1:25 p.m. January 16, 2011

Introduction

Bill Casselman
University of British Columbia
cass@math.ubc.ca

I'll begin by stating Siegel's formula, along with whatever is required to make it comprehensible. Then I'll do one simple example. In the last section I'll discuss very briefly how it relates to other theorems in number theory, and in particular volumes on adele groups. All these topics will be covered later on in much detail.

Contents

1. Siegel's formula

Suppose for the moment R to be an arbitrary ring. In these notes, a quadratic form with coefficients in R is any homogeneous function of degree two $Q(x_1, \dots, x_d)$ on a free module V, say of rank d. Any quadratic form determines an associated symmetric bilinear form

$$B_Q(x, y) = B(x, y) = Q(x + y) - Q(x) - Q(y)\,.$$

This is not a universal convention—sometimes one starts with a symmetric bilinear form β and defines $Q(x) = \beta(x, x)$. In these circumstances, $B = 2\beta$. If 2 is invertible in R, the definitions are essentially equivalent, but otherwise not. The difference in definitions is seen already in the origins of the subject—for Gauss, a binary integral quadratic form was required to be of the form $Ax^2 + 2Bxy + Cy^2$ with A, B, C all integers. Thus $x^2 + xy + y^2$, which is perfectly acceptable according to my definition, did not qualify.

I am not sure exactly why number theorists in the nineteenth century followed. Even in modern times his condition is common if by no means universal—for example, [Cassells:1968] follows Gauss, but [Grosswald:1985] does not. One possible convenience of Gauss' definition is that, given a basis (e_i) of the R-module V, quadratic forms associated to symmetric bilinear forms β are in bijection with symmetric $n \times n$ matrices—we can associate to Q the matrix $M_Q = (q_{i,j})$ in which

$$q_{i,j} = \beta(e_i, e_j)\,.$$

In this case, we can write

$$Q(x) = {}^t x\, M_Q\, x = \sum_{i,j} q_{i,j} x_i x_j\,,$$

where x is a column matrix of coordinates with respect to the e_i. This makes notation mildly more convenient than in the other convention. What's really going on here is that there are are two separate concepts involved—symmetric integral bilinear forms and integral quadratic forms. There are contexts in which the bilinear forms are the natural objects—for example in the topology of even-dimensional manifolds when defining a form by the intersection of cycles—but number theory is not one of them.

Two integral quadratic forms are said to be **equivalent** if there exists an invertible linear transformation taking one into the other; and on R^n, with the standard basis, they are **properly** equivalent if this can be done by a unimodular transformation. If 2 is invertible, then two forms Q_1, Q_2 are equivalent if and only if we have a matrix equation

$$M_{Q_1} = {}^tX M_{Q_2} X$$

for some invertible $d \times d$ matrix X. The determinant $\det(M_Q)$ of the equivalence class of the form is hence well defined in R modulo multiplication by elements of $(R^\times)^2$. Again here, my convention is not universal—sometimes the image modulo $(R^\times)^2$ is called the discriminant, but I'll use this term for something else.

A form is called non-degenerate if B is a non-degenerate bilinear form—that is to say, if $B(x, y) = 0$ for all y in V implies that $x = 0$. But here, too, there is some variety in terminology, especially in characteristic 2. According to this definition, x^2 is degenerate in characteristic 2.

If $R = \mathbb{R}$, then two quadratic forms are equivalent if and only if they have the same signature—that is to say, every real quadratic form is equivalent to a unique one with an expression

$$x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2 \,,$$

in which case it is said to have signature (p, q). A real quadratic form is positive definite if $Q(x) > 0$ for all $x \neq 0$, or in other words if $p = d$. The different equivalence classes of non-degenerate forms are the connected components of the complement of the closed subset in which the determinant vanishes, in the space of all symmetric $d \times d$ matrices.

If Q is a quadratic form with rational coefficients, it determines as well a real form and one rational over each $\mathbb{Q}_p$. If two forms are equivalent over Q then each pair of these local quadratic forms are also equivalent. The Hasse principle (due in this case to Minkowski) asserts that the converse is true. It is not difficult to tell whether two local forms are equivalent, so this gives a practical criterion for equivalence of $\mathbb{Q}$-rational forms.

I now suppose that $V = \mathbb{Z}^d$, with the standard coordinate system. An integral quadratic form Q is one with has integral coefficients. In this situation it determines a real quadratic form, as well as one in each $(\mathbb{Z}_p)^d$. But it is no longer true that two forms that are equivalent over $\mathbb{R}$ as well as all $\mathbb{Z}_p$ are properly equivalent over $\mathbb{Z}$. We shall see later on how to construct a multitude of counter-examples. The distinction is measured by defining two integral forms are said to be in the same class if they are equivalent over $\mathbb{Z}$, and in the same genus if they are equivalent over $\mathbb{R}$ and each $\mathbb{Z}_p$, so the claim is that a given genus may fail to contain a unique class. The failure is not so bad, however:

[intro-reduction] **Theorem 1.1.** *The number of equivalence classes in a given genus of non-degenerate integral quadratic forms is finite.*

If R is a field, then the orthogonal group possesses reflections, so equivalence and proper equivalence are the same.

The integral forms Q_1, Q_2 are equivalent over $\mathbb{Z}$ if and only if there exists an integral matrix X whose inverse is integral transforming one to the other. The determinant of X must be a unit in $\mathbb{Z}$, hence ± 1. If two forms Q_1, Q_2, Q_3 are all equivalent, then at least two of them must be properly equivalent, so the map from proper classes to classes is at most two-to-one. There is therefore not much to be lost by restricting ourselves to the stricter notion of equivalence, and as we'll see there is something to be gained. From now on I'll consider only proper equivalence.

I'll also assume from now on that all integral quadratic forms in view are positive definite. In this case, if n is a natural number I define

$$N_{\mathbb{Z}}(Q, n) = \text{ the number of } x \text{ in } \mathbb{Z}^d \text{ such that } Q(x) = n \ .$$

This is necessarily finite, since it is the intersection of a compact ellipse with the discrete set $\mathbb{Z}^d$.

The history of number theory is tied up intimately with the problem of finding a formula for $N_{\mathbb{Z}}(Q,n)$. One of the simplest and oldest cases is that when $Q(x) = x_1^2 + x_2^2$, which we'll see treated in several ways, one of them going back to Fermat and the origins of modern number theory. I can at least state the result in very simple terms. First of all, a necessary condition can be found by considering the equation as a congruence condition—if x is an integer than $x^2 \equiv 0$ or $1 \bmod 4$, so n cannot be a sum of two squares if $n \equiv 3 \bmod 4$.

The orthogonal group of an integral quadratic form is the group of integral matrices preserving it. A simple observation is that the integral orthogonal group of this Q has 8 elements. These observations are at least consistent with this theorem:

[intro-twosquares-1] **Theorem 1.2.** *Suppose*

$$n = p_1^{n_1} \dots p_m^{n_m}$$

is the prime factorization of n, and

$$Q(x,y) = x^2 + y^2 \,.$$

Then $N_{\mathbb{Z}}(Q,n) \neq 0$ if and only if n_i is even for every $p_i \equiv 3 \bmod 4$.

[intro-twosquares-2] **Theorem 1.3.** *Suppose the condition of the previous Theorem to be satisfied, and let*

$$n_1 = \prod_{p_i \equiv 1 \bmod 4} p_i^{n_i} \,.$$

Then

$$N_{\mathbb{Z}}(Q,n) = N_{\mathbb{Z}}(Q,n_1) = \ 4 \text{ times the number of divisors of } n_1 \,.$$

In particular if p is a prime then n is the sum of squares if and only if $p \equiv 1 \bmod 4$, and in that case $N_{\mathbb{Z}}(Q,p) = 8$, which means in essentially one way, since from one expression 7 others may be derived immediately.

The theory of integral quadratic forms was taken up to some extent by Fermat, Euler, Lagrange, and Legendre, but it had a new and thorough foundation laid in the book *Disquisitiones Arithmeticae* by Gauss. Subsequently in the nineteenth century this problem and related ones passed through the hands of Jacobi, Dirichlet, Eisenstein, and Minkowski, among others, but it was only in the mid nineteen-thirties that the most intriguing contribution, rooted in earlier discoveries by these nineteenth century men, was made by Carl Ludwig Siegel. Before I state his theorem, I need a preliminary result. For every prime power q, define

$$N_q(Q,n) = \text{ the number of solutions of } Q(x) \equiv n \bmod, q \ .$$

Continue to let d be the dimension of Q.

[intro-hensel1] **Theorem 1.4.** *For any prime p, the sequence of ratios*

$$\frac{N_{p^k}(Q,n)}{(p^k)^{(d-1)}}$$

becomes constant for $k \gg 0$.

Let it be $\alpha_p(Q,n)$. Let $N_{\mathbb{Z}}(Q)$ be the size of the integral orthogonal group of Q. Let $\alpha_{\mathbb{R}}(Q,n)$ be the limiting ratio of

$$\lim_{U \to \{n\}} \frac{\operatorname{meas} Q^{-1}(U)}{\operatorname{meas} U} \,.$$

[intro-real-factor] **Theorem 1.5.** *We have*

$$\alpha_{\mathbb{R}}(Q,n) = \frac{n^{d/2-1} d\, V_d(1)}{2\,|\det M_Q|}\,,$$

where $V_d(1)$ is the volume of the unit ball in $\mathbb{R}^d$.

[intro-siegel] **Theorem 1.6.** (Siegel's formula) *Suppose Q to be a positive definite integral quadratic form of dimension $d > 1$, and suppose the Q_i to be a complete set of representatives for the proper classes of quadratic forms in the same genus as Q. Then*

$$\frac{\displaystyle\sum_i \frac{N_{\mathbb{Z}}(Q_i,n)}{N_{\mathbb{Z}}(Q_i)}}{\displaystyle\sum_i \frac{1}{N_{\mathbb{Z}}(Q_i)}} = f\cdot\alpha_{\mathbb{R}}(Q,n)\cdot\prod_p \alpha_p(Q,n)$$

where $f = 1/2$ if $d = 2$, and 1 otherwise.

2. An example

To get the flavour of this formula, let's look at the case $Q = x^2 + y^2$ and n is a prime $q \equiv 1 \bmod 4$.

♣ [intro-twosquares-2] Theorem 1.3 tells us that $N_{\mathbb{Z}}(Q,q) = 8$, but let's see what Siegel's formula tells us.

[intro-sums-of-squares] **Theorem 2.1.** *The form $Q(x,y) = x^2 + y^2$ is the only one in its genus.*

(1) Therefore the left hand side of Siegel's formula is just $N_{\mathbb{Z}}(Q,q)$.

♣ [intro-real-factor] (2) According to Theorem 1.5 the real factor is π.

(3) So it remains to compute each $\alpha_p(Q,n)$. The tool required is Hensel's Lemma, which I'll formulate in two versions. First the simplest one:

[intro-hensel2] **Theorem 2.2.** (Hensel's Lemma, non-singular form) *Suppose $f(x)$ to be a function of d variables with coefficients in $\mathbb{Z}_p$, and suppose that x_0 is solution of $f \equiv 0$ modulo p such that*

$$\nabla f = \Big(\frac{\partial f}{\partial x_i}\Big)$$

evaluated at x_0 is a non-zero vector modulo p. If x_n is a solution of $f \equiv 0$ modulo p^n with $x_n \equiv x_0$ modulo p, then there exist exactly p^{d-1} solutions $x_{n+1} \equiv x_n$ modulo p^n.

The consequence is that $N_{p^k}(f,0)/(p^k)^{d-1}$ is the same for all $k \geq 1$.

This will handle the calculations of $\alpha_p(Q,q)$ for $p \neq 2, q$. In the field $\mathbb{Z}/p$, suppose $x^2 + y^2 = q$. The gradient is $[2x, 2y]$. Under the assumption that $p \neq 2, q$, this cannot vanish, so we may apply the Theorem, and have $\alpha_p(Q,q) = N_p(Q,q)/p$. So we ask, how many solutions of $x^2 + y^2 = q$ are there if $p \neq 2, q$?

There are now two cases:

(a) Suppose $p \equiv 1 \bmod 4$. In this case, -1 is a square in $\mathbb{Z}/p$, say $i^2 = -1$. Then $x^2+y^2 = (x-iy)(x+iy)$ and 2 is invertible modulo p, the number of solutions is the same as the number of solutions of $xy = q$, which is $p - 1$.

(b) Suppose $p \equiv 3 \bmod 4$. In this case, let F be the field obtained from $\mathbb{Z}/p$ by adjoining $i = \sqrt{-1}$. The norm map from F to $\mathbb{Z}/p$ takes $x + iy$ to $x^2 + y^2$. The norm map is surjective onto the units of $\mathbb{Z}/p$ and its kernel has order $p + 1$. If (x,y) is one solution of $x^2 + y^2 = q$, all others are obtained by multiplying $x + iy$ by an element of norm 1, so the number of solutions is $p + 1$.

For the two remaining cases, we need this version of Hensel's Lemma:

[intro-hensels2] **Theorem 2.3.** (Hensel's Lemma, crude singular form) *Suppose $f(x)$ to be a function of d variables with coefficients in $\mathbb{Z}_p$, and suppose that x_m is a solution of $f \equiv 0$ modulo p^m such that*

$$\nabla f = \Big(\frac{\partial f}{\partial x_i}\Big)$$

evaluated at x_m is a non-zero vector modulo p^{m-N}, with $m > 2N$. The ratio

$$\frac{N_{p^k}(f,0)}{(p^k)^{d-1}}$$

is constant for all $k \geq m$.

For $p = 2$ and $f(x,y) = x^2 + y^2 - q$ we may take $m = 3$, $N = 2$. Modulo 8 there are 16 solutions.

For $p = q$, the solutions modulo q breaks up into two types, those where $\nabla \neq (0,0)$ and those where $\nabla = (0,0)$. To the first we may apply the simple form of Hensel's Lemma. We are essentially solving $xy = 0$ modulo q, after a coordinate change. The solution set is the union of x- and y-axes, so there are $2(q-1)$ non-singular solutions. For the second, we have to see what happens modulo q^2. It turns out there are no new solutions. So $N_{q^k}(Q,q) = 2(q-1)\cdot q^{k-1}$ for all $k \geq 1$.

Summarizing:

$$\alpha_p = \begin{cases} 1 - 1/p & \text{if } p \equiv 1 \bmod 4,\ p \neq q \\ 1 + 1/p & \text{if } p \equiv 3 \bmod 4 \\ 2(1-1/p) & \text{if } p = q \\ 2 & \text{if } p = 2. \end{cases}$$

If I set

$$\chi(p) = \begin{cases} 1 & \text{if } p \equiv 1 \bmod 4 \\ -1 & \text{if } p \equiv 3 \bmod 4 \end{cases}$$

the right hand side of Siegel's formula is therefore

$$\begin{aligned} \frac{1}{2}\alpha_{\mathbb{R}}\cdot\alpha_2\cdot\alpha_q \prod_{p\equiv 1, p\neq q} \alpha_p &= \frac{1}{2}\cdot\pi\cdot 2\cdot 2\cdot \prod_{p\equiv 1,3}\Big(1 - \frac{\chi(p)}{p}\Big) \\ &= 8 \end{aligned}$$

since according to a formula of Leibniz and an observation of Euler

$$\prod_{p\equiv 1,3}\Big(1 - \frac{\chi(p)}{p}\Big) = \frac{1}{1 - 1/3 + 1/5 - 1/7 + \cdots} = \frac{4}{\pi}\,.$$

Siegel's formula relates to many different topics in mathematics. Cassells' book does not discuss it in the main text, but in an appendix on analytic methods. Just understanding what it says, without even contemplating its proof, will require some elementary calculus, a discussion of explicit values of Dirichlet L-functions, some elementary algebraic geometry, something from the reduction theory of positive definite quadratic forms, and also an examination of quadratic forms over finite fields.

3. History

Gauss, Jacobi, Dirichlet, Eisenstein, Minkowski, Weil, Tamagawa, Langlands, Kottwitz.

Binary forms, the upper half plane, SL_n, modular forms.

4. References

1. Emil Artin, **Geometric algebra**, Wiley, 1957.

<p> Sizes of finite spheres.

2. Z. I. Borevitch and I. R. Shafarevitch, **Number theory**, Academic Press, 1967.

Binary forms and quadratic extensions of $\mathbb{Q}$, Bernoulli numbers and values of L-functions.

3. J. W. S. Cassels, **Rational quadratic forms**, Academic Press, 1978. Reprinted by Dover in 2008.

Chapter 4 is an introduction to integral quadratic forms in general, and sums of three squares in particular. Appendix B is a good summary of Siegel's formula and related methods.

4. Alex Eskin, Zeev Rudnik, and Peter Sarnak, 'A proof of Siegel's weight formula', *International Mathematics Research Notices* (1991), 65–69.

5. Carl Friedrich Gauss, **Disquisitiones arithmeticae**, Yale University, 1966. Republished later by Springer.

This is the only English translation of Gauss' original Latin. §§266–292 deal with ternary quadratic integral forms, and give Gauss' formula for $r_3(n)$. Much of this is explained more clearly in Grosswald's book, but everyone should try to read Gauss sometime.

6. Emil Grosswald, **Representations of numbers as sums of squares**, Springer, 1985.

Chapter 3 is about sums of three squares.

7. Robert Gunning, **Lectures on modular forms** Princeton University Press, 1962.

Perhaps the simplest introduction to the relationship between theta functions and integral quadratic forms.

8. R. E. Kottwitz, 'Tamagawa numbers', *Annals of Mathematics* **127** (1988), 629–626.

9. King Fai Lai, *Tamagawa numbers of reductive algebraic groups*, *Compositio Mathematicae* **41** (1980), 153-188.

10. R. P. Langlands, 'The volume of the fundamental domain for some arithmetical subgroups of Chevalley groups', in **Algebraic groups and discontinuous subgroups**, the proceedings of the Boulder conference, *Symposia in Pure Mathematics* **9**, 1966. Edited by A. Borel and D. Mostow.

11. Hermann Minkowski, 'Grundlagen für eine Theorie quadratischen Formen mit ganzahligen Koefficienten', in **Gesammelte Abhandlungen**, 3–145.

This is Minkowski's prize essay, written when he was 18 years old.

12. Hermann Minkowski, 'Diskontinuitätsbereich für arithmetische Äquivalenz', *Crelles Journal für die reine und angewandte Mathematik* **129** (1905), 220–274.

This was Minkowski's last paper in number theory. It describes in some detail the action of $\mathrm{SL}_n(\mathbb{Z})$ on the space of positive definite quadratic forms.

13. Rudolf Scharlau, 'Martin Kneser's work on quadratic forms and algebraic groups'.

This is a set of slides presented at a conference in Chile. It can be found at `http://inst-mat.utalca.cl/qfc2007/Talks/scha`
It includes a brief history of the theory of integral quadratic forms. One thing it mentions is that Martin Kneser seems to have been the first to point out the connection between Siegel's formula and adèle groups (in 1956). Weil's Bourbaki talk already mentions this, but more vaguely. It is hard now to tell exactly what Kneser had in mind.

14. Rudolf Scharlau, 'Kneser's work on quadratic forms and algebraic groups', in the **Quadratic Forms–Algebra, Arithmetic, and Geometry**, edited by Ricardo Baeza et al., published by the A. M. S.

This is the final published text of his talk listed above.

15. Carl Ludwig Siegel, 'Über die analytischen Theorie der quadratischer Formen' I, II, III, in the *Annals of Mathematics* 1935–1937. Also in Siegel's **Gesammelte Abhandlungen**.

16. Carl Ludwig Siegel, 'The volume of the fundamental domain for some infinite groups', *Transactions of the American Mathematical Society* **39** (1936), 209–218.

He finds a simple proof of the volume calculation of $\mathrm{SL}(n)\backslash X_n$, the subject of the last paper of Minkowski, and generalizes it.

17. Carl Ludwig Siegel, **Lectures on the analytical theory of quadratic forms** (notes in English by Morgan Ward), Peppmüller, 1963.

This is the main reference in English for Siegel's original work on quadratic forms.

18. H. J. Smith, **Collected mathematical papers**, Oxford Press, 1894.

Available online from a link on Wikipedia's entry on 'Henry John Stephen Smith'. His long 'Report on the theory of numbers' is easy reading.

19. Tsuneo Tamagawa, 'Adèles', in **Algebraic groups and discontinuous subgroups**, the proceedings of the Boulder conference, *Symposia in Pure Mathematics* **9**, 1966. Edited by A. Borel and D. Mostow.

Maybe the only place where the relationship between Siegel's formula and the volume of adèle quotients is explained in relatively elementary terms. The entire proceedings are available for free download from the AMS web page:

`http://www.ams.org/bookstore-getitem/item=PSPUM-9-E`

20. V. E. Voskresenskii, 'Adèle groups and Siegel-Tamagawa formulas', *Journal of Mathematical Sciences* **73** (1995), 47–67.

This is a general survey of Tamagawa numbers, with one group of examples being representation by sums of squares. There is even a long, more or less self-contained, account of Langlands' computation of Tamagawa numbers in terms of Eisenstein series, a very ambitious effort. Overall, it's fairly good, although local analysis is not done well, the English is shaky, and there are a few fine points about integral quadratic forms he doesn't get quite right. The bibliography is thorough, although recent work of Kottwitz has been missed.

21. André Weil, 'Adèles et groupes algébriques', *Séminaire Bourbaki*, exposé 186, 1959.

The first published account relating Siegel's formula to adèle groups.

22. André Weil, 'Sur la théorie des forme quadratiques', 9–22 in **Colloque sur la théorie des groupes algébriques**, C. B. R. M., Bruxelles, 1962.

23. André Weil, **Adèles and algebraic groups**, Birkhäuser, 1982.

Originally a set of mimeographed notes from the Institute in Princeton from around 1965. Computes adelic volume formulas for several algebraic groups.