11:20 a.m. February 5, 2011

Quadratic forms over finite fields

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

The aim of this essay is to classify non-degenerate quadratic forms over finite fields in arbitrary characteristic, and to derive a few properties, such as the sizes of spheres.

I begin with a basic discussion of bilinear forms and quadratic forms. The distinction between the two is particularly important for us, since I am not going to assume 2 is invertible. Then I go on to prove versions of theorems about quadratic forms that are independent of characteristic, and finish up with quadratic forms over finite fields.

For analyzing forms over arbitrary fields, I follow [Elman et al.:2008]. The theory in odd characteristic is well known, but the source of most of the theory in characteristic 2 seems to be [Arf:1941]. For sizing spheres, I follow Minkowski's exposition in his prize essay of 1883 (available in [Minkowski:1911]).

Contents

1. Bilinear and quadratic forms

A **bilinear form** of dimension d over a field F is a function ∇ on $F^d \times F^d$ separately linear in each factor. It is **symmetric** if $\nabla(x,y) = \nabla(y,x)$. Given a coordinate system, a symmetric bilinear form has an expression

$$\nabla(x,y) = \sum\nolimits_{i,j} a_{i,j} x_i y_j$$

with $a_{i,j} = a_{j,i}$, all coefficients in F, and if M_∇ is the matrix $(a_{i,j})$ then

$$\nabla(x,y) = {}^t x\, M_\nabla\, y\,.$$

A bilinear form on a vector space V determines a map from V to its dual $\widehat{V}$, and the matrix of B is the matrix of that linear transformation. The bilinear form is said to be **non-degenerate** if this transformation—or, equivalently, its matrix—is invertible. I'll write the map from V to $\widehat{V}$ as $v \mapsto \nabla_v$, so

$$\langle \nabla_v, u\rangle = \nabla(v,u)\,.$$

Any map from $f\colon V \to \widehat{V}$ determines a transpose map $\widehat{f}$ of duals, and ∇ is symmetric if and only if $\widehat{f} = f$.

A **quadratic form** of dimension d is a function Q defined on some F^d by an expression

$$Q(x) = \sum\nolimits_{i\le j} a_{i,j}x_ix_j\,.$$

There is a close relationship between the two notions, one that can lead to some confusion. First of all, every bilinear form ∇ gives rise to a quadratic form

$$Q_\nabla(x) = \nabla(x,x)\,.$$

If the matrix of ∇ is $(a_{i,j})$ the formula for Q_∇ is

$$Q_\nabla(x) = \sum_i a_{i,i}x_i^2 + \sum_{i<j} 2a_{i,j}x_ix_j\,.$$

As you can see, the quadratic forms that arise in this way are special—the coefficients of the cross terms are always even.

On the other hand, every quadratic form Q determines a bilinear form

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

If $Q(x) = \sum_{i\le j} a_{i,j}x_ix_j$ its formula is

$$\begin{aligned}\nabla_Q(x,y) &= \sum_{i\le j} a_{i,j}\big((x_i+y_i)(x_j+y_j) - x_ix_j - y_iy_j\big)\\ &= \sum_{i\le j} a_{i,j}(x_iy_j + x_jy_i)\\ &= \sum_i 2a_{i,i}x_iy_i + \sum_{i<j} a_{i,j}x_iy_j + \sum_{i>j} a_{j,i}x_iy_j\,.\end{aligned}$$

Again, only certain bilinear forms arise in this way from quadratic forms.

It is possible to define a quadratic form independently of a choice of coordinates as a function $Q(x)$ such that (a) $Q(cx) = c^2Q(x)$ for c in F and (b) the function $\nabla_Q(x,y) = Q(x+y) - Q(x) - Q(y)$ is bilinear.

If we start with a bilinear form ∇, construct $Q = Q_\nabla$, then go on to construct ∇_Q, we get 2∇. In a diagram, the composite map

$$\text{symmetric bilinear forms} \overset{\nabla\mapsto Q_\nabla}{\longrightarrow} \text{quadratic forms} \overset{Q\mapsto\nabla_Q}{\longrightarrow} \text{symmetric bilinear forms}$$

amounts to multiplication by two. Hence if 2 is invertible, the form Q is always defined in terms of a bilinear form, namely $(1/2)\nabla_Q(x,y)$, since

$$\nabla_Q(x,x) = Q(2x) - 2Q(x) = 2Q(x), \quad Q(x) = (1/2)\nabla_Q(x,x)\,.$$

All these distinctions are unimportant if the characteristic of F is odd, but if it is two they are crucial.

The bilinear form associated to a quadratic form is what is called in calculus its gradient or derivative, since

$$Q(x+y) = Q(x) + \nabla_Q(x,y) + Q(y)\,.$$

Thus if $F = \mathbb{R}$

$$\lim_{t\to 0}\left[\frac{Q(x+ty) - Q(x)}{t}\right] = \nabla_Q(x,y)\,.$$

In the literature there is some confusion about exactly what qualifies as a quadratic form. During much if not all of the nineteenth century, starting with Gauss and running through Minkowski, integral quadratic forms were taken to be only the ones defined in terms of a bilinear form, hence with a factor of 2 in all coefficients of cross terms $x_i x_j$. This is often the case even in modern times, for example in the book [Cassells:1978]. It is not clear to me why this tradition has persisted in number theory. For example, excluding the integral quadratic form $x^2 + xy + y^2$, which is the norm form on the ring of algebraic integers in $\mathbb{Q}(\sqrt{-3})$, seems rather eccentric. But nowadays there are many applications in which it is important to work with symmetric bilinear forms, for example in considering the intersection of cycles in the middle dimension on a manifold. Integral bilinear forms share much of the life of integral quadratic forms, but have a path of their own.

Anyway, this essay will be about quadratic forms—I shall not in general assume the cross-term coefficients to be even, although doing so will play a role elsewhere in the process of interpreting Minkowski in modern terms.

2. Non-degenerate quadratic forms

The **radical** of a bilinear form ∇ is the subspace

$$\mathfrak{rad}_\nabla = \big\{ v \in V \,\big|\, \nabla(v, V) = 0 \big\}$$

and the radical of the quadratic space (V, Q) is

$$\mathfrak{rad}_Q = \big\{ v \in \mathfrak{rad}_\nabla \,\big|\, Q(v) = 0 \big\}.$$

Thus $\mathfrak{rad}_Q \subseteq \mathfrak{rad}_{\nabla_Q}$.

For example, in odd characteristic both radicals of the one-dimensional quadratic form x^2 are trivial, while if the characteristic is two $\mathfrak{rad}_\nabla = F$ but $\mathfrak{rad}_Q = 0$. In odd characteristic both radicals of the two-dimensional form $x^2 + y^2$ are trivial, but in even characteristic $\mathfrak{rad}_\nabla$ is the whole space and $\mathfrak{rad}_Q$ is the line $x + y = 0$, since $x^2 + y^2 = (x + y)^2$.

The bilinear form ∇ determines a well defined bilinear form $\overline{\nabla}$ on $V/\mathfrak{rad}_\nabla$, since if u, v lie in V and x, y in $\mathfrak{rad}_\nabla$ then $\nabla(u + x, v + y) = \nabla(u, v)$. The bilinear form $\overline{\nabla}$ is non-degenerate.

[rad-Q] Proposition 2.1. *If P_∇ is the canonical projection from V to $V/\mathfrak{rad}_\nabla$ then $\nabla(u, v) = \overline{\nabla}\big(P_\nabla(u), P_\nabla(v)\big)$.*

The quadratic form Q determines a quadratic form $\overline{Q}$ on $V/\mathfrak{rad}_Q$ since if v lies in V and x in $\mathfrak{rad}_Q$ then $Q(v + x) = Q(v)$.

[rad-Q] Proposition 2.2. *If P_Q is the projection from V to $V/\mathfrak{rad}_Q$ then $Q(v) = \overline{Q}\big(P_Q(v)\big)$.*

Following [Elman et al.:2008] loosely, I'll call (V, Q) **weakly non-degenerate** if $\mathfrak{rad}_Q = 0$; **non-degenerate** if $\mathfrak{rad}_Q = 0$ and the dimension of $\mathfrak{rad}_{\nabla_Q}$ is at most one; and **strictly non-degenerate** if $\mathfrak{rad}_{\nabla_Q} = 0$.

[disc-odd] Proposition 2.3. *If F has odd characteristic, then $\mathfrak{rad}_Q = \mathfrak{rad}_\nabla$.*

Proof. This is immediate. □

For the moment suppose that F has characteristic two. Let (V, Q) be a weakly non-degenerate quadratic space, U a linear complement to $\mathfrak{rad}_\nabla$. If u lies in U and x in $\mathfrak{rad}_\nabla$, then $Q(u + x) = Q(v) + Q(x)$, so Q is completely determined by its restrictions to U and $\mathfrak{rad}_\nabla$. Its restriction to U is strictly non-degenerate. As for its restriction to $\mathfrak{rad}_\nabla$, the following is just a matter of definition:

[char2-degen] **Lemma 2.4.** *If F has characteristic 2 and (V, Q) is a weakly non-degenerate quadratic space of dimension m with $V = \mathfrak{rad}_\nabla$, then in any coordinate system*

$$Q(v) = \sum_{i=1}^{m} c_i x_i^2$$

with all $c_i \neq 0$.

[disc-even] **Proposition 2.5.** *Assume F to be a perfect field of characteristic two. Every weakly non-degenerate quadratic space is non-degenerate.*

Proof. I recall that a perfect field of characteristic 2 is one for which $x \mapsto x^2$ is an automorphism. In particular, all finite fields $\mathbb{F}_{2^n}$ are perfect.

Suppose (V, Q) to be a weakly non-degenerate quadratic space over F. If u, v are linearly independent in $\mathfrak{rad}_\nabla$, then

$$Q(au + bv) = a^2 Q(u) + b^2 Q(v)\,.$$

By assumption $Q(u)$, $Q(v) \neq 0$ and F is perfect, so we may solve $Q(au + bv) = 0$ by setting $b = 1$, $a = \sqrt{Q(v)/Q(u)}$. Since $au + bv$ is in $\mathfrak{rad}_\nabla$, this contradicts the definition of weak non-degeneracy. ∎

If U is a subspace of V then I define $U^\perp$ to be the subspace orthogonal to it with respect to ∇_Q.

[orthogonal-decomp] **Proposition 2.6.** *If (V, Q) is a quadratic space over F and U a subspace of V such that the restriction of Q to U is strictly non-degenerate, then $V = U \oplus U^\perp$.*

In these circumstances, I call U a strictly non-degenerate subspace of (V, Q).

Proof. We want to define a projection P from V onto U such that $v - P(v)$ lies in $U^\perp$. Let (e_i) be basis of U, let $M_\nabla = \nabla(e_i, e_j)$ be the matrix of $\nabla|U$. By assumption it is non-singular. Given v, we are looking for $u = \sum c_i e_i$ such that

$$\nabla\Big(v - \sum c_i e_i, e_j\Big) = 0, \quad \sum c_i \nabla(e_i, e_j) = \nabla(v, e_j)$$

for all j. But this is a system of equations for the unknowns c_i with invertible coefficent matrix. ∎

One strictly non-degenerate quadratic space that exists for all fields is the hyperbolic plane (F^2, H) for which $H(x, y) = xy$.

A vector v is called **anisotropic** if $Q(v) \neq 0$ and **isotropic** if $v \neq 0$ but $Q(v) = 0$. In H the isotropic vectors are those on the x- and y-axes.

[isotropic-char] **Corollary 2.7.** *If (V, Q) is a non-degenerate quadratic space and v is an isotropic vector in V, then there exists a hyperbolic plane in V containing v.*

Proof. Since $Q(v) = 0$ and $\mathfrak{rad}_Q = 0$, the vector v cannot lie in $\mathfrak{rad}_\nabla$. There thus exists $u \in V$ with $\nabla(u, v) = 1$. Then

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

and we may solve this to find a vector w with $Q(w) = 0$, but now

$$\nabla(w, v) = \nabla(u, v) + c\nabla(v, v) = 1$$

and the plane $\langle\langle w, v\rangle\rangle$ they span is a hyperbolic plane. ∎

A **totally isotropic** subspace of a quadratic space (V, Q) is a subspace U such that $Q(u) = 0$ for all u in U. If u, v lie in U then also

$$\nabla(u, v) = Q(u + v) - Q(u) - Q(v) = 0\,.$$

[hyperbolic-ndg] **Corollary 2.8.** *Every non-degenerate quadratic space containing a totally isotropic subspace of dimension n is isomorphic to nH plus an orthogonal complement.*

[q-q] **Corollary 2.9.** *If Q is a strictly non-degenerate quadratic form of dimension n then $Q \oplus -Q$ is isomorphic to nH.*

Proof. Because $x_i = y_i$ is a maximal isotropic subspace. □

[anisotropic-summand] **Corollary 2.10.** *Any non-degenerate quadratic space (V, Q) may be expressed as the orthogonal sum of copies of H and an anisotropic subspace.*

3. The orthogonal group

Let (V, Q) be a quadratic space. The **isometry group** or **orthogonal group** $O(Q)$ is the group of linear maps of V to itself preserving Q.

At this point we know almost nothing about isometries. If σ is an isometry and $\sigma u = v$ then $Q(u) = Q(v)$. But what about the converse? *Suppose $Q(u) = Q(v)$. Does there exist an isometry taking u to v?* This is the question I shall investigate next.

• **Reflections.** I start with a simple result that we shall see used several times.

[u-v-eq] **Lemma 3.1.** *Suppose u, v two vectors in V, with $Q(v) \neq 0$. Then $Q(u) = Q(u - tv)$ if and only if $t = 0$ or $t = -\nabla(u, v)/Q(v)$.*

Proof. Because

$$Q(u + tv) = Q(u) + t\nabla(u, v) + t^2 Q(v) = Q(u)$$

if and only if $t\nabla(u, v) = -t^2 Q(v)$. □

If v is anisotropic, the linear map

$$r_v\colon x \longmapsto x - \frac{\nabla(x, v)}{Q(v)}\, v\,.$$

is therefore an isometry. It fixes a vector x if and only if $\nabla(x, v) = 0$. If $y = r_v x$ then

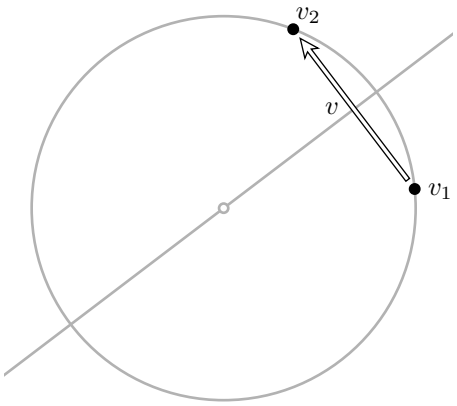
$$\nabla(y, v) = \nabla(x, v) - \frac{\nabla(x, v)}{Q(v)}\,\nabla(v, v) = -\nabla(x, v)$$
$$r_v^2 x = r_v y = y - \frac{\nabla(y, v)}{Q(v)}\, v = x\,,$$

so r_v has order two. It takes v to $-v$, and in odd characteristic we cannot have $\nabla(v, v) = 0$, so it is a reflection in the hyperplane $\nabla(x, v) = 0$. In even characteristic $\nabla(v, v) = 2Q(v) = 0$ so v always lies in the plane $\nabla(x, v) = 0$, and r_v is a shear parallel to that hyperplane. Nonetheless, I'll call it a reflection in all cases.

There is another way to state the Lemma:

[unique-reflection] **Lemma 3.2.** *If $\nabla(u, v) \neq 0$ the vector $r_v u$ is the unique vector w other than u on the line $t \mapsto u + tv$ with $Q(w) = Q(u)$.*

We'll find this useful for visualization in just a moment. What does this have to do with the problem of finding an isometry that takes v_1 to v_2? Suppose for the moment that $R = \mathbb{R}$ and $Q(x, y) = x^2 + y^2$ on the Euclidean plane. Given v_1 and v_2 of the same length, we can reflect v_1 in the line between them and get v_2.



images/real-reflection.eps

This line is the line perpendicular to $v = v_2 - v_1$, and the reflection subtracts from x the projection of x onto the line through v. In the standard notation of dot products, this projection is $(v_1 \bullet v)/(v \bullet v)v$, and the formula for a reflection is therefore

$$x \longmapsto x - 2\left(\frac{v \bullet x}{v \bullet v}\right) v$$

But the Euclidean norm is that determined by the dot-product, so we have here $B_Q(u, v) = 2(u \bullet v)$, and this formula says that in our terminology $r_v v_1 = v_2$. This is a general fact:

[reflection-lemma] **Proposition 3.3.** *Suppose v_1, v_2 to be two vectors with $Q(v_1) = Q(v_2)$. If $v = v_2 - v_1$ is anisotropic then $r_v v_1 = v_2$.*

♣ [u-v-eq] *Proof.* An immediate consequence of Lemma 3.1. □

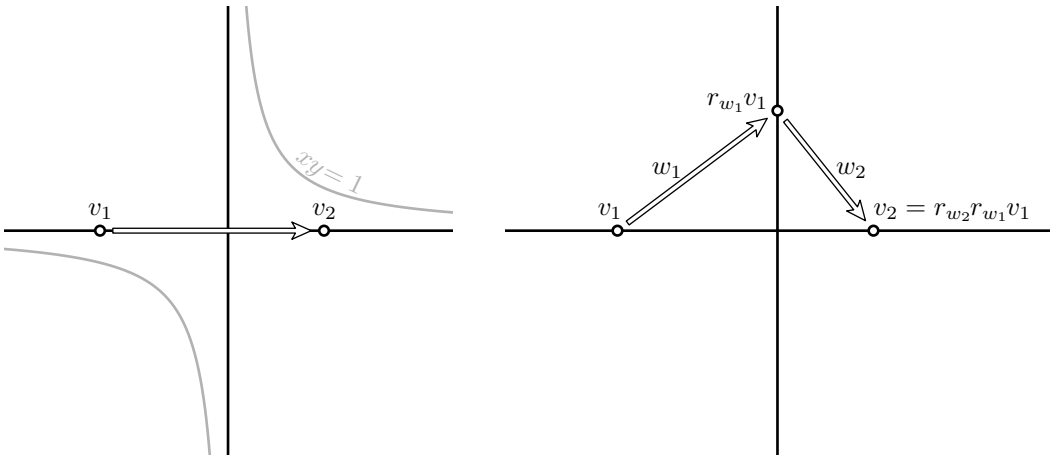
The following is trivial, but it will be useful for me to be able to refer to:

[trivial-lemma] **Lemma 3.4.** *If $Q(v_1) = Q(v_2)$ and $v = v_2 - v_1$ then $\nabla(v, v_2) = -\nabla(v, v_1) = Q(v)$.*

Proof. We have

$$\begin{aligned}
\nabla(v, v_1) &= \nabla(v_2 - v_1, v_1) \\
&= \nabla(v_2, u_1) - \nabla(v_1, v_1) \\
&= \nabla(v_2, u_1) - 2Q(v_1) \\
&= -Q(v_2) + \nabla(v_2, v_1) - Q(v_1) \\
&= -Q(v)\,. \qquad \square
\end{aligned}$$

• **Composites of reflections.** But now suppose that $Q(v_1) = Q(v_2)$, $v = v_2 - v_1$, but $Q(v) = 0$. We cannot reflect v_1 into v_2 in one shot. The Euclidean model will not suggest anything, because V possesses isotropic vectors. Instead, we take as our model the real hyperbolic plane with $Q(x, y) = xy$. As the first picture below should make clear, in this example if $Q(v_1) = Q(v_2)$ and $Q(v_2 - v_1) = 0$ then we must have $Q(v_1) = Q(v_2) = 0$. As the second picture should make clear, we can obtain v_2 by a composite of two reflections, with respect to anisotropic vectors w_1, w_2, with w_1 chosen more or less randomly, and w_2 then chosen so as to move $r_{w_1} v_1$ to v_2. Not quite randomly—one condition is that the reflection r_{w_1} must actually move v_1 so as to get w_2 anisotropic. This is not quite sufficient.



images/ne-sphere.eps
images/ne-reflection.eps

♣ **[reflection-lemma]** According to Proposition 3.3, we may choose $w_2 = v_2 - r_{w_1}v_1$. But then

$$\begin{aligned} w_2 &= v_2 - \left(v_1 - \frac{\nabla(w_1, v_1)}{Q(w_1)}\, w_1\right) \\ &= v + \frac{\nabla(w_1, v_1)}{Q(w_1)}\, w_1 \\ Q(w_2) &= Q(v) + \frac{\nabla(w_1, v_1)\nabla(w_1, v_1)}{Q(w_1)} + \frac{\nabla(w_1, v_1)\nabla(w_1, v_2 - v_1)}{Q(w_1)} \\ &= \frac{\nabla(v_1, w_1)\nabla(v_2, w_1)}{Q(w_1)}\,. \end{aligned}$$

Hence we have proved the following:

flection-product-lemma] **Lemma 3.5.** *Suppose v_1, v_2 to be two vectors with $Q(v_1) = Q(v_2)$. Let $v = v_2 - v_1$ and suppose that $Q(v) = 0$. If w_1 is an anisotropic vector that is not perpendicular to either v_1 or v_2, then $w_2 = v_2 - r_{w_1}v_1$ is also anisotropic and $r_{w_2}r_{w_1}v_1 = v_2$.*

• **Shears.** Now to define a different type of orthogonal transformation, one that exists only for quadratic spaces with sufficiently many isotropic vectors.

The pair u, v is called totally isotropic if $Q(u)$, $Q(v)$, and $\nabla(u, v)$ all vanish, or equivalently if the span $\langle\langle u, v\rangle\rangle$ of u and v is totally isotropic.

Define $\tau_{u,v}$ to be the linear transformation

$$\tau_{u,v}\colon x \longmapsto x + \nabla(x, v)u - \nabla(x, u)v\,.$$

If u and v span a line this is just the identity, and otherwise (a) it fixes vectors on the linear space $\{\nabla(x, u) = 0 \cap \nabla(x, v) = 0\}$, and (b) shifts vectors parallel to the plane spanned by u and v. If u, v are totally isotropic, it is a **shear** or **transvection**.

[transvection-lemma] **Lemma 3.6.** *If u and v are a totally isotropic pair, $\tau_{u,v}$ is an isometry.*

Proof. Because

$$\begin{aligned} Q(\tau_{u,v}x) = Q(x) &\\ &+ \nabla(x, v)\nabla(x, u) - \nabla(x, u)\nabla(x, v) \\ &\quad - \nabla(x, u)\nabla(x, v)\nabla(u, v) + \nabla(x, u)^2 Q(v) + \nabla(x, v)^2 Q(u) \\ = Q(x)\,. \end{aligned}$$

□

The following is a version of the extension theorem due to Ernst Witt in odd characteristic.

[witt-transitive] **Theorem 3.7.** *Suppose (V, Q) to be any quadratic space over F with bilinear form $\nabla = \nabla_Q$. If U_1, U_2 are subspaces of V such that $U_1 \cap \mathfrak{rad}_\nabla = U_2 \cap \mathfrak{rad}_\nabla = 0$, any isometry $\sigma\colon U_1 \to U_2$ may be extended to an isometry of V.*

In other words, $O(Q)$ acts transitively on certain embedded quadratic subspaces forming a Zariski-open subset of the relevant Grassmannian. Some restriction is necessary, since $O(Q)$ preserves $\mathfrak{rad}_\nabla$.

Proof. This is 8.3 of [Elman et al.:2008]. In odd characteristic this is well known and relatively easy, but in even characteristic more difficult. For the most part I follow the proof found in [Elman et al.:2008], except that I have modified their proof to make it (in principle) constructive. The proof is rather intricate. We start with this:

◇ *We are given an isometry $\sigma\colon U_1 \to U_2$. These subspaces satisfy the condition $U_i \cap \mathfrak{rad}_\nabla = 0$ for $i = 1, 2$. We wish to extend σ to an isometry of V.*

Step 1. The proof goes by induction on the common dimension of U_1 and U_2. The result is trivial when this dimension is 0. So now assume this dimension to be $n > 0$, and assume the result to be true in dimension $n - 1$. Let W_1 be a subspace of U_1 of codimension 1, $W_2 = \sigma(W_1)$. By the induction hypothesis, we may find an isometry of V extending $\sigma|W_1$. Replacing U_1 by $\sigma(U_1)$, we may now assume:

◇ *The intersection of U_1 and U_2 is a subspace W of codimension 1 in each, and there exists an isometry $\sigma\colon U_1 \to U_2$ which is I on W. We wish to find an extension of σ to all of V.*

Step 2. Choose u_1 in $U_1 - W$, and set $u_2 = \sigma(u_1)$. Let $u = u_2 - u_1$. If $u = 0$, $u_1 = u_2$ and we are done.

◇ *We have vectors u_i spanning U_i/W, with $u_2 = \sigma(u_1)$, $u = u_2 - u_1 \neq 0$.*

Step 3. For w in W

$$\nabla(u_1, w) = \nabla(\sigma u_1, \sigma w) = \nabla(u_2, w)$$

so $u \in W^\perp$. If u is anisotropic then the reflection r_w takes u_1 to u_2 and fixes all vectors in W. We are
♣ [trivial-lemma] again through. Otherwise, by Lemma 3.4, we may now assume:

◇ *We have $\nabla(u, u_1) = \nabla(u, u_2) = Q(u) = 0$.*

Step 4. At this point I call on the assumption that $U_i \cap \mathfrak{rad}_\nabla = 0$.

[elman] **Lemma 3.8.** *If U is a vector subspace of the quadratic space (V, Q) such that $U \cap \mathfrak{rad}_{\nabla_Q} = 0$, the map taking v to the restriction of ∇_v to U is surjective onto the linear dual of U.*

Proof of the Lemma. We have in general the exact sequence

$$0 \longrightarrow U \cap \mathfrak{rad}_\nabla \longrightarrow U \overset{u \mapsto \nabla_u}{\longrightarrow} \widehat{V}$$

whose transpose diagram, since ∇_Q is symmetric, is

$$V \overset{v \mapsto \nabla_v}{\longrightarrow} \widehat{U} \longrightarrow (U \cap \mathfrak{rad}_\nabla)\widehat{\ } \longrightarrow 0\,.$$

The claim follows, since $U \cap \mathfrak{rad}_\nabla = 0$. ∎

Because of the Lemma, we can find v_i in V such that $\nabla(v, W) = 0$ and $\nabla(v_i, u_i) \neq 0$. Thus:

◇ *Each subspace H_i of $W^\perp$ where $\nabla_{u_i} = 0$ is a proper subspace of $W^\perp$.*

flection-product-lemma] **Step 5.** Note that u lies in their intersection. According to Lemma 3.5, if we can find x anisotropic in $W^\perp$ that lies neither in H_1 nor H_2, we can find a composite of reflections takes u_1 to u_2. This may not be possible, but when it si not we can use a transvection instead.

◇ *At this point we look separately at two alternatives. Either (1) $u^\perp \cap W^\perp \subseteq H_1$ or (2) there exists x in $u^\perp \cap W^\perp$ not in H_1.*

Step 6. Suppose (1) $H = u^\perp \cap W^\perp \subseteq H_1$. Then in fact $H = H_1 = H_2$. Choose an arbitrary x in $W^\perp - H$ such that $\nabla(x,u) \neq 0$. If $Q(x) \neq 0$, we are done. Otherwise $Q(x) = 0$ and

$$Q(x+u) = Q(x) + \nabla(x,u) + Q(u) = \nabla(x,u) \neq 0\,.$$

But $x+u$ is again in $W^\perp - H$, and we are again done.

Step 7. Suppose (2) $u^\perp$ intersects $W^\perp - H_1$. Suppose (a) we can find x in this which is isotropic. We may scale it so that $\nabla(x,u_1) = \nabla(x,u_2) = 1$. Then x, u form a totally isotropic pair. The transvection $\tau_{u,w}$ fixes w in W and takes

$$u_1 \longmapsto u_1 + \nabla(w,u_1)u + \nabla(u,u_1)w = u_1 + u = u_2\,.$$

and we are done.

Step 8. Otherwise (b) no vector in $u^\perp - H_1$ is isotropic. If x lies in this, then necessarily $\nabla(x,u_2) \neq 0$ as well, and we are again done. The proof of Theorem 3.7 is finally complete. ∎

[witt-transitive]

[transitive-sphere] **Corollary 3.9.** *Suppose (V,Q) a non-generate quadratic space, u_1 and u_2 in V with $Q(u_1) = Q(u_2)$, neither in $\mathfrak{rad}_\nabla$. There exists an isometry of V taking u_1 to u_2.*

For example, if Q is a strictly non-degenerate quadratic form of dimension n in characteristic 2, then $Q \oplus x_{n+1}^2$ is non-degenerate. The space $\mathfrak{rad}_\nabla$ is spanned by $\varepsilon = (0,0,\dots,1)$. The unit sphere is the union of two orbits under the orthogonal group, ε and its complement.

[witt-transitive] *Proof.* This is just a special case of Theorem 3.7. ∎

[anisotropic-isomorphic] **Corollary 3.10.** *Any two decompositions of (V,Q) into a multiple of H plus an anisotropic subspace are equivalent by an isometry of V.*

Proof. Suppose $V = n_1H \oplus U_1 = n_2H \oplus U_2$, with U_1, U_2 anisotropic. Suppose $n_1 \leq n_2$. By Witt's Theorem we may find an isometry of V taking n_1H into n_2H. But then $(n_2 - n_1)H \oplus U_2 \cong U_1$, so $n_2 = n_1$ and $U_1 \cong U_2$. ∎

4. Binary forms

One possible non-degenerate quadratic form of dimension two is the hyperbolic plane H, and any non-degenerate plane with an isotropic vector is isomorphic to it. So to classify all binary forms, we have only to classify the anisotropic ones.

There is one simple way to get one: Let K be a separable quadratic extension of F, and let $N_{K/F}(x) = x\overline{x}$ be the norm map from K to F. It is a quadratic form on K considered as a vector space of dimension two over F. Related forms are the $aN_{K/F}$, with a in $F^\times$, and $aN_{K/F}$ and $bN_{K/F}$ are equivalent if and only if a/b lies in the image of $N_{K/F}K^\times$ in $F^\times$.

[binary-ndg] **Proposition 4.1.** *Every non-degenerate quadratic space of dimension 2 is isomorphic either to H or to some $aN_{K/F}$.*

Proof. Any quadratic form in dimension two has a formula $Q(x,y) = Ax^2 + Bxy + Cy^2$. If both A and C are 0, this is the hyperbolic plane. Otherwise, swapping x and y if necessary we may assume $A \neq 0$, and now the form factors as $A(x - \alpha y)(x - \beta y)$ over an algebraic closure of F. If α and β are in F, we

may change variables to get this of the form $Ax(x-\gamma)$. If $\gamma = 0$, the form will be degenerate. Otherwise, we can change variables again to make it Axy, so once more we have the hyperbolic plane.

We may now assume $\alpha \neq \beta$ to be conjugates in a quadratic extension K/F, and this is $AN_{K/F}$. ∎

[binary-sub] **Proposition 4.2.** *Suppose (V, Q) to be a non-degenerate quadratic space. If its dimension is even, it is the direct sum of even non-degenerate binary subspaces. If it is odd, it is the sum of non-degenerate binary subspaces and a single term cx^2.*

Proof. Suppose first the characteristic to be odd. Let v be such that $Q(v) \neq 0$. Then V is the direct sum of $F \cdot v$ and its orthogonal complement. Apply induction.

Now suppose the characteristic to be two. If all cross terms are 0, the form must be of dimension one. Otherwise, there exists a cross term Bxy with $B \neq 0$. The terms in x, y will give us a two-dimensional non-degenerate quadratic sub-space. The space V will be the direct sum of this and its orthogonal complement. Apply induction. ∎

5. Classification over finite fields

Throughout this section, suppose $q = p^n$ to be a prime power, $F = \mathbb{F}_q$, and Q to be a non-degenerate quadratic form on $V = F^d$.

[norm-onto] **Lemma 5.1.** *The norm $N_{K/F}$ from any finite extension K/F is surjective.*

Proof. We must have $K = \mathbb{F}_{q^m}$ for some m. The Galois group of K/F is cyclic, generated by $x \mapsto x^q$. The norm of x is

$$x^{1+q+\cdots+q^{m-1}} = x^{(q^{m-1}-1)/(q-1)}$$

and because the multiplicative group of $K^\times$ is cyclic, this is a surjective map onto elements of order $q-1$, which is $F^\times$. ∎

[classify-aniso-bin] **Lemma 5.2.** *The only anisotropic binary quadratic form over a finite field is $N_{K/F}$, with K/F the unique quadratic extension of F.*

♣ [binary-ndg] *Proof.* From the previous Lemma, by Proposition 4.1. ∎

[classify-irr] **Theorem 5.3.** *Any anisotropic quadratic form over a finite field is either one-dimensional or equivalent to $N_{K/F}$, with K/F the unique quadratic extension of F.*

Proof. If the dimension is one or two, this follows immediately from the previous Lemma. If it is three
♣ [binary-sub] or more, by Proposition 4.2 we can represent it as the sum of the space is the orthogonal sum of $N_{K/F}$
♣ [norm-onto] and some non-degenerate subspace. But by Lemma 5.1 we can then find an isotropic vector. ∎

[classification-even] **Corollary 5.4.** *A quadratic space of even dimension is a unique orthogonal sum $nH \oplus N_{K/F}$. A quadratic space of odd dimension is a unique orthogonal sum $nH \oplus cx^2$.*

If the characteristic is odd there are two distinct cases in odd dimension, according to whether c is a square or not. If it is even we may choose $c = 1$.

[odd-classification] **Proposition 5.5.** *Over finite fields of odd characteristic, a non-degenerate form is characterized by dimension and determinant.*

To be precise, if the dimension is $2n$ then it is isomorphic to nH is the $\det(M_Q)/(-1)^n$ is a square, and otherwise $(n-1)H \oplus N_{K/F}$. In dimension $2n+1$ it is isomorphic to $nH \oplus cx^2$ if $\det(M_Q)/(-1)^n = c$.

6. Arithmetic of Galois extensions

For the moment, suppose F to be any field, K/F a Galois extension with Galois group $\mathcal{G}$. I recall the basic technical Lemma of Galois theory.

[lin-ind-gal] **Proposition 6.1.** *The automorphisms in $\mathcal{G}$ are linearly independent over F.*

This is well known, but I'll include a proof here.

Proof. It is to be proved that if S is any finite subset of $\mathcal{G}$ and

$$\sum_S a_s x^s = 0$$

for all x in K, then all $a_s = 0$. The proof will be by induction on the size of S. The claim is trivial for $|S| = 1$, so now we assume it is true for any subset smaller than S. Suppose $|S| = n$ and that

$$\sum_{i=1}^{n} a_{s_i} x^{s_i} = 0$$

for all x in K. Then also

$$\sum_i a_{s_i} (xy)^{s_i} = \sum_S a_{s_i} x^{s_i} y^{s_i} = 0$$

for all x, y. Multiply the first equation by y^{s_n} to get

$$\sum_i a_{s_i} x^{s_i} y^{s_n} = 0$$

and then subtract to get

$$\sum_{i=1}^{n-1} a_{s_i} (y^{s_n} - y^{s_i}) x^{s_i} = 0\,.$$

We may apply the induction hypothesis to deduce that

$$a_{s_i} (y^{s_n} - y^{s_i}) = 0$$

for all i and all y. But s_n is different from all the s_i, so for each i we may find y with $y^{s_n} \neq y^{s_i}$. ∎

[trace-cor] **Corollary 6.2.** *The trace map*

$$\operatorname{trace}\colon K \to F, \quad x \longmapsto \sum_s x^s$$

is surjective.

Proof. It is an F-linear map, and by the Proposition its image is not 0. ∎

7. 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
♣ [trace-cor] 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 by Corollary 6.2 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:

[char-sum-null] **Lemma 7.1.** *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

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

[finite-FT] **Proposition 7.2.** *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:

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

♣ [char-sum-null] *Proof.* An application of Lemma 7.1. ∎

[plancherel-F] **Lemma 7.3.** *Irm (Plancherel formula) For f in $\mathbb{C}(F)$*

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

♣ [char-sum-null] *Proof.* Also an application of Lemma 7.1. ∎

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 $y \neq 0$

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

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.

In other words, the Fourier transform of χ is, up to a constant, the character χ^{-1}. The Fourier transform applied twice gives us $\chi(-x)$, which tells us that

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

and if $\chi = \mathrm{sgn}$

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

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

We have

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

To say that it has magnitude $\sqrt{q}$ is to say that the summands (which are complex numbers of absolute value 1) behave roughly like q steps of unit length in a random walk on the complex plane.

8. The sizes of spheres

Let (V, Q) be a non-degenerate quadratic space of dimension d over $F = \mathbb{F}_q$. The method I am going to use to find the sizes of the 'spheres' in V 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. (1) The function ν_Q can be recovered from it by the inverse Fourier transform:

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

(2) It can be easily calculated, since

[gammaq-sum] **Proposition 8.1.** *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 are going to 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.

Now let's look at the possibilities. classified basically by dimension d and the parity of q.

Example 1. Say first that $d = 1$, $Q(x) = ax^2$, q odd. We can see directly that

$$\nu_Q(x) = \begin{cases} 1 & \text{if } x = 0 \\ 2 & \text{if } x/a \text{ is a square} \\ 0 & \text{otherwise.} \end{cases}$$

If

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

then in all cases $\nu_Q = 1 + \operatorname{sgn}(x/a)$, and if $y \neq 0$

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

I recall that $|\mathfrak{G}| = \sqrt{q}$.

Example 2. Say $d = 1$, $Q(x) = x^2$, q even. Then $\nu_Q(x) = 1$ for all x and for $y \neq 0$

$$\gamma_Q(y) = \sum_x \psi(-xy) = 0$$

Example 3. Say now $d = 2$, $Q = H$. We can compute explicitly that

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

and for $y \neq 0$

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

♣ [norm-onto] **Example 4.** Say $d = 2$, $Q(x) = N_{K/F}(x)$. Then by Lemma 5.1

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

so 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 ♣ [gn:binary-sub] $|\gamma_Q(y)| = q^{d/2}$ for $d = 1$ or 2. But then by Proposition 4.2 and Proposition 8.1 this holds for all d. 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 forall $q \geq 2$, we have a second proof that a quadratic space over a finite field of dimension 3 or more always has isotropic vectors.

Example 5. If $d = 2n$, $Q = nH$, 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

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

so

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

Example 6. If $d = 2n$, $Q = (n-1)H + N_{K/F}$

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

Here

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

so

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

Example 7. If $d = 2n+1$, $Q = nH + ax^2$, q odd:

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

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

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

Example 8. If $d = 2n+1$, $Q = nH + x^2$, q even:

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

and $\nu_Q(x) = q^{2n}$ for all n

9. Concluding remarks

The functions ν_Q and γ_Q can be defined for all local fields F, 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 $\mathbb{Z}/p^n$, and may be used even more neatly for the classification over the p-adic integers. That story will be told elsewhere.

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

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.