The Shaping of Arithmetic

after C. F. Gauss's Disquisitiones Arithmeticae

Catherine Goldstein
Norbert Schappacher
Joachim Schwermer

Editors

Springer

The Shaping of Arithmetic after C. F. Gauss's Disquisitiones arithmeticae, edited by Catherine Goldstein, Norbert Schappacher and Joachim Schwermer, Berlin, Heidelberg, New York: Springer, 2007, is available at http://www.springer.com

Editions of Carl Friedrich Gauss's *Disquisitiones Arithmeticae*

The *Disquisitiones Arithmeticae* has been omitted from the list of references of the individual chapters: we list underneath its various editions. Throughout this book, passages from Gauss's *Disquisitiones Arithmeticae* are referred to only by the article number. The title of Gauss's work is routinely abbreviated as "D.A." For all works, a mention of [Author 1801a] refers to the item "AUTHOR. 1801a" in the bibliography, a mention of [Author 1801/1863] refers to the 1863 edition in this item.

1801. *Disquisitiones Arithmeticae*. Leipzig: Fleischer. Repr. Bruxelles: Culture et civilisation, 1968. Repr. Hildesheim: Olms, 2006. Rev. ed. in *Werke*, vol. 1, ed. Königliche Gesellschaft zu Göttingen [E. Schering]. Göttingen: Universitäts-Druckerei, 1863; 2nd rev. ed., 1870; repr. Hildesheim: Olms, 1973.

http://gallica.bnf.fr
http://dz-srv1.sub.uni-goettingen.de/cache/toc/D137206.html

1807. *Recherches arithmétiques*. French transl. A.-C.-M. Poullet-Delisle. Paris: Courcier. Repr. Sceaux: Gabay, 1989.

http://gallica.bnf.fr

1889. *Arithmetische Untersuchungen*. German transl. H. Maser. In *Untersuchungen über höhere Arithmetik*, pp. 1–453. Berlin: Springer. Repr. New York: Chelsea, 1965; 2nd ed., 1981.

http://dz-srv1.sub.uni-goettingen.de/cache/toc/D232699.html

1959. *Arifmetičeskie issledovaniya*. Russian transl. V. B. Dem'yanov. In *Trudi po teorii čisel* [Works on number theory], ed. I. M. Vinogradov, B. N. Delone, pp. 7–583. Moscow: Academy of Sciences USSR.

1966. *Disquisitiones Arithmeticae*. English transl. A. A. Clarke. New Haven: Yale University Press. Rev. ed. W. C. Waterhouse. New York: Springer, 1986.

1995. *Disquisitiones Arithmeticae*. Spanish transl. H. Barrantes Campos, M. Josephy, A. Ruiz Zùñiga. Colección Enrique Pérez Arbelaez 10. Santa Fe de Bogotá: Academia Colombiana de Ciencias Exactas, Fisicas y Naturales.

1995. *Seisuu ron*. Japanese transl. Takase Masahito. Tokyo: Asakura-Shoten.

1996. *Disquisicions aritmètiques*. Catalan transl. G. Pascual Xufré. Barcelona: Institut d'Estudis Catalans, Societat Catalana de Matemàtiques.

II.2

Composition of Binary Quadratic Forms and the Foundations of Mathematics

HAROLD M. EDWARDS

Writing to Leopold Kronecker on June 14, 1846, Ernst Kummer said of his newly created theory of ideal prime factors:

> Dirichlet strongly urged me to work the theory out completely and submit it to Crelle for publication as soon as possible. He also told me and showed me, from oral and written indications of Gauss himself, that when Gauss was completing the section of *Disqu. arith.* on composition of forms he had something similar to ideal factors for his private use, but that he had not put it on firm ground. Specifically, Gauss says in a note to his treatise on the factorization of integral rational functions into linear factors something like: "Had I been willing to use imaginaries in the way that earlier mathematicians did, I would have been able to simplify substantially one of my researches which, as it is, is quite difficult." Gauss later told Dirichlet that the reference here was to the composition of forms.[1]

William Waterhouse has convincingly argued that Gauss was referring in the footnote Kummer mentions *not* to the composition of forms in sec. 5 but to the unfinished sec. 8 of *Disquisitiones Arithmeticae*.[2] One should not, however, allow

1. [Kummer 1846/1910]: *Dirichlet hat mich sehr ermahnt die Theorie bald fertig auszuarbeiten und Crelle zum Drucke zu übergeben. Auch hat er mir erzählt und gezeigt, nämlich aus mündlichen und schriftlichen Aeußerungen von Gauss, daß Gauss schon bei Anfertigung des Abschnittes* de compositione formarum *aus den* Disqu. arith. *etwas ähnliches wie ideale Factoren zu seinem Privatgebrauche gehabt hat, daß er dieselben aber nicht auf sicheren Grund zurückgeführt hat, er sagt nämlich in einer Note seiner Abhandlung über die Zerfällung der ganzen rat. Functionen in lineäre Factoren ohngefähr so: „Wenn ich hätte auf dieselbe Weise verfahren wollen wie die früheren Mathematiker mit dem imaginären, so würde eine andere meiner Untersuchungen die sehr schwierig ist sich auf sehr leichte Weise haben machen lassen." Daß hier die* compositio formarum *gemeint ist, hat Dirichlet später mündlich von Gauss erfahren.*

2. See [Waterhouse 1984]. The eighth section is discussed by G. Frei in chap. II.4 of the

this important correction to cancel the remaining, more interesting part of Kummer's assertion. Although one of the three men, Gauss or Dirichlet or Kummer, appears to have misremembered or misunderstood what had occasioned a footnote published 47 years earlier, they all seem to have thought in 1846 that Gauss used "something similar to ideal prime factors" for his own calculations of compositions of forms when he was composing the *Disquisitiones*, but that he had not put it on "firm ground." Consideration of such a possibility raises an interesting question about the *Disquisitiones*: What was Gauss's conception of "firm ground" in 1801, and – regardless of what he might have left out – what firm ground underlay the theories that he did include?

There are no statements about the foundations of mathematics in the *Disquisitiones.* A glimpse of Gauss's views appears in his statement in the preface that all of mathematical analysis is the study of general properties and relations of numerical[3] quantities, whereas number theory (arithmetic) studies just *whole* numbers. This attitude implies that he thought of mathematics as being founded on the notion of "number," but he seems never to have discussed, in the *Disquisitiones* or elsewhere, his conception of "numbers."[4] In sec. 7 he certainly computes with irrational numbers – for example, the values of the trigonometric functions for arguments of the form $2\pi p/q$ with integral p and q in art. 336 – but he gives no explanation of them. He does not even justify his use of them in a book on arithmetic other than to say that "the exposition will make abundantly clear that this subject is linked to higher arithmetic in an intimate connection."[5] I infer from these few remarks that Gauss's view of mathematics was that it deals with *computations* with *numbers,* and that, like many other mathematicians since, his interest lay in pursuing mathematics itself, not in investigating its metaphysical underpinnings in the notion of number.

1. The Composition of Forms in the *Disquisitiones*

The difficult theory of composition of forms in sec. 5 is indeed closely related to Kummer's ideal prime factors, so it is not surprising that Kummer, Dirichlet and Gauss would have discussed connections between the two. Kummer explicitly mentioned binary quadratic forms in his first paper on ideal prime factors, [Kummer 1847], saying that the theory of numbers of the form $x+y\sqrt{D}$ leads to a theory of ideal factors, and that the natural way of partitioning these ideal factors into equivalence classes corresponds exactly to Gauss's way of partitioning binary quadratic forms into equivalence classes. He saw this as a powerful validation of his theory because the

present book [Editors' note].

3. Gauss does not refer specifically to *numerical* quantities, but I am told that Maser's use of this term in his 1889 German translation correctly describes the way in which Gauss's contemporaries would have understood his phrase.
4. A small note by Gauss on his conception of magnitudes, "Zur Metaphysik der Mathematik," published in vol. XII of his *Werke*, is discussed by J. Boniface in chap. V.1 of the present book [Editors' note].
5. Gauss's *Disquisitiones Arithmeticae*, art. 335: *Tractatio ipsa abunde declarabit, quam intimo nexu hoc argumentum cum arithmetica sublimiori coniunctum sit.*

Gaussian classification of forms, although it appeared artificial from the standpoint of the theory of forms, had been demonstrated by Gauss to be more fruitful than the obvious classification. Unfortunately, Kummer gave no detailed explanation, and he never returned to the subject of ideal prime factors of numbers $x + y\sqrt{D}$ and binary quadratic forms.

A great obstacle for modern students of Gauss's theory of composition of forms (arts. 234–251) is Gauss's use of the word "composition" to denote an operation that is *not a binary operation.* Modern treatments normally ignore the composition of forms altogether and deal only with the composition of *equivalence classes* of forms, which *is* a binary operation. Even André Weil, [Weil 1984], p. 334, says the Gaussian theory was a "stumbling-block" until Dirichlet "restored its simplicity," without noting that Dirichlet only composed forms that satisfy certain *additional conditions* (conditions of "concordance"). Dirichlet in fact made no attempt to compose forms, as Gauss had done, but instead focussed on the question of determining which numbers were represented by which forms; in this study, it is natural to replace a form by an equivalent form whenever it is convenient to do so, and that is what Dirichlet did. In other words, he did not compose the two given forms, but instead replaced them, when necessary, with equivalent forms in order to find forms that were easy to compose. In this way, he solved the problems that interested him and avoided the complications of Gauss's theory, but he left aside the challenging problem Gauss had successfully solved, the problem of composing arbitrarily given forms. (See [Dirichlet 1851] or §146 of [Dirichlet-Dedekind 1879].)

Composition of forms is an elaboration of the ancient formula

$$(x^2 - Dy^2)(u^2 - Dv^2) = (xu + Dvy)^2 - D(xv + yu)^2, \tag{0}$$

where D is a specified integer. Given three binary quadratic forms f, ϕ, and F (in the ancient example, all three are the form $X^2 - DY^2$), a *transformation* formula is a formula $f(x, y)\phi(u, v) = F(X, Y)$ where X and Y are linear functions

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} \begin{bmatrix} xu \\ xv \\ yu \\ yv \end{bmatrix}$$

of the four monomials xu, xv, yu and yv. (Gauss did not, of course, write the transformation equations in the matrix form used here.) A transformation is a *composition* (art. 235) if (1) the six 2×2 minors of the matrix $[a_{ij}]$ have greatest common divisor 1 and (2) the first two minors, that is, $a_{11}a_{22} - a_{21}a_{12}$ and $a_{11}a_{23} - a_{21}a_{13}$, are both positive. In the example (0), the transformation is given by the matrix

$$\begin{bmatrix} 1 & 0 & 0 & D \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

so it is a composition because the six minors are 1, 1, 0, 0, $-D$ and $-D$. Actually, Gauss only imposes the technical condition (2) at a later stage in the development of the theory (art. 239).

— 341 —

$$App + 2Bpq + Cqq = aa' \quad [1]$$
$$Ap'p' + 2Bp'q' + Cq'q' = ac' \quad [2]$$
$$Ap''p'' + 2Bp''q'' + Cq''q'' = ca' \quad [3]$$
$$Ap'''p''' + 2Bp'''q''' + Cq'''q''' = cc' \quad [4]$$
$$App' + B(pq' + qp') + Cqq' = ab' \quad [5]$$
$$App'' + B(pq'' + qp'') + Cqq'' = ba' \quad [6]$$
$$Ap'p''' + B(p'q''' + q'p''') + Cq'q''' = bc' \quad [7]$$
$$Ap''p''' + B(p''q''' + q''p''') + Cq''q''' = cb' \quad [8]$$
$$A(pp''' + p'p'') + B(pq''' + qp''' + p'q'' + q'p'') + C(qq''' + q'q'') = 2bb' \quad [9]$$

Sint determinantes formarum F, f, f' resp. D, d, d'; diuisores communes maximi numerorum A, $2B$, C; a, $2b$, c; a', $2b'$, c' resp. M, m, m' (quos omnes positiue acceptos supponimus). Porro determinentur sex numeri integri $\mathfrak{A}, \mathfrak{B}, \mathfrak{C}$, $\mathfrak{A}', \mathfrak{B}', \mathfrak{C}'$ ita vt sit $\mathfrak{A}a + 2\mathfrak{B}b + \mathfrak{C}c = m$, $\mathfrak{A}'a' + 2\mathfrak{B}'b' + \mathfrak{C}'c' = m'$. Denique designentur numeri $pq' - qp'$, $pq'' - qp''$, $pq''' - qp'''$, $p'q'' - q'p''$, $p'q''' - q'p'''$, $p''q''' - q''p'''$ resp. per P, Q, R, S, T, U, sitque ipsorum diuisor communis maximus positiue acceptus $= k$. — Iam ponendo

$$App''' + B(pq''' + qp''') + Cqq''' = bb' + \Delta \quad [10]$$

fit ex aequ. 9

$$Ap'p'' + B(p'q'' + q'p'') + Cq'q'' = bb' - \Delta \quad [11]$$

Ex his vndecim aequationibus 1 ... 11, sequentes nouas euoluimus *):

*) Origo harum aequationum haec est: 12 ex 5. 5 — 1. 2; 13 ex 5. 9 — 1. 7 — 2. 6; 14 ex 10. 11 — 6. 7; 15 ex 5. 8 + 5. 8 + 10. 10 + 11. 11 — 1. 4 — 2. 3 —

Y 3

Fig. II.2A. Computations at the core of Gauss's composition of forms: an extract from art. 235 in the 1801 edition of the *Disquisitiones arithmeticae*.

Gauss sets down in equations [1]-[9] the equations that describe a composition formula and begins the long solution of the problem, "Given two forms, determine whether there is a third form that composes them, and, if so, find all such forms."

Note that this definition rests on the firm ground of computation with whole numbers. The forms f, ϕ and F are described by triples of integers, and a composition is described by equations of the form given above. Composition is not a binary operation, but a ternary *relation.* Given f and ϕ, there may not be any F at all for which there is a composition formula, so it is meaningless to talk about "*the* composite of two given forms."[6] To make matters worse, if there *is* an F there are certainly *infinitely many* of them, because X and Y can be subjected to an arbitrary unimodular change of variables.

Gauss says (art. 234) "thus far no one has considered this topic." Small wonder. Ever since Gauss, mathematicians have struggled with it. The first theorem he proves states: If f and ϕ can be composed, the ratio of their determinants must be a ratio of squares. The proof is a display of algebraic virtuosity that occupies a few pages. Just as demanding and lengthy is his proof of the converse: if the ratio of the determinants of two forms is a ratio of squares, a third form can be written as a composite of them. His proof of this theorem is of course a construction; given two forms, and given that the ratio of their determinants is equal to a ratio of squares, his proof (art. 236) is an algorithm for constructing a third form and a 2×4 matrix that fulfills the conditions that define compositions. (In fact, in his masterly fashion, he shows how to construct *all possible* compositions of them.)

This is lengthy and daunting, but Gauss has only begun to do what he needs to do. Earlier (art. 158) he has defined what it means for two forms to be *equivalent*. In the main this definition is the natural one – each form can be obtained from the other by a change of variables with integer coefficients – but, as Gauss was the first to realize, addition of the seemingly unnatural requirement for the determinant of the change of variables to be *positive* improves the theory. He next proves that compositions of equivalent forms are equivalent. More precisely, if a form F can be expressed as a composite of two forms f and f', and if f'' is a form equivalent to f', then there is a form F' that is a composite of f and f'', and any such F' is equivalent to F. Note that the second statement implies, when $f' = f''$, that two compositions of the same pair of forms are equivalent.

And more: He needs to prove that this binary operation on equivalence classes is *associative,* a theorem that requires several more pages of work. Because he deals with composition of forms rather than of classes of forms, his statement of this associative law in art. 240 needs to be rather lengthy:

> If from the forms f, f' the form F is composed, from F, f'' the form $\mathfrak{F}$, from f, f'' the form F' and from F', f' the form $\mathfrak{F}'$, then the forms $\mathfrak{F}$ and $\mathfrak{F}'$ are equivalent.

Finally, after eight long and difficult articles (arts. 234–241) dealing with composition of forms in complete generality, Gauss turns in art. 242 to the specific problem of computing composites of given forms. Again I would like to emphasize that he composes *forms,* not equivalence classes. For example, in art. 243 he gives himself the problem of finding a form that is a composite of the forms (3, 1, 47), (4, 0, 35),

6. Both the English and the German translations of the *Disquisitiones* wrongly translate the theorem of art. 249 when they use definite articles rather than indefinite ones; the original Latin of course has no articles.

solus $\mathfrak{P}'''$ ingreditur, qui est valor expr. $\frac{h^\nu}{b+b'}$ (mod. h^λ). Si *e. g.* quaeritur forma composita ex (16, 3, 19) et (8, 1, 37), est $h = 2$, $\varkappa = 4$, $\lambda = 5$, $\nu = 2$. Hinc $A = 8$, $\mathfrak{P}'''$ valor expr. $\frac{4}{4}$ mod. 8), qualis est 1, vnde $B = 8k - 73$, adeoque faciendo $k = 9$, $B = -1$ atque $C = 37$, siue (8, — 1, 37) forma quaesita.

Propositis itaque formis quotcunque, quarum termini initiales omnes sunt potestates numerorum primorum, circumspiciendum erit, num aliquarum termini antecedentes sint potestates *eiusdem* numeri primi, atque hae inter se respectiue per regulam modo traditam componendae. Hac ratione prodibunt formae, quarum termini primi etiamnum erunt potestates numerorum primorum, sed omnino diuersorum; forma itaque ex his composita per obseru. tertiam definiri poterit. *E. g.* propositis formis (3, 1, 47), (4, 0, 35), (5, 0, 28), (16, 2, 9), (9, 7, 21), (16, 6, 11), ex prima et quinta conflatur forma (27, 7, 7); ex secunda et quarta confit (16, — 6, 11, ex hac et sexta (1, 0, 140), quae negligi potest. Supersunt itaque (5, 0, 28), (27, 7, 7), ex quibus producitur (135, — 20, 4), cuius loco assumi potest proprie aequiualens (4, 0, 35). Haec itaque est resultans ex compositione sex propositarum.

Ceterum ex hoc fonte plura alia artificia in applicatione vtilia hauriri possunt; sed ne nimis

Fig. II.2B. A composite of 6 forms:
an extract from art. 243 in the 1801 edition of the *Disquisitiones arithmeticae*.

(5, 0, 28), (16, 2, 9), (9, 7, 21) and (16, 6, 11), all of which have determinant −140. The form he finds is (135, −20, 4).

He goes on to mention that (135, −20, 4) is equivalent to (4, 0, 35) – a fact that follows easily from the presence of 4 in both and the divisibility of the middle terms in both by 4 – perhaps because (4, 0, 35) is a simpler representative of the equivalence class of the result, and, as was noted above, if F is a composite of f and f', then any form equivalent to F is also a composite of f and f'. The result (4, 0, 35) can also be found in the following way: as Gauss states, (27, 7, 7) is a composite of the first and fifth forms, (3, 1, 47) and (9, 7, 21); and, as is easily found (see below), (4, 0, 35) is a composite of the fourth and sixth forms, (16, 2, 9), and (16, 6, 11). Since (20, 0, 7) is a composite of (4, 0, 35) and (5, 0, 28) – easily found because 4 and 5 are relatively prime – a composite of all six forms but the second is found by using the fact that 27 and 20 are relatively prime to conclude that (540, −20, 1) is a composite of (27, 7, 7) and (20, 0, 7) and therefore is a composite of the five forms other than (4, 0, 35). This form (540, −20, 1) is equivalent to the principal form (1, 0, 140), as the last coefficient 1 shows, so any composite of all six forms must be equivalent to the composite of (1, 0, 140) and (4, 0, 35) and must therefore be equivalent to (4, 0, 35).

My main point is that *computations* of this sort are the core of Gauss's theory of composition of forms. Gauss has gone to great lengths to describe in full generality the ways in which they may be done and the properties they have. His immediate purpose is, as the following sections of the *Disquisitiones* show, the proof of the law of quadratic reciprocity, which he extracts from simple facts about composition of primitive equivalence classes of forms for various determinants.[7] This marvellous proof leaves the reader with an impression that the theory is a powerful tool that will open the way to other realms of arithmetic, as indeed it has.

2. Revisiting the Composition of Forms

I would now like to describe a simple method of accomplishing the composition of forms that I hope will give some insight into the operations involved and into the way in which Gauss's approach, cumbersome as it is, does place the theory on the firm ground of computations with integers in an admirable and rather natural way.

Let an integer D, not a square, be fixed. I will take the addition and multiplication of numbers $x + y\sqrt{D}$, where x and y are integers, for granted. By a *module* of numbers of the form $x + y\sqrt{D}$, where x and y are integers, I will mean a list of (a finite number of) such numbers written between square brackets, $[x_1 + y_1\sqrt{D}, x_2 + y_2\sqrt{D}, \ldots, x_n + y_n\sqrt{D}]$. The term "module" is motivated by the following definition: a module is *congruent to zero* modulo another module, written

$$[x_1 + y_1\sqrt{D}, \ldots, x_n + y_n\sqrt{D}] \equiv 0 \bmod [x'_1 + y'_1\sqrt{D}, \ldots, x'_m + y'_m\sqrt{D}]$$

7. A form is said to be primitive if the coefficients (a, b, c) have no common divisor; if there is a primitive form in an equivalence class, all the forms of the class are primitive and the class is said to be primitive.

if each of its entries is a sum of multiples of entries of the other module in the sense that for each $i = 1, 2, \ldots, n$ there are integers $u_1, u_2, \ldots, u_m, v_1, v_2, \ldots, v_m$ such that

$$x_i + y_i\sqrt{D} = \sum_{\sigma=1}^{m} (u_\sigma + v_\sigma\sqrt{D})(x'_\sigma + y'_\sigma\sqrt{D}).$$

Two modules are by definition *equal* if each is congruent to zero modulo the other. As is easily seen, two modules are equal if and only if each can be transformed into the other by a sequence of operations of three types: (1) Rearrange terms. (2) Annex or delete zeros. (3) Add a multiple of one entry in the module to another entry – the multiplier being a number of the form $x + y\sqrt{D}$. We can then find a simple, uniquely determined representation of a given module:

Theorem. Let an integer D, not a square, be fixed, and let a list $x_1 + y_1\sqrt{D}$, $x_2 + y_2\sqrt{D}$, …, $x_n + y_n\sqrt{D}$ of numbers of the form $x + y\sqrt{D}$ be given. Provided at least one of the listed numbers is not zero, there are nonnegative integers e, f and g for which $ef \neq 0$, $g < f$, $g^2 \equiv D \bmod f$ and

$$[x_1 + y_1\sqrt{D}, x_2 + y_2\sqrt{D}, \ldots, x_n + y_n\sqrt{D}] = [ef, eg + e\sqrt{D}].$$

Two modules in this form $[ef, eg + g\sqrt{D}]$, where e, f and g are nonnegative integers, $ef \neq 0$, $g < f$, and $g^2 \equiv D \bmod f$, are equal only if they are identical.

Proof. Let a module be called *full* if $\sqrt{D}$ times any entry in the module can be written as a sum of *integer* multiples of entries in the module. Every module is equal to a full module, as one can prove as follows: Double the length of the module by annexing to the end a number of zeros equal to the number of terms in the module. To each of the zeros in the second half, add $\sqrt{D}$ times the corresponding term in the first half. Then $\sqrt{D}$ times any term in the second half is equal to D times the corresponding term in the first half, so the new module is both full and equal to the original one.

Since every module is equal to a full module, it will suffice to prove that every full module that is not equal to [0] is equal to one in the required form $[ef, eg + g\sqrt{D}]$. This will be done in two stages.

Stage one: Because reversing the sign of an entry in a module obviously gives an equal module, one can assume without loss of generality that the coefficient of $\sqrt{D}$ in each term listed in the module is nonnegative. Because the module is assumed to be full and not equal to [0], at least one entry must contain $\sqrt{D}$ with a positive coefficient. (Use is made here of the assumption that $\sqrt{D}$ is not an integer.) If only one entry does, pass to stage two. Otherwise, choose an entry in which the coefficient of $\sqrt{D}$ is positive but otherwise as small as possible and subtract this entry from each other entry in which the coefficient of $\sqrt{D}$ is positive. The new module obtained in this way is equal to the old one, and the coefficients of $\sqrt{D}$ are all nonnegative. The new module is also full. (A number $x + y\sqrt{D}$ is a sum of integer multiples of the entries in the old module if and only if it is a sum of integer multiples of entries in the new module, and $\sqrt{D}$ times an entry in the new module is either $\sqrt{D}$ times an

entry in the old module or it is a difference of two such, so it is certainly a sum of integer multiples of entries in either the old or the new module.) Such a step reduces the total of the coefficients of $\sqrt{D}$ (by $(k-1)$ times the smallest of the nonzero coefficients, where k is the number of nonzero coefficients), so repetition of the step eventually reaches a full module in which all entries but one are integers, and the one entry that is not an integer contains $\sqrt{D}$ with a positive coefficient, at which point one passes to stage two.

Stage two: Given a full module in which all terms but one are integers, one can again reverse signs, if necessary, to find an equal full module in which all terms but one are *nonnegative* integers. Delete all zeros from the module. At least one positive integer remains, because the module is assumed to be full and not equal to [0] (and D is not a square, so $(x+y\sqrt{D})\sqrt{D} = \mu(x+y\sqrt{D})$ is impossible for integer μ). If only one remains, a module of the form $[a, b+\sqrt{D}]$ that is full and equal to the original module has been reached. Otherwise, among the integers in the module (now all positive) choose one that is as small as possible, subtract it from each of the other integers in the module, and delete all zeros that result. Since each step of this type reduces the total of the integers in the module, repetition of it eventually results in a full module of the form $[a, b+c\sqrt{D}]$ equal to the original module.

Thus, given any module that is not equal to [0], one can construct a full module of the form $[a, b+c\sqrt{D}]$ that is equal to it. Moreover, one can assume without loss of generality that a and c are positive. Because $a\sqrt{D} = \mu\cdot a + \nu\cdot(b+c\sqrt{D})$ where μ and ν are integers, $a = \nu\cdot c$. Moreover, $b\sqrt{D}+cD = \sigma\cdot a+\tau\cdot(b+c\sqrt{D})$, so $b = \tau\cdot c$ and $cD = \sigma a+\tau b = \sigma\nu c+\tau^2 c$. Thus, with $e=c$, $f=\nu$ and $g=\tau$, the module is $[ef, eg+e\sqrt{D}]$, where $g^2+\sigma f = D$, so $g^2 \equiv D \bmod f$. Since eg can be changed by any multiple of ef, g can be replaced by any integer congruent to it mod f, and one can assume without loss of generality that $0 \le g < f$, in which case the module has the required form.

Suppose now that both $[ef, eg+e\sqrt{D}]$ and $[e'f', e'g'+e'\sqrt{D}]$ have the required form, and suppose they are equal. Since $[ef, eg+e\sqrt{D}]$ is full, the statement that $[e'f', e'g'+e'\sqrt{D}] \equiv 0 \bmod [ef, eg+e\sqrt{D}]$ implies $e'f' = \mu\cdot ef+\nu\cdot(eg+e\sqrt{D})$ and $e'g'+e'\sqrt{D} = \sigma\cdot ef+\tau\cdot(eg+e\sqrt{D})$. Since ν must be zero, $e'f'$ must be a multiple of ef. By symmetry, ef must also be a multiple of $e'f'$. Since they are both positive integers, $ef = e'f'$. Similarly, since $e' = \tau\cdot e$ and, by symmetry, e is also a multiple of e' and both are positive, $e = e'$. Thus, $f = f'$. Since τ must then be 1, $eg' = e'g' = \sigma\cdot ef+eg$, so g' must be congruent to g mod f. Since both are nonnegative and less than f, $g = g'$, and the proof is complete.

A module of this form $[ef, eg+\sqrt{D}]$ will be said to be in *canonical form*. (The integers e, f and g are nonnegative, $ef \neq 0$, $g < f$, and $g^2 \equiv D \bmod f$.) The Theorem solves the problem: "Given two modules, determine whether they are equal." Each is equal to one in canonical form, and two in canonical form are equal only if they are identical.

Modules can be *multiplied* in a natural way: the entries of the product module are the products of two factors in which the first factor is from the first module and the second factor is from the second. This definition depends, of course, on the fact

that it is consistent with the definition of equality of modules, which is to say that if one of the two modules is replaced by an equal module, the product module is replaced by an equal module. This is easy to prove.

Thus, every module can be written as a product $[e][f, g+\sqrt{D}]$ in which the first factor is $[e]$ for a positive integer e and the second factor is $[f, g+\sqrt{D}]$, where f is a positive integer and g is a square root of D mod f. Multiplication of any module by $[e]$ is easy, so the multiplication of two modules in canonical form, say $[e][f, g+\sqrt{D}]$ and $[e'][f', g'+\sqrt{D}]$ comes down to the computation of $[f, g+\sqrt{D}][f', g'+\sqrt{D}]$ which is to say the reduction of $[ff', f(g'+\sqrt{D}), f'(g+\sqrt{D}), gg'+D+(g+g')\sqrt{D}]$ to canonical form. *This operation contains the essence of the idea of the composition of forms.*

For example, Gauss's statement, mentioned above, that $(27, 7, 7)$ is a composite of $(3, 1, 47)$ and $(9, 7, 21)$ follows from

$$
\begin{aligned}
&[3, 1+\sqrt{-140}][9, 7+\sqrt{-140}] \\
&\qquad = [27, 3(7+\sqrt{-140}), 9(1+\sqrt{-140}), -133+8\sqrt{-140}] \\
&\qquad = [27, 21+3\sqrt{-140}, 9+9\sqrt{-140}, 2+8\sqrt{-140}] \\
&\qquad = [27, 21+3\sqrt{-140}, 7+\sqrt{-140}, 2+8\sqrt{-140}] \\
&\qquad = [27, 0, 7+\sqrt{-140}, -54] = [27, 7+\sqrt{-140}].
\end{aligned}
$$

(These two forms $(3, 1, 47)$ and $(9, 7, 21)$ are concordant in Dedekind's sense, which is to say that the greatest common divisor of $a = 3$, $\alpha = 9$ and $b+\beta = 1+7$ is 1. Therefore the composite $(27, 7, 7)$ is determined, as Dedekind showed, by the fact that $B = 7$ must be 1 mod 3 and 7 mod 9 and must be a square root of -140 mod 27.)

Similarly, the above statement that $(4, 0, 35)$ is a composite of $(16, 2, 9)$ and $(16, 6, 11)$ follows from

$$
\begin{aligned}
&[16, 2+\sqrt{-140}][16, 6+\sqrt{-140}] \\
&\qquad = [256, 16(6+\sqrt{-140}), 16(2+\sqrt{-140}), -128+8\sqrt{-140}] \\
&\qquad = [8][64, 12+2\sqrt{-140}, 4+2\sqrt{-140}, -16+\sqrt{-140}] \\
&\qquad = [8][64, 8, 4+2\sqrt{-140}, -16+\sqrt{-140}] \\
&\qquad = [8][8, 4+2\sqrt{-140}, \sqrt{-140}] = [8][4, \sqrt{-140}].
\end{aligned}
$$

(This is a composition of forms that are not concordant in Dedekind's sense, which is to say that the greatest common divisor of $a = 16$, $\alpha = 16$, and $b+\beta = 8$ is not equal to 1. Therefore, Dedekind's method does not produce a composite.)

If f and f' are relatively prime, the product $[f, g+\sqrt{D}][f', g'+\sqrt{D}]$ of two modules in canonical form with $e = 1$ is simply $[ff', G+\sqrt{D}]$, where G is determined mod ff' by $G \equiv g \bmod f$ and $G \equiv g' \bmod f'$. This is a consequence of the fact that there are integers σ and τ for which $\sigma f + \tau f' = 1$; since both

$\sigma f(G+\sqrt{D})$ and $\tau f'(G+\sqrt{D})$ are zero mod $[f, G+\sqrt{D}][f', G+\sqrt{D}]$, so is their sum $G+\sqrt{D}$, and

$$\begin{aligned}[f, g+\sqrt{D}][f', g'+\sqrt{D}] &= [f, G+\sqrt{D}][f', G+\sqrt{D}] \\ &= [ff', f(G+\sqrt{D}), f'(G+\sqrt{D}), (G+\sqrt{D})^2] \\ &= [ff', f(G+\sqrt{D}), f'(G+\sqrt{D}), (G+\sqrt{D})^2, G+\sqrt{D}] \\ &= [ff', G+\sqrt{D}].\end{aligned}$$

Explicitly, multiplication of modules can be used to construct composites of given forms[8] in the following way (assuming, of course, that the ratio of their determinants is a ratio of squares):

Theorem. Let $ax^2+2bxy+cy^2$ and $\alpha u^2+2\beta uv+\gamma v^2$ be given forms, and suppose that the ratio of their determinants is a ratio of squares, but that the determinants themselves are not squares. An explicit composition formula

$$(ax^2+2bxy+cy^2)(\alpha u^2+2\beta uv+\gamma v^2) = AX^2+2BXY+CY^2 \qquad (1)$$

can be constructed as follows. Choose positive integers s and σ for which $s^2(b^2-ac)$ and $\sigma^2(\beta^2-\alpha\gamma)$ are equal. Let D denote their common value, which is by assumption not a square. Put the module $[sa, sb+\sqrt{D}][\sigma\alpha, \sigma\beta+\sqrt{D}]$ in canonical form, say $[sa, sb+\sqrt{D}][\sigma\alpha, \sigma\beta+\sqrt{D}] = [E][F, G+\sqrt{D}]$. Then formula (1) holds for

$$AX^2+2BXY+CY^2 = \pm\frac{m\mu}{M}(FX^2+2GXY+HY^2) \qquad (2)$$

when the sign is the sign of $a\alpha$, when $H=(G^2-D)/F$, when m, μ and M are the positive integers defined by $[m]=[a, 2b, c]$, $[\mu]=[\alpha, 2\beta, \gamma]$ and $[M]=[F, 2G, H]$ (in short, they are the "contents" of the forms $ax^2+2bxy+cy^2$, $\alpha u^2+2\beta uv+\gamma v^2$ and $FX^2+2GXY+HY^2$) and when X and Y are the linear functions of xu, xv, yu and yv determined implicitly by

$$(sax+(sb+\sqrt{D})y)(\sigma\alpha u+(\sigma\beta+\sqrt{D})v) = E(FX+(G+\sqrt{D})Y). \qquad (3)$$

Proof. Reversing the sign of either of the given forms merely reverses the signs of both sides of (1) (the sign of the right side is reversed because the sign of (2) is reversed), so there is no loss of generality in assuming that a and α are both positive. Neither a nor α can be zero because the determinants are by assumption not squares. The definition (3) of X and Y obviously implies $saxv+\sigma\alpha yu+(sb+\sigma\beta)yv = EY$, after which it implies $EFX = sa\sigma\alpha xu+sa\sigma\beta xv+\sigma\alpha sbyu+(sb\sigma\beta+D)yv-$

8. The degenerate case in which the given forms factor over the rationals – which is to say that their determinants are squares – will be ignored.

$G(saxv + \sigma\alpha yu + (sb + \sigma\beta)yv)$ so the explicit expression of X and Y in terms of xu, xv, yu and yv is

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \frac{sa\sigma\alpha}{EF} & \frac{sa(\sigma\beta - G)}{EF} & \frac{\sigma\alpha(sb-G)}{EF} & \frac{sb\sigma\beta + D - G(sb+\sigma\beta)}{EF} \\ 0 & \frac{sa}{E} & \frac{\sigma\alpha}{E} & \frac{sb+\sigma\beta}{E} \end{bmatrix} \begin{bmatrix} xu \\ xv \\ yu \\ yv \end{bmatrix}. \quad (4)$$

Multiplication of the defining equation (3) by its conjugate $(sax + (sb - \sqrt{D})y)$ $(\sigma\alpha u + (\sigma\beta - \sqrt{D})v) = E(FX + (G - \sqrt{D})Y)$ – which is the same statement as (3) – gives $((sax + sby)^2 - (s^2b^2 - s^2ac)y^2)((\sigma\alpha u + \sigma\beta v)^2 - (\sigma^2\beta^2 - \sigma^2\alpha\gamma)v^2) = E^2((FX + GY)^2 - (G^2 - FH)Y^2)$, that is,

$$s^2\sigma^2 a\alpha(ax^2 + 2bxy + cy^2)(\alpha u^2 + 2\beta uv + \gamma v^2) = E^2F(FX^2 + 2GXY + HY^2).$$

Division by $s^2\sigma^2 a\alpha$ gives equation (1) with $A = \frac{E^2F^2}{s^2\sigma^2 a\alpha}$, $2B = \frac{2E^2FG}{s^2\sigma^2 a\alpha}$ and $C = \frac{E^2FH}{s^2\sigma^2 a\alpha}$. Thus, the theorem will be proved[9] if $\frac{E^2F}{s^2\sigma^2 a\alpha}$ is shown to be equal to $\frac{m\mu}{M}$, if the entries of the matrix in (4) are shown to be integers and if the greatest common divisor of the 2×2 minors of this matrix is shown to be 1. (The first two minors $\frac{(sa)^2\sigma\alpha}{E^2F}$ and $\frac{sa(\sigma\alpha)^2}{E^2F}$ are positive because s, σ, a, α and F are all positive.)

By definition, $[sa, sb + \sqrt{D}][\sigma\alpha, \sigma\beta + \sqrt{D}] = [E][F, G + \sqrt{D}]$. Thus, E, F and G are found by putting $[sa\sigma\alpha, sa\sigma\beta + sa\sqrt{D}, \sigma\alpha sb + \sigma\alpha\sqrt{D}, sb\sigma\beta + D + (sb + \sigma\beta)\sqrt{D}]$ in canonical form. Let $\mathfrak{P}'$, $\mathfrak{P}''$, and $\mathfrak{P}'''$ (Gauss's notation) be such that $\mathfrak{P}'sa + \mathfrak{P}''\sigma\alpha + \mathfrak{P}'''(sb + \sigma\beta)$ is the greatest common divisor of sa, $\sigma\alpha$, and $sb + \sigma\beta$, call it d. Then d clearly divides both coefficients of all four numbers in the product module $[sa\sigma\alpha, sa\sigma\beta + sa\sqrt{D}, \sigma\alpha sb + \sigma\alpha\sqrt{D}, sb\sigma\beta + D + (sb + \sigma\beta)\sqrt{D}]$ with the possible exception of $sb\sigma\beta + D$, and it divides this coefficient as well because $D \equiv (sb)^2 \bmod sa$, so $sb\sigma\beta + D \equiv sb\sigma\beta + (sb)^2 \equiv sb(\sigma\beta + sb) \equiv 0 \bmod d$. Let G_0 be defined by the equation $d(G_0 + \sqrt{D}) = \mathfrak{P}'sa(\sigma\beta + \sqrt{D}) + \mathfrak{P}''\sigma\alpha(sb + \sqrt{D}) + \mathfrak{P}'''(sb\sigma\beta + D + (sb + \sigma\beta)\sqrt{D})$. In other words,

$$G_0 = \frac{1}{d}(\mathfrak{P}'sa\sigma\beta + \mathfrak{P}''\sigma\alpha sb + \mathfrak{P}'''(sb\sigma\beta + D)).$$

Then

$[sa, sb + \sqrt{D}][\sigma\alpha, \sigma\beta + \sqrt{D}]$

$= [sa\sigma\alpha, sa\sigma\beta + sa\sqrt{D}, \sigma\alpha sb + \sigma\alpha\sqrt{D}, sb\sigma\beta + D + (sb + \sigma\beta)\sqrt{D}, d(G_0 + \sqrt{D})]$

$= [sa\sigma\alpha, sa\sigma\beta - saG_0, \sigma\alpha sb - \sigma\alpha G_0, sb\sigma\beta + D - (sb + \sigma\beta)G_0, d(G_0 + \sqrt{D})]$.

This module is full because it can be found by starting with the full module $[sa, sb + \sqrt{D}][\sigma\alpha, \sigma\beta + \sqrt{D}]$ (full because each factor is full), annexing a zero, and adding

9. By the definition of M, the coefficients A, $2B$, C of the composite form are integers. For (1) to be a composition in the Gaussian sense, $2B$ must be proved to be *even,* but, for reasons explained below, this point will be ignored and (1) will be accepted as a "composition" without a proof that $2B$ is even.

integer multiples of other entries to this zero – namely, adding $\mathfrak{P}'$, $\mathfrak{P}''$ and $\mathfrak{P}'''$ times the appropriate entries. Let F_0 be defined by

$$[dF_0] = [sa\sigma\alpha, sa\sigma\beta - saG_0, \sigma\alpha sb - \sigma\alpha G_0, sb\sigma\beta + D - (sb + \sigma\beta)G_0].$$

Then the module $[dF_0, dG_0 + d\sqrt{D}]$ can be reached by successively subtracting integer multiples of entries of $[sa\sigma\alpha, sa\sigma\beta - saG_0, \sigma\alpha sb - \sigma\alpha G_0, sb\sigma\beta + D - (sb + \sigma\beta)G_0, d(G_0 + \sqrt{D})]$ to reduce the integers in this module to their greatest common divisor dF_0 while leaving $d(G_0 + \sqrt{D})$ unchanged, so it is full, as well as equal to the product module.

Because $[dF_0, dG_0 + d\sqrt{D}]$ is full and d and F_0 are positive, it is in canonical form, except that G_0 may not be in the range $0 \leq G_0 < F_0$. Because $[dF_0, dG_0 + d\sqrt{D}] = [E][F, G + \sqrt{D}]$, it follows that $E = d$, $F = F_0$, and $G \equiv G_0 \bmod F$.

In particular, E is the greatest common divisor of sa, $\sigma\alpha$ and $sb + \sigma\beta$, which shows that the entries in the second row of the matrix in (4) are integers. Moreover, $EF = dF_0$ is the greatest common divisor of $sa\sigma\alpha$, $sa\sigma\beta - saG_0$, $\sigma\alpha sb - \sigma\alpha G_0$ and $sb\sigma\beta + D - (sb + \sigma\beta)G_0$, from which it follows, because $G_0 \equiv G \bmod F$, that the entries in the first row of the matrix in (4) are integers.

The product of $[sa, sb + \sqrt{D}]$ with its conjugate $[sa, sb - \sqrt{D}]$ is $[sa, sb + \sqrt{D}][sa, sb - \sqrt{D}] = [(sa)^2, sa(sb - \sqrt{D}), sa(sb + \sqrt{D}), s^2b^2 - s^2(b^2 - ac)] = [sa][sa, sb + \sqrt{D}, sb - \sqrt{D}, sc] = [sa][sa, 2sb, sc, sb + \sqrt{D}] = [sa][sm, sb + \sqrt{D}]$, where m is as in the statement of the theorem. Similar calculations apply to the other two modules in the equation $[sa, sb+\sqrt{D}][\sigma\alpha, \sigma\beta+\sqrt{D}] = [E][F, G+\sqrt{D}]$, so multiplication of this equation by its conjugate gives

$$[sa][sm, sb + \sqrt{D}][\sigma\alpha][\sigma\mu, \sigma\beta + \sqrt{D}] = [E^2][F][M, G + \sqrt{D}].$$

Since $sb \equiv -sb \bmod sm$, $\sigma\beta \equiv -\sigma\beta \bmod \sigma\mu$ and $G \equiv -G \bmod M$, the modules in this equation are self-conjugate and multiplication of the equation with its conjugate amounts to squaring and results in

$$[sa\sigma\alpha]^2[sm][sm, sb + \sqrt{D}][\sigma\mu][\sigma\mu, \sigma\beta + \sqrt{D}] = [E^2F]^2[M][M, G + \sqrt{D}]$$

which combines with the previous equation to give

$$[sa\sigma\alpha][sm\sigma\mu][E^2F][M, G + \sqrt{D}] = [E^2F]^2[M][M, G + \sqrt{D}].$$

When G is reduced mod M (the result must be 0 or $\frac{1}{2}M$) the modules on either side are in canonical form, and $sa\sigma\alpha \cdot sm\sigma\mu = E^2F \cdot M$ follows. In other words, $\frac{E^2F}{s^2\sigma^2a\alpha} = \frac{m\mu}{M}$, as was to be shown.

Finally, let Δ_{ij} for $0 < i < j \leq 4$ be the minor of the matrix in (4) that uses columns i and j. It remains to show that the greatest common divisor of the Δ_{ij} is equal to 1. By direct computation,

$$\begin{aligned}[E^2F][&\Delta_{12}, \Delta_{13}, \Delta_{14}, \Delta_{23}, \Delta_{24}, \Delta_{34}] \\ &= [sa\sigma\alpha][sa, \sigma\alpha, sb + \sigma\beta, \sigma\beta - sb, sc, \sigma\gamma].\end{aligned}$$

(The calculation is simplified when one observes that adding G times the second row to the first does not change the minors.) What is to be proved, then, is that $[E^2F] = [sa\sigma\alpha][sa, \sigma\alpha, sb + \sigma\beta, \sigma\gamma, sc]$, which is to say

$$\begin{aligned}[E][sa\sigma\alpha, sa(\sigma\beta - G_0), \sigma\alpha(sb - G_0), sb\sigma\beta + D - (sb + \sigma\beta)G_0] \\ = [sa\sigma\alpha][sa, \sigma\alpha, sb + \sigma\beta, \sigma\beta - sb, \sigma\gamma, sc].\end{aligned}$$

The first three terms on the right – that is, $sa\sigma\alpha$ times sa, $\sigma\alpha$ and $sb + \sigma\beta$ – are all divisible by $Esa\sigma\alpha$ and are therefore zero modulo the module on the left. The remaining three are zero modulo the module on the left by virtue of

$$sa\sigma\alpha(\sigma\beta - sb) = \frac{\sigma\alpha}{E} \cdot Esa(\sigma\beta - G_0) - \frac{sa}{E} \cdot E\sigma\alpha(sb - G_0)$$

$$\begin{aligned}sa\sigma\alpha\sigma\gamma &= \frac{sb + \sigma\beta}{E} \cdot Esa(\sigma\beta - G_0) - \frac{sa}{E} \cdot E(sb\sigma\beta + D - (sb + \sigma\beta)G_0) \\ sa\sigma\alpha sc &= \frac{sb + \sigma\beta}{E} \cdot E\sigma\alpha(sb - G_0) - \frac{\sigma\alpha}{E} \cdot E(sb\sigma\beta + D - (sb + \sigma\beta)G_0).\end{aligned}$$

Finally, the four terms in the module on the left are zero modulo the module on the right by virtue of

$$\begin{aligned}Esa\sigma\alpha &= \mathfrak{P}'s^2a^2\sigma\alpha + \mathfrak{P}''sa\sigma^2\alpha^2 + \mathfrak{P}'''(sb + \sigma\beta)sa\sigma\alpha \\ Esa(\sigma\beta - G_0) &= \mathfrak{P}''sa\sigma\alpha(\sigma\beta - sb) + \mathfrak{P}'''sa\sigma\alpha\sigma\gamma \\ E\sigma\alpha(sb - G_0) &= \mathfrak{P}'sa\sigma\alpha(sb - \sigma\beta) + \mathfrak{P}'''sa\sigma\alpha sc \\ E(sb\sigma\beta + D - (sb + \sigma\beta)\sigma\beta) &= \mathfrak{P}'sa\sigma\alpha\sigma\gamma + \mathfrak{P}''sa\sigma\alpha sc,\end{aligned}$$

(the last three equations are obtained by eliminating one of the $\mathfrak{P}$s from $E = \mathfrak{P}'sa + \mathfrak{P}''\sigma\alpha + \mathfrak{P}'''(sb + \sigma\beta)$ and $EG_0 = \mathfrak{P}'sa\sigma\beta + \mathfrak{P}''\sigma\alpha sb + \mathfrak{P}'''(sb\sigma\beta + D)$) and the proof is complete.

Allowing forms to have odd middle coefficients permits the theorem to take the more natural form

$$\frac{ax^2 + 2bxy + cy^2}{m} \cdot \frac{\alpha u^2 + 2\beta uv + \gamma v^2}{\mu} = \frac{FX^2 + 2GXY + HY^2}{M} \qquad (5)$$

of a composition of two *primitive* forms (forms in which the greatest common divisor of the coefficients is 1) in which the composite is also primitive. One can obviously compose *arbitrary* forms if one can compose *primitive* forms, so it is natural to restate the theorem in the form: Given two primitive forms $ax^2 + bxy + cy^2$ and $\alpha u^2 + \beta uv + \gamma v^2$, if the ratio of $b^2 - 4ac$ to $\beta^2 - 4\alpha\gamma$ is a ratio of squares, the obvious modification of the construction of the theorem gives a composition formula for them in which the composite form is primitive.

For example, to compose $x^2 + xy - y^2$ with itself, the theorem replaces it with $2x^2 + 2xy - 2y^2$ and computes $[2, 1 + \sqrt{5}][2, 1 + \sqrt{5}] = [2][2, 1 + \sqrt{5}]$ to find the composition $(2x^2 + 2xy - 2y^2)(2u^2 + 2uv - 2v^2) = 2(2X^2 + 2XY - 2Y^2)$, where

X and Y are defined by $(2x + (1+\sqrt{5})y)(2u + (1+\sqrt{5})v) = 2(2X + (1+\sqrt{5})Y)$, which is to say $X = xu + vy$ and $Y = xv + yu + yv$. The more natural way to state this composition formula is of course

$$(x^2 + xy - y^2)(u^2 + uv - v^2) = X^2 + XY - Y^2.$$

Formula (5) can be used to construct a composition of any two forms, when the ratio of their determinants is a ratio of squares, whether or not they are primitive and whether or not their middle coefficients are even. Once a *single* composition is known, all others are obtained by taking unimodular changes of variables $X' = pX + qY$, $Y' = rX + sY$, where p, q, r and s are integers with $ps - qr = 1$.

From Gauss's point of view the theorem does not provide a composition of two given forms until B is proved to be an integer, or, in the terms of the modified theorem, until it is proved that if the middle coefficients of the given primitive forms are both even, the middle coefficient of the composite form given by the theorem is even. This statement is true, as follows from Gauss's construction of art. 236, but it is a matter of little importance unless there is a reason to restrict consideration to forms with even middle coefficients, a point on which I and many other of Gauss's readers remain unconvinced.

3. Conclusion

I hope that the use of module multiplication in some measure simplifies Gauss's theory of composition of forms. For example, it clarifies the difficult associativity property described and proved by Gauss in art. 240. Multiplication of modules is *obviously* an associative binary operation, and this property easily translates into the property Gauss uses.[10]

But, more importantly, I hope that by focussing attention on Gauss's composition of actual forms – as opposed to equivalence classes of forms as in the modern theory – I have highlighted Gauss's great achievement in giving a rigorous treatment of the composition of binary quadratic forms in the greatest possible generality.

His theory is "rigorous," not only in the usual sense that it is mathematically convincing, but also in the literal sense that it makes great demands on the reader. The second kind of rigor has caused succeeding generation of mathematicians to devote some of their best efforts to avoiding it. But it is the first kind of rigor that makes Gauss the great master. It is based on his mastery of the computational structure of his subject and his ability to explain that structure in the most general circumstances. While they may seek to avoid the difficulties of Gauss's theory, succeeding generations should never cease to admire it.

10. Since multiplication of modules can be used to establish the theory of composition of forms, Gauss's proof of quadratic reciprocity using composition of forms can be deduced in this way. However, quadratic reciprocity can be proved working directly with multiplication of modules. Therefore, if the goal is quadratic reciprocity, one can dispense with quadratic forms altogether. Other aspects of Gauss's theory can be revisited in a similar way. See [Edwards 2005].

References

Dirichlet, Johann Peter Gustav Lejeune-. 1851. *De Formarum Binariarum Secundi Gradus Compositione*. Berlin: Akademie Verlag. Repr. with changes in *Journal für die reine und angewandte Mathematik* 47 (1854), 155–160. Repr. in *Werke*, ed. L. Fuchs and L. Kronecker, vol. 2, pp. 107–114. Berlin: Reimer, 1889-1897.

———. 1879. *Vorlesungen über Zahlentheorie*, ed. with supplements by R. Dedekind. 3rd ed. Braunschweig: Vieweg.

Edwards, Harold M. 2005. *Essays in Constructive Mathematics*. New York: Springer.

Kummer, Ernst Eduard. 1846. Letter to Leopold Kronecker, June 14, 1846. *Festschrift zur Feier des 100 Geburtstages Eduard Kummer mit Briefen an seine Mutter und an Leopold Kronecker*, ed. K. Hensel. Berlin and Leipzig: Teubner, 1910. Repr. in [Kummer 1975], vol. 1, p. 68.

———. 1847. Zur Theorie der complexen Zahlen. *Journal für die reine und angewandte Mathematik* 35, pp. 319–326. Repr. in [Kummer 1975], vol. 1, pp. 203–210.

———. 1975. *Collected Papers*, ed. A. Weil. 2 vols. New York: Springer.

Waterhouse, William. 1984. A note by Gauss referring to ideals? *Historia Mathematica* 11, 142–146.

Weil, André. 1984. *Number Theory. An Approach Through History from Hammurapi to Legendre*. Boston: Birkhäuser.