Quadratic forms over finite fields

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

Suppose V to be a module over a ring R. A **quadratic form** on V is a function Q with values in R such that (1) $Q(cv) = c^2 Q(v)$ for all scalars c; (2) the function

$$\nabla = \nabla_Q\colon\ (u, v) \longmapsto Q(u + v) - Q(u) - Q(v)$$

is bilinear (and of course symmetric). If a coordinate system has been chosen, it will be expressed as

$$Q(x) = \sum\nolimits_{i \le j} a_{i,j} x_i x_j \quad (a_{i,j} \in R)\,,$$

and then

$$\nabla(x, y) = \sum\nolimits_i 2a_{i,i} x_i y_i + \sum\nolimits_{i<j} a_{i,j}(x_i x_j + x_j y_i) = {}^t x\, M\, y\,,$$

with

$$M_{i,j} = \begin{cases} 2a_{i,i} & \text{if } i = j \\ a_{i,j} & \text{if } i < j \\ a_{j,i} & \text{otherwise.} \end{cases}$$

The matrix M is also that of the linear transformation $V \to \widehat{V}$ induced by ∇.

If R is a field, I call the quadratic space (V, Q) **strictly nondegenerate** if ∇_Q induces an isomorphism of V with $\widehat{V}$. I want to classify all strictly nondegenerate quadratic forms over finite fields, and to derive some of their properties—for example, the size of spheres.

The principal motivation for this topic is a matter of number theory. Suppose Q to be a positive definite quadratic form of dimension d over $\mathbb{Z}$ that is strictly nondegenerate over $\mathbb{Q}$. If $L = \mathbb{Z}^d$, then Q takes integral values on L. Let

$$L^\perp = \{v \in \mathbb{Q}^d \mid \nabla(v, L) \subseteq \mathbb{Z}\}\,.$$

Thus $L^\perp = M^{-1}(L)$. It is a lattice in which L has index equal to $\det(M)$, called the **discriminant** of Q. It can be understood better by applying the principal divisor theorem to $L \subseteq L^\perp$. For each prime p, the integral form Q induces a quadratic form on $(\mathbb{Z}/p)^d$, whch will be strictly nondegenerate if and only if p does not divide $\det(M)$. A classic formula due to Siegel relates the number of x in $\mathbb{Z}^d$ with $Q(x) = n$ to a product $\prod_p \mu_p$ over all p, in which

$$\mu_p = \frac{\left|\left\{v \in (\mathbb{Z}/p)^d \,\middle|\, Q(v) \equiv n\right\}\right|}{p^{d-1}}$$

for p not dividing n or the discriminant. As a consequence of later computations (together with quadratic reciprocity), this will be seen to be up to a finite number of factors an Euler product associated to a certain Dirichlet L function.

There are several possible approaches to this topic. One standard reference is [Dickson:1902]. But by and large, I follow Minkowski's exposition in his prize essay of 1883 (available in [Minkowski:1911]), which applies the finite Fourier transform. This turns out to be a good model for dealing with quadratic forms over local fields.

Contents

1. Introduction

For the moment, let F be an arbitrary field.

1.1. Proposition. *Suppose (V, Q) to be a strictly nondegenerate quadratic space over F. Then:*
(a) if the characteristic of F is odd, (V, Q) is a direct sum of one dimensional quadratic spaces;
(b) if the characteristic of F is two, (V, Q) is the direct sum of two-dimensional quadratic spaces.

Of course each constituent will also be strictly non-degenerate.

Proof. The first is straightforward.

As for the second, suppose the characteristic of F to be two, and argue by induction on dimension, starting at dimension 0. If $x \neq 0$, then because ∇ is non-degenerate, there must exist y such that $\nabla(x, y) = 1$. It cannot be a multiple of x, since $\nabla(x, x) = 2\,Q(x) = 0$. Therefore x and y span a plane H. (This excludes already the possibility that V have dimension one.) If $v = ax + by$, then

$$\nabla(x, v) = b, \quad \nabla(y, v) = a$$

so that the restriction of Q to H is strictly non-degenerate. Therefore $H \cap H^\perp = 0$, so V is the orthognal sum $H \oplus H^\perp$. The restriction of Q to $H^\perp$ is also strictly nondegenerate, so we may apply induction.

The meaning of this for odd characteristic is obvious. But in even characteristic, we must ask, *what are the possible strictly nondegenerate quadratic spaces of dimension 2?* I'll answer this question more generally.

BINARY FORMS. Continue to let F be an arbitrary field.

There are two natural ways to construct strictly nondegenerate quadratic forms of dimension two.

• *The hyperbolic plane H.* Let a, b be a basis (a, b) of F^2, and set

$$Q(xa + yb) = xy\,.$$

• *Quadratic norms.* Suppose K/F to be a separable quadratic extension with conjugation $x \mapsto \overline{x}$. For c in $F^\times$, define

$$Q(x) = c\,N_{K/F}(x) = c\,(x\,\overline{x})\,.$$

1.2. Proposition. *Any strictly nondegenerate quadratic form of dimension two is one of these two types. In the second case, two scalars determine isomorphic forms if and only if their ratio lies in the image of $N_{K/F}$.*

Proof. Suppose (V, Q) to be any strictly nondegenerate quadratic space of dimension two.

(a) If there exists u such that $Q(u) = 0$ (it is said to be **isotropic**), let v be such that $\nabla(u, v) = 1$. For c in F

$$Q(v + cu) = Q(v) + c\,\nabla(u, v)$$

so that there exists a vector $w = v + cu$ linearly independent of u such that $Q(w) = 0$. The pair u, w span a hyperbolic plane.

(b) We may now assume that there are no isotropic vectors (the form is said to be **anisotropic**). Choose a basis (a, b), and suppose that

$$Q(xa + yb) = Ax^2 + Bxy + Cy^2\,.$$

If $A = 0$ there exist isotropic vectors. So we may write

$$Q(xa + yb) = A(x^2 + (B/A))xy + (C/A)y^2) = A(x^2 + Bx(y/A) + (AC)(y/A)^2)\,.$$

That is to say, we may change coordinates to make our form look like

$$A(x^2 + Bxy + Cy^2)\,.$$

Here B cannot be 0, because of nondegeneracy. The expression in parentheses therefore factors as

$$(x - \alpha)(x - \beta)$$

over a separable closure of F. If either α or β lies in F, then we are looking at the hyperbolic plane, otherwise at the norm form of a separable quadratic extension of F.

This discussion is capped by:

1.3. Lemma. *If F is a finite field, there is up to isomorphism exactly one separable quadratic extension.*

Proof. The cases in which the characteristic of F is two has to be treated differently from the rest. In that case, any separable quadratic extension is generated by the roots of an **Artin-Schreier** polynomial

$$x^2 + x + \gamma = 0$$

The extension depends only on γ modulo the image of the $\mathbb{F}_2$-linear map $\mathfrak{P}\colon x \mapsto x^2 + x$. This image has index 2 in F, since its kernel is $\mathbb{F}_2$. The case $\gamma = 0$ gives rise to $F \oplus F$, and the non-trivial coset to the unique quadratic extension.

In odd characteristic, every quadratic extension is by a square root. A similar discussion takes place with $F^\times/(F^\times)^2$ replacing $F/\mathfrak{P}(F)$.

From now on, let N be the norm on this unique extension.

1.4. Proposition. *If F is a finite field, there are up to isomorphism exactly two binary quadratic forms, H and N.*

Proof. This amounts to showing that all forms aN are isomorphic, which follows from:

1.5. Lemma. *If K/F is a finite extension of the finite field F, both the norm and trace maps are surjective.*

Proof. Say $F = \mathbb{F}_q$, with $K = \mathbb{F}_{q^n}$. The norm map takes x to $x^{q^{n-1}+\cdots+1}$. The following sequence is exact:

$$1 \longrightarrow \mathrm{KER}(N) \longrightarrow K^\times \longrightarrow \mathrm{IM}(N) \longrightarrow 1\,.$$

The size of the kernel is at most $(q^n - 1)/(q - 1)$, since a polynomial of degree r can have at most r roots. The size of the image is at most $q - 1$. Therefore

$$|\mathrm{KER}| \cdot |\mathrm{IM}| \leq q^n - 1\,.$$

But because the sequence is exact, the product is $q^n - 1$. Therefore the image has size $q - 1$.

A similar argument works for the trace.

1.6. Theorem. *Every strictly nondegenerate quadratic space of a finite field is isomorphic to exactly one of the following:*

(a) in dimension $2n + 1$ and odd characteristic, $nH \oplus cx^2$;
(b) in dimension $2n$ and any characteristic, nH or $(n - 1)H + N$.

Proof. In view of earlier results, in order to show that every quadratic space is isomorphic to one of these, it suffices to show that $N \oplus N$ and $N \oplus cx^2$ contains an isotropic vector. I leave this as an exercise.

Uniqueness in odd characteristic follows from an easy calculation of discriminants. In characteristic two, there is an analogue of the discriminant called the Arf invariant that will do the job, but I haven't introduced it. Instead, I'll deal with the question later, by means of the Fourier transform.

The following is basic in analyzing the structure of the orthogonal group of a quadratic space.

1.7. Theorem. (Witt's Theorem) *Suppose given two subspaces U_1, U_2 and an isometry σ of the restrictions of Q to each. Then there exists an isometry on all of V extending σ.*

In odd characteristic this is straightforward, but not in characteristic two. I'll not prove it here, and instead refer to [Elman et al.:2008].

2. The Fourier transform on finite fields

Suppose $F = \mathbb{F}_q$, with $q = p^n$. This is a Galois extension of $\mathbb{F}_p$, which may be identified with $\mathbb{Z}/p$. The Galois group is cyclic with generator $\mathfrak{F}\colon x \mapsto x^p$. Let τ be the trace map from F to $\mathbb{F}_p$, which is surjective, and define

$$\psi\colon\ x \longmapsto e^{2\pi i \tau(x)/p}\,,$$

which is a character of the additive group of F. For every y in F the function

$$\psi_y\colon x \longmapsto \psi(xy)$$

is also a character, and if ψ_y is the trivial character then $y = 0$. As a consequence, since a non-trivial character sums to 0:

2.1. Lemma. *If y lies in F then*

$$\sum_x \psi(xy) = \begin{cases} q & \text{if } y = 0 \\ 0 & \text{otherwise.} \end{cases}$$

We can define a kind of Fourier transform from $\mathbb{C}(F)$ to itself by the formula

$$\widehat{f}(y) = \frac{1}{\sqrt{q}} \sum_{x \in F} f(x)\psi(-xy)\,.$$

2.2. Proposition. *The map taking f to $\widehat{f}$ is an isomorphism of $\mathbb{C}(F)$ with itself. The inverse map takes φ to*

$$f(x) = \frac{1}{\sqrt{q}} \sum_{y \in F} \varphi(y)\psi(xy)\,.$$

One way to phrase this is to say that the Fourier transform applied twice is:

$$\widehat{\widehat{f}}(x) = f(-x)\,.$$

Proof. An application of Lemma 2.1. ∎

2.3. Lemma. (Plancherel formula) *For f in $\mathbb{C}(F)$*

$$\sum_{x \in F} \left|f(x)\right|^2 = \sum_{y \in F} \left|\widehat{f}(y)\right|^2\,.$$

Proof. Also an application of Lemma 2.1. ∎

The whole point of the $\sqrt{q}$ factor in the definition of the Fourier transform is to make it a unitary transformation.

If χ is a multiplicative character of $F^\times$, extend it to all of F by setting $\chi(0) = 0$. Define the corresponding **Gauss function** to be its Fourier transform:

$$G_{\psi,\chi}(y) = \frac{1}{\sqrt{q}} \sum_x \chi(x)\psi(-xy)\,.$$

We have

$$G_{\psi,\chi}(0) = \begin{cases} \sqrt{q} & \text{if } \chi = 1 \\ 0 & \text{otherwise.} \end{cases}$$

While if $y \neq 0$

$$\begin{aligned} G_{\psi,\chi}(y) &= \frac{1}{\sqrt{q}} \sum_x \chi(x)\psi(-xy) \\ &= \frac{\chi^{-1}(y)}{\sqrt{q}} \sum_x \chi(x)\psi(-x) \\ &= \chi^{-1}(y) G_{\psi,\chi}(1)\,. \end{aligned}$$

In other words, the Fourier transform of χ is, up to a constant, the character χ^{-1}.

Let

$$\mathfrak{G}_\chi = \sqrt{q}\, G_{\psi,\chi}(-1) = \sum_x \chi(x)\psi(-x)\,.$$

The Plancherel formula implies that $|\mathfrak{G}_\chi| = \sqrt{q}$ if χ is not the trivial character. The Fourier transform applied twice gives us $\chi(-x)$, which tells us that

$$\mathfrak{G}_\chi \mathfrak{G}_{\chi^{-1}} = \chi(-1)\, q\,.$$

Remark. If q is odd and $\chi = \mathrm{sgn}$, the unique non-trivial character of order two, then

$$\mathfrak{G}_{\mathrm{sgn}}^2 = \mathrm{sgn}(-1)\, q\,.$$

There is an extremely interesting story to be told about which square root occurs, but I'll not tell it here.

Remark. The formula

$$\mathfrak{G} = \sum_x \chi(x)\psi(-x)\,.$$

is a sum of terms of complex magnitude 1. To say that it has magnitude $\sqrt{q}$ is to say that the summands behave roughly like q steps of unit length in a random walk on the complex plane. The point is that the additive and multiplicative structures of F interact more or less independently.

3. Quadratic forms and the Fourier transform

Let (V, Q) be a non-degenerate quadratic space of dimension d over $F = \mathbb{F}_q$. The method I am now going to introduce is due to Hermann Minkowski (who found it when he was about 17 years old).

For x in F, let

$$\nu_Q(x) = \text{ size of the 'sphere' } \{v \in V \mid Q(v) = x\},$$

and let $\gamma_Q(y)$ be a slight variation of its Fourier transform:

$$\gamma_Q(y) = \sum_{x \in F} \nu_Q(x)\psi(-xy) = \sum_{v \in V} \psi\big(-Q(v)y\big).$$

One value can be calculated immediately:

$$\gamma_Q(0) = q^d.$$

There are two points to working with γ_Q. • The function ν_Q can be recovered from it by the inverse Fourier transform:

$$\nu_Q(x) = \left(\frac{1}{q}\right) \sum_{y \in F} \gamma_Q(y)\psi(xy).$$

• It can be easily calculated, since

3.1. Proposition. *If $(V, Q) = (V_1, Q_1) \oplus (V_2, Q_2)$ then $\gamma_Q = \gamma_{Q_1} \cdot \gamma_{Q_2}$.*

Proof. A straightforward calculation. ∎

The way things will go should be easy to predict. When V has dimension one or two, we can calculate ν_Q easily, and then get γ_Q from it. For dimensions three or more, we'll go the other way, from γ_Q to ν_Q.

I'll now summarize results. Proofs will come a bit later. In odd characteristic, define the function on F:

$$\operatorname{sgn}(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } x \neq 0 \text{ is a square} \\ -1 & \text{otherwise.} \end{cases}$$

(1) $d = 1$, $Q(x) = cx^2$, q odd.

$$\nu_Q(x) = \begin{cases} 1 & \text{if } x = 0 \\ 1 + \operatorname{sgn}(x/c) & \text{otherwise.} \end{cases}$$

(2) $d = 2$, $Q = H$.

$$\nu_Q(x) = \begin{cases} 2q - 1 & \text{if } x = 0 \\ q - 1 & \text{otherwise.} \end{cases}$$

(3) $d = 2$, $Q = N$.

$$\nu_Q(x) = \begin{cases} 1 & \text{if } x = 0 \\ q + 1 & \text{otherwise} \end{cases}$$

(4) $d = 2n$, $Q = nH$,

$$\nu_Q(x) = \begin{cases} q^{2n-1} + q^n - q^{n-1} & \text{if } x = 0 \\ q^{2n-1} - q^{n-1} & \text{otherwise.} \end{cases}$$

(5) $d = 2n$, $Q = (n-1)H + N$

$$\nu_Q(x) = \begin{cases} q^{2n-1} - q^n + q^{n-1} & \text{if } x = 0 \\ q^{2n-1} + q^{n-1} & \text{otherwise.} \end{cases}$$

(6) $d = 2n+1$, $Q = nH + cx^2$, q odd:

$$\nu_Q(x) = \begin{cases} q^{2n} & \text{if } x = 0 \\ q^{2n} + q^n \mathrm{sgn}(-x/c) & \text{otherwise.} \end{cases}$$

Now for details.

(1) $d = 1$, $Q(x) = cx^2$, q odd. The formula for ν is immediate. As for γ_Q, if $y \neq 0$ then

$$\begin{aligned} \gamma_Q(y) &= \sum_x \nu(x)\psi(-xy) \\ &= \sum_x \big(1 + \mathrm{sgn}(x/c)\big)\psi(-xy) \\ &= \sum_x \mathrm{sgn}(x/c)\psi(-xy) \\ &= \sqrt{q}\, G_{\psi,\mathrm{sgn}}(cy) \\ &= \mathrm{sgn}(cy)\, \mathfrak{G}\,. \end{aligned}$$

(2) $d = 2$, $Q = H$. We can calculate ν_Q directly. For $y \neq 0$

$$\gamma_Q(y) = (2q+1) + (q-1)\sum_{x\neq 0} \psi(-xy) = (2q-1) - (q-1) = q\,.$$

(3) $d = 2$, $Q = N$. We can calculate ν_Q directly. For $y \neq 0$

$$\gamma_Q(y) = 1 + (q+1)\sum_{x\neq 0} \psi(-xy) = 1 - (q+1)\big) = -q\,.$$

Remark. I can summarize briefly the results so far by saying that $\gamma_Q(0) = q^d$ and that if $y \neq 0$ then $|\gamma_Q(y)| = q^{d/2}$ for $d = 1$ or 2. But then by Proposition 3.1 this holds for all dimensions. In all cases

$$\nu_Q(0) = \frac{1}{q}\Big(q^d + \sum_{y\neq 0} \gamma_Q(y)\psi(xy)\Big), \quad \big|\nu_Q(0) - q^{d-1}\big| \leq (1-1/q)q^{d/2}\,.$$

This implies that $\nu_Q(0) > q^{d-1} - (1-1/q)q^{d/2}$ for $d \geq 3$. Since this is greater than 1 for all $q \geq 2$, we have a second proof that a quadratic space of dimension 3 or more over a finite field always has isotropic vectors.

(4) $d = 2n$, $Q = nH$. We reverse the procedure we have used so far—we now use the rule for calculating γ_Q when Q is an orthogonal sum, and deduce ν_Q from γ_Q.

$$\gamma_Q(y) = \begin{cases} q^{2n} & \text{if } y = 0 \\ q^n & \text{otherwise.} \end{cases}$$

Hence

$$\begin{aligned} \nu(x) &= \frac{1}{q}\sum_y \gamma(y)\psi(xy) \\ &= \frac{1}{q}\Big(q^{2n} + \sum_{y\neq 0} q^n \psi(xy)\Big) \\ &= q^{2n-1} + q^{n-1}\sum_{y\neq 0} \psi(xy) \end{aligned}$$

(5) $d = 2n$, $Q = (n-1)H + N$

$$\gamma_Q(x) = \begin{cases} q^{2n} & \text{if } x = 0 \\ -q^n & \text{otherwise.} \end{cases}$$

Here

$$\begin{aligned} \nu(x) &= \frac{1}{q} \sum_y \gamma(y)\psi(xy) \\ &= \frac{1}{q} \Big(q^{2n} - \sum_{y \neq 0} q^n \psi(xy) \Big) \\ &= q^{2n-1} - q^{n-1} \sum_{y \neq 0} \psi(xy) \Big) \end{aligned}$$

(6) $d = 2n+1$, $Q = nH + ax^2$, q odd.

$$\gamma_Q(y) = \begin{cases} q^{2n+1} & \text{if } x = 0 \\ q^n \cdot \operatorname{sgn}(-ay)\, \mathfrak{G} & \text{otherwise.} \end{cases}$$

$$\nu_Q(x) = \frac{1}{q} \Big(q^{2n+1} + q^n \sum_{y \neq 0} \operatorname{sgn}(ay) \mathfrak{G}\, \psi(xy) \Big)$$

4. Concluding remarks

The functions ν_Q and γ_Q can be defined for all local fields, such as $\mathbb{Q}_p$ and $\mathbb{R}$. In the case of $\mathbb{R}$ we get something related to the Fresnel integrals

$$\int_{\mathbb{R}} e^{\pi i x^2}\, dx\,.$$

and for p-adic fields they allow a very neat classification of non-degenerate quadratic forms. The functions γ_Q alo play an important role in the representations of $\mathrm{SL}_2(F)$ first defined by André Weil.

Minkowski's method was also used by him to classify quadratic forms over the rings $\mathbb{Z}/p^n$. It works well for odd p, and may be used even more neatly for the classification over the p-adic integers. That story will be told elsewhere.

5. References

1. Cahit Arf, 'Untersuchungen der quadratischen Formen in Körpen de Charakteristik 2', *Journal für die reine and angewandte Mathematik* **183** (1941), 148–167.

2. Leonard Eugene Dickson, **Linear groups**, Teubnaer, 1902.

3. Richard Elman, Nikita Karpenko, and Alexander Merkurjev, **The algebraic and geometric theory of quadratic forms**, A. M. S., 2008.

4. Hermann Minkowski, 'Grundlagen für eine Theorie quadratischen Formen mit ganzahligen Koeffizienten', in **Gesammelte Abhandlungen**, 3–145. Originally published in 1911, now available in the Chelsea series, published by the American Mathematical Society.

5. André Weil,'Sur certains groupes d'opérateurs unitaires', *Acta Mathematica* **111** (1964), 143–211.