SOLUBILITY OF SYSTEMS OF QUADRATIC FORMS

GREG MARTIN

It has been known since the last century that a single quadratic form in at least five variables has a nontrivial zero in any $p$-adic field, but the analogous question for systems of quadratic forms remains unanswered. It is plausible that the number of variables required for solubility of a system of quadratic forms simply is proportional to the number of forms; however, the best result to date, from an elementary argument of Leep [6], is that the number of variables needed is at most a quadratic function of the number of forms. The purpose of this paper is to show how these elementary arguments can be used, in a certain class of fields including the $p$-adic fields, to refine the upper bound for the number of variables needed to guarantee solubility of systems of quadratic forms. This result partially addresses Problem 6 of Lewis’ survey article [7] on Diophantine problems.

By a nontrivial zero of a system of forms $f_1, \ldots, f_t \in F[x_1, \ldots, x_n]$, we mean a nonzero element $a$ of $F^n$ such that $f_j(a) \equiv 0$ simultaneously for $1 \leq j \leq t$. We let $u_F(t)$ denote the supremum of those positive integers $n$ for which there exist $t$ quadratic forms over $F$ in $n$ variables with no nontrivial zero. In other words, assuming $u_F(t) < \infty$, any set of $t$ quadratic forms in $F[x_1, \ldots, x_n]$, with $n > u_F(t)$, will have a nontrivial zero (equivalently, a projective zero, since the forms are homogeneous), while this property does not hold for $n = u_F(t)$. We may now state our main theorem.

**Theorem 1.** Let $F$ be a field, and suppose that for some positive integer $m$, we have

$$u_F(m) = mu_F(1).$$  

(1)

Then

$$u_F(t) \leq \frac{1}{4}(t(t - m + 2) + \tau(m - \tau))u_F(1),$$  

(2)

where $\tau$ is the unique integer satisfying $1 \leq \tau \leq m$ and $\tau \equiv t \pmod{m}$.

We remark that for any $1 \leq r \leq t$, we always have the lower bound

$$u_F(t) \geq u_F(r) + u_F(t - r),$$  

(3)

for if $f(x_1, \ldots, x_{u_F(r)})$ ($1 \leq i \leq r$) and $g(y_1, \ldots, y_{u_F(t-r)})$ ($1 \leq j \leq t-r$) are systems of quadratic forms with no nontrivial zeros, then we can combine the two systems and the two sets of variables to yield a system of $t$ quadratic forms in $u_F(r) + u_F(t-r)$ variables with no nontrivial zeros. In particular, equation (3) readily implies that for all $t \geq 1$, we have

$$u_F(t) \geq tu_F(1).$$  

(4)

Thus the hypothesis (1) of Theorem 1 is a natural one, representing the best-possible situation for systems of $m$ quadratic forms.

\[Received 23 April 1996.\]

1991 Mathematics Subject Classification 11D72.

In fact, if $F$ is a local field (a finite extension either of $\mathbb{Q}_p$ for some prime $p$, or of $k((T))$ for some finite field $k$), Hasse [4] has shown that $u_\mathbb{F}(1) = 4$ (see Lam [5] for an exposition), and Demjanov [3] has shown that $u_\mathbb{F}(2) = 8$ (a simpler proof has been provided by Birch, Lewis and Murphy [2]). Thus the following corollary of Theorem 1 is immediate.

**Corollary 1.1.** Let $F$ be a local field. Then

$$u_\mathbb{F}(t) \leq \begin{cases} 2t^2 + 2, & t \text{ odd}, \\ 2t^2, & t \text{ even}. \end{cases}$$

It has also been shown by Birch and Lewis [1], with a correction and refinement by Schuur [8], that whenever $p \geq 11$, we have $u_{\mathbb{Q}_p}(3) = 12$. Therefore we can again apply Theorem 1 to obtain the following corollary, which is superior to Corollary 1.1 for these primes.

**Corollary 1.2.** Let $p \geq 11$ be prime. Then

$$u_{\mathbb{Q}_p}(t) \leq \begin{cases} 2t^2 - 2t + 4, & t \not\equiv 0 \pmod{3}, \\ 2t^2 - 2t, & t \equiv 0 \pmod{3}. \end{cases}$$

The methods employed in this paper are a modest refinement of those of Leep [6], who has shown that $u_F(t) \leq \frac{1}{2}(t+1)u_F(1)$ for arbitrary fields $F$, and also that $u_{\mathbb{Q}_p}(t) \leq 2t^2 + 2t - 4$ (for $t \geq 2$) for every prime $p$. Because the argument is brief and completely elementary, we may provide an essentially self-contained proof of Theorem 1.

It is a pleasure to thank Trevor Wooley and Hugh Montgomery for their suggestions on improving this paper and for their guidance in general. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship.

## 1. Preliminary lemmas

Let $u^{d^u}(t)$ denote the supremum of those positive integers $n$ for which there exist $t$ quadratic forms over $F$ in $n$ variables whose set of solutions contains no $(d + 1)$-dimensional subspace of $F^n$. In other words, any set of $t$ quadratic forms in $F[x_1, \ldots, x_n]$, with $n > u^{d^u}(t)$, will have a $(d + 1)$-dimensional subspace of simultaneous zeros (or, equivalently, a $d$-dimensional subspace of projective zeros), while this property does not hold for $n \leq u^{d^u}(t)$. For instance, we have $u^{d^u}(t) = u_d(t)$.

The following two lemmas can be found in Leep [6]; we provide proofs for the sake of completeness.

**Lemma 2.** For any field $F$, and for all positive integers $k < t$, we have

$$u_F(t) \leq u_F^{u_F(k)}(t-k).$$

**Proof.** Let $n > u^{u_F(k)}(t-k)$, and let $f_1, \ldots, f_t$ be quadratic forms over $F$ in $n$ variables. To establish the lemma, it suffices to show that these forms have a nontrivial zero in $F^n$. By the definition of $u^{u_F(k)}(t-k)$, the system $f_1, \ldots, f_{t-k}$ of $t-k$ quadratic forms has a $(u_F(k)+1)$-dimensional subspace $S$ of zeros. By parametrizing
S with variables \(y_1, \ldots, y_{n+1}\), we may consider the restrictions of the forms \(f_1, \ldots, f_t\) to \(S\) as quadratic forms in \(u_\rho(k)+1\) variables. Now by the definition of \(u_\rho(k)\), these forms have a nontrivial zero in \(S\), and so the forms \(f_1, \ldots, f_t\) have a nontrivial zero in \(F^n\).

**Lemma 3.** For any field \(F\), and for all positive integers \(t\) and \(d\), we have

\[
u^{(d^t)}_\rho(t) \leq u^{(d-1)}_\rho(t) + t + 1.
\]

**Proof.** Let \(n > u^{(d-1)}_\rho(t) + t + 1\), and let \(f_1, \ldots, f_t\) be quadratic forms over \(F\) in \(n\) variables. To establish the lemma, it suffices to show that \(F^n\) contains a \((d+1)\)-dimensional subspace of zeros for these forms. Since \(n > u^{(d-1)}_\rho(t) \geq u_\rho(t)\), we can certainly find a nontrivial zero for the forms \(f_1, \ldots, f_t\), which generates a 1-dimensional subspace \(T\) of zeros of these forms. By making a linear change of variables, we may assume that \(T\) is spanned by the vector \((0, \ldots, 0, 1)\). For each \(1 \leq j \leq t\), we may write

\[
 f_j(x_1, \ldots, x_n) = x_n^2 f_j(0, \ldots, 0, 1) + x_n L_j(x_1, \ldots, x_{n-1}) + Q_j(x_1, \ldots, x_{n-1}), \tag{6}
\]

where the \(L_j\) and \(Q_j\) are linear and quadratic forms, respectively, in \(n-1\) variables (here we are identifying \(T^1\) with \(F^{n-1}\)). But we are working under the assumption that each \(f_j(0, \ldots, 0, 1)\) equals 0, and elementary linear algebra allows us to find a subspace \(S\) of \(F^{n-1}\) of codimension \(t\) on which the \(t\) linear forms \(L_1, \ldots, L_t\) all vanish identically. Again we parametrize \(S\) by variables \(y_1, \ldots, y_{n-t-1}\) and consider the restrictions of the forms \(Q_1, \ldots, Q_t\) to \(S\) as quadratic forms in \(n-1+t > u^{(d-1)}_\rho(t)\) variables. By the definition of \(u^{(d-1)}_\rho(t)\), we may find a \(d\)-dimensional subspace \(U\) of \(S\) consisting of zeros of the forms \(Q_1, \ldots, Q_t\). We now see from (6) that \(U \oplus T\) is a \((d+1)\)-dimensional subspace of zeros of the original forms \(f_1, \ldots, f_t\).

2. **Proof of Theorem 1**

We begin by making some remarks that hold in any field \(F\), without the hypothesis (1) of Theorem 1. Using Lemma 2 together with several applications of Lemma 3, we see that

\[
u_\rho(t) \leq u^{(d^t)}_\rho(t) \leq u_\rho(t-k) + (t-k+1) u_\rho(k).
\]

Therefore, for any positive integer \(r\) such that \(rk < t\), we have

\[
u_\rho(t) \leq u_\rho(t-rk) + \sum_{i=1}^{r} (t-ik+1) u_\rho(k).
\]

(7)

Thus we have established a bound for \(u_\rho(t)\) in terms of \(u_\rho(j)\) for small values of \(j\). In fact, this is precisely the approach in Leep [6], with the choices \(k = 1\) and \(r = t-1\), so that the final bound is in terms of \(u_\rho(1)\) alone. One can also choose \(r = t-2\) and obtain a bound for \(u_\rho(t)\) in terms of \(u_\rho(1)\) and \(u_\rho(2)\), which will be better if the value of \(u_\rho(2)\) is known to be small.

However, for fields \(F\) that satisfy the hypothesis (1) for some positive integer \(m\), it turns out to be more beneficial to take \(k = m\) in the bound (7). We choose \(r\) to make \(t-rk\) as small as possible while still positive: if we let \(\tau\) be the integer satisfying \(1 \leq \tau \leq m\) and \(\tau \equiv t \pmod{m}\), then \(r = (t-\tau)/m\). With these choices, equation (7) becomes

\[
u_\rho(t) \leq u_\rho(\tau) + \frac{t-\tau}{2m}(t-m+\tau+2) u_\rho(m).
\]

(8)
We claim that \( u_F(m) = mu_F(1) \) forces \( u_F(\tau) = \tau u_F(1) \) as well, since by the lower bounds (3) and (4), we have

\[
\tau u_F(1) \leq u_F(\tau) \leq u_F(m) - u_F(m - \tau) \\
\leq mu_F(1) - (m - \tau) u_F(1) = \tau u_F(1).
\]

Substituting these expressions in the bound (8) gives us

\[
u_F(t) \leq \tau u_F(1) + \frac{t - \tau}{2m} (t - m + \tau + 2) mu_F(1),
\]

which is the same as the bound (2). This establishes the theorem.

References


Department of Mathematics
University of Michigan
Ann Arbor, MI 48109-1003
USA