, and Clifford algebra in an even broader setting — this is the motivation for our present survey of algebra. The treatment of PH curves begins in earnest in Part IV.

2.3 Al–Jabr wa'l–Muqabala

The etymological origins of the term *algebra*, as the descriptor of a particular style of mathematical methodology, can be traced to the *Kitab al–mukhtasar fi hisab al–jabr wa'l–muqabala* [273,380], a treatise in Arabic by the 9th–century Persian mathematician Muhammad ibn Musa al–Khwarizmi (or Muhammad, son of Moses, of Khwarizm). A copy of this manuscript, dated A. H. 743 (A. D. 1342), is housed in the Bodleian Library of Oxford University: see Fig. 2.6.

In rough translation, the phrase *al–jabr wa'l–muqabala* means "restoration and balancing" — in reference to the rearrangements of terms in an equation, so as to determine its solution.[7] Khwarizmi's book was translated into Latin in 1145 by the Englishman Robert of Chester, while living in Segovia (Spain), as the *Liber algebrae et almucabola* — hence the discipline *algebra*. The term

[7] Another use of *algebra* was in the sense of "reunion of broken parts," in reference to the surgical process of setting fractured bones. According to a 1565 quotation in the *Oxford English Dictionary*, "This Araby worde Algebra sygnifyeth as well fractures of bones, etc. as sometyme the restauration of the same."

Fig. 2.6. Opening page (folio 1a) of MS. Huntington 214 in the Bodleian Library, University of Oxford — a compilation of mathematical treatises including the *Kitab al–mukhtasar fi hisab al–jabr wa'l–muqabala* by Muhammed ibn Musa al–Khwarizmi and several related works by other authors. Reproduced with permission.

algorithm, prevalent in modern computer science, arose from a corruption of al–Khwarizmi's name through the title of the translation[8] of another treatise, dealing with the Hindu numeral system: the *Algoritmi de numero Indorum*.

Another famous medieval Persian algebraist (but more famous as a poet) was Omar Khayyam (1048–1131), or Ghiyath al–Din Abu'l–Fath Umar ibn Ibrahim al–Nisaburi al–Khayyami to be more precise, where the moniker al–Nisaburi identifies his place of origin as the town of Nishapur in Khurasan, and al–Khayyami reveals the family profession, namely, tent–makers. Among his diverse mathematical, astronomical, musical, and poetical writings is the *Risala fi'l–barahin 'ala masa'il al–jabr wa'l–muqabala* (or *Treatise on Proofs in Problems of Algebra*) written c. 1070 under, by his own account, difficult circumstances of political upheaval [274]. In it he proclaims

> *I say, with God's help and good guidance, that the art of al–jabr and al–muqabala is a mathematical art, whose subject is pure number and mensurable quantities in as far as they are unknown, added to a known thing with the help of which they may be found; and that*

[8] Possibly by Adelard of Bath [75] c. 1130: the translation was discovered by Baron Baldassarre Boncompagni in Cambridge, and published in 1857 — see [420].

> *thing is either a quantity or a ratio, so that no other is like it, and the thing is revealed to you by thinking about it. And what is required in it are the coefficients which are attached to its subject–matter in the manner stated above. And the perfection of the art is knowing the mathematical methods by which one is led to the manner of extracting the numerical and mensurable unknowns.*[9]

This has been regarded as one of the first definitions of *algebra*, as a clearly–identified and articulated field of mathematical study [478].

Among his diverse scientific accomplishments, Khayyam was engaged in a refinement of the calendar by measuring the length of the year in days to an accuracy of five decimal places (the true value actually varies in the sixth decimal place over a human lifespan), and he also developed methods to solve specific types of cubic equations "geometrically" in terms of the intersections of conic curves. For example, he solved cubics of the form

$$x^3 + a^2x = a^2b \qquad \text{and} \qquad x^3 + ax^2 = b^3 \tag{2.9}$$

in terms of the intersections of conics (see Fig. 2.7). In the former case, he drew the parabola $x^2 = ay$ and the circle $x^2 + y^2 - bx = 0$. If P is their point of intersection (other than the origin), and we drop a perpendicular from it to the point Q on the x–axis, the unique real root is given by OQ. In the latter case, he invoked the parabola $y^2 = b(x + a)$ and rectangular hyperbola $xy = b^2$. Dropping a perpendicular from P (their intersection point in the right half–plane) to Q on the x–axis, the desired positive root is OQ.

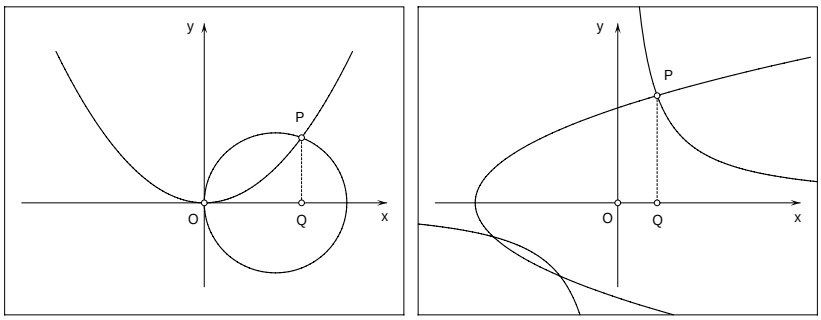

Fig. 2.7. Omar Khayyam's solution of the cubic equations (2.9), in terms of the parabola $x^2 = ay$ and circle $x^2+y^2-bx = 0$ on the left, and the parabola $y^2 = b(x+a)$ and hyperbola $xy = b^2$ on the right. In each case, the length OQ is the desired root.

Khayyam knew that some cubics possess more than one real root, and he aspired to a method for solving general cubics. But this was not achieved until more than 400 years later, using complex numbers, in Renaissance Italy. Today, he is more renowned as a poet, for his famous *Ruba'iyat* (quatrains),

[9] As translated in S. H. Nasr, *Science and Civilization in Islam* [341].

popularized by Edward FitzGerald's translation/interpretation of 1859. These stanzas — alternately mystical and sensual, optimistic and fatalistic — offer a fascinating glimpse into the complexity and subtlety of Khayyam's mind:

> *The moving finger writes, and, having writ,*
> *Moves on: nor all thy piety nor wit*
> *Shall lure it back to cancel half a line,*
> *Nor all thy tears wash out a word of it.*

It has been said of the *Ruba'iyat* that "No other book of poetry has appeared in so many guises, from the edition de luxe to the penny pamphlet" [121] — it has even been rendered as a musical score, for voice and orchestra, by the composer Alan Hovhaness in 1975 (opus 282).

Of course, in the time of al–Khwarizmi and Khayyam, algebraic deductions were conducted entirely in prose: the use of symbolic methods in algebra came much later. The universal symbol x for the unknown quantity in an algebraic equation is thought to be derived through Spanish from the Arabic word *shay'* for "thing" — by which al–Khwarizmi and Khayyam referred to the unknown.

2.4 Fields, Rings, and Groups

Beginning with the "natural" numbers (i.e., the positive integers), which arise directly from physical experience, the development of algebra is characterized by a steadily increasing level of abstraction in the concept of *number*. Despite the absurdity of a negative number of cows or sheep, the *negative numbers* are simply too useful in calculations to be disqualified on philosophical grounds. Elementary geometrical problems soon lead to confrontations with *irrational numbers*, such as $\sqrt{2}$, and even *transcendental numbers* like π. The desire to systematically solve non–linear algebraic equations obliges us to introduce the "two–dimensional" *complex numbers* $a + \mathrm{i}\, b$, where $\mathrm{i} = \sqrt{-1}$. Despite lingering doubts over their "existence," the complex numbers prove immensely valuable in contexts that greatly exceed their original purpose (see Chap. 4).

The quaternions, which resulted from Hamilton's attempt to construct a "three–dimensional number" system, are a turning point in this development: aspects of the familiar rules of arithmetic, formerly considered inviolable, were for the first time relinquished — the result of multiplying two or more of these entities depends on the *order* in which they are specified. This led to a certain loss of inhibition among algebraists: the laws of algebra were no longer viewed as immutable expressions of the natural order that governs the physical world, but as more–or–less arbitrary rules (or *axioms*) that one can posit at will, in order to investigate their logical consequences. Although this has incurred an explosion in the variety and complexity of algebraic systems that have been subject to detailed scrutiny, it has been convincingly argued by Morris Kline [282] that the resulting detachment of mathematics from the "natural world" has not been an unequivocally beneficial development.

Since we will be working with algebraic systems such as the real numbers, complex numbers, quaternions, polynomials, and rational functions, it is useful to briefly review some of the basic principles used to categorize them. Suppose a, b, c are elements of some set S, and let $+$ and $\times$ be two binary operations that, acting on any pair of elements from S, generate another element of S. We postulate a set of possible rules for these operations, as follows:

A_1. $a + b = b + a$

A_2. $(a + b) + c = a + (b + c)$

A_3. there exists $z \in S$ such that $a + z = a$ for all $a \in S$

A_4. for all $a \in S$ there exists $-a \in S$ such that $(-a) + a = z$

M_1. $a \times b = b \times a$

M_2. $(a \times b) \times c = a \times (b \times c)$

M_3. there exists $u \in S$ such that $a \times u = a$ for all $a \in S$

M_4. for all $a \in S$, except z, there exists $a^{-1} \in S$ such that $a^{-1} \times a = u$

D_1. $a \times (b + c) = (a \times b) + (a \times c)$

The binary operations $+$ and $\times$ on pairs of elements in S are called *addition* and *multiplication*. Rules A_1 and M_1 specify the *commutative law* for sums and products, which requires the result to be independent of the *order* of the two operands. Similarly, A_2 and M_2 specify the *associative law* for sums and products: this states that the result is independent of the *grouping* of terms in a sum or product of three (or more) elements. Rules A_3 and M_3 guarantee that an *additive identity* and *multiplicative identity* exist as elements of S. In all the sets that interest us, these elements of are simply $z = 0$ and $u = 1$. Furthermore, rules A_4 and M_4 ensure that each element of S has an *additive inverse* and (except z) a *multiplicative inverse*. Finally, the *distributive law* D_1 states that the product of an element with a sum equals the sum of the products of that element with each of the summands.

Rules A_4 and M_4 allow us to introduce *inverses* $-$ and $\div$ to the operations $+$ and $\times$. Specifically, we set $a - b = a + (-b)$ and $a \div b = a \times (b^{-1})$, and the existence of the additive and multiplicative inverse for every element of S ensures *closure* under these operations, called *subtraction* and *division*.

A *field* is a set S whose elements are subject to a pair of operations $+, \times$ that satisfy *all* of the rules A_1–A_4, M_1–M_4, and D_1. Some familiar fields are the rational numbers (i.e., fractions) $\mathbb{Q}$, real numbers $\mathbb{R}$, complex numbers $\mathbb{C}$, and rational functions (i.e., ratios of polynomials) with real coefficients $\mathbb{R}(t)$. All these sets exhibit closure under the operations of addition, subtraction, multiplication, and division. Moreover, sums and products in these systems are commutative and associative, and they obey the distributive law.

A *ring* is a set S whose elements are subject to a pair of operations $+, \times$ that satisfy the rules A_1–A_4, M_2, and D_1. Rule M_1 may or may not be also satisfied — if it is, we have a *commutative ring*, otherwise a *non–commutative ring*. In other words, addition, subtraction, and multiplication (which may or

may not be commutative) are always possible within S, although division is not. Familiar examples of rings are the integers $\mathbb{Z}$, and the polynomials $\mathbb{R}[t]$ with real coefficients in some variable t. We can add, subtract, and multiply integers or polynomials, and the result is always an integer or polynomial. However, we cannot in general divide integers or polynomials, and expect the result to always be an integer or polynomial.

The integers and polynomials are commutative rings, in which the order of terms in a product does not matter. An example of a non–commutative ring is $\mathbb{R}^{n\times n}$, the set of $n \times n$ matrices with real entries. Matrix products do not, in general, commute — $BA \neq AB$ for general matrices $A, B \in \mathbb{R}^{n\times n}$ so M_1 is not satisfied. Also, matrices must be non–singular to have an inverse, so in general they do not satisfy M_4 (although M_3 is satisfied).

Some systems that concern us lie "between" a ring and a field in terms of their algebraic structure — i.e., they obey all the laws of a ring, but not quite all the laws of a field. Many commutative rings that interest us also satisfy M_3 but not M_4. A system that obeys all the laws of a field *except* M_4 is an *integral domain*. The integers $\mathbb{Z}$ are, of course, the archetypal example of such systems. Another example is the polynomials with real coefficients $\mathbb{R}[t]$ in a variable t. We can construct a field from an integral domain by extending membership of the set S to include all ratios a/b of elements a and $b \neq z$. Such *quotient fields* include the rational numbers (obtained from the integers) and rational functions (obtained from the polynomials).

A system that obeys all the laws of a field except M_1 is a *division ring* (or a *skew field* or *non–commutative field*). The example of primary interest to us here is the *quaternions* $\mathbb{H}$. We defer a detailed treatment of them to Chap. 5 and simply observe now that, although every quaternion has a multiplicative inverse, the non–commutative nature of quaternion products requires us to make a careful distinction between the processes of "left–multiplication" and "right–multiplication" in manipulating quaternion expressions.

Table 2.2 summarizes these classifications. However, not every system with the two binary operations $+$ and $\times$ will fall neatly into one of these categories. Consider, for example, the case of *interval arithmetic* — which is concerned [332, 333] with sets of real values t, of the form $[\,a, b\,] = \{\, t \mid a \leq t \leq b \,\}$. The result of an arithmetic operation $* \in \{\, +, -, \times, \div \,\}$ on interval operands $[\,a, b\,]$ and $[\,c, d\,]$ is the set of values obtained by applying $*$ to pairs of values drawn from each of the two intervals:

$$[\,a, b\,] * [\,c, d\,] \;=\; \{\, x * y \mid x \in [\,a, b\,] \text{ and } y \in [\,c, d\,] \,\}\,.$$

From this definition, one may infer that

$$\begin{aligned}
[\,a, b\,] + [\,c, d\,] &= [\,a + c, b + d\,]\,,\\
[\,a, b\,] - [\,c, d\,] &= [\,a - d, b - c\,]\,,\\
[\,a, b\,] \times [\,c, d\,] &= [\,\min(ac, ad, bc, bd), \max(ac, ad, bc, bd)\,]\,,\\
[\,a, b\,] \div [\,c, d\,] &= [\,a, b\,] \times [\,1/d, 1/c\,]\,,
\end{aligned} \tag{2.10}$$

Table 2.2. Summary of rules observed ($*$) or not observed ($-$) by the two binary operations $+$ and $\times$ in canonical algebraic systems, together with some examples.

	ring	commutative ring	integral domain	division ring	field
A_1	$*$	$*$	$*$	$*$	$*$
A_2	$*$	$*$	$*$	$*$	$*$
A_3	$*$	$*$	$*$	$*$	$*$
A_4	$*$	$*$	$*$	$*$	$*$
M_1	$-$	$*$	$*$	$-$	$*$
M_2	$*$	$*$	$*$	$*$	$*$
M_3	$-$	$-$	$*$	$*$	$*$
M_4	$-$	$-$	$-$	$*$	$*$
D_1	$*$	$*$	$*$	$*$	$*$
example	$\mathbb{R}^{n\times n}$	$\mathbb{Z}$, $\mathbb{R}[t]$	$\mathbb{Z}$, $\mathbb{R}[t]$	$\mathbb{H}$	$\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}(t)$

where division is usually defined only for denominators such that $0 \notin [\,c, d\,]$. This system may be employed to model the propagation of errors in numerical computations, or calculations with uncertain input values (see §12.3.4).

It can be verified that addition and multiplication are commutative and associative, and the degenerate[10] intervals $[\,0, 0\,]$ and $[\,1, 1\,]$ define the additive and multiplicative identities. However, non–degenerate intervals $[\,a, b\,]$ do not have additive or multiplicative inverses ($-$, $\div$ are *not* the inverses to $+$, $\times$). Furthermore, multiplication does not in general distribute over addition — instead, we have the *sub–distributive law*

$$[\,a, b\,] \times ([\,c, d\,] + [\,e, f\,]) \;\subseteq\; ([\,a, b\,] \times [\,c, d\,]) \,+\, ([\,a, b\,] \times [\,e, f\,])\,.$$

Thus, interval arithmetic has a rather unusual algebraic structure — it obeys the rules A_1–A_3 and M_1–M_3, but not A_4, M_4, and D_1.

We conclude by briefly mentioning the simpler algebraic structure known as a *group*. This is a set S equipped with just a single binary operation. This operation obeys the associative law, and the set exhibits closure under it — if the group operation also obeys the commutative law, we have a *commutative* (or *Abelian*) group, otherwise a *non–commutative* group. S also includes an identity element with respect to the group operation, and each element of S has a corresponding inverse in S. An important example is $\mathrm{SO}(n)$, the set of *special orthogonal* real $n \times n$ matrices. A matrix is *orthogonal* if its inverse is identical to its transpose, and it is *special* if its determinant is unity. Since the product of two special orthogonal matrices is always a special orthogonal matrix, such matrices constitute a (non–commutative) group under matrix multiplication. The geometrical significance of the matrices in the group $\mathrm{SO}(n)$ is that they describe *rotations* in the Euclidean space $\mathbb{R}^n$ (see §5.7).

[10] By including degenerate elements, interval arithmetic subsumes the real numbers.