Pairs of homogeneous additive equations

Michael P. Knapp

Abstract

In 1966, Davenport & Lewis published their paper *Notes on congruences III*, in which they proved that under some mild conditions a system of two additive forms of equal degrees must have a nonsingular simultaneous zero modulo any prime number. In their paper, they ask whether the theorem is true in general finite fields, and point out that one of their key lemmas is no longer true in this situation. In this paper we answer their question in the affirmative, proving that under the same conditions a system of two additive forms over any finite field must have a nonsingular simultaneous zero. We then apply this result to obtain an upper bound on the number of variables required to ensure that a system of two additive forms of equal degrees has a nontrivial zero in a $\mathfrak{p}$-adic field.

1. Introduction

A special case of a well-known theorem due to Chevalley [Che35] states that if $f_1(x_1, \dots, x_s)$ and $f_2(x_1, \dots, x_s)$ are polynomials defined over a finite field $\mathbb{F}_q$ which have degree k and no constant terms, and if $s > 2k$, then the system of equations

$$f_1(\mathbf{x}) = f_2(\mathbf{x}) = 0$$

has a nontrivial solution in $\mathbb{F}_q$. That is, there exists a solution with at least one variable nonzero. Sometimes it is necessary, however, to know that a system of equations over a finite field has a *nonsingular* solution. For example, to show that a system over a $\mathfrak{p}$-adic field has a solution, one

---

2000 Mathematics Subject Classification 11D72, (11T06, 11E95)

Work supported by NSF grant DMS-0344082.

often reduces modulo the maximal ideal to a finite field. If the resulting system has a nonsingular solution, then Hensel's Lemma guarantees that there exists a solution over the $\mathfrak{p}$-adic field.

A theorem of Davenport & Lewis [DL66] gives conditions under which a system of two diagonal forms has a nonsingular zero over a finite field $\mathbb{F}_p$, where p is prime. Specifically, they consider the system

$$\begin{aligned} a_1x_1^k + \cdots + a_sx_s^k &= 0 \\ b_1x_1^k + \cdots + b_sx_s^k &= 0 \end{aligned} \tag{1}$$

and show that under certain conditions the system possesses a nonsingular solution. That is, there is a solution such that the expression

$$k^2x_i^{k-1}x_j^{k-1}(a_ib_j - a_jb_i) \tag{2}$$

is nonzero (in $\mathbb{F}_p$) for some variables x_i and x_j.

In [DL66], Davenport & Lewis ask whether their theorem holds over general finite fields. They point out that one of their lemmas is no longer true in this case, and say that they were unable to determine whether the theorem should be true. The main purpose of the present paper is to answer this 40-year-old question in the affirmative. In particular, we prove the following theorem.

Theorem 1. *Let p be a prime number and let $q = p^f$ for some positive integer f. Consider the system (1) where the coefficients are in the finite field $\mathbb{F}_q$ and the coefficients a_i and b_i are never both zero. Suppose that $p \nmid k$, that $s \geqslant 2k+1$, and that any nontrivial linear combination of the two forms contains at least $k+1$ variables with nonzero coefficients. Then the system (1) has a nonsingular solution.*

If $f = 1$, so that $q = p$, then Theorem 1 reduces to the Davenport-Lewis theorem. The condition $p \nmid k$ is necessary, as otherwise the expression (2) would be zero for any pair x_i, x_j since $\mathbb{F}_q$ has characteristic p. The condition $s \geqslant 2k+1$ is included to guarantee (by Chevalley's theorem) that (1) has a nontrivial solution. The final condition guarantees that every nontrivial linear combination of the forms has a nontrivial zero. As Davenport & Lewis show in [DL66], if any nontrivial linear combination of the forms has only the trivial zero then the system (1) has only singular solutions.

It is worth mentioning that these conditions on numbers of variables cannot be replaced by the corresponding conditions on zeros of forms, even over fields with a prime number of elements. For example, consider the system

$$\begin{aligned} F_1(\mathbf{x}) = \quad x_1^6 - x_2^6 + x_3^6 + x_4^6 + x_5^6 &= 0 \\ F_2(\mathbf{x}) = \quad x_3^6 + x_4^6 - 2x_5^6 &= 0 \end{aligned}$$

over $\mathbb{F}_7$. Since the only values of x^6 in $\mathbb{F}_7$ are 0 and 1, it is not too hard to check that all of the nontrivial solutions of this system have the form $\mathbf{x} = (x_1, x_2, 0, 0, 0)$ where x_1 and x_2 are both nonzero, and that these solutions are all singular. Moreover, if one considers any nontrivial linear combination $mF_1 + nF_2$, one finds that if $m \neq 0$ then $\mathbf{x} = (1, 1, 0, 0, 0)$ is a nontrivial zero of this form and if $m = 0$ then $\mathbf{x} = (0, 0, 1, 1, 1)$ is a nontrivial zero.

As an application of Theorem 1, standard methods allow us to prove the following theorem about zeros of forms over $\mathfrak{p}$-adic fields.

Theorem 2. *Let p be a prime number and let $\mathbb{L}$ be any (possibly infinite) algebraic extension of the field $\mathbb{Q}_p$. Suppose that $p \nmid k$ and consider the system (1) where the coefficients may be any elements of $\mathbb{L}$. If $s \geqslant 2k^2+1$, then the system has a nontrivial solution.*

While the condition that $p \nmid k$ is annoying, it should be noted that obtaining bounds in the case where $p | k$ is a quite difficult problem. One suspects that the bound $s \geqslant 2k^2+1$ will also hold in this situation, but this is not even known when $\mathbb{L}$ is $\mathbb{Q}_p$ itself. When $\mathbb{L}$ is an extension of $\mathbb{Q}_p$, the problem is even harder due to the possibility of ramification.

The proof of Theorem 1 is split into two cases depending on the field generated by the k^{th} powers in $\mathbb{F}_q$ under addition. If the k^{th} powers generate $\mathbb{F}_q$ under addition, then the ideas of Davenport & Lewis suffice to prove the theorem, although one of their key lemmas seems to require a much different proof over general finite fields. If the k^{th} powers generate a proper subfield of $\mathbb{F}_q$, then we replace the system (1) by another system of equations with coefficients in this subfield. This increases the number of equations we need to solve, but it turns out that we are able to simultaneously lower the degrees of the equations. Then the aforementioned result of Chevalley suffices to prove the theorem. Although it is not needed in our proof, we point out that Bhaskaran [Bha66, Theorem G] (see [Bha71] for a corrected proof) has characterized the exponents k which fall into each case.

Once Theorem 1 is proven, Theorem 2 can be proved using techniques which are now fairly standard. First we apply a normalization procedure which allows us to restrict to forms which have certain desirable properties. Then, as indicated above, we reduce the forms modulo the maximal ideal of the field $\mathbb{L}$ and find a nonsingular solution of the system in the residue field. Finally, we use a version of Hensel's Lemma to lift this solution to a solution in $\mathbb{L}$.

2. Notation and Preliminaries

First we fix some notation for the proof of Theorem 1. We define p, q, f, k and s as in the statement of the theorem. Note that since the set of k^{th} powers in $\mathbb{F}_q$ is the same as the set of $(k, q-1)$-st powers, we may assume without loss of generality that $k|(q-1)$, and we do so throughout the proof. Let F be the set of nonzero k^{th} powers in $\mathbb{F}_q$, and let $\mathbb{K}$ be the subfield of $\mathbb{F}_q$ generated by F under addition. Define r so that $\mathbb{K}$ contains p^r elements, and note that $r \geqslant 1$ and that $\mathbb{F}_q$ is a vector space over $\mathbb{K}$ of dimension $d = f/r$.

In both parts of the proof, we use the notion of colored variables developed by Brüdern & Godinho in [BG02]. Consider the vectors

$$\mathbf{e}_\infty = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{e}_v = \begin{pmatrix} v \\ 1 \end{pmatrix}, \quad v \in \mathbb{F}_q.$$

Note that for any variable x_i, the vector of coefficients of x_i can be written as

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} = a\mathbf{e}_v$$

for some $a \in \mathbb{F}_q$ and $v \in \mathbb{F}_q \cup \{\infty\}$. In this case we call v the *color* of the variable x_i. We say that a color v is *zero-representing* if the system (1) has a nontrivial solution in which only variables of color v are nonzero.

The utility of this definition is clear upon considering the nonsingularity condition. One can see that there are variables x_i and x_j such that (2) is nonzero if and only if $p \nmid k$ and the matrix

$$\begin{pmatrix} a_1x_1 & \cdots & a_sx_s \\ b_1x_1 & \cdots & b_sx_s \end{pmatrix}$$

has rank 2. Equivalently, a solution of (1) is nonsingular if and only if it contains at least two nonzero variables of different colors. This observation leads immediately to the following lemma, which (in the context of congruences modulo powers of p) is Lemma 4.1 of [BG02].

Lemma 1. *If two colors are zero-representing, then the system (1) has a nonsingular solution.*

Another consequence of this observation implicit in the work of Brüdern & Godinho is that we may assume that exactly one color is zero-representing. If two or more colors represent zero, then

Lemma 1 guarantees that the system (1) has a nonsingular solution. If no colors represent zero, then Chevalley's theorem implies that (1) has a nontrivial solution. But since no color represents zero, this solution must have nonzero variables from at least two different colors and so the solution is nonsingular. Therefore we may assume that precisely one color represents zero. Moreover, by taking linear combinations of the forms if necessary we may assume if we wish that ∞ is the zero-representing color.

3. The Proof of Theorem 1 When $\mathbb{K} = \mathbb{F}_q$

If the k^{th} powers of $\mathbb{F}_q$ actually generate $\mathbb{F}_q$ under addition, then our proof proceeds along the lines of the one given by Davenport & Lewis in [DL66]. In fact, our two preliminary lemmata are direct extensions of theirs, although our proof of Lemma 2 is completely different since theirs does not seem to extend to general finite fields. Our proof of Theorem 1 is also slightly different than theirs in order to highlight the use of colored variables, although it retains the same flavor. We begin with the lemmata.

Lemma 2. *Consider the additive form*

$$G(\mathbf{x}) = a_1 x_1^k + \cdots + a_n x_n^k$$

over $\mathbb{F}_q$, where $k|(q-1)$ and all of the coefficients are nonzero. Suppose that the k^{th} powers in $\mathbb{F}_q$ generate $\mathbb{F}_q$ under addition. Then the number of distinct nonzero elements of $\mathbb{F}_q$ represented by G is at least $\min\{\frac{n(q-1)}{k}, q-1\}$.

Proof. This proof was inspired by the proof of Theorem 1 of [Tor38]. Note that since the elements of the set of k^{th} powers generate $\mathbb{F}_q$ under addition, they are a spanning set for $\mathbb{F}_q$ when $\mathbb{F}_q$ is considered as a vector space over $\mathbb{F}_p$. Similarly, if $a \in \mathbb{F}_q$ and $a \neq 0$, then the set $\{ax^k : x \in \mathbb{F}_q\}$ spans $\mathbb{F}_q$. Hence each such set contains a basis for $\mathbb{F}_q$ as a vector space over $\mathbb{F}_p$.

Now, if $1 \leqslant i \leqslant n$, let

$$K_i = \{a_1 x_1^k + \cdots + a_i x_i^k : x_1, \ldots, x_i \in \mathbb{F}_q\}$$

and note that $K_i \subseteq K_j$ whenever $i \leqslant j$. Our first goal is to show that if $K_i = K_{i+1}$ for some i, then we must have $K_i = \mathbb{F}_q$. Since the set $\{a_{i+1} x_{i+1}^k : x_{i+1} \in \mathbb{F}_q\}$ contains an $\mathbb{F}_p$-basis for $\mathbb{F}_q$, the condition $K_i = K_{i+1}$ implies that adding any basis element to a member of K_i yields another member of K_i. Therefore adding any finite sum of basis elements to a member of K_i yields another member of K_i. Since any element of $\mathbb{F}_q$ is a finite sum of basis elements, we see that if one takes any element of K_i and adds any element of $\mathbb{F}_q$ to it, the sum is still in K_i. But this can only happen if $K_i = \mathbb{F}_q$.

Now, if $K_i = K_{i+1}$ for any i then since $K_i = \mathbb{F}_q$ the form G represents every nonzero element of $\mathbb{F}_q$, and so the lemma is true. In light of this, the lemma will follow if we can show first that K_1 contains both zero and at least $(q-1)/k$ nonzero elements and also that if $K_i \neq K_{i+1}$, then we have

$$|K_{i+1}| \geqslant |K_i| + \frac{q-1}{k}.$$

The first statement is trivial since there are $(q-1)/k$ nonzero k^{th} powers in $\mathbb{F}_q$, each giving rise to a different nonzero element of K_1 and since setting $x_1 = 0$ gives $0 \in K_1$.

To prove the second statement, suppose that $x \in K_{i+1} - K_i$. Since $0 \in K_1$, we know that $x \neq 0$. We now show that if $y \in \mathbb{F}_q$ and $y \neq 0$ then $y^k x \in K_{i+1} - K_i$. Indeed, since $x \in K_{i+1}$, we may write $x = a_1 x_1^k + \cdots + a_{i+1} x_{i+1}^k$ for some $x_1, \ldots, x_{i+1}$, whence we have $y^k x = a_1 (y x_1)^k + \cdots + a_{i+1} (y x_{i+1})^k \in K_{i+1}$. Now suppose that we had $y^k x \in K_i$. Then we could write $y^k x = a_1 x_1^k + \cdots + a_i x_i^k$ for some (different) $x_1, \ldots, x_i$. But since $y \neq 0$, we would then have $x = a_1 (y^{-1} x_1)^k + \cdots + a_i (y^{-1} x_i)^k \in K_i$, a contradiction. Again since there are $(q-1)/k$ nonzero values of y^k in $\mathbb{F}_q$, this provides $(q-1)/k$

elements of K_{i+1} which are not in K_i. The second statement, and hence the lemma, follow. $\Box$

LEMMA 3. *Let $q = p^f$ and suppose that $k|(q-1)$. Let G, H be forms defined over $\mathbb{F}_q$ of degree k in s variables, where $s > k$. Let $\gamma_1, \ldots, \gamma_\mu$ be distinct nonzero elements of $\mathbb{F}_q$, where*

$$\mu > (q-1)(2k-s)/k.$$

Then there exist $y_1, \ldots, y_s$ not all zero such that

$$\begin{aligned} G(\mathbf{y}) &= 0 \text{ or } \gamma_i \\ H(\mathbf{y}) &= 0 \end{aligned}$$

for some i.

This is the generalization to arbitrary finite fields of Lemma 2 of [DL66]. This time, the proof given by Davenport & Lewis generalizes almost without change, and so we will not give a proof here. The only nontrivial alteration involves their appeal to a "well-known principle" (see page 56 of [Dav82]) to guarantee that there is a choice of variables such that a particular auxiliary polynomial is not equal to zero. In the more general setting, this principle should be replaced by a theorem of Alon (Theorem 1.2 of [Alo99]) which generalizes the principle to polynomials over arbitrary fields.

Now we can prove the theorem when $\mathbb{K} = \mathbb{F}_q$. As mentioned earlier, we may assume that exactly one color of variables represents zero, and without loss of generality we may assume that this color is ∞. Suppose that there are n variables with this color. Then the system (1) looks like

$$\begin{aligned} a_1x_1^k + \cdots + a_nx_n^k + a_{n+1}x_{n+1}^k + \cdots + a_sx_s^k &= 0 \\ b_{n+1}x_{n+1}^k + \cdots + b_sx_s^k &= 0, \end{aligned}$$

where $b_{n+1}, \ldots, b_s \neq 0$. The hypotheses of the theorem imply that $s \geqslant 2k+1$ and $s - n \geqslant k+1$.

Since we are assuming that $k|(q-1)$, Lemma 2 tells us that the expression $a_1x_1^k + \cdots + a_nx_n^k$

takes on at least $\min\{n(q-1)/k, q-1\}$ nonzero values in $\mathbb{F}_q$. Suppose first that $n \geqslant k$, so we know that the expression takes on $q-1$ nonzero values. Since the color ∞ is zero-representing, this expression nontrivially represents every element of $\mathbb{F}_q$. Since $s-n \geqslant k+1$, there exist by Chevalley's theorem elements $y_{n+1}, \dots, y_s \in \mathbb{F}_q$ not all zero such that $b_{n+1}y_{n+1}^k + \cdots + b_s y_s^k = 0$. Suppose that $a_{n+1}y_{n+1}^k + \cdots + a_s y_s^k = A$. Then we can find $y_1, \dots, y_n$ not all zero such that $a_1 y_1^k + \cdots + a_n y_n^k = -A$. Then $\mathbf{x} = \mathbf{y}$ yields a solution of the system (1). Since this solution involves nonzero variables of different colors (one variable from $x_1, \dots, x_n$ and at least one from $x_{n+1}, \dots, x_s$), this solution is nonsingular.

Now suppose that $n < k$, so that we can only guarantee that the expression $a_1 x_1^k + \cdots + a_n x_n^k$ takes on at least $n(q-1)/k$ nonzero values. Call these values $-\gamma_1, \dots, -\gamma_\mu$, where $\mu \geqslant n(q-1)/k$. Then we want to find $y_{n+1}, \dots, y_s$ not all zero such that we have

$$\begin{aligned} &a_{n+1}y_{n+1}^k + \cdots + a_s y_s^k = 0 \text{ or } \gamma_i \text{ for some } i \\ &b_{n+1}y_{n+1}^k + \cdots + b_s y_s^k = 0. \end{aligned}$$

Now, note that the number of variables here is $s-n > k$ and that since $s > 2k$ we have

$$\begin{aligned} \mu &\geqslant \frac{n(q-1)}{k} \\ &= \frac{(q-1)}{k}(2k-(s-n)) + \frac{(q-1)}{k}(s-2k) \\ &> \frac{(q-1)}{k}(2k-(s-n)). \end{aligned}$$

Hence such values of $y_{n+1}, \dots, y_s$ exist by Lemma 3. Suppose that these values give $a_{n+1}y_{n+1}^k + \cdots + a_s y_s^k = \gamma$. Then since either $\gamma = 0$ or $\gamma = \gamma_i$ for some i, we can choose $y_1, \dots, y_n$ not all zero such that $a_1 y_1^k + \cdots + a_n y_n^k = -\gamma$. Thus $\mathbf{x} = \mathbf{y}$ is a nontrivial solution of the system (1). As before, this solution contains nonzero variables of different colors (one from $x_1, \dots, x_n$ and at least one from $x_{n+1}, \dots, x_s$), and so this solution is nonsingular. This completes the proof of Theorem 1 in the case where $\mathbb{K} = \mathbb{F}_q$. $\Box$

4. The Proof of Theorem 1 When $\mathbb{K} \neq \mathbb{F}_q$

If the k^{th} powers in $\mathbb{F}_q$ only generate a proper subfield, then the strategy used in the previous section no longer works. In particular, the assumption $\mathbb{K} = \mathbb{F}_q$ is critical in the proof of Lemma 2, and Davenport & Lewis [DL66] give an example showing that the lemma is false without this assumption. Thus we need a different strategy to deal with this situation. We begin with a lemma.

Lemma 4. *Suppose that $k|(q-1)$ and the k^{th} powers in $\mathbb{F}_q$ generate the subfield $\mathbb{K} = \mathbb{F}_{p^r}$ under addition. Then the set of k^{th} powers in $\mathbb{F}_q$ is the same as the set of m^{th} powers in $\mathbb{F}_{p^r}$, where*

$$m = \frac{k(p^r - 1)}{q - 1}.$$

Proof. Note that 0 is both a k^{th} power in $\mathbb{F}_q$ and an m^{th} power in $\mathbb{F}_{p^r}$. Since $k|(q-1)$, the set F of nonzero k^{th} powers in $\mathbb{F}_q$ forms a multiplicative subgroup of $\mathbb{F}_q^*$ of cardinality $(q-1)/k$. This set is also a multiplicative subgroup of $\mathbb{K}^*$, and so $(q-1)/k$ must evenly divide $|\mathbb{K}^*| = p^r - 1$. In other words, we have

$$\frac{q-1}{k} = \frac{p^r - 1}{m}$$

for some integer m. Now, since $\mathbb{K}^*$ is a cyclic group of order $p^r - 1$ it has a unique subgroup of order $(p^r - 1)/m$, and we have just seen that the set F forms a subgroup of $\mathbb{K}^*$ of this order. Also, since $m|(p^r - 1)$ we know that there are $(p^r - 1)/m$ nonzero m^{th} powers in $\mathbb{K}^*$, and these form a subgroup. Hence these subgroups must be the same. The lemma follows. $\Box$

Our next lemma provides the meat of the proof. It shows that when $\mathbb{K} \neq \mathbb{F}_q$, only a small number of variables are necessary to guarantee that a system such as (1) has a nontrivial (*not* a nonsingular) solution.

Lemma 5. *Suppose that $k|(q-1)$, that the k^{th} powers in $\mathbb{F}_q$ generate under addition the subfield $\mathbb{K}$ containing p^r elements and that $q \neq p^r$. If $q \neq 4$, then the system (1) has a nontrivial solution provided only that $s \geqslant k+1$. If $q = 4$, then the system (1) has a nontrivial solution provided only that $s \geqslant k+2$.*

Proof. Write $d = [\mathbb{F}_q : \mathbb{K}] = f/r$. Then we have $\mathbb{F}_q = \mathbb{K}(\alpha)$ for some $\alpha \in \mathbb{F}_q$ having degree d over $\mathbb{K}$. So any element $a_i \in \mathbb{F}_q$ can be written as

$$a_i = a_{i,0} + a_{i,1}\alpha + \cdots + a_{i,d-1}\alpha^{d-1}, \tag{3}$$

where $a_{i,0}, \ldots, a_{i,d-1} \in \mathbb{K}$. Then we can replace an equation such as

$$a_1 x_1^k + \cdots + a_s x_s^k = 0 \tag{4}$$

with coefficients in $\mathbb{F}_q$ by a system of equations

$$\begin{aligned} a_{1,0}x_1^k + \cdots + a_{s,0}x_s^k &= 0 \\ \vdots \qquad\qquad &\quad \vdots \\ a_{1,d-1}x_1^k + \cdots + a_{s,d-1}x_s^k &= 0, \end{aligned} \tag{5}$$

where the coefficients are now in $\mathbb{K}$, noting that if the system (5) has a nontrivial solution then so does the equation (4).

Similarly, the system (1) can be replaced by a system

$$\begin{aligned} a_{1,i}x_1^k + \cdots + a_{s,i}x_s^k = 0 \\ b_{1,i}x_1^k + \cdots + b_{s,i}x_s^k = 0 \end{aligned} \qquad (0 \leqslant i \leqslant d-1) \tag{6}$$

of $2d$ diagonal equations of degree k with coefficients in $\mathbb{K}$. As above, we note that if the system (6) has a nontrivial solution then so does the system (1). Since the set of k^{th} powers in $\mathbb{F}_q$ equals the set of m^{th} powers in $\mathbb{K}$ by Lemma 4, replacing all occurrences of x_i^k in (6) by x_i^m yields a new system of $2d$ equations of degree m which has nontrivial solutions over $\mathbb{K}$ if and only if the system (6) does. Thus, in order to solve (1) over $\mathbb{F}_q$, it suffices to solve this system of degree m equations with all of the variables in $\mathbb{K}$.

By Chevalley's theorem, we may do this whenever we have

$$s > 2dm = 2d\frac{k(p^r-1)}{q-1}.$$

So we need to show that

$$k+1 > 2d\frac{k(p^r-1)}{q-1} \tag{7}$$

when $q \neq 4$ and that $k+2$ is greater than the right-hand side of (7) when $q = 4$. Suppose first that $2d \leqslant p^{d-1}$. Then we have

$$\begin{aligned} 2d\frac{k(p^r-1)}{q-1} &\leqslant kp^{d-1}\frac{p^r-1}{p^f-1} \\ &= kp^{d-1}\frac{1}{p^{f-r}+p^{f-2r}+\cdots+p^r+1} \\ &\leqslant \frac{kp^{d-1}}{p^{f-r}} \\ &\leqslant k \\ &< k+1, \end{aligned}$$

as desired, where the second to last inequality holds because $f = dr$. So the lemma holds whenever $2d \leqslant p^{d-1}$.

Now, since $\mathbb{K}$ is a proper subfield of $\mathbb{F}_q$, we must have $d \geqslant 2$. It is not too hard to check that when $d \geqslant 2$ the only pairs (p,d) for which the condition $2d \leqslant p^{d-1}$ is false are $(2,3)$, $(3,2)$ and $(2,2)$. If $(p,d) = (2,3)$, then we can calculate that

$$2d\frac{k(p^r-1)}{q-1} = \frac{6k}{2^{2r}+2^r+1},$$

which is less than k for all $r \geqslant 1$. If $(p,d) = (3,2)$, then we calculate that

$$2d\frac{k(p^r-1)}{q-1} = \frac{4k}{3^r+1},$$

which is at most k for all $r \geqslant 1$. In either case, we have $k+1 > 2dk(p^r-1)/(q-1)$ for all $r \geqslant 1$.

Finally, if $(p,d) = (2,2)$, then we have

$$2d\frac{k(p^r-1)}{q-1} = \frac{4k}{2^r+1}.$$

If $r \geqslant 2$, then $4/(2^r+1) < 1$ and we obtain $k+1 > 2dk(p^r-1)/(q-1)$. If $r=1$ then this inequality does not hold, but in this case we have $f = dr = 2$ and so $q = p^f = 4$. In this case, the assumptions that $k|(q-1)$ and $\mathbb{K} \neq \mathbb{F}_q$ imply that $k=3$. Then we can easily see that

$$k+2 = 5 > 4 = 2d\frac{k(p^r-1)}{q-1},$$

and so Chevalley's theorem again implies that the system has a nontrivial solution. This concludes the proof of the lemma. $\square$

Now we can prove the theorem when $\mathbb{K} \neq \mathbb{F}_q$. As mentioned in Section 2, we can assume that there is precisely one color – call it v – of variables which is zero-representing. Also, the condition in the theorem about linear combinations of the forms implies that there are at least $k+1$ variables which are not of this color. Set all but one of the variables of color v equal to zero. Then we are left with one variable of color v and at least $k+1$ variables of other colors. By Lemma 5, there exists a nontrivial zero of the system (1) involving only these variables. Note that any nontrivial solution of the system must have at least two nonzero variables. If the variable having color v is nonzero, then this variable and any other nonzero variable are two nonzero variables of different colors, and so the solution is nonsingular. If the variable of color v is equal to zero, then all of the nonzero variables in this solution are from colors which do not represent zero. Thus it is impossible for all the nonzero variables to have the same color, and so there must again be two nonzero variables of different colors. So in this case our solution is also nonsingular. This completes the proof of the theorem. $\square$

5. The Proof of Theorem 2

In order to prove Theorem 2, we first note that it suffices to prove the theorem for finite extensions of $\mathbb{Q}_p$. Suppose that the theorem is true for finite extensions, and that $\mathbb{L}$ is an infinite algebraic extension of $\mathbb{Q}_p$. If we have a system

$$\begin{aligned} a_1x_1^k + \cdots + a_sx_s^k &= 0 \\ b_1x_1^k + \cdots + b_sx_s^k &= 0 \end{aligned} \tag{8}$$

defined over $\mathbb{L}$, then we can also think of this system as being defined over the subfield $\mathbb{K} = \mathbb{Q}_p(a_1, \ldots, a_s, b_1, \ldots, b_s)$, which is a finite extension of $\mathbb{Q}_p$. If $s \geqslant 2k^2 + 1$, then the system has a nontrivial solution in $\mathbb{K}$, which will also be a nontrivial solution in $\mathbb{L}$. Thus the truth of the theorem for finite extensions implies that it is true for any algebraic extension of $\mathbb{Q}_p$.

We begin the proof for finite extensions by applying a normalization procedure to the forms. Suppose that $\mathbb{L}$ is a finite extension of $\mathbb{Q}_p$, and let π be a generator of the maximal ideal of $\mathbb{L}$. First note that by clearing denominators if necessary, we may assume that all of the coefficients in our system lie in $\mathfrak{O}_{\mathbb{L}}$, the ring of integers of $\mathbb{L}$. For a system of equations $\mathbf{F} = (F_1(\mathbf{x}), F_2(\mathbf{x})) = \mathbf{0}$ as in (8), we define

$$\Theta(\mathbf{F}) = \prod_{1 \leqslant i < j \leqslant s} (a_ib_j - a_jb_i).$$

By a standard argument involving the compactness of $\mathfrak{O}_{\mathbb{L}}$, we may assume that $\Theta(\mathbf{F}) \neq 0$. (One may see [DL69, pp. 572-573] for an example of this argument. As with [DL69, Lemma 11], which will be quoted shortly, although this fact is written down only for the case $\mathbb{L} = \mathbb{Q}_p$, it is not hard to see that its proof extends to general $\mathfrak{p}$-adic fields by merely replacing occurrences of $\mathbb{Q}_p$ and p by $\mathbb{L}$ and π respectively.) Next, we say that two systems of additive equations with coefficients in $\mathfrak{O}_{\mathbb{L}}$ are *equivalent* if one can be obtained from the other through a combination of the following three operations:

(i) replacing a variable x_i by $\pi^\alpha x_i$ for some integer α

(ii) dividing one or more equations by an integral power of π

(iii) taking nonsingular $\mathfrak{O}_{\mathbb{L}}$-linear combinations of the equations.

A system $\mathbf{F}$ is said to be *π-normalized* if both $\Theta(\mathbf{F}) \neq 0$ and the power of π dividing $\Theta(\mathbf{F})$ is less than or equal to the power of π dividing $\Theta(\mathbf{G})$ for all systems $\mathbf{G}$ equivalent to $\mathbf{F}$. Since any system $\mathbf{F}$ with $\Theta(\mathbf{F}) \neq 0$ is equivalent to a π-normalized system, it suffices to prove the theorem for π-normalized systems. The next lemma shows that π-normalized systems have some useful properties. After making the slight changes mentioned above, this lemma is [DL69, Lemma 11], specialized to two forms.

LEMMA 6. *A π-normalized system of two additive forms as in (8) can be written (after renumbering the variables) as*

$$F_1 = f_1(x_1, \ldots, x_r) + \pi g_1(x_{r+1}, \ldots, x_s)$$

$$F_2 = f_2(x_1, \ldots, x_r) + \pi g_2(x_{r+1}, \ldots, x_s),$$

where $r \geqslant s/k$ and if $1 \leqslant i \leqslant r$, then the coefficient of x_i in either f_1 or f_2 is not divisible by π. Moreover, if we form any nontrivial linear combination of F_1 and F_2, then at least $s/2k$ variables will appear in this linear combination with a coefficient not divisible by π.

Now for $i = 1, 2$, let F_i^* be the form obtained by reducing F_i modulo π, and consider the system of congruences

$$\begin{aligned} F_1^* &\equiv 0 \pmod{\pi} \\ F_2^* &\equiv 0 \pmod{\pi}. \end{aligned} \tag{9}$$

Since $\mathbb{L}$ is a finite extension of $\mathbb{Q}_p$, the residue field $\mathfrak{O}_{\mathbb{L}}/(\pi)$ is a finite field containing $q = p^f$ elements, where f is the residue degree of $\mathbb{L}$ over $\mathbb{Q}_p$. Also, since we are assuming that $s \geqslant 2k^2 + 1$, Lemma 6 tells us that (9) is explicit in at least $2k + 1$ variables, and that any nontrivial linear combination of F_1^* and F_2^* is explicit in at least $k+1$ variables. Since we are assuming that $p \nmid k$, the conditions of Theorem 1 are satisfied, and we know that (9) has a nonsingular solution. That is, (9) has a solution $\mathbf{x} = \mathbf{t}$ such that the expression

$$k^2 t_i^{k-1} t_j^{k-1} (a_i b_j - a_j b_i)$$

is nonzero modulo π for some variables t_i and t_j.

Finally, we must lift this solution of congruences modulo π to a solution in $\mathbb{L}$ of our original system. For this we use the following version of Hensel's lemma. This is a special case of [Gre69, Lemma 5.21].

Lemma 7. *Let $\mathbb{L}$ be a finite extension of $\mathbb{Q}_p$ with maximal ideal generated by π. Consider the system $\mathbf{F} = \mathbf{0}$ as in (8) in $s \geqslant 2$ variables defined over $\mathbb{L}$. Suppose that there exists $\mathbf{t} \in \mathbb{L}^s$ such that*

$$\mathbf{F}(\mathbf{t}) \equiv \mathbf{0} \pmod{\pi}$$

and that the Jacobian matrix

$$\begin{bmatrix} ka_1 t_1^{k-1} & \cdots & ka_s t_s^{k-1} \\ kb_1 t_1^{k-1} & \cdots & kb_s t_s^{k-1} \end{bmatrix}$$

has maximal rank modulo π. Then there exists $\mathbf{u} \in \mathbb{L}^s$ such that $\mathbf{F}(\mathbf{u}) = \mathbf{0}$ and $\mathbf{u} \equiv \mathbf{t} \pmod{\pi}$.

Now, we know that there exists $\mathbf{t} \in \mathbb{L}^s$ which satisfies (9). Since $\mathbf{t}$ is a nonsingular solution, the Jacobian matrix has a 2×2 submatrix with nonzero determinant modulo π, and hence has maximal rank modulo π. Therefore Lemma 7 may be applied to our nonsingular solution of (9), and this solution lifts to a solution $\mathbf{u}$ of (8). Finally, we note that since $\mathbf{t} \not\equiv \mathbf{0} \pmod{\pi}$ and $\mathbf{u} \equiv \mathbf{t} \pmod{\pi}$, we know that $\mathbf{u}$ is a nontrivial solution of (8), as desired.□

References

Alo99 N. Alon, *"Combinatorial Nullstellensatz"*, Combin. Probab. Comput. **8** (1999), 7–29.

BG02 J. Brüdern and H. Godinho, *"On Artin's conjecture, II: pairs of additive forms"*, Proc. London Math. Soc. **84** (2002), 513–538.

Bha66 M. Bhaskaran, *"Sums of m^{th} Powers in Algebraic and Abelian Number Fields"*, Arch. Math. (Basel) **17** (1966), 497–504.

Bha71 ______, *"Corrections to the paper "Sums of m^{th} powers""*, Arch. Math. (Basel) **22** (1971), 370–371.

Che35 C. Chevalley, *"Démonstration d'une hypothèse de M. Artin"*, Abh. Math. Sem. Hamburg **11** (1935), 73–75.

Dav82 H. Davenport, *The higher arithmetic: an introduction to the theory of numbers*, 5th ed., Cambridge University Press, Cambridge, 1982.

DL66 H. Davenport and D. J. Lewis, *"Notes on congruences (III)"*, Quart. J. Math. Oxford Ser. (2) **17** (1966), 339–344.

DL69 ______, *"Simultaneous equations of additive type"*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **264** (1969), 557–595.

Gre69 M. J. Greenberg, *Lectures on forms in many variables*, W. A. Benjamin and Co., New York, 1969.

Tor38 L. Tornheim, *"Sums of n-th powers in fields of prime characteristic"*, Duke Math. J. **4** (1938), 359–362.

Michael P. Knapp mpknapp@loyola.edu

Mathematical Sciences Department, Loyola College, 4501 North Charles Street, Baltimore, MD 21210-2699