Astérisque

WOLFGANG M. SCHMIDT

Congruences and equations

Astérisque, tome 94 (1982), p. 165-173

<http://www.numdam.org/item?id=AST_1982__94__165_0>

CONGRUENCES AND EQUATIONS

by

Wolfgang M. SCHMIDT

-:-:-:-

Part I. - Congruences

Suppose $\mathfrak{F}_1(x_1, \dots, x_s), \dots, \mathfrak{F}_r(x_1, \dots, x_s)$ are cubic forms with integer coefficients. We propose to study systems of congruences

$$\mathfrak{F}_i(x_1, \dots, x_s) \equiv 0 \pmod{p^{\ell}} \quad (i = 1, \dots, r)$$

where p is a prime, or in vector notation,

$$\mathfrak{F}(\underline{\underline{x}}) \equiv \underline{\underline{0}} \pmod{p^{\ell}} \tag{1}$$

where $\underline{\underline{x}} = (x_1, \dots, x_s)$ and $\underline{\underline{\mathfrak{F}}} = (\mathfrak{F}_1, \dots, \mathfrak{F}_r)$. A solution $\underline{\underline{x}}$ is called *primitive* if it is $\not\equiv \underline{\underline{0}} \pmod{p}$.

THEOREM 1. - *There is a primitive solution if*

$$s > 50,000 \,.\, r^3 \;.$$

It is well known that a system of homogeneous congruences has a primitive solution modulo each power of p if and only if the corresponding system of equations

$$\underline{\underline{\mathfrak{F}}}(\underline{\underline{x}}) = \underline{\underline{0}}$$

has a nontrivial solution in the p-adic field $\mathbb{Q}_p$. Hence our result is equivalent to

THEOREM 1'. - *Given* r *cubic forms with* p-*adic coefficients, they always have a common nontrivial* p-*adic zero if the number of variables exceeds* $50,000 \,.\, r^3$.

The map $\underline{\underline{x}} \mapsto \underline{\underline{\mathfrak{F}}}(\underline{\underline{x}})$ induces a map $X_\ell \to A_\ell$, where X_ℓ is the group of s-tuples $\underline{\underline{x}} \pmod{p^\ell}$, and A_ℓ is the group of r-tuples $\underline{\underline{a}} = (a_1, \dots, a_r) \pmod{p^\ell}$. In any finite commutative group G of order $|G|$ we have

$$\sum_\chi \chi(g) = \begin{cases} |G| & \text{if } g = 0, \\ 0 & \text{otherwise,} \end{cases}$$

where χ runs through the characters of G. The characters of A_ℓ are

$$\chi_{\underline{\underline{\alpha}}}(\underline{\underline{a}}) = e(\underline{\underline{a}} \cdot \underline{\underline{\alpha}}),$$

where $e(z) = e^{2\pi i z}$ and where $\underline{\underline{\alpha}}$ runs through the set $\mathfrak{U}_\ell$ consisting of the points of the unit cube $0 \leqq \alpha_i < 1$ $(i = 1, \dots, r)$ whose components are rational with denominator p^ℓ. Combining these facts one obtains the following well known relation for the number ν_ℓ of solutions of (1):

$$\nu_\ell = p^{-\ell r} \sum_{\underline{\underline{\alpha}} \in \mathfrak{U}_\ell} \Big(\sum_{\underline{\underline{x}} \in X_\ell} e(\underline{\underline{\alpha}} \cdot \underline{\underline{\mathfrak{F}}}(\underline{\underline{x}})) \Big). \tag{2}$$

We are interested in ν_ℓ as a function of ℓ. It is reasonable to conjecture that in most cases

$$\nu_\ell \approx |X_\ell| / |A_\ell| = p^{\ell(s-r)}.$$

Hence we put

$$\mu_\ell = \nu_\ell \, p^{-\ell(s-r)},$$

and we hope to show that under some mild assumptions $\mu_\ell \approx 1$.

From the above expression for ν_ℓ it follows easily that

$$\mu_\ell = 1 + \lambda_1 + \dots + \lambda_\ell \tag{3}$$

where

$$\lambda_j = p^{-js} \sum_{\underline{\underline{\alpha}} \in \mathfrak{U}_j - \mathfrak{U}_{j-1}} \; \sum_{\underline{\underline{x}} \in X_j} e(\underline{\underline{\alpha}} \cdot \underline{\underline{\mathfrak{F}}}(\underline{\underline{x}})).$$

Here the summand 1 comes from the term $\underline{\underline{\alpha}} = \underline{\underline{0}}$ in (2), λ_1 comes from $\underline{\underline{\alpha}}$ with least denominator p, etc. The hope is that the summand 1 in (3) is the main term, more generally that $\underline{\underline{\alpha}}$ with small denominator give the main contribution. This is in fact often the case. The situation is similar as in the Circle Method, where reals close to rationals with small denominator give the main contribution.

In order to carry out our program, we need to know about exponential sums. Suppose we have a sum

$$\sum_{\underline{\underline{x}} \ (\mathrm{mod}\ m)} e(f(\underline{\underline{x}})/m) \ ,$$

where f is a polynomial with integer coefficients. It is convenient to think of the map $\underline{\underline{x}} \mapsto f(\underline{\underline{x}})/m$ as a map $X \to \mathbb{Q}/\mathbb{Z}$, where X is the group of s-tuples $\underline{\underline{x}} \ (\mathrm{mod}\ m)$. At this point it is useful to generalize. A map

$$\mathfrak{F} : G \to H \ ,$$

where G, H are Abelian groups, will be called a polynomial of degree $\leqq k$ if its "k-th symmetrization"

$$\mathfrak{F}_k(g_1, \dots, g_k) = \sum_{\varepsilon_1=0}^{1} \dots \sum_{\varepsilon_k=0}^{1} \mathfrak{F}(\varepsilon_1 g_1 + \dots + \varepsilon_k g_k)$$

is a multilinear form, i.e., if it is a homomorphism in each variable. When G is finite and $H = \mathbb{Q}/\mathbb{Z}$, we may introduce the exponential sum

$$S(\mathfrak{F}) = \sum_{g \in G} e(\mathfrak{F}(g)) \ .$$

In the case where $\mathfrak{F}$ is a quadratic polynomial, i.e., a polynomial of degree $\leqq 2$, the sums are essentially Gauss sums, and

$$|S(\mathfrak{F})|^2 = \begin{cases} |G| \ |N| & \text{or} \\ 0 \ , & \end{cases}$$

where N is the period subgroup, consisting of $n \in G$ such that $\mathfrak{F}(g+n) = \mathfrak{F}(g)$ for all $g \in G$. Thus qualitatively, *either* $|S(\mathfrak{F})|$ *is small*, *or the subgroup* N *is large*. In the cubic case our result is of a similar nature :

THEOREM 2. - *Suppose* $\mathfrak{F} : G \to \mathbb{Q}/\mathbb{Z}$ *is a polynomial of degree* $\leqq 3$, *where the order of* G *is a power of a prime* p. *Let* α *be real and* a *natural*. *Then either* $|S(\mathfrak{F})| \leqq |G| \, p^{-\alpha}$, *or there is a subgroup* K *of* G *which is so large that the subgroup* $p^{a-1}(G/K)$ *of* G/K *has order*

$$|p^{a-1}(G/K)| \leqq p^{\gamma \max(a, \alpha^2/a)} \tag{4}$$

where γ *is an absolute constant*, *and such that* $\mathfrak{F}$ *is constant on some coset of* K.

This may sound a bit complicated, but the case $a = 1$ of (4) is simply $|G/K| \leqq p^{\gamma \alpha^2}$. Also, when $p > 3$ and $\mathfrak{F}$ is a cubic form, it is possible to conclude that $\mathfrak{F}$ vanishes on K. In the case where G is a sum of groups of order p, a somewhat stronger result was established by Davenport and Lewis [4]. They also gave an example which shows that the situation is rather different for quartic polynomials.

Now let us return to the congruences. We clearly cannot expect that μ_ℓ is close to 1 is $\underline{\underline{\mathfrak{F}}}$ is identically zero, or if all the coefficients of the forms involved are highly divisible by p. In this case it is simpler to deal with $p^{-1}\underline{\underline{\mathfrak{F}}}$, and we write $\underline{\underline{\mathfrak{F}}} \succ p^{-1}\underline{\underline{\mathfrak{F}}}$. More generally, when both $\underline{\underline{\mathfrak{F}}}$ and $\underline{\underline{\mathfrak{F}}}'$ are systems of forms of degree d with integer coefficients, we write

$$\underline{\underline{\mathfrak{F}}} \succ \underline{\underline{\mathfrak{F}}}'$$

if the following holds : $\underline{\underline{\mathfrak{F}}}'(\underline{\underline{x}}) = T^{-1}\,\underline{\underline{\mathfrak{F}}}(\tau(\underline{\underline{x}}))$, where $T : \mathbb{Q}^r \to \mathbb{Q}^r$ and $\tau : Q^s \to Q^s$ are nonsingular linear transformations which map integer points into integer points and which have

$$|\det T|_p^{s/dr} < |\det \tau|_p ,$$

where $|\ldots|_p$ is the p-adic absolute value. The exponent s/dr can be remembered by the remark that the transformations T, τ with $T(\underline{y}) = p^d \underline{y}$, $\tau(\underline{\underline{x}}) = p\underline{\underline{x}}$ have $\underline{\underline{\mathfrak{F}}} = T^{-1}\,\underline{\underline{\mathfrak{F}}}(\tau(\underline{\underline{x}}))$ and $|\det T|_p^{s/dr} = |\det \tau|_p$.

We call $\underline{\underline{\mathfrak{F}}}$ reduced if there is no $\underline{\underline{\mathfrak{F}}}'$ with $\underline{\underline{\mathfrak{F}}} \succ \underline{\underline{\mathfrak{F}}}'$. We call $\underline{\underline{\mathfrak{F}}}$ bottomed if it is reduced or if there is a reduced system $\underline{\underline{\mathfrak{F}}}^*$ with $\underline{\underline{\mathfrak{F}}} \succ \underline{\underline{\mathfrak{F}}}^*$. Otherwise we call $\underline{\underline{\mathfrak{F}}}$ bottomless ; in this case there is an infinite chain $\underline{\underline{\mathfrak{F}}} \succ \underline{\underline{\mathfrak{F}}}^{(1)} \succ \ldots$. It turns out that the bottomless systems form an algebraic sub set of the space of systems, and in fact a proper algebraic subset when $s > r$. Thus the bottomed systems are "everywhere dense".

THEOREM 3. (A) - Suppose $d = 2$ and $s > 4r^2 + 4r$. Then if the system is reduced, we have

$$|\mu_\ell - 1| < p^{r+1-(s/4r)} < 1 .$$

In this case, as well as in the more general case when the system is bottomed, we have

$$c_1 p^{\ell(s-r)} < \nu_\ell < c_2 p^{\ell(s-r)} \tag{5}$$

with positive constants c_1, c_2.

(B) - Suppose $d = 3$ and $s > Cr^3$ where C is an absclute constant. Then if the system is reduced, we have

$$|\mu_\ell - 1| < p^{r - \sqrt{s/Cr}} < 1 .$$

In this case, as well as in the case when the system is bottomed, (5) holds.

It can be deduced that a bottomed system of r quadratic forms has a p-adic zero when $s > 4r^2 + 4r$. Since bottomed systems are everywhere dense, a simple limit argument shows that in fact a common zero always exists when $s > 4r^2 + 4r$. Similarly, a system of r cubic forms has a p-adic zero when $s > Cr^3$.

Let $v_{d,r}$ be the smallest number (if such numbers exist) such that a system of r forms of degree d in more than $v_{d,r}$ variables always has a p-adic zero for each prime p . It follows e.g. from work of R. Brauer [2] that all the $v_{d,r}$ exist, but it had usually been assumed that the elementary inductive arguments employed by Brauer led to totally unpractical bounds. However, this is not so, and Leep [5] used such an argument to show that

$$v_{2,r} \leq 2r^2 + 2r ,$$

which is stronger than the estimate $v_{2,r} \leqq 4r^2 + 4r$ implied above. Recently Leep and the author [6] used elementary methods to show that

$$v_{3,r} < \frac{81}{2} r^4 .$$

The analytic method above gives $v_{3,r} \leqq Cr^3$, and with a little extra effort the bound

$$v_{3,r} \leqq 50,000 . r^3$$

of Theorem 1' follows.

Incidentally, Leep and the author also found that

$$v_{d.1} \ll e^{(d!)^2 (1+\varepsilon)^d}$$

for each $\varepsilon > 0$. This is admittedly a very poor bound, but I don't remember seeing any bound in the literature.

Part II. - Equations

Let $w_{d,r}$ be the smallest number (if such numbers exist) such that a system of r forms of degree d with rational coefficients has a nontrivial rational zero if the number of variables exceeds $w_{d,r}$. It is clear that $w_{d,r}$ does not exist when d is even, but Birch [1] has shown that it does exist when d is odd. There does not seem to be an elementary induction argument which yields reasonable bounds for such $w_{d,r}$.

Davenport [3] used analytic methods to show that $w_{3,1} \leqq 15$. It had been known for some time that $w_{d,r} \geqq d^2 r$, whence that $w_{3,1} \geqq 9$. Now we have

THEOREM 4. - $w_{3,r} \leqq (10r)^5$.

That is, a system of r *cubic forms has a nontrivial rational zero if the number of variables exceeds* $(10r)^5$.

In many cases it is possible to give an asymptotic formula for the number of integer zeros in certain boxes :

THEOREM 5. - *Let* $\mathfrak{F}_1, \dots, \mathfrak{F}_r$ *be cubic forms in* s *variables, with integer coefficients. Suppose that no form of the rational pencil generated by* $\mathfrak{F}_1, \dots, \mathfrak{F}_r$ *vanishes on a subspace of* $\mathbb{Q}^s$ *of*

$$\text{codim} \leqq 10\, r^2 + 6\, r \,. \tag{6}$$

Given a box $\mathfrak{B} \subset \mathbb{R}^s$ *with sides parallel to the coordinate axes, write*

$$z_P = z_P(\underline{\underline{\mathfrak{F}}}, \mathfrak{B})$$

for the number of common integer point zeros of our forms in the "blown up" box $P\mathfrak{B}$.

Then as $P \to \infty$,

$$z_P = \mathfrak{S}\, P^{s-3r} + O(P^{s-3r-\delta}) \tag{7}$$

where $\delta > 0$ *and* $\mathfrak{S} = \mathfrak{S}(\underline{\underline{\mathfrak{F}}}, \mathfrak{B})$ *are independent of* P.

No bound for s is explicitly postulated, but (6) implies that $s > 10\, r^2 + 6r$. The relation (7) gives an asymptotic formula for z_P only when $\mathfrak{S} > 0$. So what about $\mathfrak{S}$? This constant is given as an infinite product

$$\mathfrak{C} = \chi(\infty)\ \chi(2)\ \chi(3)\ \chi(5)\ \dots\ .$$

Here $\chi(\infty)$ is connected with the "density of zeros" of our forms in the real field $\mathbb{R} = \mathbb{Q}_\infty$, while $\chi(p)$ for a prime p may be called the "density" of p-adic zeros, i.e. zeros in $\mathbb{Q}_p$. Only the factor $\chi(\infty)$ depends on the box $\mathfrak{B}$. This factor is usually called the "Singular Integral" and denoted by $\mathfrak{J}$, while the product $\chi(2)\ \chi(3)\dots$ is called the "Singular Series" and denoted by $\mathfrak{S}$. Two additional results are required in order to guarantee that $\mathfrak{C} > 0$.

FIRST SUPPLEMENT. - $\mathfrak{J} > 0$ if the manifold of real zeros of $\underline{\underline{\mathfrak{F}}}$ in the interior of $\mathfrak{B}$ has dimension $\geqq s - r$. In particular, $\mathfrak{J} > 0$ when $\mathfrak{B}$ contains the origin in its interior.

The second statement of this supplement follows from topology, e.g. from the Borsuk-Ulam Theorem.

SECOND SUPPLEMENT. - $\mathfrak{S} > 0$ if no form in the rational pencil generated by $\underline{\underline{\mathfrak{F}}}$ vanishes on a rational subspace of

$$\text{codim} \leqq 8r\, v_{3,r} + 2r^2 - 2r\ . \tag{8}$$

Notice that neither of our supplements needs the usual hypothesis about the existence of a nonsingular zero in the local fields. Theorem 5 together with its supplements gives

$$w_{3,r} \leqq w_{3,r-1} + 8r\, v_{3,r} + 2r^2 - 2r\ , \tag{9}$$

as we now proceed to show. For by the first supplement, we can make $\mathfrak{J} > 0$ by the proper choice of $\mathfrak{B}$. And $\mathfrak{S} > 0$ by the second supplement if the condition with (8) holds, which is stonger than (6). On the other hand if the condition involving (8) is violated, then we may in fact suppose without loss of generality that $\mathfrak{F}_r$ vanishes on a rational subspace S with (8). Now if the number of variables exceeds the right hand side of (9), then $\dim S > w_{3,r-1}$, and by definition the forms $\mathfrak{F}_1, \dots, \mathfrak{F}_{r-1}$ have a common zero on S .

Theorem 4 follows from estimates of $v_{3,r}$ and from (9) by induction on r .

In terms of the quantities $\mu_\ell = \mu_\ell(p)$ and $\lambda_j = \lambda_j(p)$ defined at the beginning, the local density $\chi(p)$ is given by

$$\chi(p) = \lim_{\ell \to \infty} \mu_\ell = 1 + \lambda_1 + \lambda_2 + \dots\ . \tag{10}$$

It is a fairly elementary fact that

$$1 + \lambda_1 + \ldots + \lambda_\ell = \mu_\ell \gg p^{\ell(r - v_{3,r})} .$$

On the other hand the hypothesis of the second supplement implies that

$$\lambda_j \ll p^{j(r - v_{3,r} - \delta)}$$

where $\delta > 0$, and combining these two estimates one finds that $\chi(p) > 0$.

With a cubic form $\mathfrak{J}$ there is associated a symmetric trilinear form $\mathfrak{T}(\underline{\underline{x}}, \underline{y}, \underline{\underline{z}})$ such that $\mathfrak{J}(\underline{\underline{x}}) = \mathfrak{T}(\underline{\underline{x}}, \underline{\underline{x}}, \underline{\underline{x}})$. Consider the set $\mathfrak{B}$ of pairs $\underline{\underline{x}}, \underline{y}$ such that $\mathfrak{T}(\underline{\underline{x}}, \underline{y}, \underline{\underline{z}})$ vanishes identically in $\underline{\underline{z}}$. The proof of Theorem 5 depends on the fact (here stated only in vague form) that *if* $\mathfrak{B}$ *is large*, *then there is a large subspace* S *on which* $\mathfrak{J}$ *is zero*. Probably subsequent improvements of our results will come from a better understanding of this phenomenon. To deal with forms of higher degree, it will be necessary to discover corresponding phenomena for multilinear forms in general.

A discussion of bottomed and bottomless systems of forms, and part A of Theorem 3 on quadratic forms, may be found in [7]. The details on cubic forms will appear in [8].

-:-:-:-

REFERENCES

[1] B. J. BIRCH, *Homogeneous forms of odd degree in a large number of variables*, Mathematika 4 (1962), 102-105.

[2] R. BRAUER, *A note on systems of homogeneous algebraic equations*, Bull. A.M.S. 51 (1945), 749-755.

[3] H. DAVENPORT, *Cubic forms in* 16 *variables*, Proc. Royal Soc. A, 272 (1963), 285-303.

[4] H. DAVENPORT and D. J. LEWIS, *Exponential sums in many variables*, Amer. J. Math. 84 (1962), 649-665.

[5] D. LEEP, *Systems of quadratic forms*, (in preparation).

[6] D. LEEP and W. M. SCHMIDT, *Systems of homogeneous equations*, (in preparation).

[7] W. M. SCHMIDT, Simultaneous p-adic zeros of quadratic forms, Monatsh. Math. 90 (1980), 45-65.

[8] W. M. SCHMIDT, On cubic polynomials I-IV, Monatsh. Math. (to appear).

-:-:-:-

Wolfgang M. SCHMIDT
University of Colorado
Boulder