The Local-to-Global Principle for Quadratic Forms

Jennifer Sinnott

May 24, 2004

Final Paper for Math 129: Topics in Number Theory, Prof. William Stein

In this paper we are interested in zeros of polynomials whose coefficients lie in $\mathbb{Q}$: we will ask, what does local information about the zeros of a polynomial tell us about its zeros globally? Local behavior amounts to checking whether the polynomial has zeros in the field $\mathbb{Q}$ completed with respect to some valuation: thus, we check whether a polynomial has a zero in $\mathbb{Q}_p$ for each prime $p \in \mathbb{Z}$, where $\mathbb{Q}_p$ is the field $\mathbb{Q}$ completed with respect to the p-adic valuation; and also whether it has a zero in $\mathbb{Q}_\infty = \mathbb{R}$. The main result of this paper will be that when $f(x_1, x_2, \ldots, x_n)$ is a homogeneous quadratic polynomial, the existence of a nontrivial zero in $\mathbb{Q}_p$ for each p and a nontrivial zero in $\mathbb{R}$ implies that the polynomial has a nontrivial zero in $\mathbb{Q}$. To prove this result, which is called the Hasse-Minkowski Theorem, I will follow closely Jean-Pierre Serre's discussion of the question in *A Course in Arithmetic*; as the treatment there was fairly condensed I found it worthwhile to expand on some of the details, which is primarily what I have done in what follows.

This ability to move from local information to global information begins to fail when we consider homogeneous cubic polynomials. A well-known example is due to Ernst Selmer: $3x^3 + 4y^3 + 5z^3$ has a zero in $\mathbb{R}$ and a zero in $\mathbb{Q}_p$ for each prime p, but no solutions in $\mathbb{Q}$.

At the end of this paper I will also mention an application of Hasse-Minkowski that will describe which integers can be written as the sum of three squares; a corollary will show that every integer can be written as the sum of four squares. Both of these can be proven using basic number theory, but they are also easy corollaries of the main result of this paper. But there's a lot of ground to cover before we get there.

We assume that all fields have characteristic $\neq 2$, and we assume that all vector spaces are finite dimensional over their field of scalars.

Quadratic Forms: Definitions and Basic Properties

Definition 1. Let V be a vector space over a field F. Let $Q : V \to F$ be a function from the vector space into the field satisfying the following two conditions:

(i) $Q(av) = a^2 Q(v)\ \forall\ a \in F, v \in V$

(ii) The function $\langle \cdot, \cdot \rangle : V \times V \to F$ given by $\langle v, w \rangle = \frac{1}{2}[Q(v+w) - Q(v) - Q(w)]$ is a (symmetric) bilinear form.

Then we say that Q is a *quadratic form*, and we call (V, Q) a *quadratic module*.

We will often make use of the bilinear form defined in (ii), and, in fact, on a given vector space V, there is a natural correspondence between its symmetric bilinear forms and its quadratic forms, as follows. If we are given a symmetric bilinear form, we can define $Q(v) = \langle v, v\rangle$, and get back a quadratic form (computing $Q(av) = \langle av, av\rangle = a^2\langle v, v\rangle = a^2 Q(v)$ shows that property (i) holds, and $\frac{1}{2}[\langle v + w, v + w\rangle - \langle v, v\rangle - \langle w, w\rangle] = \langle v, w\rangle$ shows that property (ii) holds). Expanding $Q(v+w)$, and solving for $\langle v, w\rangle$ shows that the formula given in the definition gets us back to the same bilinear form. Similarly, if we start with a quadratic form and define a bilinear form $\langle v, w\rangle$ by the expression given in (ii), then $\langle v, v\rangle$ is the quadratic form we started with. Thus, we get a one-to-one correspondence between quadratic forms and symmetric bilinear forms on a vector space V (when, again, we assume the scalar field has characteristic $\neq 2$).

The more familiar definition of a quadratic form is a polynomial in variables $x_1, \ldots, x_n$ where each term has degree 2, so an expression of the form

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j.$$

This description can be derived from the more abstract definition given above by choosing a basis for the vector space V. Let $e_1, \ldots e_n$ be a basis for V. For $x \in V$, we write $x = \sum_{i=1}^{n} x_i e_i$. Then

$$\begin{aligned} Q(x) = Q(\sum_{i=1}^{n} x_i e_i) &= \langle \sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} x_j e_j\rangle \\ &= \sum_{i=1}^{n} x_i \langle e_i, \sum_{j=1}^{n} x_j e_j\rangle \quad \text{(bilinearity in the first component)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \langle e_i, e_j\rangle \quad \text{(bilinearity in the second component)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j, \end{aligned}$$

where in the last expression $a_{ij} = \langle e_i, e_j\rangle$. We can in turn think of this expression as $x^t A x$, where x is a vector with coordinates (x_i) in the basis (e_i), and $A = (a_{ij})$. Note that A is symmetric, since $\langle e_i, e_j\rangle = \langle e_j, e_i\rangle$. We call A the matrix associated to the quadratic form; notice that the same matrix defines the associated bilinear form: $\langle x, y\rangle = x^t A y$.

It would be nice to associate with a quadratic form Q the determinant of the matrix A which represents the form, but A depends on the choice of basis, so we won't in general be able to associate a single number to Q in this way. But we will try to anyway: we define the *discriminant* of Q to be the determinant of its matrix A with respect to some basis of V. If we recall from linear algebra that changing the basis of V changes the matrix of the bilinear form to $A' = X^t A X$, where X is some invertible matrix, then we see that $\det A' = \det A(\det X)^2$; thus, the discriminant

is determined up to multiplication by the square of a nonzero field element. Thus, if we are working over $\mathbb{Q}$ or $\mathbb{R}$, for example, then it makes sense to say that the discriminant is positive, negative, or 0, because those distinctions are not affected by squares. Regardless of the field F, we can view the discriminant as a uniquely determined element of $\{0\} \cup F^*/F^{*2}$, the elements of the field modulo the square elements of the field.

The distinction between nonzero and 0 discriminant is of course always valid. If Q has discriminant 0, we say it is a *degenerate* form; if Q has nonzero discriminant, it is *nondegenerate*. We can say more to distinguish degenerate forms from nondegenerate ones, but we need to define orthogonality first:

Definition 2. Let (V, Q) be a quadratic module. We say $x, y \in V$ are *orthogonal* if $\langle x, y\rangle = 0$. If $H \subset V$, then define the *orthogonal complement* $H^0 = \{y \in V : \langle x, y\rangle = 0 \ \forall \ x \in H\}$; note that H^0 forms a vector subspace. Finally, if V_1 and V_2 are two vector subspaces of V, then they are *orthogonal* if $x \in V_1, y \in V_2 \implies \langle x, y\rangle = 0$. That means $V_1 \subset V_2^0$ and $V_2 \subset V_1^0$.

Q is a nondegenerate form $\iff$ the orthogonal complement of the whole space V with respect to the form Q is trivial: if $V^0 = 0$. This equivalence is easy to see: letting A be the matrix of Q with respect to some basis, we see that a vector $v \in V^0$ would have the property that $w^t Av = 0$ for every $w \in V$, so we would have to have $Av = 0$, which would mean $v \in \ker A$; thus, there's a nontrivial vector in $V^0 \iff$ there's a nontrivial vector in the kernel of $A \iff \det A = 0$.

A very useful technique when dealing with a quadratic module (V, Q) is going to be breaking it up into pairwise orthogonal subspaces $U_1, \ldots U_m$, whose direct sum is equal to the whole space V. We call V the *orthogonal direct sum* of the U_i, and use the notation $\hat{\oplus}$ to signify the pairwise orthogonality of the U_i :

$$V = U_1 \hat{\oplus} \cdots \hat{\oplus} U_m.$$

A useful property of this construction is that if you take an element x and write it as a sum of $x_i \in U_i$ for each i, then $Q(x) = Q_1(x_1) + \cdots + Q_m(x_m)$, where Q_i is the restriction of Q to U_i; if, on the other hand, you have a collection of quadratic modules (U_i, Q_i), you can construct a quadratic module (V, Q) where $V = \oplus U_i$ and $Q(x) = Q_1(x_1) + \cdots + Q_m(x_m)$, for $x = (x_1, \ldots, x_m) \in V$; this means that V is an orthogonal direct sum of the U_i.

Suppose that f and g are two forms written out with respect to a basis: so $f(x_1, x_2, \ldots, x_n)$ and $g(y_1, y_2, \ldots y_m)$. We will write $f \dot{+} g$ to denote the form in $n + m$ variables defined by: $f(x_1, x_2, \ldots, x_n) + g(x_{n+1}, x_{n+2}, \ldots, x_{n+m})$. This $\dot{+}$ operation corresponds to orthgonal sum of the quadratic modules associated to f and g. This operation will be very useful to us later.

Definition 3. Let (V, Q) be a quadratic module. An element $x \in V$ is said to be *isotropic* if $Q(x) = 0$. A vector subspace $U \subset V$ is said to be *isotropic* if $Q(x) = 0$ $\forall x \in U$.

Proposition 4. Let (V, Q) be a quadratic module; let U be a vector subspace. Then:

$$U \text{ is isotropic} \iff U \subset U^0 \iff Q|_U = 0.$$

Pf. If U is isotropic, then for $x, y \in U$ $\langle x, y\rangle = \frac{1}{2}[\langle x+y, x+y\rangle - \langle x, x\rangle - \langle y, y\rangle] = 0$, so $U \subset U^0$; the other equivalences are clear from the definitions. □

Definition 5. We call a quadratic module U a hyperbolic plane if it has basis x, y, where x and y are each isotropic, but $\langle x, y\rangle \neq 0$. Scaling y by $\frac{1}{\langle x,y\rangle}$, we can assume $\langle x, y\rangle = 1$. Note that we can tell already that U is a nondegenerate module: using the definition of the defining matrix A as having entries $a_{ij} = \langle e_i, e_j\rangle$, we can compute the matrix of Q to be $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ which has nonzero determinant.

Proposition 6. Let (V, Q) be a nondegenerate quadratic module. Then any nonzero isotropic element x is contained in a hyperbolic plane $U \subset V$.

Pf. Since V is nondegenerate we can find some $z \in V$ so that $\langle x, z\rangle \neq 0$; scale z so that $\langle x, z\rangle = 1$. Then, letting $y = z - \frac{1}{2}\langle z, z\rangle x$, we have:

$$\begin{aligned}\langle x, y\rangle &= \langle x, z - \frac{1}{2}\langle z, z\rangle x\rangle \\ &= \langle x, z\rangle - \frac{1}{2}\langle z, z\rangle\langle x, x\rangle \\ &= \langle x, z\rangle = 1.\end{aligned}$$

Also,

$$\begin{aligned}\langle y, y\rangle &= \langle z - \frac{1}{2}\langle z, z\rangle x, z - \frac{1}{2}\langle z, z\rangle x\rangle \\ &= \langle z, z\rangle - \langle z, z\rangle\langle x, z\rangle + \frac{1}{4}\langle z, z\rangle^2\langle x, x\rangle = 0\end{aligned}$$

So the space U spanned by x and y is the desired hyperbolic plane. □

Representation of Field Elements by Quadratic Forms

We say a form Q *represents* an element a in the scalar field F if there is a nonzero element x so that $Q(x) = a$; if the form is written out as $f(x_1, x_2, \ldots, x_n)$, then f represents an element a if there's an n-tuple $(x_1, x_2, \ldots, x_n) \neq (0, 0, \ldots, 0)$ so that $f(x_1, x_2, \ldots, x_n) = a$.

We say that two forms f and f', with coefficient matrices A and A' respectively, are *equivalent* if there exists an invertible matrix X so that X^tAX. It is easy to see that two equivalent forms represent the same values: for, if f represents a, then there exists some $x \neq 0$ so that $x^tAx = a$; we find $y = X^{-1}x$, so that $x = Xy$, so $a = x^tXx = (Xy)^tAXy = y^t(X^tAX)y = y^tA'y$; thus, this follows from the invertibility of the change of basis matrix X. We write $f \sim f'$ when f and f' are equivalent.

Recalling the definition of a hyperbolic plane, we now define a hyperbolic form:

Definition 7. $f(x_1, x_2)$ is called hyperbolic if $f \sim x_1x_2 \sim x_1^2 - x_2^2$. This means the quadratic module (F^2, f) is a hyperbolic plane: for, if f is hyperbolic, then it is equivalent to x_1x_2, so that the matrix of the form is equivalent to $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which gives us some basis e_1, e_2 so that $\langle e_i, e_i \rangle = 0$ and $\langle e_1, e_2 \rangle \neq 0$. Conversely, if we have a hyperbolic plane with isotropic basis vectors x, y, we saw that the matrix of the form was $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$; scaling y by $\frac{1}{2}$, the matrix becomes $\begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$; multiplying out the form we get exactly xy. Finally, diagonalizing the matrix gets us the form $x_1^2 - x_2^2$.

From this we get the following proposition:

Proposition Juniper. Let f be a nondegenerate quadratic form that represents 0. Then we have $f \sim f_2 \dot{+} g$, where f_2 is hyperbolic. Furthermore, f represents all elements of the field F.

Pf. Proposition 6 lets us break our nondegenerate module (V, Q) corresponding to f into a hyperbolic plane and its orthogonal complement. (One can check that if U is a nondegenerate subspace of V, that $V = U \hat{\oplus} U^0$.) Then, the equivalence established in definition 7 gives us that $f \sim f_2 \dot{+} g$, where f_2 is hyperbolic. f represents all values of F because a hyperbolic plane (V, Q) has the property that $Q(V) = F$, for $Q(x + \frac{a}{2}y) = a$. □

Finally note the following important theorem:

Theorem 8. Let f be a quadratic form in n variables. Then there exist $a_1, a_2, \ldots, a_n \in F$ so that $f \sim a_1x_1^2 + \cdots + a_nx_n^2$.

Pf. This follows from the fact that we can find an orthogonal basis for any quadratic module (V, Q): if all elements are isotropic, then any basis is orthogonal; otherwise, choose a nonisotropic vector e_1, let it be your first basis element, and let H be its orthogonal complement, which has dimension 1 smaller than V; repeating this process, we get an orthogonal basis.

So, there's some basis of V with respect to which the form has matrix

$$A = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_n \end{pmatrix},$$

and so $f \sim a_1x_1^2 + \cdots + a_nx_n^2$. □

We now pursue the question of when a form represents 0: the above theorem allows us to assume $f(x_1, \ldots, x_n) = a_1x_1^2 + \cdots + a_nx_n^2$.

First, we need to talk a little bit about something called the Hilbert symbol.

Hilbert Symbols

In this section we assume that F is either $\mathbb{R}$ or $\mathbb{Q}_p$ for some prime $p \in \mathbb{Z}$. We define the *Hilbert symbol* (a, b) of two elements $a, b \in F^*$ by:

· $(a, b) = 1$ if $z^2 - ax^2 - by^2$ represents 0;

· $(a, b) = -1$ otherwise.

We will often write $(a, b)_v$ to emphasize that the symbol is dependent on the field we are talking about — $\mathbb{Q}_v$ for some valuation v. If we multiply a or b by a square, that square can be absorbed, so to speak, by x^2 or y^2, so we can think of a and b both as elements of the multiplicative group F^*/F^{*2}; this is often more useful. Thus, we can think of the Hilbert symbol as a map $F^*/F^{*2} \times F^*/F^{*2} \to \{\pm 1\}$.

Here is an easy proposition that will come in handy:

Proposition 9. Let $a, b \in F^*$, and let $\beta = \sqrt{b}$. Then the Hilbert symbol $(a, b) = 1 \iff a$ is the norm of an element in $F(\beta)$.

Pf. ($\Longleftarrow$) If b is already a square in F, say $b = c^2$, then $F(\beta) = F$ so requiring that a is the norm of an element is just requiring $a \in F^*$, at which point we get that $(a, b) = 1$ by letting $(x, y, z) = (0, 1, c)$. If b is not a square, then $F(\beta)$ is a quadratic extension of F, and every element α can be written $\alpha = s + \beta t$ for $s, t \in F$; every norm is of the form $s^2 - bt^2$, so if $a = s^2 - bt^2$, then $(1, t, s)$ is a solution to $z^2 - ax^2 - by^2 = 0$.

($\Longrightarrow$) If $(a, b) = 1$, then we have some $(x, y, z) \neq (0, 0, 0)$ with $z^2 - ax^2 - by^2 = 0$; if $x = 0$, then $z^2 = by^2$, so b must be a square; that means $F = F(\beta)$ as above, so again a is a norm because it is in F^*. If $x \neq 0$, we can solve for a: we see $a = \frac{z^2}{x^2} - b\frac{y^2}{x^2}$, so that $a = Norm_{F(\beta)/F}(\frac{z}{x} - \beta\frac{y}{x})$. □

Proposition 10. Let a and b be two elements of F^*, where F is $\mathbb{Q}_p$ or $\mathbb{R}$. The Hilbert symbol (a, b) satisfies the following:

(i) $(a, b) = (b, a)$ (Symmetry)

(ii) $(a, c^2) = 1$

(iii) $(a, -a) = 1$

(iv) $(a, 1 - a) = 1$

(v) $(a, b) = (a, -ab)$

(vi) $(a, bb') = (a, b)(a, b')$ (Bilinearity)

Pf. The first expression is obvious from the definition of Hilbert symbol; for (ii), if $b = c^2$, then as in the proof of proposition 9, $(0, 1, c)$ is a nontrivial solution; for (iii), if $b = -a$, then $(1, 1, 0)$ is a nontrivial solution; for (iv), if $b = 1 - a$, then $(1, 1, 1)$ is a nontrivial solution. Finally, (vi) follows from theorem Paperclip below by inspection; then (v) follows from (vi) combined with (iii). □

Letting $(\frac{a}{p})$ denote the Legendre symbol (so $(\frac{a}{p}) = 1$ if a is a square mod p, and -1 otherwise) the following theorem concerning the Hilbert symbol can be shown:

Theorem Paperclip. When $F = \mathbb{Q}_\infty = \mathbb{R}$, $(a, b) = 1$ if at least one of a and b is positive, and $(a, b) = -1$ if they're both negative.

When $F = \mathbb{Q}_p$, write a and b as $a = p^\alpha u$ and $b = p^\beta v$, where u and v are p-adic units. Then when $p \neq 2$ we have the following relation:

$$(a, b) = (-1)^{\alpha\beta\varepsilon(p)}(\frac{u}{p})^\beta(\frac{v}{p})^\alpha,$$

where $\varepsilon(u)$ denotes the equivalence class modulo 2 of $\frac{u-1}{2}$. When $p = 2$, we have:

$$(a, b) = (-1)^{\varepsilon(u)\varepsilon(v)+\alpha\omega(v)+\beta\omega(u)},$$

where $\omega(u)$ is the class modulo 2 of $\frac{u^2-1}{8}$.

In particular (this being a fact we will later use), for $p \neq 2$, if a and b are both p-adic units, $(a, b) = 1$.

Pf. See [Serre, p. 20]. □

The Hilbert symbol satisfies a product formula which looks similar to the product formula for valuations. In the following, given $a, b \in \mathbb{Q}^*$, we use the symbol $(a, b)_v$ to denote the Hilbert symbol of the images of a and b in the completed field $\mathbb{Q}_v$ for $v \in V = \{p \text{ prime in } \mathbb{Z}\} \cup \{\infty\}$.

Theorem 11. Let $a, b \in \mathbb{Q}^*$. Then $(a, b)_v = 1$ for almost all $v \in V$, and $\prod_{v \in V}(a, b)_v = 1$.

Pf. See [Serre, p. 23]. □

Finally, we assume the following "approximation theorem" for Hilbert symbols: we can take a finite set of elements a_i and find an $x \in \mathbb{Q}^*$ so that $(a_i, x)_v$ is whatever we like at each valuation v, so long as our whim meets certain obvious constraints:

Theorem 12. Let $\{a_i\}_{i \in I}$ be a finite set of elements of $\mathbb{Q}^*$. Let $\{\delta_{i,v}\}_{i \in I, v \in V}$ be a collection of numbers equal to ± 1. Then there exists some $x \in \mathbb{Q}^*$ with $(a_i, x)_v = \delta_{i,v}$ for every $i \in I$ and every $v \in V$ exactly when the following three conditions hold:

(i) Almost all the $\delta_{i,v}$ are equal to 1.

(ii) For each $i \in I$, $\prod_{v \in V} \delta_{i,v} = 1$.

(iii) For every $v \in V$, $\exists\ x_v \in \mathbb{Q}_v^*$ with $(a_i, x_v)_v = \delta_{i,v}$ for every $i \in I$.

Pf. See [Serre, p. 24]. □

The Invariants of a Form

We continue to assume that F is $\mathbb{Q}_v$, the completion of $\mathbb{Q}$ with respect to v, which is either the archimedean valuation or a p-adic valuation. Now we can define two invariants of a form which will help us determine when a form can represent 0:

Let f be a nondegenerate quadratic form over F. We let $d_v(f)$ be the form's discriminant. If the form is diagonalized, i.e. written as $f(x_1, \ldots, x_n) = a_1x_1^2 + \cdots a_nx_n^2$, we have clearly $d_v(f) = a_1 \cdots a_n$.

We also define an invariant $\varepsilon_v(f) = \prod_{i<j}(a_i, a_j)_v$. We have $\varepsilon_v = \pm 1$.

Since d_v is taken as an element of F^*/F^{*2} and the determinant of the matrix of the form only varies by squares, it is easy to see that d_v is an invariant of the equivalence class of the form. One must put more effort into showing that ε_v is invariant. See [Serre, p. 35].

A criterion for squares follows from definitions: a number x is a square in $F^* \iff x \equiv 1$ in F^*/F^{*2}.

Theorem 13. Let $F = \mathbb{Q}_v$ be a completion of $\mathbb{Q}$. Let f be a nondegenerate quadratic form in n variables. Let $d = d_v(f)$ and $\varepsilon = \varepsilon_v(f)$. Then f represents 0 $\iff$

(i) $d = -1 \quad (n = 2)$;

(ii) $(-1, -d) = \varepsilon \quad (n = 3)$;

(iii) either $d \neq 1$ or $d = 1$ and $\varepsilon = (-1, -1) \quad (n = 4)$;

(iv) $v = p$ or $v = \infty$ and f is indefinite $\quad (n \geq 5)$;

where equality is equivalence up to squares.

Pf. There are the four cases to check, and each is a matter of observation and manipulation until we get the formulas above; so here I will show (i) and (ii) which indicate the sorts of arguments involved; for (iii) and (iv), see [Serre, p. 37-8]. Write f as $a_1x_1^2 + \cdots + a_nx_n^2$.

(i) $(n = 2)$ If (x_1, x_2) is going to be a solution to $a_1x_1^2 + a_2x_2^2 = 0$, then for one thing neither x_1 nor x_2 can be 0. Then f represents 0 $\iff$ $a_1x_1^2 = -a_2x_2^2$ $\iff$ $-\frac{a_1}{a_2} = (\frac{x_2}{x_1})^2$, so if and only if $-\frac{a_1}{a_2}$ is a square. That happens if and only if $-\frac{a_1}{a_2} \equiv 1$ in F^*/F^{*2}. Since $-\frac{a_1}{a_2} \equiv -a_1a_2 = -d$ in F^*/F^{*2}, $-\frac{a_1}{a_2} \equiv 1 \iff d = -1$.

(ii) ($n = 3$) Multiplying through by a_3 and absorbing squares, we get that f represents $0 \iff a_1a_3x_1^2 + a_2a_3x_2^2 + x_3^2$ represents 0. By definition of the Hilbert symbol, that form represents $0 \iff (-a_1a_3, -a_2a_3) = 1$. Expanding this using bilinearity of the Hilbert symbol, we eventually end up with the requirement specified above. The details are as follows:

$$\begin{aligned}
&(-a_1a_3, -a_2a_3) \\
&= (-1, -a_2a_3)(a_1, -a_2a_3)(a_3, -a_2a_3) \\
&= (-1,-1)(-1,a_2)(-1,a_3)(a_1,-1)(a_1,a_2)(a_1,a_3)(a_3,-a_3)(a_3,a_2) \\
&= (-1,-1)(-1,a_1)(-1,a_2)(-1,a_3)(a_1,a_2)(a_1,a_3)(a_2,a_3) \\
&= (-1, -a_1a_2a_3)\prod_{i<j}(a_i,a_j) \\
&= (-1,-d(f))\varepsilon(f)
\end{aligned}$$

which equals 1 exactly when $(-1,-d) = \varepsilon$, so our claim is proven. □

We also prove the following proposition about $F = \mathbb{Q}_p$ for some prime p; it follows directly from the above theorem. Here we are concerned with when a form represents an element $a \in F^*/F^{*2}$.

Proposition Goldfish. Let $F = \mathbb{Q}_p$ for p prime. Let $a \in F^*/F^{*2}$. Then $f = a_1x_1^2 + \cdots + a_nx_n^2$ represents $a \iff$

(i) $a = d \quad (n = 1)$;

(ii) $(a, -d) = \varepsilon \quad (n = 2)$;

(iii) either $a \neq -d$ or $a = -d$ and $(-1,-d) = \varepsilon \quad (n = 3)$;

(iv) no conditions $\quad (n \geq 4)$.

Pf. Let $a \in F^*/F^{*2}$. Defining $\tilde{f} = f \dot{-} az^2$, then clearly $\tilde{f}$ represents $0 \iff f$ represents a: for, if f represents a, it is clear that $\tilde{f}$ represents 0; conversely, if $\tilde{f}$ represents 0, then either $z = 0$ so f represents 0, at which point it represents all of F, or $z \neq 0$, at which point $f(x_1/z, \ldots, x_n/z)$ represents a.

Furthermore, $d(\tilde{f}) = -ad(f)$ (recalling that d is the product of the coefficients of the form); and $\varepsilon(\tilde{f}) = (-a, d(f))\varepsilon(f)$, since ε is the product of the Hilbert symbols of each pair of the coefficients, so the addition of a new coefficient will lead to multiplication of the original $\varepsilon(f)$ by $(-a,a_1)(-a,a_2)\cdots(-a,a_n)$.

So in each case we will have that the form f represents $a \iff \tilde{f}$ represents 0. We proceed by translating the conditions on $\tilde{f}$ from theorem 13 into conditions on f and a by using the relations just established between $\varepsilon(f)$ and $\varepsilon(\tilde{f})$, and between $d(f)$ and $d(\tilde{f})$.

(i) When $n = 1$, $\tilde{f}$ represents 0 if $d(\tilde{f}) = -1$ by part (i) of the above theorem, so when $-ad(f) = -1$, so when $a = d(f)$.

(ii) When $n = 2$,

$$\begin{aligned}
\tilde{f} \text{ represents } 0 &\iff (-1, -d(\tilde{f})) = \varepsilon(\tilde{f}) \\
&\iff (-1, ad(f)) = (-a, d(f))\varepsilon(f) \\
&\iff (-1, a)(-1, d(f)) = (-1, d(f))(a, d(f))\varepsilon(f) \\
&\iff (-1, a) = (a, d(f))\varepsilon(f) \\
&\iff (a, -1)(a, d(f)) = \varepsilon(f) \\
&\iff (a, -d(f)) = \varepsilon(f)
\end{aligned}$$

(iii) When $n = 3$, $\tilde{f}$ represents 0 if $d(\tilde{f}) \neq 1$, or if $d(\tilde{f}) = 1$ and $\varepsilon(\tilde{f}) = (-1, -1)$; for the first case

$$d(\tilde{f}) \neq 1 \iff -ad(f) \neq 1 \iff -a \neq d(f);$$

for the second case,

$$d(\tilde{f}) = 1 \iff -ad(f) = 1 \iff a = -d(f);$$

in this case, we have $\varepsilon(\tilde{f}) = (-1, -1)$, which happens

$$\begin{aligned}
&\iff (-a, d(f))\varepsilon(f) = (-1, -1) \\
&\iff (-1, d(f))(a, d(f))\varepsilon(f) = (-1, -1);
\end{aligned}$$

since in this case $(a, d(f)) = (-d(f), d(f)) = 1$, we have $(-1, -d(f)) = \varepsilon(f)$.

(iv) Finally, when $n \geq 4$, $\tilde{f}$ always represents 0, so f always represents a. □

Remark. When $F = \mathbb{R}$, (i), (ii), and (iii) of proposition Goldfish are still true; when $n \geq 4$, we always get a solution when not all of the coefficients have the same sign (i.e., the form is indefinite); if the form is positive definite (all its coefficients are positive) we get a solution exactly when a is positive; and if the form is negative definite (all its coefficients are negative) we get a solution exactly when a is negative.

The Main Result

We are now interested in the local-to-global principle: letting f be a quadratic form with rational coefficients, we will show that f represents 0 in $\mathbb{Q}_v$ for all v exactly when f represents 0 in $\mathbb{Q}$. We will sometimes use the notation f_v when we want to be clear that we are viewing the form over $\mathbb{Q}_v$.

Theorem 14. (Hasse-Minkowski) A quadratic form f represents 0 in $\mathbb{Q} \iff$ f represents 0 in $\mathbb{Q}_v \; \forall \; v \in V = \{p \text{ prime in } \mathbb{Z}\} \cup \{\infty\}$.

Pf. ($\Longrightarrow$) Since $\mathbb{Q} \subset \mathbb{Q}_v$ for each v, a representation of 0 in $\mathbb{Q}$ is also a representation of 0 in $\mathbb{Q}_v$.

($\Longleftarrow$) As we have seen earlier, we can assume that one of the coefficients in our form $f = a_1x_1^2 + \cdots + a_nx_n^2$ is 1 by multiplying through; say by $a_1 : a_1 f = a_1^2x_1^2 + \cdots + a_1a_nx_n^2$, which clearly represents 0 exactly when the original form does. We split up into four cases:

($n = 2$)

We can assume that $f = x_1^2 + ax_2^2$; since f_∞ represents 0, we know that $a > 0$. We can factor $a \in \mathbb{Q}^*$ as a product of primes; the power of each p in the factorization is exactly the valuation $v_p(a)$ (including when p doesn't divide a and so the valuation is 0) so we can write a as a product in all primes: $a = \prod_p p^{v_p(a)}$, where $v_p \in \mathbb{Z}$, and for most p, $v_p(a) = 0$.

We know by assumption that in each p-adic field $\mathbb{Q}_p$, f_p represents 0. Thus, in each field $\mathbb{Q}_p$, solving for a we find that a is a square: $0 = x_1^2 - ax_2^2 \implies a = (\frac{x_1}{x_2})^2$ in $\mathbb{Q}_p$. Thus, $v_p(a)$ is even for each p. Thus, it is clear that a is in fact a square in $\mathbb{Q}$; therefore, f represents 0, since $x_1 = \sqrt{a}$ and $x_2 = 1$ will work.

($n = 3$)

When $n = 3$, the form looks like $f = x_1^2 - ax_2^2 - bx_3^2$, where $a, b \in \mathbb{Q}^*$ (we don't assume anything about their sign). If a and b have square factors, we can absorb them into x_2 and x_3, so we can assume they are square free, and we can assume they are integers by multiplying by p^2 if p is in the denominator. That is, $v_p(a)$ and $v_p(b)$ are 0 or 1 for each prime p.

We can also assume without loss that $|a| \leq |b|$; we let $m = |a| + |b|$ and induct on the size of m:

Base case: $m = 2$. We have $f = x_1^2 \pm x_2^2 \pm x_3^2$. Since f_∞ represents 0, we know that not all signs are positive, and so it is easy to find solutions in $\mathbb{Q}$.

Now the case when $m > 2$. Then the larger coefficient $|b| \geq 2$. Write $b = \pm p_1 \cdots p_k$ for distinct primes p_i. Then we claim that a is a square $\pmod{p_i}$. If $a \equiv 0 \pmod{p_i}$, this is obvious; otherwise a is a unit in $\mathbb{Q}_{p_i}$. We know that f represents 0 in $\mathbb{Q}_{p_i}$, so find nontrivial $(x, y, z) \in \mathbb{Q}_{p_i}^3$ so that $z^2 - ax^2 - by^2 = 0$; by homogeneity we can assume that (x, y, z) are in $\mathbb{Z}_{p_i}^3$, and not all divisible by p_i. Reducing modulo p_i, we see that $z^2 - ax^2 \equiv 0 \pmod{p_i}$. If $x \equiv 0 \pmod{p_i}$, then $z \equiv 0 \pmod{p_i}$, so that in the original expression $z^2 - ax^2 - by^2 = 0$, we find $z^2 - ax^2 = by^2$ we see that $p_i^2 | by^2$; but p_i only divided b once, so $p_i | y$, and the triplet (x, y, z) is no longer primitive, contradiction. Thus, we cannot have $x \equiv 0 \pmod{p_i}$, so a is a square in $\mathbb{Z}/p_i\mathbb{Z}$.

Now, we apply the Chinese remainder theorem to see that $\mathbb{Z}/b\mathbb{Z} \cong \prod_{p_i|b} \mathbb{Z}/p_i\mathbb{Z}$, since the p_i are assumed to be distinct, a is in fact a square $\pmod{b}$. So we can find integers t and b' so that $t^2 = a + bb'$; we can choose t so that $|t| \leq \frac{|b|}{2}$. Solving for bb', we see that $bb' = t^2 - a$, so that bb' is a norm from $F(\sqrt{a})$, when F is either $\mathbb{Q}$ or $\mathbb{Q}_v$. By proposition 9, we have $f = x_1^2 - ax_2^2 - bx_3^2$ represents $0 \iff b$ is

a norm from $F(\sqrt{a}) \iff b'$ is a norm from $F(\sqrt{a}) \iff f' = x_1^2 - ax_2^2 - b'x_3^2$ represents 0. Since bb' is a norm and norms are multiplicative, either b and b' are both norms or neither one is; this is how we got the middle equivalence. Thus f' represents 0 in each $\mathbb{Q}_v$ since f does; and it will suffice to show that f' represents 0 in $\mathbb{Q}$, since then f will.

Factoring b' into a square-free factor and a square factor $b' = b''u^2$, we see that f' is equivalent to $f'' = x_1^2 - ax_2^2 - b''x_3^2$, and we get to use our inductive hypothesis because $|b''| < |b|$:

$$|b''| = |\frac{b'}{u^2}| \leq |b'| = |\frac{t^2 - a}{b}| \leq |\frac{t^2}{b}| + |\frac{a}{b}| \leq \frac{(\frac{|b|}{2})^2}{|b|} + 1 = \frac{|b|}{4} + 1 < |b|$$

with the last inequality holding because $|b| \geq 2$. Thus f'' has $m'' = |a| + |b''| < m = |a| + |b|$; since f' and hence f'' represents 0 in each $\mathbb{Q}_v$, f'' represents 0 in $\mathbb{Q}$ by inductive hypothesis. Now, f'' represents 0 in $\mathbb{Q}$, so its equivalent f' represents 0, and we saw that happens exactly when f represents 0, so we're done.

$(n = 4)$

We rewrite our form in 4 variables as the difference of two forms in 2 variables, so that we can use results about forms with 2 variables that we have already shown. so let $f = ax_1^2 + bx_2^2 - (cx_3^2 + dx_4^2)$. We are assuming that f_v represents 0 for each valuation $v \in V$. So let (t_1, t_2, t_3, t_4) be a solution in $\mathbb{Q}_v$; then we have $at_1^2 + bt_2^2 = ct_3^2 + dt_4^2$.

We'd like to say that this means $ax_1^2 + bx_2^2$ and $cx_3^2 + dx_4^2$ both represent the same nonzero element of $\mathbb{Q}_v$; we argue as follows. Suppose we have $at_1^2 + bt_2^2 = ct_3^2 + dt_4^2 = 0$, then this means that at least one of these binary quadratic forms represents 0 (in the sense of having a *nontrivial* solution); say, without loss, that this is $ax_1^2 + bx_2^2$. Then by proposition Juniper, the form represents all elements of the field F; so, in particular, we choose any nonzero represented by $cx_3^2 + dx_4^2$ and get that it is also represented by $ax_1^2 + bx_2^2$. So we conclude that we can find in each complete field $\mathbb{Q}_v$ a nonzero quadruple (t_1, t_2, t_3, t_4) so that so that $at_1^2 + bt_2^2 = ct_3^2 + dt_4^2 = t_v \neq 0$.

Now we apply proposition Goldfish, which says that when $n = 2$, f_v represents a if and only if $(a, -d_v(f_v))_v = \varepsilon_v(f_v)$. Thus, since both binary forms represent t_v, we get $(t_v, -ab)_v = (a, b)_v$ and $(t_v, -cd)_v = (c, d)_v$ for every $v \in V$.

Now, we will use theorem 12 to find a single element $x \in \mathbb{Q}^*$ satisfying the above equalities — i.e., an $x \in \mathbb{Q}^*$ so that $(x, -ab)_v = (a, b)_v$ and $(x, -cd)_v = (c, d)_v$ for all $v \in V$. Explicitly, we have $a_1 = -ab$ and $a_2 = -cd$; $\delta_{1,v} = (a, b)_v$ and $\delta_{2,v} = (c, d)_v$; the product formula for Hilbert Symbols (theorem 11) guarantees that (i) and (ii) of theorem 12 are met; and the existence of the t_v above gives (iii); and then we get this single x with $(x, -ab)_v = (a, b)_v$ and $(x, -cd)_v = (c, d)_v$ for all $v \in V$.

Then, working backwards, there's some $x \in \mathbb{Q}^*$ so that for each valuation v we can find $(t_1, t_2, t_3, t_4) \in \mathbb{Q}_v^4$ so that $at_1^2 + bt_2^2 = ct_3^2 + dt_4^2 = x \neq 0$. Then, the form

$ax_1^2 + bx_2^2 - xz^2$ represents 0 (using our already chosen t_1, t_2, and letting $z = 1$) in $\mathbb{Q}_v$ for each v; we conclude by part (ii) above that this means $ax_1^2 + bx_2^2 - xz^2$ represents 0 in $\mathbb{Q}$; similarly, the form $cx_3^2 + dx_4^2 - xz^2$ represents 0 in each field $\mathbb{Q}_v$, so by part (ii) it represents 0 in $\mathbb{Q}$; finally, since both represent 0 in $\mathbb{Q}$, the binary forms $ax_1^2 + bx_2^2$ and $cx_3^2 + dx_4^2$ each represent x in $\mathbb{Q}$, so that their difference $ax_1^2 + bx_2^2 - (cx_3^2 + dx_4^2)$ represents 0 in $\mathbb{Q}$, as desired.

($n \geq 5$)

We will induct on n: write f as $f = h \dot{-} g$, where $h = a_1x_1^2 + a_2x_2^2$ and $g = a_3x_3^2 + \cdots + a_nx_n^2$.

We construct a finite set S of valuations V as follows: let $S = \{2\} \cup \{\infty\} \cup \{p \in \mathbb{Z} \mid v_p(a_i) \neq 0 \text{ for } 3 \leq i \leq n\}$. So let $v \in S$. f_v represents 0 by assumption, so we can argue as we did in the second paragraph of the case $n = 4$ that there is some *nonzero* element of $a_v \in \mathbb{Q}_v^*$ which is represented by both h and g in $\mathbb{Q}_v$; in other words, we can find $x_{v,i} \in \mathbb{Q}_v$ so that $a_v = h(x_{v,1}, x_{v,2}) = g(x_{v,3}, \ldots, x_{v,n})$. One can show that the squares of $\mathbb{Q}_v$ form an open set, so $\exists$ some real $\varepsilon \geq 0$ so that $|1 - u|_v \leq \varepsilon$ implies u is a square. We now use the Weak Approximation Theorem to find $x_1, x_2 \in \mathbb{Q}$ so that $a = h(x_1, x_2)$ has the property that $|a - a_v|_v \leq \varepsilon |a_v|_v$ for each $v \in S$. Then $|\frac{a}{a_v} - 1|_v \leq \varepsilon$, so that $\frac{a}{a_v}$ is a square in $\mathbb{Q}_v$ for each $v \in S$.

So now we consider the form $f_1 = g - az^2$. When $v \in S$, we know $\frac{a}{a_v}$ is a square by the above reasoning, so g represents a in $\mathbb{Q}_v$ because g represents a_v; so f_1 represents 0 in $\mathbb{Q}_v$. When $v \notin S$, then $v \neq 2$ or ∞, and by construction of S we get that the coefficients of g, $a_3, \ldots, a_n$ are all units in $\mathbb{Q}_v$; thus $d_v(g)$ is also a unit. $(-1, -d_v(g))_v = 1$ because both are units; this follows from theorem Paperclip; by the same theorem, each Hilbert symbol $(a_i, a_j)_v = 1$ because they're all units, so $\varepsilon_v(g) = \prod_{i<j}(a_i, a_j)_v = 1$; thus, by proposition Goldfish (iii) and (iv) (which we can apply because $v \neq \infty$), we get that g represents a (whether or not $a = -d_v(g)$). Thus, f_1 represents 0 in $\mathbb{Q}_v$ for all v; f_1 is a form in $n-1$ variables, so by inductive hypothesis f_1 represents 0 in $\mathbb{Q}$; so g represents a in $\mathbb{Q}$. Since what we did at first was finding the x_1, x_2 so that h represented a, we now can conclude that $f = h \dot{-} g$ represents 0.

And now the theorem is proved for all n...! □

An Application of Hasse-Minkowski

We conclude with a nice application of Hasse-Minkowski.

Theorem 15. Let $n \in \mathbb{Z}^+$. Then n can be written as the sum of three square integers $\iff$ n is not of the form $4^a(8b - 1)$.

Pf. We will need the following lemma, a result about 2-adic numbers:

Lemma 16. Let $x \in \mathbb{Q}_2^*$; write $x = 2^n u$, where $v_2(u) = 0$. Then x is a square $\iff$ n is even and $n \equiv 1 \pmod 8$.

Pf. See [Serre, p. 18]. □

From this lemma, we see that if there exist $a, b \in \mathbb{Z}$ so that $n = 4^a(8b-1)$, this means exactly that $-n = 2^{2a}(1-8b)$ is a square in $\mathbb{Q}_2$.

Now, it is easy to see that the Hasse-Minkowski theorem implies that for any $a \in \mathbb{Q}^*$, a is representable by a form in $\mathbb{Q}$ $\iff$ a is representable by that form in $\mathbb{Q}_v$ for each valuation v; this follows from applying the theorem to the form $f \dot{-} az^2$. So we will try to show that n is representable by the form $f = x_1^2 + x_2^2 + x_3^2$ in each $\mathbb{Q}_v$.

We need two more lemmas, the first of which I will prove, the second of which I will skip. The first establishes conditions on an arbitrary $a \in \mathbb{Q}^*$ so that it is representable by the form f; the second allows us to move from representations in $\mathbb{Q}$ to representations in $\mathbb{Z}$.

Lemma 17. Let $a \in \mathbb{Q}^*$. Then a can be written as the sum of three squares in $\mathbb{Q}$ $\iff$ a is positive and $-a$ not a square in $\mathbb{Q}_2$.

Pf. Looking at $\mathbb{Q}_\infty = \mathbb{R}$, a is the sum of three squares in $\mathbb{R}$ exactly when it's positive, so that takes care of the positivity bit.

Now we use proposition Goldfish to see when a is the sum of three squares in $\mathbb{Q}_p$ for each p: part (iii) tells us that a is representable by f so long as either $a \neq -d_p(f)$, or $a = -d_p(f)$ and $(-1, -d_p(f))_p = \varepsilon_p(f)$. For each p, the discriminant $d_p = 1$; also, it is easy to see that $(1,1)_p = 1$, so $\varepsilon_p(f) = 1$ as well. When $p > 2$, we have $(-1,-1)_p = 1$ by theorem Paperclip which said that when a and b are units and $p \neq 2$, $(a,b)_p = 1$. Thus, when $p > 2$, we get $(-1, -d_p)_p = (-1,-1)_p = 1 = \varepsilon_p$; so a is the sum of three squares in $\mathbb{Q}_p$ for each $p > 2$.

When $p = 2$, however, $(-1, -d_2)_2 = (-1,-1)_2 = -1$, because the form $x_1^2 + x_2^2 + x_3^2$ does not represents 0 in, say, $\mathbb{Z}/8\mathbb{Z}$ (we can assume by homogeneity of the form that the x_i are integers and at least one of them is odd, we must have 2 of them odd, and then the only possibilities mod 8 for $x_1^2 + x_2^2 + x_3^2$ are 2, 3, and 6). Thus, $(-1, -d_2)_2 = (-1,-1)_2 = -1 \neq \varepsilon_2$. Thus, for f to represent 0 in $\mathbb{Q}_2$, we must have $a \neq -d_p = -1$ in $\mathbb{Q}_2^*/\mathbb{Q}_2^{*2}$, which means exactly that $-a$ is not a square in $\mathbb{Q}_2$. □

Lemma 18. If f is a positive definite quadratic form with integer coefficients, and if for each $(x_1, \ldots, x_m) \in \mathbb{Q}^m$, there is $(y_1, \ldots, y_m)$ in $\mathbb{Z}^m$ with $f(x - y) < 1$, then for any $n \in \mathbb{Z}$ is representable by f over $\mathbb{Q}$, we get that n is also representable by f over $\mathbb{Z}$.

Pf. See [Serre, p. 46]. □

Now, we can prove the theorem. The form f is certainly positive definite, which just means it is always positive on nonzero triples (x_1, x_2, x_3). It also satisfies the

other hypothesis of lemma 18, for given a triplet of rationals (x_1, x_2, x_3), we can choose a triplet of integers (y_1, y_2, y_3) so that each component y_i is within $\frac{1}{2}$ of its corresponding components x_i; then we have $f(x-y) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 \leq \frac{3}{4} < 1$. Thus, if $n \in \mathbb{Z}$ is representable by f over $\mathbb{Q}$, it is representable over $\mathbb{Z}$. And, of course, n is representable by f over $\mathbb{Q}$ exactly when n is positive and $-n$ is not a square by lemma 17; and by lemma 15, this is exactly when n is not of the form $4^a(8b-1)$. So, the theorem holds. □

And we conclude with the four squares theorem:

Theorem 19. (Legendre) Every positive integer is the sum of four squares.

Pf. Let $n \in \mathbb{Z}^+$. Write $n = 4^a m$, where m is not divisible by 4. Then m is not congruent to 0 or 4 (mod 8); if $m \equiv 1, 2, 3, 5, 6 \pmod 8$, then n is the sum of three squares by theorem 15, and thus the sum of four. If $m \equiv 7 \pmod 8$, then $m - 1 \equiv 6 \pmod 8$, so $m-1$ is the sum of three squares: $m - 1 = x_1^2 + x_2^2 + x_3^2$; so m is the sum of four squares: $m = x_1^2 + x_2^2 + x_3^2 + 1$, so n is the sum of four squares too: $n = (2^a x_1)^2 + (2^a x_2)^2 + (2^a x_3)^2 + (2^a)^2$. □

Bibliography

Artin, Michael. *Algebra.* Upper Saddle River: Prentice Hall, 1991.

Serre, Jean-Pierre. *A Course In Arithmetic.* New York: Springer-Verlag, 1973.

Stein, William. Lectures and lecture notes from Math 129, Harvard, Spring 2004.

Notes

As I said at the beginning, I followed Serre's book extremely closely, going through parts of chapters II, III, and IV (pp. 11-47). My other listed sources were consulted for algebraic and number theoretic background. A draft was read over by Warren Sinnott, who pointed out errors and helped me clean up some of the proofs. My work was primarily expanding Serre's write-up on the subject.