ACTA ARITHMETICA
Online First version

On the values of $\Gamma^*(k,p)$ and $\Gamma^*(k)$

by

Hemar Godinho (Brasília), Michael P. Knapp (Baltimore, MD),
Paulo H. A. Rodrigues (Goiânia) and
Daiane Veras (Valparaíso de Goiás)

1. Introduction. For $k \in \mathbb{N}$ and p a prime number, define $\Gamma^*(k,p)$ to be the smallest $n \in \mathbb{N}$ such that every diagonal form $f(x_1,\dots,x_s) = a_1x_1^k + \cdots + a_sx_s^k$ with integer coefficients has a nontrivial zero over $\mathbb{Q}_p$ whenever $s \geq n$. Define also

$$\Gamma^*(k) = \max_{p\,\text{prime}} \Gamma^*(k,p).$$

In 1963 Davenport and Lewis [4] proved that $\Gamma^*(k) \leq k^2+1$ and $\Gamma^*(p-1,p) = (p-1)^2+1$, and in 1967 Dodson [5] published an improved bound, $\Gamma^*(k) \leq \frac{49}{64}k^2+1$ if $k+1$ is composite. For specific values of k, a number of exact values of $\Gamma^*(k)$ have been computed. Lewis [9] showed that $\Gamma^*(3) = 7$. The combined work of Gray [6] and Chowla [3] shows that $\Gamma^*(5) = 16$. Bierstedt [1] appears to have been the first to show that $\Gamma^*(7) = 22$ and $\Gamma^*(11) = 45$. These values were independently discovered by Norton [10], who also gave the value $\Gamma^*(9) = 37$. The values of $\Gamma^*(7)$ and $\Gamma^*(9)$ were also discovered independently by Dodson [5]. Bovey [2] showed that $\Gamma^*(8) = 39$, and recently Knapp [7, 8] has determined the exact values of $\Gamma^*(k)$ for all remaining $k \leq 32$ with $k+1$ composite. In unpublished work, two undergraduate students of Knapp, Christopher Broll and Jessica Jennings, have pushed this bound to $k \leq 39$.

To state our main results we need some notation. For a fixed prime p, write $k = p^\tau k_0$, with $\gcd(p,k_0) = 1$. Also, define

$$\gamma = \gamma(k) = \begin{cases} 1 & \text{if } \tau = 0, \\ \tau+1 & \text{if } \tau > 0 \text{ and } p > 2, \\ \tau+2 & \text{if } \tau > 0 \text{ and } p = 2, \end{cases} \tag{1}$$

and write $k = \gamma q + r$, with $q, r \in \mathbb{Z}$, $0 \leq r < \gamma$.

2010 *Mathematics Subject Classification*: Primary 11D72; Secondary 11D79, 11D88.
Key words and phrases: Artin's conjecture, diagonal forms.
Received 13 June 2018; revised 7 December 2018.
Published online *.

DOI: 10.4064/aa180613-4-1

Theorem 1.1. *With the notation above, we have*

$$\Gamma^*(k,p) \le (p^\gamma - 1)q + p^r, \tag{2}$$

and equality holds whenever $p-1$ divides k.

Theorem 1.2. *We have* [1] $\Gamma^*(54) = 1049$.

The $p=2$ case of Theorem 1.1 is due to Knapp, and our proof proceeds along the same lines. Suppose that f is a diagonal form in the number of variables given in Theorem 1.1. First, we apply a "normalization" procedure due to Davenport & Lewis which shows that we may write $f = F_0 + pF_1 + p^2F_2 + \cdots + p^{k-1}F_{k-1}$, where each of $F_1,\ldots,F_{k-1}$ is a diagonal form whose coefficients are not divisible by p, and we may make certain assumptions about the number of variables in each of the forms. The proof now proceeds inductively in the following manner. At each stage, we use a trivial extension of a lemma due to Bovey [2] to show that either f has a nontrivial zero or else we are able to find a set $\{F_a, F_{a+1},\ldots,F_{a+\tau}\}$ of subforms, disjoint from the sets found in any previous stages, which has a "small" number of variables. After repeating the inductive step q times, we are left with only r remaining subforms, and these subforms must have a relatively large number of variables between them. We are then able to use Bovey's lemma on these final subforms, along with the structure of the disjoint sets produced earlier, to show that f has a nontrivial zero. This shows that $\Gamma^*(k,p)$ is at most the bound given in the theorem. If it happens that $(p-1)\mid k$, then we are able to explicitly give a form in one fewer variable that has no nontrivial p-adic zeros, showing that the number of variables given in the theorem is the smallest that can guarantee that f has a nontrivial zero. This shows that $\Gamma^*(k,p)$ equals the value in the theorem when $(p-1)\mid k$.

The proof of Theorem 1.2 involves a combination of theory and brute-force calculations. Suppose that f is a diagonal form of degree 54 in 1049 variables. We need to show that f has a nontrivial p-adic zero for all primes p. For $p = 2,3$, we are done by Theorem 1.1. Then we use Chevalley's theorem and results of Dodson [5] to prove our result for all p except those for which $163 \le p \le 4159$ and $54\mid(p-1)$. We then follow ideas of Bovey to develop a function $Q(k,p,t)$ with the property that (for our purposes) if $Q(54,p,20)<1$, then f has a nontrivial p-adic zero. A brute-force computation shows that this inequality holds for all primes except $p=163$. Finally, we perform a direct calculation to show that every diagonal form of degree 54 in 1049 variables must have a nontrivial 163-adic zero. We note that the only obstruction to using these methods to compute other values of $\Gamma^*(k)$ is the computing time necessary for evaluating $Q(k,p,t)$ and for doing the direct calculation at the end. We hope to come back to these ideas in a future article.

[1] This value was also discovered independently by Jessica Jennings (unpublished).

It is appropriate here to briefly discuss the quality of Theorem 1.1. First, we examine the situations in which Theorem 1.1 gives an equality. Suppose that $p-1$ divides k, and write $k = p^\tau(p-1)k_1$, with $p \nmid k_1$. If $p > 2$ and either $\tau \geq 2$ or $k_1 \geq 2$, then the result that our expression equals the value of $\Gamma^*(k,p)$ is new. (The result with $p = 2$ is due to Knapp [8]. The results with $k_1 = 1$ and $\tau = 0, 1$ are given by Davenport and Lewis [4] and Dodson [5], respectively.)

In the cases where we do not claim that our bound is an equality, we can compare our results with the bounds for $\Gamma^*(k)$ given by Davenport & Lewis and Dodson above. We can also compare it with Dodson's bound (see [5])

$$\Gamma^*(k,p) \leq \left\lfloor \frac{k}{\gamma}(p^\gamma - 1) \right\rfloor + 1.$$

If $\gamma \mid k$, then our bound equals the bound of Dodson displayed above, and our bound is better in all other cases. Moreover, our bound trivially implies

$$\Gamma^*(k,p) \leq \frac{p}{\gamma k_0} k^2 + k, \tag{3}$$

and so our bound is better than the bounds for $\Gamma^*(k)$ mentioned above whenever γk_0 is suitably large compared to k. However, the situation is better than this, since the $pk^2/(\gamma k_0)$ term in (3) is obtained under the assumption that $r = 0$ (among others), while the k term is obtained under the assumption that $r = \tau$ (among others). So our bound will in fact always be somewhat smaller than the one in (3).

As mentioned above, the case $p = 2$ in Theorem 1.1 was proved by Knapp in [8], and the case $\gamma = 1$ (i.e., $p \nmid k$) is a direct consequence of [4, Lemma 1] and [5, Lemmas 4.2.1 and 4.2.2]. Therefore, from this point on we will assume each of the following:

(4)
- $p \geq 3$;
- $p \mid k$, so that $\gamma = \tau + 1 \geq 2$;
- $\gamma < k$ (this follows from the previous assumptions);
- $s \geq (p^\gamma - 1)q + p^r$.

The proof of Theorem 1.1 starts with the simple observation that any diagonal form $f(x_1, \ldots, x_s)$ can be written as

$$f = \sum_{i=0}^{m} p^i f_i$$

for some m, where the subforms f_i are diagonal forms of degree k whose coefficients are all coprime to p. If $m \geq k$, then for each l with $k \leq l \leq m$ we write $l = \delta k + \rho$, with $\rho, \delta \in \mathbb{Z}$ and $0 \leq \rho < k$. Suppose that x is a variable in f_l with coefficient a. Then the term in f involving x is

$$p^l a x^k = p^\rho a (p^\delta x)^k = p^\rho a y^k,$$

where $y = p^\delta x$. Repeating this argument for every variable in every subform f_l with $k \le l \le m$, we obtain a new equivalent form f^*,

$$f^* = f_0^* + pf_1^* + \cdots + p^{k-1} f_{k-1}^*, \tag{5}$$

of degree k. Note that f^* has a nontrivial zero in $\mathbb{Q}_p$ if and only if f does. Write v_i^* for the number of variables in f_i^*, and note that $\sum_{i=0}^{k-1} v_i^* = s$.

Lemma 1.3 (Davenport–Lewis [4, Lemma 2]). *Let $n_0, \ldots, n_{k-1} \in \mathbb{R}$ and put $n_{k+j} = n_j$ for all $j \ge 0$. Let $n_0 + \cdots + n_{k-1} = s$. Then there exists a number r such that*

$$n_r + \cdots + n_{r+t-1} \ge ts/k \quad \textit{for } t = 1, \ldots, k.$$

Define now $f_{k+j}^* = f_j^*$ and $v_{k+j}^* = v_j^*$, for $j \ge 0$. It follows from Lemma 1.3 that there exists a number $r < k$ such that

$$v_r^* + \cdots + v_{r+t-1}^* \ge ts/k$$

for $t = 1, \ldots, k$. Replacing all the variables x of $f_0^*, \ldots, f_{r-1}^*$ by px we obtain (see (5)) the form

$$G = p^k f_0^* + p^{k+1} f_1^* + \cdots + p^{k+r-1} f_{r-1}^* + p^r f_r^* + \cdots + p^{k-1} f_{k-1}^*$$

and hence

$$p^{-r} G = f_r^* + pf_{r+1}^* + \cdots + p^{k-r-1} f_{k-1}^* + p^{k-r} f_0^* + \cdots + p^{k-1} f_{r-1}^*. \tag{6}$$

We can now rename $p^{-r}G$ as F and write

$$F = F_0 + pF_1 + \cdots + p^{k-1} F_{k-1}. \tag{7}$$

If we now write v_i for the number of variables in F_i, then our work shows that $v_0 + \cdots + v_{t-1} \ge ts/k$ for $t = 1, \ldots, k$. Clearly, if F has a nontrivial p-adic solution, so does f.

Lemma 1.4. *With the notation above, suppose that for some $n \in \mathbb{N} \cup \{0\}$ there is a vector $\boldsymbol{z} \in \mathbb{Z}^s$ such that*

$$F(\boldsymbol{z}) \equiv 0 \pmod{p^{n+\gamma}}$$

and at least one of the entries of $\boldsymbol{z}$ corresponding to a variable in F_j, $j \le n$, is coprime to p. Then $\boldsymbol{z}$ can be lifted to a nontrivial p-adic zero of F.

Proof. This is [8, Lemma 2.1]. The proof of this Hensel-type lemma follows from [4, Lemma 4] and, if necessary, a cyclic permutation of $F = \sum_{i=0}^{k-1} p^i F_i$ as in (6). ∎

2. Proof of Theorem 1.1. We start by presenting a series of lemmas that will be used in the proof of Theorem 1, especially the next lemma whose proof is a straightforward modification of [2, proof of Lemma 1].

Lemma 2.1. *Let $0 \le n \le k-1$ and write $F_i = \sum_{j=1}^{v_i} a_{ij} x_{ij}^k$, where none of the coefficients a_{ij} are divisible by p. Suppose*

$$\begin{aligned} v_0 &\ge p, \\ v_0 + v_1 &\ge p^2, \\ &\vdots \\ v_0 + \cdots + v_{n-1} &\ge p^n. \end{aligned} \tag{8}$$

Then, for $N > n$, the form $F = \sum_{i=0}^n p^i F_i$ represents at least $\min(\sum_{i=0}^n v_i, p^N)$ distinct residue classes modulo p^N with $x_{ij} = 0$ or 1, and at least one x_{0j} equal to 1.

The next two lemmas will be used to prove that inequalities such as the ones in (8) hold.

Lemma 2.2. *Suppose that r is a real number and a is a positive integer such that $r > p-1$ and $ar > p^a - 1$. Then $tr > p^t - 1$ whenever $2 \le t \le a-1$.*

Proof. Let $f(t) = tr - p^t + 1$ and consider t as a real variable. The lemma follows if we prove that $f(t) > 0$ for $1 \le t \le a$. By hypothesis, this is true for $t = 1$ and $t = a$. We have $f'(t) = r - p^t \log p$, so there exists only one critical point of f, say t_0. Moreover, $f'(t) > 0$ for $t < t_0$ and $f'(t) < 0$ for $t > t_0$, which implies that t_0 is a maximum point. If $f(t) < 0$ for some t between 1 and a, then $f(t)$ would have a minimum point, a contradiction. ■

Lemma 2.3. *Let p be an odd prime, $m, l \in \mathbb{N}$, $m \ge 2$ and $r \in \mathbb{Z}$ with $0 \le r < m$. Then*

$$\left\lceil \frac{t((p^m - 1)l + p^r)}{ml + r} \right\rceil \ge p^t \quad \text{for } 1 \le t \le m-1,$$

where $\lceil \cdot \rceil$ is the ceiling function.

Proof. By Lemma 2.2, it suffices to prove the conclusion for $t = 1$ and $t = m-1$. When $t = 1$, we will show that

$$\frac{(p^m - 1)l + p^r}{ml + r} \ge p,$$

which clearly implies the inequality in the lemma. To see this, we note that since $m \ge 2$ and r is a nonnegative integer, we have $p^m - 1 \ge 2(p-1)m$ and $p^r > (p-1)r$. So,

$$\begin{aligned} \frac{(p^m - 1)l + p^r}{ml + r} &> \frac{2(p-1)ml + (p-1)r}{ml + r} = p - 1 + \frac{(p-1)ml}{ml + r} \\ &\ge p - 1 + \frac{2ml}{ml + (m-1)} > p - 1 + 1 = p. \end{aligned}$$

Now suppose that $t = m-1 \geq 2$ (i.e. $m \geq 3$). We will prove that

$$(m-1)((p^m-1)l+p^r) > (p^{m-1}-1)(ml+r),$$

which again implies the assertion. The above inequality is equivalent to

$$(m-1)p^r + l + r + p^{m-1}\big((p(m-1)-m)l - r\big) > 0.$$

Since $(m-1)p^r + l + r > 0$, it suffices to prove that $l(p(m-1)-m) \geq r$. We will show that

$$l(p(m-1)-m) \geq m-1,$$

which is sufficient since $m-1 \geq r$. By hypothesis we have

$$l(p(m-1)-m) \geq p(m-1)-m \geq 3(m-1)-m = 2m-3 \geq m-1,$$

where the last inequality follows since $m \geq 2$. ■

At this point let us recall that $F = F_0 + pF_1 + \cdots + p^{k-1}F_{k-1}$ is a diagonal form of degree $k = p^\tau k_0 = \gamma q + r$, $0 \leq r < \gamma$, in $s \geq (p^\gamma - 1)q + p^r$ variables, where $\gcd(p, k_0) = 1$. Moreover, the analysis following Lemma 1.3 shows that (perhaps after making a change of variables) we may assume that we have the system of inequalities

$$(9) \qquad v_0 + \cdots + v_{t-1} \geq \frac{t}{k}((p^\gamma - 1)q + p^r) \qquad \text{for } t = 1, \ldots, k.$$

As $p \geq 3$, we see that $k = p^\tau k_0 \geq \tau + 1 = \gamma$. This implies $q > 0$.

Since $\gamma \leq k$, the inequalities in (9) hold for $t = 1, \ldots, \gamma - 1$, and therefore Lemma 2.3 gives us

$$\begin{aligned} v_0 &\geq p, \\ v_0 + v_1 &\geq p^2, \\ &\vdots \\ v_0 + \cdots + v_{\gamma-2} &\geq p^{\gamma-1}. \end{aligned}$$

First, let us assume that $r = 0$. In this case, $s \geq (p^\gamma - 1)q + 1$, and (9) gives us

$$v_0 + \cdots + v_{\gamma-1} \geq \frac{\gamma[(p^\gamma-1)q+1]}{\gamma q} = p^\gamma - 1 + \frac{1}{q} > p^\gamma - 1.$$

Since $v_0 + \cdots + v_{\gamma-1}$ must be an integer, we obtain $v_0 + \cdots + v_{\gamma-1} \geq p^\gamma$. Together with the above system of inequalities, Lemma 2.1 shows that F represents at least $\min(\sum_{i=0}^{\gamma-1} v_i, p^\gamma) = p^\gamma$ distinct residue classes modulo p^γ (and in particular represents 0) with $x_{ij} = 0$ or 1 and $x_{0j} = 1$ for some j. By Lemma 1.4, F has a nontrivial p-adic zero.

Now, assume instead that $r > 0$. In this situation, all the inequalities in the displayed system above continue to hold, but Lemma 1.3 no longer guarantees that $v_0 + \cdots + v_{\gamma-1} \geq p^\gamma$. However, if this final inequality does hold, then F has a nontrivial p-adic zero as above. So from now on, we will assume that $\sum_{i=0}^{\gamma-1} v_i < p^\gamma$.

Let $I = \{0,1,\dots,k-1\}$ and define

$$\begin{aligned} T_0 &= \{0,1,\dots,\gamma-1\},\\ T_1 &= \{1,2,\dots,\gamma\},\\ &\ \ \vdots\\ T_{k-\gamma} &= \{k-\gamma,\dots,k-1\}. \end{aligned}$$

Let $\mathfrak{T} = \{T_0, T_1, \dots, T_{k-\gamma}\}$ and suppose that there exist pairwise disjoint sets $S_1,\dots,S_h \in \mathfrak{T}$ such that for each i we have $S_i = \{s_i, s_i+1,\dots,s_i+\tau\}$, where

$$\begin{aligned} &\bullet\ v_{s_i} + \dots + v_{s_i+t-1} \ge p^t \text{ for } t = 1,\dots,\gamma-1,\\ &\bullet\ v_{s_i} + \dots + v_{s_i+\tau} < p^\gamma. \end{aligned} \tag{10}$$

Observe that $0 \le h \le q$ since $k = \gamma q + r$ and $0 \le r < \gamma$. In fact, since T_0 satisfies (10), we can take $S_1 = T_0$ and hence may assume that $h \ge 1$.

Consider the 1-1 correspondence between the nonempty subsets $\mathfrak{U} \subset I$ and the subforms of $F = F_0 + pF_1 + \dots + p^{k-1}F_{k-1}$ given by

$$\mathfrak{U} = \{u_1,\dots,u_s\} \leftrightarrow F_{\mathfrak{U}} = p^{u_1}F_{u_1} + \dots + p^{u_s}F_{u_s}.$$

Let $J = I - S_1 \cup \dots \cup S_h$ and write $J = \{j_1,\dots,j_m\}$, where $m = |J| = \gamma(q-h)+r$. Consider the subform F_J, in $N \ge (p^\gamma-1)(q-h)+p^r$ variables, so that $v_{j_1} + \dots + v_{j_m} = N$. By Lemma 1.3, there exists l such that

$$v_{j_l} + \dots + v_{j_{l+t-1}} \ge tN/m \tag{11}$$

for $t = 1,\dots,m$, where as usual we define $v_{j_{m+i}} = v_{j_i}$.

Lemma 2.4. *With the notation above, suppose that $h = q$. Then there exists a p-adic zero of F.*

Proof. Since $h=q$, we have $m=r$ and $N \ge p^r$. Combining this with (11), we see that

$$v_{j_l} + \dots + v_{j_{l+t-1}} \ge tp^r/r$$

for $t = 1,\dots,r$. Since $v_{j_l} \ge p^r/r \ge p$ and $v_{j_l} + \dots + v_{j_{l+r-1}} \ge p^r$, it follows from Lemma 2.2 that

$$v_{j_l} + \dots + v_{j_{l+t-1}} \ge p^t \tag{12}$$

for $1 \le t \le r$. Let z be the smallest integer such that $j_{l+z+1} \ne j_{l+z}+1$, so that $j_{l+i} = j_l + i$ for $i = 0,1,\dots,z$. Observe that if $z = r-1$, then the condition $j_{l+r} \ne j_{l+r-1}+1$ is empty. Suppose first that $j_{l+z} \ne k-1$. Then $j_{l+z}+1 \in S_i$ for some i. Since $S_i = \{s_i,\dots,s_i+\tau\}$ and $j_{l+z} \notin S_1 \cup \dots \cup S_h$, we get $j_{l+z}+1 = s_i$. On the other hand, $j_l,\dots,j_{l+z}$ are consecutive. As a consequence, (12) implies that

$$v_{j_l} + \dots + v_{j_{l+t}} \ge p^{t+1}$$

for $t = 0, 1, \ldots, z$. By Lemma 2.1, the congruence

$$F^* = \sum_{i=0}^{z} p^i F_{l+i} \equiv 0 \pmod{p^{z+1}}$$

has a solution ξ whose entries are 0 or 1 and at least one variable from F_l equals 1. Then $H(\xi) = p^{j_l} F^*(\xi)$ is congruent to 0 modulo $p^{j_l+z+1} = p^{s_i}$. Now, by Lemma 2.1, the form F_{S_i} represents all multiples of p^{s_i} modulo $p^{s_i+\tau}$. Since $H(\xi) \equiv 0 \pmod{p^{s_i}}$, there exists a solution to the congruence $H + F_{S_i} \equiv 0 \pmod{p^{s_i+\tau}}$ satisfying the conditions of Lemma 1.4, which gives a nontrivial p-adic solution for F.

If it should happen that $j_{l+z} = k - 1$, then the above idea works with a slight modification. Suppose that $x_1, \ldots, x_t$ are the variables involved in the subforms $F_0, \ldots, F_\tau$ and consider the form

$$G = F(px_1, \ldots, px_t, x_{t+1}, \ldots, x_s).$$

This has the effect of transforming the original form F in (7) into

$$G = p^{\tau+1} F_{\tau+1} + \cdots + p^{k-1} F_{k-1} + p^k F_0 + p^{k+1} F_1 + \cdots + p^{k+\tau} F_\tau.$$

We can now consider S_1 to be the set $\{k, k+1, \ldots, k+\tau\}$, so that $s_1 = k$. Then the same reasoning as above shows that G has a nontrivial p-adic solution, and hence F does as well. ■

The next lemma completes the proof of the first assertion of Theorem 1.1.

LEMMA 2.5. *With the same conditions as above, if $h < q$ then either F has p-adic zeros or there exists another set S_{h+1}, disjoint from $S_1 \cup \cdots \cup S_h$, satisfying* (10).

Proof. By (11) we have

$$v_{j_l} + \cdots + v_{j_{l+t-1}} \geq t \frac{(p^\gamma - 1)(q-h) + p^r}{\gamma(q-h) + r} \tag{13}$$

for $t = 1, \ldots, \gamma - 1$, and since $q > h$ we see that $\gamma < m$. By Lemma 2.3,

$$v_{j_l} + \cdots + v_{j_{l+t-1}} \geq p^t \quad \text{for } t = 1, \ldots, \gamma - 1.$$

As before, let z be the smallest positive integer such that $j_{l+z+1} \neq j_{l+z} + 1$, so that $j_{l+i} = j_l + i$ for $i = 0, 1, \ldots, z$. Observe that if $j_{l+z} \neq k - 1$, then $j_{l+z} + 1 = s_i \in S_i = \{s_i, \ldots, s_i + \tau\}$ for some i. Now, $j_l, j_{l+1}, \ldots, j_{l+z}$ are consecutive. If $z < \gamma - 1$ then we can proceed as in Lemma 2.4 (in which $z \leq r - 1 \leq \gamma - 2$) and obtain a p-adic zero for F. Note that even if $j_{l+z} = k - 1$, the proof of Lemma 2.4 shows that F has a p-adic zero.

Assume then that $z \geq \gamma - 1 = \tau$. If

$$v_{j_l} + \cdots + v_{j_{l+\gamma-1}} \geq p^\gamma,$$

then together with (13) we can apply Lemma 2.1 and obtain a solution of $F_J \equiv 0 \pmod{p^{j_l+\gamma}}$ with at least one variable from F_{j_l} nonzero. Then

Lemma 1.4 shows that F has a nontrivial p-adic zero. On the other hand, if

$$v_{j_l} + \cdots + v_{j_{l+\gamma-1}} < p^\gamma,$$

then we may set $S_{h+1} = \{j_l, j_l + 1, \ldots, j_l + \tau\}$, and then S_{h+1} is disjoint from $S_1 \cup \cdots \cup S_h$ and satisfies the properties in (10). ∎

In order to see that this completes the proof of the first assertion of Theorem 1.1, note that the initial work we did after assuming that $r > 0$ shows that we may take $S_1 = T_0$, so that we may have $h = 1$. If $1 < q$, then Lemma 2.5 shows that either F (and hence f) has a nontrivial p-adic zero or else we may take $h = 2$. Repeating this process, we see that either f has a p-adic zero or else we may take $h = q$. If this last possibility occurs, then Lemma 2.4 shows that f has a nontrivial p-adic zero.

We now prove the second assertion of Theorem 1.1 by giving an example of a form of degree $k = p^\tau(p-1)k_1$, $p \nmid k_1$, in $s = (p^\gamma - 1)q + p^r - 1$ variables with no nontrivial p-adic zeros.

First note that if $p \mid x$ then $x^k \equiv 0 \pmod{p^\gamma}$, and otherwise $x^k \equiv 1 \pmod{p^\gamma}$. The second case follows from the fact that $\varphi(p^\gamma) = p^\tau(p-1)$, where φ is the Euler φ-function.

Following the discussion above, consider the additive form

(14)
$$F = \sum_{i=1}^{p^\gamma-1} x_i^k + p^\gamma \sum_{i=p^\gamma}^{2(p^\gamma-1)} x_i^k + \cdots + p^{(q-1)\gamma} \sum_{i=(q-1)(p^\gamma-1)+1}^{q(p^\gamma-1)} x_i^k + p^{q\gamma} \sum_{i=q(p^\gamma-1)+1}^{q(p^\gamma-1)+p^r-1} x_i^k.$$

The next lemma is a trivial consequence of Euler's theorem, and so we will not give a proof.

LEMMA 2.6. *Let $H = \sum_{i=1}^{p^t-1} x_i^k$ with $(p-1) \mid k$. If $t \le \gamma$, then the congruence $H \equiv 0 \pmod{p^t}$ has no nontrivial solution.*

Now, let F be as in (14) and suppose that $F(\xi) = 0$ for some nonzero $\xi \in \mathbb{Z}_p$. By the homogeneity of F, we may suppose that some coordinate of ξ is a unit in $\mathbb{Z}_p$, and in particular this coordinate is coprime to p. Let $m = p^\gamma - 1$ for convenience, and rename the variables of F so that

- $F_j = F_j(x_{1,j}, \ldots, x_{m,j})$ for $j = 0, \ldots, q-1$,
- $F_q = F_q(x_{1,q}, \ldots, x_{p^r-1,q})$,
- $\xi = (\xi_{1,0}, \ldots, \xi_{m,0}, \ldots, \xi_{1,q-1}, \ldots, \xi_{m,q-1}, \eta_1, \ldots, \eta_{p^r-1})$.

Suppose that the first coordinate of ξ coprime to p corresponds to a variable in F_j for some fixed $j < q$. Without loss of generality we may suppose that this coordinate is $\xi_{1,j}$. Then $\xi_{u,v} = p\xi_{u,v}^*$ for $v \le j-1$. Therefore

$$F(\xi) = p^k H + p^{\gamma j} F_j(\xi_{1,j}, \ldots, \xi_{m,j}) + p^{\gamma(j+1)} G = 0,$$

where we have written

$$H = \sum_{i=0}^{j-1} p^{i\gamma} F_i(\xi_{1,i}^*, \dots, \xi_{m,i}^*),$$

$$G = p^{-\gamma(j+1)}\Big(\sum_{i=j+1}^{q-1} p^{i\gamma} F_i(\xi_{1,i}, \dots, \xi_{m,i}) + p^{q\gamma} F_q(\eta_1, \dots, \eta_{p^r-1})\Big).$$

Hence

$$p^{-\gamma j} F(\xi) = F_j(\xi_{1,j}, \dots, \xi_{m,j}) + p^{\gamma} G + p^{k-\gamma j} H = 0,$$

and in particular

$$p^{-\gamma j} F(\xi) \equiv F_j(\xi_{1,j}, \dots, \xi_{m,j}) \equiv 0 \pmod{p^\gamma},$$

which contradicts Lemma 2.6.

Suppose instead that the first coordinate coprime to p is one of the p-adic integers $\eta_1, \dots, \eta_{p^r-1}$. In the same way as above we obtain

$$p^{-q\gamma} F(\xi) = \sum_{i=1}^{p^r-1} \eta_i^k + p^{k-\gamma q} G' = 0.$$

In particular, since $k - \gamma q = r$ we have

$$p^{-q\gamma} F(\xi) \equiv \sum_{i=1}^{p^r-1} \eta_i^k \equiv 0 \pmod{p^r},$$

which is again impossible by Lemma 2.6.

3. Proof of Theorem 1.2. In this section we use Theorem 1.1 and some computations to prove Theorem 1.2. By Theorem 1.1 we have $\Gamma^*(54,2) = 127$ and $\Gamma^*(54,3) = 1049$. Suppose now that $p > 3$.

Let $d = \gcd(54, p-1)$. It is well known that the congruence $x^{54} \equiv a \pmod{p}$ has a solution if, and only if, the congruence $x^d \equiv a \pmod{p}$ has a solution. This implies that the set of 54th powers and the set of dth powers in $\mathbb{F}_p$ are the same. Since p does not divide 54, we have $\gamma = 1$, and without loss of generality we may replace 54 by d in any congruence $a_1x_1^{54} + \dots + a_nx_n^{54} \equiv 0 \pmod{p}$. Since p is odd, we have $2 \mid (p-1)$, and so it is sufficient to consider $d \in \{2, 6, 18, 54\}$.

By Chevalley's theorem, if $d = 2$, 6, or 18, then solubility is ensured if (in the notation of Section 2) $v_0 \geq 3$, 7, or 19, respectively. These values follow from Lemma 1.3 when we have $s \geq 109$, $s \geq 325$, or $s \geq 973$ variables, respectively. So

- $\Gamma^*(54, p) \leq 109$ if $\gcd(54, p-1) = 2$;
- $\Gamma^*(54, p) \leq 325$ if $\gcd(54, p-1) = 6$;
- $\Gamma^*(54, p) \leq 973$ if $\gcd(54, p-1) = 18$.

Now we need to analyze $d = 54$, and from this point onwards only consider primes $p \equiv 1 \pmod{54}$. Assuming we have 1049 variables, we deduce from Lemma 1.3 and the remarks afterwards that $v_0 \geq 20$. The next lemma is a trivial consequence of (the proof of) Lemma 2.4.1 of [5].

LEMMA 3.1. *Consider the congruence*

$$a_1x_1^k + \cdots + a_tx_t^k \equiv 0 \pmod{p}.$$

If p does not divide either k or any of the coefficients a_i, then the congruence above has a solution with at least one variable nonzero modulo p whenever

$$p > (d-1)^{(2t-2)/(t-2)} \quad \textit{where} \quad d = (k, p-1).$$

When $k = d = 54$, this lemma ensures the desired solubility for $p > (53)^{38/18}$. That is, we have $\Gamma^*(54, p) \leq 1049$ for each $p > 4367$.

For the remaining primes, we proceed as in [2, Lemmas 4 and 5]. It is well-known that the number of solutions of the polynomial congruence $f(\mathbf{x}) \equiv 0 \pmod{p}$ is given by

$$\frac{1}{p} \sum_{\mathbf{x} \,(\mathrm{mod}\, p)} \sum_{t=0}^{p-1} e_p(tf(\mathbf{x})),$$

where $e_p(x) = \exp(2\pi i x/p)$. Applying this to the congruence

$$c_1x_1^k + \cdots + c_sx_s^k \equiv 0 \pmod{p},$$

we see that if this congruence has only the trivial solution, then

$$\sum_{x_1,\ldots,x_s=0}^{p-1} \sum_{t=0}^{p-1} e_p(c_1x_1^k t) \cdots e_p(c_sx_s^k t) = p.$$

This is equivalent to

$$\sum_{t=1}^{p-1} S(c_1t) \cdots S(c_st) = p - p^s,$$

where we write

$$S(b) = \sum_{x=0}^{p-1} e_p(x^k b).$$

Using Hölder's inequality, and also the fact that if t runs through the nonzero residue classes modulo p then so does c_it, we obtain

$$\sum_{t=1}^{p-1} |S(t)|^s \geq p^s - p.$$

If we now define, for $s > 1$,

$$Q(k,p,s) = \frac{1}{p^s - p} \sum_{t=1}^{p-1} |S(t)|^s,$$

we see that if our congruence has only the trivial solution for even one choice of coefficients, then $Q(k,p,s) \geq 1$. This immediately leads to the following lemma.

Lemma 3.2. *If $Q(k,p,s) < 1$ then every congruence*

$$c_1x_1^k + \cdots + c_sx_s^k \equiv 0 \pmod{p}$$

has a nontrivial solution.

We now use Lemma 3.2 to begin the computational study of the remaining primes with $p \equiv 1 \pmod{54}$. Using Maple to compute $Q(54,p,20)$ for these primes, we obtain the values in the table below.

Table 1. Table of values of $Q(54,p,20)$

$Q(54,p,20)$ for $p \leq 4367$ with $p \equiv 1 \pmod{54}$				
Prime p	109	163	271	379
$Q(54,p,20)$	15.246914	4.839553	0.210817	0.013238
Prime p	433	487	541	757
$Q(54,p,20)$	0.163891	0.000753	0.024424	0.000560
Prime p	811	919	1297	1459
$Q(54,p,20)$	$1.70 \cdot 10^{-5}$	$6.52 \cdot 10^{-5}$	$2.17 \cdot 10^{-5}$	$1.03 \cdot 10^{-8}$
Prime p	1567	1621	1783	1999
$Q(54,p,20)$	$2.08 \cdot 10^{-7}$	$3.40 \cdot 10^{-6}$	$9.83 \cdot 10^{-10}$	$1.14 \cdot 10^{-8}$
Prime p	2053	2161	2269	2377
$Q(54,p,20)$	$2.36 \cdot 10^{-6}$	$2.01 \cdot 10^{-5}$	$3.17 \cdot 10^{-3}$	$1.57 \cdot 10^{-8}$
Prime p	2539	2593	2647	2917
$Q(54,p,20)$	$9.60 \cdot 10^{-10}$	$2.19 \cdot 10^{-8}$	$1.30 \cdot 10^{-10}$	$1.54 \cdot 10^{-8}$
Prime p	2971	3079	3187	3457
$Q(54,p,20)$	$7.49 \cdot 10^{-10}$	$1.81 \cdot 10^{-9}$	$8.98 \cdot 10^{-10}$	$7.30 \cdot 10^{-9}$
Prime p	3511	3673	3727	3889
$Q(54,p,20)$	$6.11 \cdot 10^{-12}$	$6.09 \cdot 10^{-9}$	$3.83 \cdot 10^{-11}$	$3.74 \cdot 10^{-11}$
Prime p	3943	4051	4159	
$Q(54,p,20)$	$3.58 \cdot 10^{-10}$	$3.86 \cdot 10^{-11}$	$1.35 \cdot 10^{-11}$	

From the table, we see that

$$Q(54,p,20) < 1, \qquad \forall p \neq 109, 163.$$

Hence having $v_0 = 20$ is sufficient for $p \neq 109, 163$. As noted above, this condition is guaranteed whenever F has at least 1049 variables, and hence $\Gamma^*(54,p) \leq 1049$ for these primes.

The next lemma is a direct consequence of [5, Lemmas 2.2.1 and 4.2.1].

Lemma 3.3. *Suppose that* $\gcd(k,p-1)=\frac{1}{2}(p-1)$. *Then*

$$\Gamma^*(k,p)\le k\cdot\left\lfloor\frac{\log p}{\log 2}\right\rfloor+1.$$

It follows from this lemma that

$$\Gamma^*(54,109)\le 325.$$

Finally, we evaluate $\Gamma^*(54,163)$. The only 54th powers modulo 163 are 0, 1, 58, and 104. Using Maple, we can verify that every congruence $ax+by+cz\equiv 0 \pmod{163}$ has a solution with $x,y,z\in\{1,58,104\}$. It follows that solubility is ensured whenever $v_0\ge 3$, which we have seen occurs whenever F has at least 109 variables. Hence $\Gamma^*(54,163)\le 109$.

While this is enough to complete the proof of Theorem 1.2, we can actually go a bit further here. Note that if a is not a 54th power (mod 163), then the congruence $ax^{54}-y^{54}\equiv 0 \pmod{163}$ has no nontrivial solution. Then an argument similar to the one following Lemma 2.6 shows that the form

$$(ax_1^{54}-y_1^{54})+163(ax_2^{54}-y_2^{54})+\cdots+163^{53}(ax_{54}^{54}-y_{54}^{54})$$

has 108 variables and no nontrivial 163-adic zeros. This gives us the exact value

$$\Gamma^*(54,163)=109,$$

and completes the proof of Theorem 1.2.

Acknowledgements. The authors would like to thank the anonymous referee for making constructive comments which have improved the quality of this article. During the preparation of this article, Daiane Veras was supported by grant 141366/2014-1 from CNPq. The results in this article make up part of her Ph.D. thesis from the Universidade de Brasília.

References

[1] R. G. Bierstedt, *Some problems on the distribution of kth power residues modulo a prime*, PhD thesis, Univ. of Colorado at Boulder, 1963.

[2] J. D. Bovey, *$\Gamma^*(8)$*, Acta Arith. 25 (1974), 145–150.

[3] S. Chowla, *On a conjecture of J. F. Gray*, Norske Vid. Selsk. Forh. Trondheim 33 (1960), 58–59.

[4] H. Davenport and D. J. Lewis, *Homogeneous additive equations*, Proc. Roy. Soc. Ser. A 274 (1963), 443–460.

[5] M. Dodson, *Homogeneous additive congruences*, Philos. Trans. Roy. Soc. London Ser. A 261 (1967), 163–210.

[6] J. F. Gray, *Diagonal forms of prime degree*, PhD thesis, Univ. of Notre Dame, 1958.

[7] M. P. Knapp, *Exact values of the function $\Gamma^*(k)$*, J. Number Theory 131 (2011), 1901–1911.

[8] M. P. Knapp, *2-Adic zeros of diagonal forms*, J. Number Theory 193 (2018), 37–47.

[9] D. J. Lewis, *Cubic congruences*, Michigan Math. J. 4 (1957), 85–95.

[10] K. K. Norton, *On homogeneous diagonal congruences of odd degree*, PhD thesis, Univ. of Illinois at Urbana-Champaign, 1966.

Hemar Godinho
Departamento de Matemática
Universidade de Brasília
Brasília, DF 70910-900, Brazil
E-mail: hemar@mat.unb.br

Michael P. Knapp
Department of Mathematics and Statistics
Loyola University Maryland
Baltimore, MD 21210-2699, U.S.A.
E-mail: mpknapp@loyola.edu

Paulo H. A. Rodrigues
Instituto de Matemática e Estatística
Universidade Federal de Goiás
Goiânia, GO 74690-900, Brazil
E-mail: paulo_rodrigues@ufg.br

Daiane Veras
Instituto federal de Goiás
Avenida Saia Velha, Km 6, BR-040, s/n
Parque Esplanada V
Valparaíso de Goiás, GO 72876-601, Brazil
E-mail: daianemat2@gmail.com

Abstract (will appear on the journal's web site only)

For $k \in \mathbb{N}$ and p a prime number, define $\Gamma^*(k,p)$ to be the smallest $n \in \mathbb{N}$ such that every diagonal form $a_1x_1^k + \cdots + a_sx_s^k$ with integer coefficients has a nontrivial zero over $\mathbb{Q}_p$ whenever $s \geq n$. Define also

$$\Gamma^*(k) = \max_{p\,\mathrm{prime}} \Gamma^*(k,p).$$

We prove an upper bound for $\Gamma^*(k,p)$ and show that it is equal to $\Gamma^*(k,p)$ whenever $p-1$ divides k. We also find the exact value of $\Gamma^*(54)$.