

Introduction

My announced intention was to reach a stage where I could introduce ideal numbers, which are then divided into classes, and could then describe the relation between the problem of counting these classes and the zeta function of Riemann. In one way or another almost all the outstanding problems of modern number theory, and there are a tremendous number of them, are part of a program to use analytical methods to evaluate numbers of various kinds that characterize the size of objects, such as the classes of ideal numbers, that arise in the analysis of diophantine equations, thus equations with integral or rational coefficients for which integral or rational solutions are wanted. These analytical methods are methods that can be easily implemented on a computer and in which there is no uncertainty, no possibility of embarking on an endless search for numbers that may or may not exist, as might happen if one undertook a direct search for solutions of such an equation.

The zeta function will probably be given short shrift, not because it cannot be introduced at the same level as the other material, but because there is not enough time in the eight lectures to treat everything. So the zeta function will be replaced by a pale reflection, the Bernoulli numbers.

The ideal numbers and their division into classes were introduced by Kummer during his attempt to prove the Fermat theorem, as was the relation between the number of classes and the Bernoulli numbers, easily calculated rational numbers. Kummer's success with Fermat's theorem was limited, but he took the subject a long way. Moreover his notion of ideal number and in general his investigation of complex numbers constructed, in a manner now familiar to us, from the solutions of equations in one variable with rational coefficients, was one of two principal starting points, the other being the analysis of symmetries associated to the name of Galois, for the development during the rest of the nineteenth century and the course of the twentieth of the theory of algebraic numbers, without which, I stress, Fermat's theorem could not have been solved, and of which indeed the recent proof has to be regarded as a part.

As preparation for Kummer, we shall spend a considerable amount of time with the work done on Fermat's theorem by Euler and others, between 1750 and 1840.

Puzzlement

Before dealing with ideal numbers, especially ideal primes, we had best provide ourselves with the basic notion of ordinary primes and their uses. Since these are treated by Euclid in Book VII – indeed the basic technique, the euclidean algorithm, is given there – it seemed to the point to continue the procedure of the fall lectures, and to begin with Euclid. This is I shall do, even though Euclid’s treatment of everything but the basic algorithm is fundamentally different from the modern treatment.

I find Book VII in general puzzling, as I shall explain in the course of presenting some of the propositions. What seems to me the enunciation of the basic proposition (Proposition 4) to be deduced from the algorithm is obscurely formulated and its proof even more badly explained. Almost all of the succeeding propositions, many of which we shall need, are almost immediate consequences of Proposition 4, but this is not clearly explained. If this were a modern book, I would be tempted to suggest that the author did not understand the material, but it is a book with the patina of more than two millennia, so that this would be impertinent.

I had hoped to find some enlightenment in Heath’s comments, which presumably reflect the state of historical research in 1925. I was disappointed. In contrast to his commentary on the earlier books, that on Book VII is perfunctory, consisting largely of translations of the statements into modern notation and language.

It may be that the clue to the obscurity is that Euclid was translating essentially arithmetical arguments, thus the kind of arguments that might have been preferred in earlier, pythagorean times into the geometric arguments preferred after the crisis created by the discovery of the irrational, but I do not know. I have searched the literature, although not thoroughly, and made inquiries of specialists, but so far have found very little that is pertinent.

The search is, however, not without its amusing aspects. There is an article – presumably written by an historian of mathematics with no pretensions as a professional mathematician – in which elderly mathematicians who have taken up the history of their subject are upbraided for their Whiggish tendencies and in which it is insinuated that their excursions (or incursions) into this foreign territory are little more than embarrassingly public confessions of the mathematical impotence reputed to appear among us with advancing age and there is a response from one of these same elderly mathematicians, impotent or not, who, with a childish or primitive belief that denying someone his name also robs him of his dignity, castigates an anonymous but clearly identified Z, who is left to infer on his own, although he is guided carefully through each step of the exercise, that he is a “would-be historian” and a “parasite”.

Instructive as such articles, in their own way, are, they do not suggest that the study of Euclid since the time of Heath has made great strides, but I continue to look.

There is, however, a remark of the professional historian (S. Unguru) that I would like to cite, as it is possible that it is a clue to the puzzling form of Book VII.

“It seems to me that it is a considerably more appealing (and certainly historically more defensible) thesis that Greek mathematics, as found in the Elements, is an outgrowth of PYTHAGOREAN mathematics, the arithmetical discreteness of the former (with all its accompanying inherent weaknesses) having been replaced in the former by the continuity of geometrical magnitude; thus in EUCLID numbers are not collections of points anymore, but segments of straight lines, etc.”

Even before seeing this remark, I had decided that the propositions of Book VII and their proofs would be much easier to follow if I replaced the line segments of Euclid by points, and I have done so. Thus I have changed the accompanying figures without changing the arguments. I observe that I cannot judge to what extent the figures appearing in Euclid are a response to the initial medium or to what extent they are Euclid’s own, and not those of subsequent editors.

In contrast to the historian’s suggestion, a statement by the professional mathematician (A. Weil) appears to be questionable.

“In EUCLID’s books VII, VIII and IX there is no trace of geometry, nor even of so-called ‘geometrical algebra.’”

If the geometric traces, namely the line segments, are entirely factitious. a result of self-imposed logical or pedagogical constraints, and therefore to be disregarded – a doubtful historical procedure – then the two statements can be reconciled

Further comments

On further reflection there is a possible explanation for the uneasiness felt on reading Book VII. Implicit in the argument, as presented by me or by Euclid, is the assumption that counting the number of units we always arrive at the same result. For us, with an *a priori* notion of number, for example, to be less than precise mathematically, of the number 8 as a linearly ordered set of eight points, there is something to be proved: it must be proved that two different ways of counting a given collection (or *multitude* in the language of Heath's Euclid) always leads to the same result. This is proved explicitly in texts on set theory.

On the other hand, it is only implicit in Euclid's Book VII. If we return to the very beginning of Book I, we find two of the *common notions* affirming that the whole is greater than the part and that equals subtracted from equals yield equals. Geometrically, this would mean, for example, that removing a unit length from the middle of a longer length and splicing the two remaining ends leads to a length that is the same as if we had simply snipped the unit length from one end. This principle when elaborated could remove the vagueness that appears in Euclid's notion of number – to which we shall come immediately.

It may have been that Euclid, and other Greek mathematicians, had more confidence in the immutability of length than in the immutability of a more abstract notion of number. Indeed as we have already seen, their confidence in the notion of area allowed them to use it in their arguments with a freedom that modern mathematicians felt obliged to justify with the help of their own, recently introduced, notions of number. For example, a modern mathematician is compelled to show that not all numbers are equal.

Comment

Despite strictures about the flaws of Whig history, the principal purpose for which a mathematician pursues the history of his subject is inevitably to acquire a fresh perception of the basic themes, as direct and immediate as possible, freed of the overlay of succeeding elaborations, of the original insights as well as an understanding of the source of the original difficulties. His notion of basic will certainly reflect his own, and therefore contemporary, concerns. For the period that begins with Fermat and ends just before Kummer, there is, so far as I know, no better treatment than that of the elderly mathematician André Weil, whom I have just cited in a more polemical mood. The pertinent book is

Number Theory – An approach through history.

I recommend it to you.

For the early Greek theorems about numbers, there seems to be no adequate reference.

Euclid: Book VII

Some Definitions.

1. A **unit** is that by virtue of which each of the things that exist is called one.
2. A **number** is a multitude composed of units.
3. A number is a **part** of a number, the less of the greater, when it measures the greater;
4. but **parts** when it does not measure it.
5. The greater number is a **multiple** of the less when it is measured by the less.
11. A **prime number** is that which is measured by an unit alone.
12. Numbers **prime to another** are those which are measured by an unit alone as a common measure.
13. A **composite number** is that which is measured by some number.
14. Numbers **composite to one another** are those which are measured by some number as common measure.
15. A number is said to **multiply** a number when that which is multiplied is added to itself as many times as there are units in the other, and thus some number is produced.
20. Numbers are **proportional** when the first is the same multiple, or the same part, or the same parts, of the second that the third is of the fourth.

Observe that for Euclid the unit is not a number, thus there is no number 1. The notion of **parts** is obscure and means more than the definition suggests. This is clear from Definition 20.

Some examples

Primes.

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31

Composite numbers.

4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30

Numbers prime to one another.

10 and 21.

Numbers composite to one another.

15 and 27 – measured by 3

Since 15 is the multiple 5 of 3, the number 3 is to be thought of as one 5th part of 15. On the other hand 10 is parts of 22 and as we shall see this parts is to be thought of as 5 of 11 parts.

The notion of parts will be explained further when we verify the following proposition.

Proposition VII.4 *Any number is either a part or parts of any number, the less of the greater.*

From the definitions, this proposition seems to say nothing at all. This is not so. It is in fact important, and can only be established after an important technique has been introduced, one that will appear in other contexts.



Five measures fifteen

Euclid's Algorithm: First Case

Proposition VII.1 *Two unequal numbers being set out, and the less being continually subtracted in turn from the greater, if the number which is left never measures the one before it until an unit is left, the original numbers will be prime to one another.*

An example. Take the numbers 22 and 105. Then

$$\begin{aligned} 105 - 22 &= 83 & 83 - 22 &= 61, & 61 - 22 &= 39, \\ 39 - 22 &= 17 & 22 - 17 &= 5 & 17 - 5 &= 12 \\ 12 - 5 &= 7 & 7 - 5 &= 2 & 5 - 2 &= 3 & 3 - 2 &= 1. \end{aligned}$$

Another example. Take 35 and 10. Then

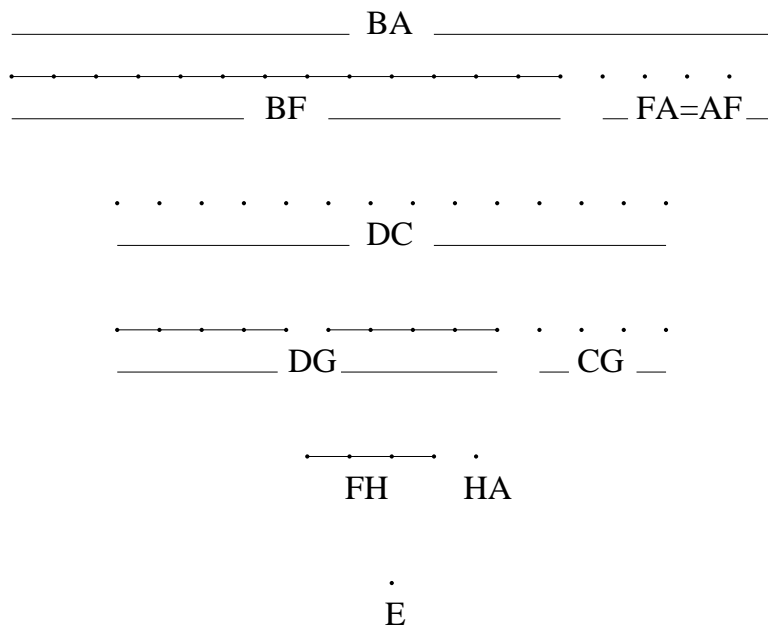
$$35 - 10 = 25 \quad 25 - 10 = 15 \quad 15 - 10 = 5$$

and 5 measures 10.

Proof

We prove the proposition with Euclid's notation. Take the two numbers to be AB and CD . Suppose that they are not prime to one another. Then they are measured by E . If for example, CD is less than AB , subtract CD continually from AB , obtaining BF which is measured by CD and a remainder FA which is less than CD . Let then AF measure GD leaving CG which is less than itself. Then let then GC measure FH leaving HA . Since E measures CD and CD measures BF , therefore E also measures BF . But it measures the whole BA . Therefore it also measures AF . But AF measures DG , so that E also measures DG . Since it also measures DC , it measures CG . But CG measures FH , so that E also measures FH . Since it also measures FA , it measures the unit AH . This is impossible.

Proposition 1



Second Case

Proposition VII.2 *Given two numbers not prime to one another, to find their greatest common measure.*

We repeat the previous process. If the smaller of the two CD measures AB , then it is the greatest common measure. Otherwise, we repeatedly subtract CD from AB until we have found EB measured by CD and a remainder AE smaller than CD . It will not be the unit, because AB and CD are not prime to one another. Either this remainder AE measures CD and therefore also AB or it does not and it will measure FD leaving the remainder CF smaller than itself, which will again not be a unit. If CF measures AE , then it measures FD and also CD , and therefore also AB . Thus it will be a common measure of AB and CD .

It is also the greatest. If not let G be a number greater than CF that measures both AB and CD . Then G measures both BE and AB so that it measures AE . Continuing, we see that it measures CD and FE , so that it also measures CF . Since it was supposed greater than CF , this is impossible.

A consequence of the algorithm drawn be Euclid

Proposition VII.30. *If two numbers by multiplying one another make some number, and any prime number measure the product, it will also measure one of the original numbers.*

This proposition, stated in modern notation, means that if the prime number p divides ab then it divides either a or b , supposed of course to be integers. In conjunction with the following, easy proposition, it implies that every positive integer is the product in a unique way up to order of primes, although this fact does not appear in Euclid, who does not explicitly mention such a possibility, perhaps because he is not inclined to consider the product of several numbers.

Proposition VII.32 *Any number either is prime or is measured by some prime number.*

A modern consequence

Granted the two propositions we start from a positive number n . It is either prime, and then $n = p$ or it is divisible by a prime p and then $n = pm$. Continuing with m , we continually divide by a prime, the quotient growing smaller and smaller, until it is 1. The uniqueness follows because, as a result of Proposition 30, if p divides abc then it divides one of a, b, c, \dots . Thus if

$$p_1^{a_1} \dots p_k^{a_k} = q_1^{b_1} \dots q_l^{b_l},$$

with all of a_1, \dots and b_1, \dots positive, then p_1 is one of q_1, q_2, \dots . So we divide both sides, one by p_1 and one by $q_j = p_1$ and continue with at least one a_i and at least one b_j smaller. Arriving finally at an expression in which one side is 1, and therefore in which both sides are 1.

Observe that this is a direct consequence of Proposition 30. Presumably it was not drawn by Euclid because he had no way of expressing the product of large numbers of integers, or even of more than three integers.

In a modern treatment, the unit would be an integer, so that Propositions 1 and 2 would be formulated as a single proposition, somewhat differently stated from which Proposition 30 would follow directly. In Euclid this is not so and there are a large number of propositions in Book VII between the first two and the thirtieth. Their purpose is, so far as I can see, largely to explain the notion of greatest common measure, a notion which for us has become, by the time we come to discuss primes, completely intuitive. I begin by discussing Proposition VII.4

In other words, as so often with Euclid, the difficulty does not lie in the proof of Proposition 30, it lies elsewhere, in earlier propositions, where the meaning of **part** and of **parts** is explained and connected with the notion of ratio. Proposition 32 is pretty much a direct consequence of the definitions.

An immediate consequence drawn by Euclid

This is the unicity of the greatest common measure. Euclid's conclusion is formulated as a porism.

Porism. *From this it is manifest that, if a number measure two numbers, it will also measure their greatest common measure.*

Explanation and Proof of Proposition VII.4

I use Euclid's notation. Let A and BC be two numbers and let BC be the less. He unfortunately distinguishes the case that A and BC are prime to each other from the case that they are not.

If they are prime to one another, then if BC is divided into the units in it, each unit of those in BC will be a part of A , so that BC is parts of A . In other words, A is a certain number of units; and BC is some smaller number of those units, and therefore makes up a proper fraction of the total number of units in A .

If they are not prime to one another, then either BC measures A and is therefore a part of A (In modern terminology, BC over A is a fraction $1/n$ with numerator 1) or BC does not measure A and there is a greatest common measure D that can be found by Proposition 2. Then D measures BC , so that BC can be divided into numbers BE, EF, FC equal to D .

Interpolation *Thus BC is made up of several pieces, say m of size D .*

As D measures A , D is a part of A .

Interpolation *Thus A is made up of several pieces of size D , say n . Moreover m is less than n . We conclude that m of those n parts makes CD **parts** of A .*

Euclid's version: But D is equal to each of the numbers BE, EF, FC ; therefore each of the numbers BE, EF, FC is also a part of A ; so that BC is parts of A .

What this proposition says is that, using the Euclidean algorithm, we can find two integers, m and n , necessarily prime to one another such that $BC : A = m : n$, but this appears to be very difficult for Euclid to articulate. Notice that if BC measures A then $m = 1$.

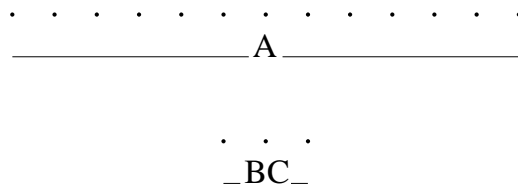
Notice also that we can break BC up into m parts all equal to each other and to each of the n parts into which we break up A .

This pair of numbers m and n is what defines the **part** or **parts** of the pair BC and A and Definition 20 makes two pairs proportional if they have the same parts.

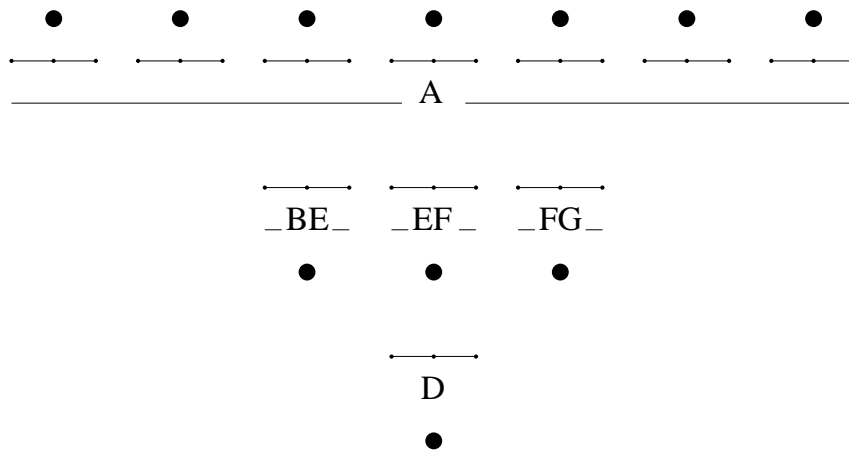
With this proposition in hand, I am going to give some consequences, using the arguments of Euclid and formulating the conclusions as propositions following him. Since the language of these propositions is, however, obscure, preference will be given to clear formulations in everyday language.

Proposition 4

Example 1.



Example 2.



Further Propositions

The first thing to observe is that if $A : B = C : D$ in the sense of Definition 20, then $A : C = B : D$. The first equality means that we can divide A and B into n and m parts respectively of equal size E and then C and D into the same numbers n and m of another size F . Then we apply Proposition VII.4 to the pair E and F , dividing them into pieces of equal (integral!) size k and l . It is clear then that the part or parts of the pair A and C is k/l as is that of the pair B and D .

This fact is expressed by Euclid as Proposition VII.13

Proposition 13. *If four numbers be proportional, they will also be proportional alternately.*

He also shows that if A , B and C are numbers, then

$$B : C = AB : AC.$$

He formulates this as two propositions.

Proposition 17. *If a number by multiplying two numbers make certain numbers, the numbers so produced will have the same ratio as the numbers multiplied.*

Proposition 18. *If two numbers by multiplying any number make certain numbers, the numbers so produced will have the same ratio as the multipliers.*

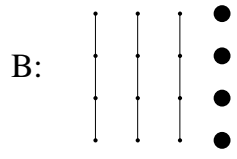
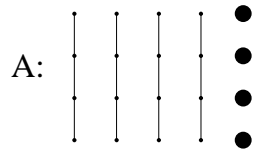
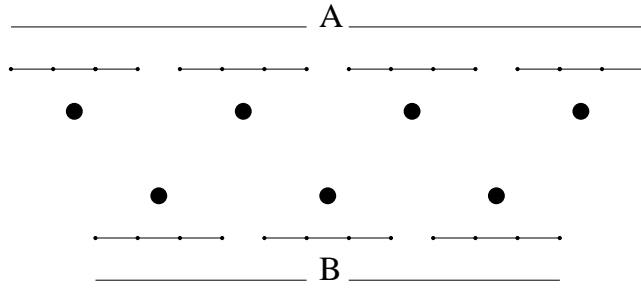
I observe that he proves these propositions separately, even though he has just proved that multiplication is commutative. Thus there is some redundancy. The proof is best presented as a diagram.

He also shows that $A : B = C : D$ in the sense of Definition 20 if and only if $AD = BC$. Here is the relevant proposition.

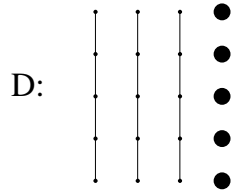
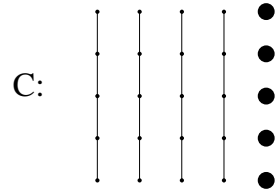
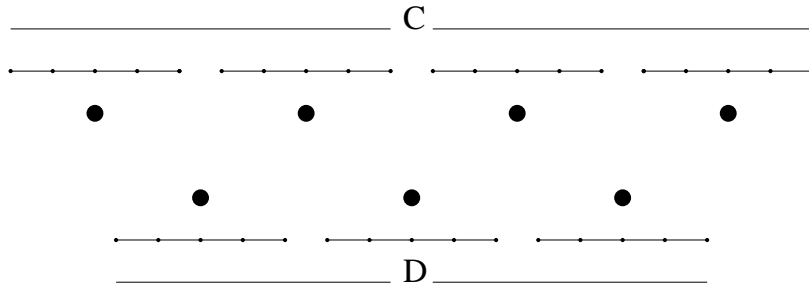
Proposition 19 *If four numbers be proportional, the number produced from the first and the fourth will be equal to the number produced from the second and the third; and if the number produced from the first and the fourth be equal to that produced from the second and the third, the four numbers will be proportional.*

To prove the first part of the proposition, let A , B , C and D be the numbers in proportion. Let A multiplying D make E and let B multiplying C be F . Finally let A multiplying C make G . Represent these three numbers and the relation $A : B = C : D$ as in the diagrams. The conclusion follows, as does the second half of the proposition.

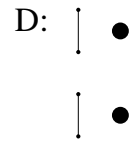
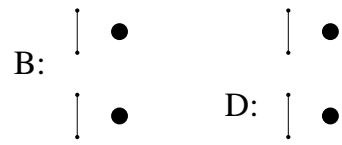
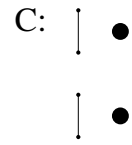
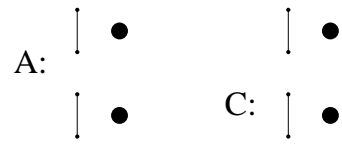
Proposition 13




Proposition 13: continued

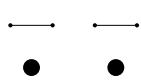


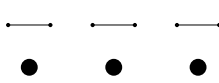
Proposition 13: a possible final step

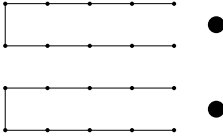


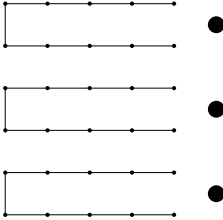
Proposition 17

A: 

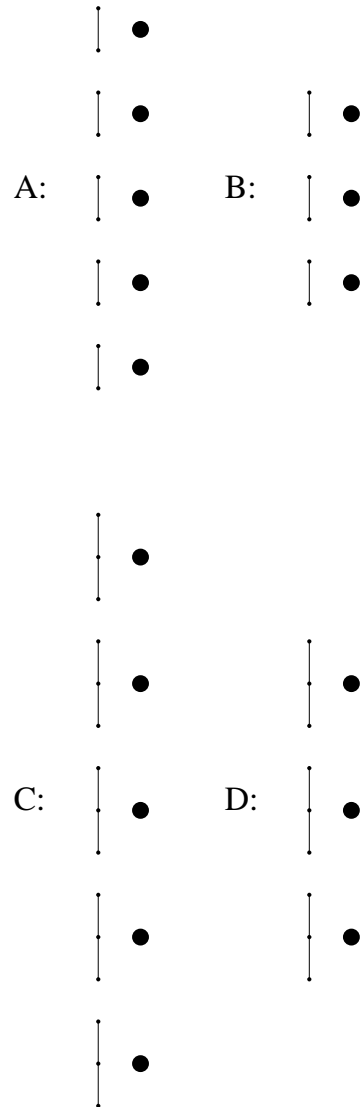
B: 

C: 

AxB 

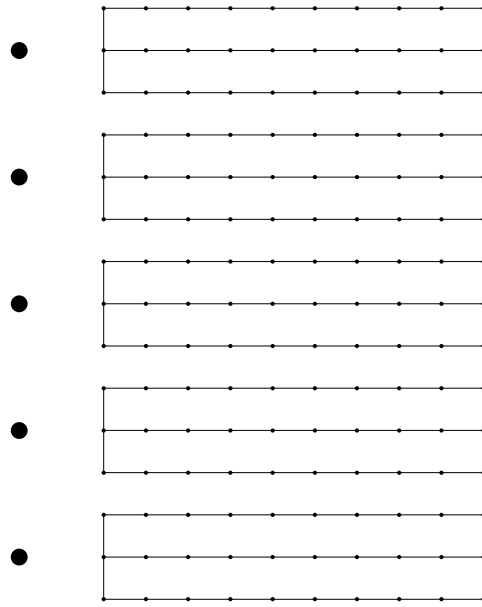
AxC 

Proposition 19: first half

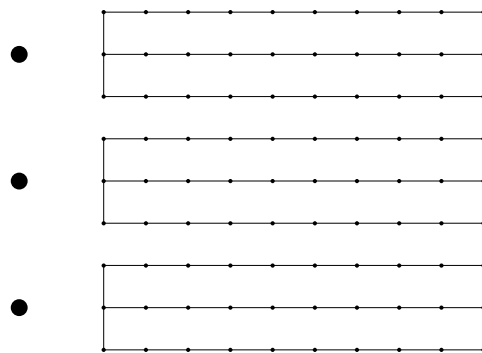


Proposition 19: first half continued.

The number produced from A and C:

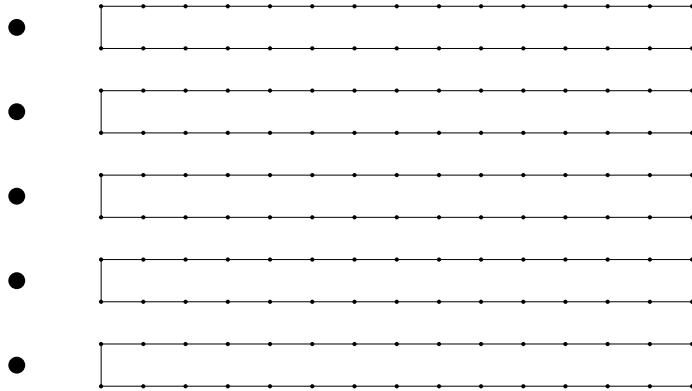


The number produced from A and D:

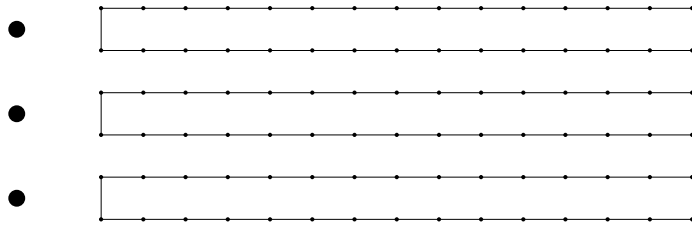


Proposition 19: first half completed.

The number produced from A and C:



The number produced from B and C:



Further Propositions: continued

The next proposition is a matter of putting a fraction in least common terms, but that has already been done in Proposition 4, so that this should be pretty much of a repeat.

Proposition 20. *The least numbers of those which have the same ratio with them measure those which have the same ratio the same number of times, the greater the greater and the less the less.*

What this means is that if $A : B$ is equal to $CD : EF$ and if CD or, what is the same if (and only if) one thinks about it, if EF is as small as possible, then A is a multiple of CD and B the same multiple of EF . It seems to me that CD and EF must obviously be the pair of integers discovered while proving Proposition 4. So the proof of Euclid appears to me unnecessary.

Proposition 21. *Numbers prime to one another are the least of those which have the same ratio with them.*

This too seems a consequence of the proof of Proposition 4.

Proofs of Propositions 30 and 32

It seems to me that this Proposition 32 is pretty much a consequence of the definitions. If a number A is prime, there is nothing to be done. Otherwise it is measured by a smaller number B , If B is prime, there is nothing further to be done. Otherwise it is measured by an even smaller C . Continuing in this way, we arrive finally at a prime.

To establish Proposition 30 following Euclid is another matter. We let A and B be two numbers and multiplying them together we obtain C . Suppose that D is a prime number that measures C . If D does not measure A then A and D are prime to each other. Suppose C is obtained by multiplying D and E . Thus

$$A \times B = D \times E$$

so that

$$A : D = E : B$$

Since A and D are prime to each other, they are the least of the numbers that have the same ratio as they do. Thus they measure E and B respectively. In particular, D measures B .

Illustration of Proposition 32

Consider the number 10780. It is not prime as it is divisible by 10. Dividing by 10, we obtain 1078, which is not prime as it is divisible by 2. Dividing by 2, we obtain 539, which is divisible by 7. To be precise $539 = 7 \times 77$. Finally 77 is divisible by either 7 or 11, both of which are primes.

A modern treatment

The unique factorization of numbers into primes will be so important to us, either as a fact or as a model, that a complete understanding of it and its proof is an essential prerequisite for all that follows. Since Euclid's treatment is much less transparent than the modern, and indeed – apart from the basic algorithm – is quite different in spirit from the modern understanding, I review quickly the modern treatment – in a modern notation.

Suppose a and b are two natural numbers, thus 1, 2, 3 and so on, where in contrast to Euclid, I have included the unit 1 as a number. This is just to avoid repetition. Consider the collection M of all the numbers m that can be represented as $ka + lb$, where k and l are integers that may not be positive. If a is larger than b and we replace a by $a - b$, then

$$m = ka + lb = k(a - b) + (l + 1)b = ka' + l'b.$$

Thus subtracting b from a to obtain a' does not change the collection M . It just changes the representation of a given m . We can repeatedly subtract b until we arrive at $c = a - nb$ that is smaller than b . If it is 0 then $a = nb$ and $m = ka + lb = (kn + l)b$ is a multiple of b . Thus M consists exactly of the multiples of b . In particular a is a multiple of b and b is the greatest common measure (called nowadays **divisor** of a and b). Otherwise we replace a by c and the pair $\{a, b\}$ by $\{b, c\}$. This does not change M . We next subtract the largest possible multiple of c from b to obtain d . If this result is 0 then M consists of multiples of c and c is the greatest common divisor of a and b because it divides a and b , both of which lie in M , and is itself in M . Notice that any number that divides both a and b divides every number in M . If d is not 0 we replace $\{b, c\}$ by $\{c, d\}$. Once again this does not change m . Since d is smaller than c , and at each stage the second element of the pair grows smaller, it must eventually be 0 and the second element at the preceding stage will be the greatest common divisor of a and b .

Table of $15k + 21\ell$, $k = -5 \dots, 5$, $\ell = -4, \dots, 4$

$m \setminus n,$	-4	-3	-2	-1	0	1	2	3	4
-5	-159	-138	-117	-96	-75	-54	-33	-12	9
-4	-144	-123	-102	-81	-60	-39	-18	3	24
-3	-129	-108	-87	-66	-45	-24	-3	18	396
-2	-114	-93	-72	-51	-30	-9	12	33	54
-1	-99	-78	-57	-36	-15	6	27	48	69
0	-84	-63	-42	-21	0	21	42	63	84
1	-69	-48	-27	-6	15	36	57	78	99
2	-54	-33	-12	9	30	51	72	93	114
3	-39	-18	3	24	45	66	87	108	129
4	-24	-3	18	39	60	81	102	123	144
5	-9	12	33	54	75	96	117	138	159

A modern treatment: continued

Example.

$$a = 151, \quad b = 27.$$

$$147 - 5 \times 27 = 147 - 135 = 12, \quad 27 - 2 \times 12 = 27 - 24 = 3, \quad 12 - 4 \times 3 = 12 - 12 = 0$$

Thus the greatest common divisor of 27 and 151 is 3.

In particular, if a and b are prime to one another or, in modern terminology, relatively prime, then there are integers k and l such that $ka + lb = 1$. Conversely, if there are such integers k and l , then a and b are relatively prime.

We return to Proposition 30. Suppose then that a and b are both relatively prime to the number d and $c = ab$. Then there are integers k and l such that $ka + ld = 1$ and integers m and n such that $mb + nd = 1$. Then

$$1 = (ka + ld)(mb + nd) = (km)(ab) + (kan + lmb + lnd)d = k'c + l'd,$$

so that c and d are also relatively prime. If d is a prime number, this is Proposition 30.

Another Proposition from Euclid

The next proposition is not one that will be necessary to the logical development of our arguments at first, but it is a very important fact, at the origin of the greatest of all problems in pure mathematics, the Riemann hypothesis and at the core of many other developments in modern mathematics that I would like, sometime, to broach. So, as it is found in Euclid, it is worth a few minutes explaining it, before we pass on to more modern material. Euclid's argument remains one of the modern arguments, the simplest. Notice that the proposition appears in Book IX and not in Book VII.

Proposition IX.20 *Prime numbers are more than any assigned multitude of prime numbers.*

Let (for example) A, B, C be the assigned prime numbers. Let the least prime number measured by A, B and C be DE and let the unit DF be added to DE . (One could also take DE to be the product of A, B and C .) Then EF is either prime or it is not. If it is prime, then the prime numbers A, B, C and EF have been found which are more than A, B, C .

Suppose then EF is not a prime. Let it be measured by the prime number G . G is different than A, B and C . If not, it measures DE because A, B and C measure DE . Thus the remainder of EF on division by G is the unit. But G measures EF . This is absurd. Thus G is not one of A, B and C and the multitude A, B, C and G is more than A, B and C .

Example.

The first few primes are 2, 3, 5, 7, 11, 13. Their product is 30030. Adding 1, we obtain 30031. It is divisible by the primes 59 and 509.

$$59 \times 509 = 30031.$$

Fermat's theorem: an introduction

I begin by recalling the statement of the theorem.

Fermat's Theorem. *If n is an integer greater than 2 there is no solution of the equation*

$$A \quad x^n + y^n = z^n, \quad xyz \neq 0$$

in integers.

Recall first that there are solutions if $n = 2$, namely

$$(a^2 - b^2)^2 + (2ab)^2 = (a^2 + b^2)^2,$$

so that we can take $x = a^2 - b^2$, $y = 2ab$, $z = a^2 + b^2$. These are, in essence, the only solutions for $n = 2$, but that is not our concern here.

Recall also that if $n = ml$ then

$$(x^l)^m + (y^l)^m = x^n + y^n = z^n = (z^l)^m,$$

so that replacing x by x^l , y by y^l and z by z^l , we obtain from a solution of (A) for n a solution for m . If the impossibility is proved for m it follows also for n . Since every integer greater than 2 is divisible either by an odd prime or by 4, it suffices to prove the impossibility for odd primes and for $n = 4$.

Since 3 is the smallest of these numbers, I start with 3. The key for us will be numbers we have seen before. Let

$$\alpha = \cos(2\pi/3) + i \sin(2\pi/3),$$

so that $\alpha^3 = 1$. We shall study numbers of the form

$$B \quad a + b\alpha$$

with a and b integers. These are of course very much like the numbers studied during the discussion of the regular pentagon and regular heptadecagon, except that we then allowed a and b to be fractions, whereas we now want to consider only numbers of this form that are *integral*, at least in a naive sense. It turns out fortunately that this naive sense, which was all that was available at first is equivalent to the more sophisticated and universally correct notion.

OBSERVATIONES DOMINI PETRI DE FERMAT.

I (p. 54).

(Ad definitionem VI Cl. Gasparis Bacheti Porismatum Libr. III.)

A duobus quibuscumque numeris formari dicitur triangulum rectangulum, quum ex aggregato et ex intervallo quadratorum ab ipsis et ex duplo plani sub ipsis numeris contenti constant latera trianguli.

A tribus numeris in proportione arithmetica possumus formare triangulum, si secundum hanc definitionem sextam formemus illud a medio et differentia. Nam solidum sub tribus ductum in differentiam faciet aream dicti trianguli, atque ideo, si differentia sit unitas, solidum sub tribus erit area trianguli.

II (p. 61).

(Ad quæstion. VIII Diophanti Alexandrini Arithmeti corum Libr. II.)

Propositum quadratum dividere in duos quadratos.

Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duas ejusdem nominis fas est dividere : cujus rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

III (p. 65).

(Ad quæstion. X Libr. II.)

Datum numerum, qui ex duobus componitur quadratis, in alios < duos > quadratos partiri.

Num verò numerum ex duobus cubis compositum dividere poterimus in alios duos cubos? Hæc quæstio difficilis sane nec Bacheto aut Vietæ

The domain $\mathbb{Z}(\alpha)$

Recall that

$$\alpha^2 + \alpha + 1 = 0,$$

so that

$$\alpha = \frac{-1 \pm \sqrt{1-4}}{2}.$$

Since $\sin(2\pi/3) > 0$, $\alpha = (-1 + i\sqrt{3})/2$.

Observe also that

$$(a + b\alpha) \times (c + d\alpha) = ac + (ad + bc)\alpha + bd\alpha^2 = (ac - bd) + (ad + bc - bd)\alpha,$$

so that two numbers of the form (B) when multiplied together yield a number of the form (B). This is also, and clearly, so when they are added or subtracted.

The basis of the proof of Fermat's theorem for $n = 3$ will be the study of primes in the domain of numbers of the form (B). Our first task, will be to prove an analogue of the euclidean algorithm. We begin with a few simple remarks.

First of all the conjugate $\bar{\alpha}$ of α is $\cos(2\pi/3) - i\sin(2\pi/3)$ and $\alpha\bar{\alpha} = 1$. Thus $\alpha^{-1} = \alpha^2$ is $\bar{\alpha}$ and is another number of the same form. The only ordinary integers n such that $1/n$ is also an integer are 1 and -1 , but in the new domain, α , $-\alpha$, $\bar{\alpha}$ and $-\bar{\alpha}$ also have this property. They are all units, not in the sense of Euclid but in a new sense: each of them divides all other numbers. If ρ' is the inverse of the unit ρ , then

$$\xi = \rho\rho'\xi = \rho\eta, \quad \eta = \rho'\xi,$$

so that ρ divides all ξ in the domain.

Observe in general that if $\xi = a + b\alpha$, then

$$N\xi = \xi\bar{\xi} = (a + b\alpha)(a + b\bar{\alpha}) = a^2 - ab + b^2 = \left(a - \frac{b}{2}\right)^2 + \frac{3b^2}{4}$$

is a positive integer and

$$N(\xi\eta) = \xi\eta\bar{\xi}\bar{\eta} = \xi\eta\bar{\xi}\bar{\eta} = (N\xi)(N\eta).$$

In particular if ξ is a unit, so that its inverse η is also in the domain, then $1 = \xi\eta$ and

$$1 = N\xi N\eta,$$

Units

1 and -1 .

$$\alpha = -\frac{1}{2} + i\frac{\sqrt{3}}{2} \quad \text{and} \quad -\alpha = \frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

$$\alpha^2 = -\frac{1}{2} - i\frac{\sqrt{3}}{2} \quad \text{and} \quad -\alpha^2 = \frac{1}{2} + i\frac{\sqrt{3}}{2}.$$

Units times λ

$$1 - \alpha, \quad -1 + \alpha.$$

$$\alpha - \alpha^2 = 1 + 2\alpha, \quad -1 - 2\alpha.$$

$$\alpha^2 - 1 = -2 - \alpha, \quad 2 + \alpha.$$

so that the positive integers $N\xi$ and $N\eta$ are both 1. Thus if $\xi = a + b\alpha$, then

$$4 = (2a - b)^2 + 3b^2.$$

For this b must be 0 or ± 1 . If b is 0 then $a = \pm 1$ and $\xi = \pm 1$. If $b = \pm 1$, then $2a - b = \pm 1$ and $a = (\pm 1 \pm 1)/2$, which is 0 or b . Then

$$b + b\alpha = -b\alpha^2.$$

We conclude that the only units are those found, $\pm 1, \pm\alpha, \pm\alpha^2$.

A number ξ in the domain $\mathbb{Z}(\alpha)$ will be called prime if in any representation

$$\xi = \eta\zeta,$$

one of η and ζ is necessarily a unit. If for example η is a unit, then

$$\zeta = \bar{\eta}\xi.$$

Some examples of primes

Observe that if $N\xi$ is a prime then ξ is a prime, because if

$$\xi = \eta\zeta,$$

then

$$N\xi = N\eta N\zeta,$$

so that either $N\eta = 1$ or $N\zeta = 1$ because $N\xi$ cannot be factored. Consider for example $\lambda = 1 - \alpha$.

$$N\lambda = (1 - \alpha)(1 - \bar{\alpha}) = (1 - \alpha)(1 - \alpha^2) = 1 - \alpha - \alpha^2 + 1 = 3,$$

so that λ is a prime and, indeed, a prime that divides 3. Moreover

$$\bar{\lambda} = 1 - \alpha^2 = -\alpha^2(1 - \alpha) = -\alpha^2\lambda = \rho\lambda,$$

where ρ is a unit. Thus

$$3 = \rho\lambda^2$$

is the factorization of 3, so that 3 is now, in essence, a square.

Examples of primes continued

In general, if n is a norm, then

$$4n = (2a - b)^2 + 3b^2.$$

I make a table of n for $a = -7, \dots, 7$, $b = 0, \dots, 7$. Since changing a to $-a$ and, at the same time, b to $-b$ does not change the norm, this table implicitly includes b from -7 to 7 .

Some norms

$a \setminus b$	0	1	2	3	4	5	6	7
-7	49	57	67	79	93	109	127	147
-6	36	43	52	63	76	91	108	127
-5	25	31	39	49	36	75	91	109
-4	16	41	28	37	48	61	76	93
-3	9	13	19	27	37	49	63	79
-2	4	7	12	19	28	39	52	87
-1	1	3	7	13	21	31	43	57
0	0	1	4	9	16	25	36	49
1	1	1	3	7	13	21	31	43
2	4	3	4	7	12	19	28	39
3	9	7	7	12	13	19	27	37
4	16	13	12	13	16	21	28	37
5	25	21	19	19	21	25	31	39
6	36	31	28	27	28	31	36	43
7	49	43	39	37	37	39	43	49

Primes continued

Apart from 3, the primes that appear in this table are 7, 13, 19, 31, 43, 61, 67, 79. In principle, numbers that have the same norm appear in groups of six, each being obtained from the other by multiplying by a unit, or, if we demand that b be positive or zero and that a be positive when $b = 0$, in groups of three. The number 0 is an exception. For example, the following numbers have norm 3:

$$\begin{aligned} 1 - \alpha, \quad -1 + \alpha, \quad \alpha - \alpha^2 = 2\alpha + 1, \\ -2\alpha - 1, \quad \alpha^2 - 1 = -2 - \alpha, \quad 2 + \alpha, \end{aligned}$$

corresponding to $(-1, 1)$, $(1, 2)$, $(2, 1)$ and their negatives.

The number 7 appears in the table six times, thus there must be at least one pair ξ and η such that η is not a unit times ξ . The first number with norm 7 is $-2 + \alpha$. We have then

$$-2 + \alpha, \quad -\alpha(-2 + \alpha) = 2\alpha - \alpha^2 = 1 + 3\alpha, \quad \alpha^2(-2 + \alpha) = 1 - 2\alpha^2 = 3 + 2\alpha,$$

accounting for three of the appearances of 7. The others are

$$-1 + 2\alpha, \quad -\alpha(-1 + 2\alpha) = \alpha - 2\alpha^2 = 2 + 3\alpha, \quad \alpha^2(-1 + 2\alpha) = 2 - \alpha^2 = 3 + \alpha.$$

Observe that the any number of the first set is relatively prime to any number of the second set.

When verifying this, we had best use the modern notion of relatively prime, because we have not yet established any kind of euclidean algorithm for the domain $\mathbb{Z}(\alpha)$. Thus to say that two numbers ξ and ζ in the domain are relatively prime means that the collection of numbers $\mu\xi + \nu\zeta$, where μ and ν are arbitrary numbers in $\mathbb{Z}(\alpha)$ contains 1. Notice that this collection does not change if we replace ξ by $\epsilon\xi$, with ϵ a unit. We have then simply to replace μ by $\mu\epsilon'$, $\epsilon' = \bar{\epsilon} = 1/\epsilon$. Take then $\xi = -2 + \alpha$ and $\zeta = 3 + \alpha$. Then $-\xi + \zeta = 5$ and $\xi\bar{\xi} = 7$. Since $3 \times 7 - 4 \times 5 = 1$, we have

$$(3\bar{\xi} + 4)\xi - 4\zeta = 1.$$

Primes continued

Primes that do not appear in the table are 2, 5, 11, 17, 23, 29. Each of these numbers is already a prime, although their norms are squares, $N 2 = 4$, $N 5 = 25$ and so on. If, for example, 2 were not a prime, we would have

$$2 = \xi\zeta, \quad 4 = N 2 = N \xi N \zeta,$$

with neither ξ nor ζ being a unit. Thus

$$2 = N \xi = N \zeta.$$

Suppose $\xi = a + b\alpha$. Then

$$2 = N \xi = a^2 - ab + b^2$$

I calculate the remainder of each side on division by 3. On the left it is 2. On the right it depends only on the remainder left by a and b .

$$(a + 3c)^2 - (a + 3c)(b + 3d) + (b + 3d)^2 = a^2 - ab + b^2 + 3e$$

with

$$e = 2ac + 3c^2 - cb - ad - 3cd + 2bd + 3d^2.$$

So there are nine possibilities.

$a \backslash b$	0	1	2
0	0	1	1
1	1	1	0
2	0	1	1

So there is no question of the two sides having the same remainder and thus no question of the sides being equal. Since 5, 11, 17, 23 and 29 all leave the remainder 2 on division by 3, the same argument applies and they too cannot be factorized.

Some factorizations

$$\xi = -2 + \alpha = -\frac{5}{2} + \frac{\sqrt{3}}{3}, \quad \bar{\xi} = -2 - 1 - \alpha = -3 - \alpha = -\frac{5}{2} + \frac{\sqrt{3}}{2}$$

$$\xi\bar{\xi} = \frac{25}{4} + \frac{3}{4} = 7.$$

Thus the prime 7 factors as $\xi\bar{\xi}$ and ξ and $\bar{\xi}$ are relatively prime.

$$\xi = -2 + 3\alpha, \quad \bar{\xi} = -5 - 3\alpha, \quad \xi\bar{\xi} = N\xi = 19.$$

Thus the prime 19 factors as $\xi\bar{\xi}$ and ξ and $\bar{\xi}$ are relatively prime. If not *and if there were unique factorization* we would have $\xi = \mu\eta$, $\bar{\xi} = \nu\eta$, where η was not a unit, and

$$19^2 = N\xi N\bar{\xi} = N\mu N\nu N\eta^2,$$

so that $N\eta = 19$ and $\bar{\xi} = \rho\xi$ is a unit times ξ .

In general, if $\xi = a + b\alpha$ so that $\bar{\xi} = a - b - b\alpha$, then

$$\bar{\xi} = \xi \implies b = 0,$$

$$\bar{\xi} = -\xi \implies a = b - a \implies \xi = a(1 + 2\alpha),$$

so that $\xi = a\alpha\lambda$ is a multiple of λ ,

$$\bar{\xi} = \alpha\xi \implies a - b - b\alpha = -b + (a - b)\alpha$$

or $a = 0$, $\xi = b\alpha$,

$$\bar{\xi} = -\alpha\xi \implies a - b - b\alpha = b + (b - a)\alpha,$$

so that $a = 2b$ and $\xi = -a\alpha^2\lambda$ is a multiple of λ ,

$$\bar{\xi} = \alpha^2\xi \implies a - b - b\alpha = b - a - a\alpha$$

or $b = a$, so that $\xi = -a\alpha^2$,

$$\bar{\xi} = -\alpha^2\xi \implies a - b - b\alpha = a - b + a\alpha,$$

and $a = -b$ and $\xi = a\lambda$ is again a multiple of λ .

Thus, in general, *provided there is a unique factorization* we can expect that an ordinary prime that leaves the remainder 1 upon division by 3 is the product of two relatively prime numbers in $\mathbb{Z}(\alpha)$.

Just as the primes that do not appear all leave the remainder 2 upon division by 3, or as one says, they are all congruent to 2 modulo 3, so do all the primes, with the exception of 3 itself, that appear leave the remainder 1. The first primes of this sort that do not appear are 73 and 97. We can guess that this is because the table is too small. Indeed further calculations show that

$$N(1 + 9\alpha) = 73, \quad N(3 + 11\alpha) = 97.$$

I shall present a proof of Fermat's theorem for $n = 3$ that depends on the possibility of factoring every element ξ of $\mathbb{Z}(\alpha)$ into primes, thus into a product

$$\xi = \epsilon \pi_1^{a_1} \pi_2^{a_2} \dots,$$

in which the factors are unique up to order and up to multiplication with a unit. The number ϵ is a unit.

Although we have some experience with ordinary primes and readily recognize the smaller of them, 2, 3, 5, 7 and so on, and can also find with no difficulty the prime factorization of small numbers,

$$98 = 2 \times 7^2,$$

this is by no means the case for larger numbers. The 1000-th prime is 7919, but it is clear to none of us without either a good deal of calculation or a knowledge of special techniques how to verify that it is prime. The situation is even worse for numbers in $\mathbb{Z}(\alpha)$, and even excellent mathematicians can be led into error when they venture into this area without previous experience. Since I, too, in these lectures am dealing with material with which I have limited experience, I give, as a cautionary tale, an example from a book of a friend, a distinguished mathematician for whom I have great respect.

It is a book about the origins of modern algebra and, in particular, about the solution of equations, so that its aims are, in part, those of these lectures.

Some identities

Since $\alpha = (-1 + \sqrt{-3})/2$, the numbers in $\mathbb{Z}(\alpha)$ are the numbers $(a + b\sqrt{-3})/2$, where a and b are both even. The book had been given to me by my friend in California and I was reading it with pleasure on the return flight when I came across the following passage.

In other words, we have the remarkable identity

$$2 = \sqrt[3]{6\sqrt{3} + 10} - \sqrt[3]{6\sqrt{3} - 10}$$

Try proving this directly!!

I had seen such identities before and had indeed found them curious, but had never stopped to reflect on them. On the airplane, however, with these lectures in mind, I stopped to reflect on the identity and quickly came to the inclusion that the ideas introduced by Kummer and Galois, to whom we shall come, strongly suggest that any such identity has to be trivial. To be more precise, they suggest that such an identity only arises when the two numbers involved, $\xi = 6\sqrt{3} + 10$ and $\zeta = 6\sqrt{3} - 10$ are both cubes,

$$\xi = \mu^3, \quad \zeta = -\nu^3,$$

and

$$\mu + \nu = 2.$$

To test this, we first calculate an analogue of the norm of ξ , multiplying ξ by $\xi' = -6\sqrt{3} + 10$,

$$\xi\xi' = (6\sqrt{3} + 10)(-6\sqrt{3} + 10) = -108 + 100 = -8,$$

which is indeed a cube, $-8 = (-2)^3$. Thus if we could find a number, whose norm (in this new sense) is -2 , we would have a start on finding μ

The norm of $a + b\sqrt{3}$ is

$$(a + b\sqrt{3})(a - b\sqrt{3}) = a^2 - 3b^2,$$

so that $a = \pm 1, b = \pm 1$ are four possibilities for a number whose norm is -2 .
Since

$$(1 + \sqrt{3})^2 = 1 + 2\sqrt{3} + 3 = 4 + 2\sqrt{3},$$

and

$$(1 + \sqrt{3})(4 + 2\sqrt{3}) = 4 + 4\sqrt{3} + 2\sqrt{3} + 6 = 10 + 6\sqrt{3},$$

one possibility for μ is $1 + \sqrt{3}$. In the same way, we can take $\nu = 1 - \sqrt{3}$.
Since

$$\mu + \nu = 2,$$

the identity is not remarkable. It just appears on first sight to be so.

Unfortunately with a tactlessness that I exhibit only too frequently, I sent off an e-mail message to my friend upon my return, expressing the hope that his students had not allowed him to pull the wool over their eyes. It was not the right response to his enthusiasm or his generosity. I hope he has forgiven me.

As an exercise, whose solution will be given next week, I let you show that for similar reasons the identity

$$1 = \sqrt[3]{2\sqrt{13} + 5} - \sqrt[3]{2\sqrt{13} - 5},$$

which appears in the same book as an exercise, is also not remarkable!

Since the book is, in spite of my teasing, an excellent book that you might well enjoy, I give the author and title:

V. S. Varadarajan, *Algebra in ancient and modern times*.

Solution to exercise

The first step is to calculate the norm of $2\sqrt{13} + 5$. It is

$$(2\sqrt{13} + 5)(-2\sqrt{13} + 5) = -52 + 25 = -27,$$

so that our first step is to find a number μ whose norm is -3 . This is easy, for

$$(7 + 2\sqrt{13})(7 - 2\sqrt{13}) = 49 - 52 = -3.$$

Since

$$(7 + 2\sqrt{13}) + (7 - 2\sqrt{13}) = 14,$$

this will not serve our purpose. Indeed we need a number of the form

$$\frac{1}{2} + b\sqrt{13},$$

so that

$$\left(\frac{1}{2} + b\sqrt{13}\right) + \left(\frac{1}{2} - b\sqrt{13}\right) = 1.$$

We need in addition,

$$\left(\frac{1}{2} + b\sqrt{13}\right)\left(\frac{1}{2} - b\sqrt{13}\right) = \frac{1}{4} - 13b^2 = -3,$$

or

$$-13b^2 = -\frac{13}{4}.$$

Thus $b = 1/2$, $\mu = (1 + \sqrt{13})/2$ and $\nu = (1 - \sqrt{13})/2$.

Since

$$\left(\frac{1}{2} + \frac{1}{2}\sqrt{13}\right)^2 = \frac{7}{2} + \frac{1}{2}\sqrt{13},$$

we have

$$\left(\frac{1}{2} + \frac{1}{2}\sqrt{13}\right)^3 = \left(\frac{1}{2} + \frac{1}{2}\sqrt{13}\right)\left(\frac{7}{2} + \frac{1}{2}\sqrt{13}\right) = 5 + 2\sqrt{13}.$$

Since $5 + 2\sqrt{13}$ looks to be an integer, it is natural to suppose that its cube root, $\mu = (1 + \sqrt{13})/2$ is also an integer. This raises the question as to when a number of some complicated form is to be regarded as integral. Although the answer is simple, it is not so easy to justify it. So I pass over the matter quickly remarking only that the numbers we shall meet will by and large be integral. I observe in addition that μ and ν satisfy the equation

$$0 = (x - \mu)(x - \nu) = x^2 - x + 3,$$

in which the coefficient of x^2 is 1 and all other coefficients ordinary integers.